

DOI:10.20176/j.cnki.nxdz.20251215

A Kind of Fast Iterative Methods With the Application Based on Diagonal Matrix Splitting

XU Qiuyan

(School of Mathematics and Statistics, Ningxia University, Yinchuan 750021, China)

Abstract: The fast solution of linear equations has always been one of the hot spots in scientific computing. A kind of the diagonal matrix splitting iteration methods are provided, which is different from the classical matrix splitting methods. Taking the decomposition of the diagonal elements for coefficient matrix as the key point, some new preconditioners are constructed. Taking the tri-diagonal coefficient matrix as an example, the convergence domains and optimal relaxation factor of the new method are analyzed theoretically. The presented new iteration methods are applied to solve linear algebraic equations, even 2D and 3D diffusion problems with the fully implicit discretization. The results of numerical experiments are matched with the theoretical analysis, and show that the iteration numbers are reduced greatly. The superiorities of presented iteration methods exceed some classical iteration methods dramatically.

Key words: iteration; matrix splitting; diffusion equation; convergence; optimal relaxation factor

CLC number: O241.6

Document code: A

1 Introduction

In the field of scientific and engineering computing, the solution of algebraic equations obtained by discretizing many mathematical models has always been a hot topic, especially with the progress of computers. Efficient numerical methods have always been a goal pursued by people, and iterative methods in numerical methods have become the mainstream solution methods. References [1–2] are two classic books, which introduced various iterative methods for solving large-scale linear algebraic equations systematically. From the Gauss–Seidel^[3] iteration to the Richardson method, the Chebyshev semi-iteration method and successive over relaxation (SOR) method in the 1950s and 1970s, the iteration methods have a long history. Hageman *et al*^[4] provided many applicable iterative methods. Axelsson^[5] introduced iterative solution methods comprehensively, including preprocessing techniques. Saad^[6] gave the iterative methods for sparse linear systems. Bai *et al*^[7–8] provided Hermitian/skew-Hermitian (HSS) splitting methods for non-Hermitian positive definite linear systems, and accelerated HSS methods for saddle-point problems. In references [9], a matrix splitting method has been proposed for special matrix and a comprehensive review of preprocessing techniques was provided. References [10] gave the application of HSS method in Sylvester equation. Li *et al*^[11] proposed a modified Gauss–Seidel type iterative methods and Jacobi type methods for Z -matrices. The interpolating preconditioners for the solution of sequence of linear systems were pro-

Received: 2025-05-09

Foundation item: The National Natural Science Foundations of China (12202219); the Natural Science Foundations of Ningxia (2024AAC02009, 2023AAC05001); the Ningxia Youth Top Talents Training Project

Biography: XU Qiuyan (1983—), female, associate professor, doctor of science, research fields: Numerical methods for partial differential equations, E-mail: qiuyanxu@nxu.edu.cn.

Citation format: A Kind of Fast Iterative Methods With the Application Based on Diagonal Matrix Splitting [J]. Journal of Ningxia University (Natural Science Edition in Chinese and English), 2026, 47(1): 1-13.

vided in references [12]. A generalized successive overrelaxation iterative method for a class of complex symmetric linear system of equations are given in references [13]. New matrix splitting iteration method for generalized absolute value equations was provided in references [14], which proposed a relaxed Newton-type matrix splitting iteration method. References [15] introduced two-parameter modified matrix splitting iteration method, established the asymptotic convergence theory and demonstrated its convergence under certain conditions. References [16] constructed a paradigm of two-step matrix splitting iteration methods and analyze its convergence property for the nonsingular and the positive-definite matrix.

Under certain conditions, the classical Jacobi, Gauss-Seidel iteration methods can solve many practical problems. Especially for the SOR^[17-19] method, which adds the optimal relaxation factor, the iteration acceleration has been greatly improved. However, we find that in the Jacobi iteration method and the Gauss-Seidel iteration method, the most primitive information are not fully utilized. They all use the initial iteration information of other points to calculate. Xu *et al*^[20-21] developed a new kind of iteration methods based on the discretization schemes with some iterative operators, which are more efficient than Jacobi, GS and SOR methods. To make more use of initial information to construct iterations, we need to split the main diagonal elements, that is, the diagonal matrix constructed by the main diagonal elements, to obtain a new kind of iterative methods. Therefore, we will propose a diagonal matrix splitting iteration methods. The classical Jacobi, Gauss-Seidel and SOR iteration methods are actually special cases of the presented methods. We have added the concept of diagonal splitter and provided the range of diagonal splitter to ensure the convergence of the generalized methods. Finally, the new methods are applied to solve linear algebraic equations and 2D and 3D diffusion problems. The experimental results show that the generalized iterative methods greatly improve computational efficiency, far surpassing classical iterative methods.

The remainder of the paper is as follows: In Sec. 2, we present new generalized iterative methods for solving linear algebraic equations; Sec. 3 provides the range of the diagonal splitter and the optimal relaxation factors of the presented iteration methods for tridiagonal linear equations. The new generalized iteration methods are applied to solve linear equations, 2D and 3D diffusion problems in Sec. 4. Finally, Sec. 5 gives the conclusions.

2 Generalized iteration methods

In this section, we provide a class of new iterative methods based on the splitting strategy, which is the extension of the classical Jacobi, Gauss-Seidel and SOR methods in fact. So we call the new methods as generalized iteration methods.

2.1 Generalized Jacobi, GS, SOR methods

Consider the following linear equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1,l-1}x_{l-1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2,l-1}x_{l-1} = b_2, \\ \vdots \\ a_{l-1,1}x_1 + a_{l-1,2}x_2 + a_{l-1,3}x_3 + \cdots + a_{l-1,l-1}x_{l-1} = b_{l-1}. \end{cases} \quad (1)$$

Which can be often written as

$$\mathbf{Ax} = \mathbf{b}. \quad (2)$$

Where $A = (a_{ij})_{(l-1) \times (l-1)}$, $\mathbf{x} = (x_j)_{(l-1) \times 1}$, $\mathbf{b} = (b_j)_{(l-1) \times 1}$, $i, j = 1, 2, \dots, l-1$. There are many methods to solve Eq. (1) or (2) like the classical Jacobi, Gauss-Seidel (GS) and SOR iterative methods and so on. We often introduce these iterative methods by the matrix splitting method, namely, the coefficient matrix can be divided into $A = D - L - U$, here $D = \text{diag}(A)$, and

$$\mathbf{L} = \begin{bmatrix} 0 & & & & \\ -a_{21} & 0 & & & \\ & & \ddots & & \\ & & & \ddots & \\ -a_{l-1,1} & -a_{l-1,2} & & & 0 \end{bmatrix}_{(l-1) \times (l-1)}, \quad \mathbf{U} = \begin{bmatrix} 0 & -a_{12} & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & -a_{l-2,l-1} & \\ & & & & 0 \end{bmatrix}_{(l-1) \times (l-1)}. \quad (3)$$

We will provide a new splitting iterative method, which can accelerate the iteration dramatically. In fact, the classical Jacobi, GS and SOR iterative methods can be extended.

First, we divide the elements of diagonal matrix D into two parts,

$$D=B+C, \tag{4}$$

namely, $a_{ii} = b_{ii} + c_{ii}, i = 1, 2, \dots, l - 1$, here

$$B = \begin{bmatrix} b_{11} & & & & \\ & b_{22} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & b_{l-1,l-1} \end{bmatrix}_{(l-1) \times (l-1)}, \quad C = \begin{bmatrix} c_{11} & & & & \\ & c_{22} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & c_{l-1,l-1} \end{bmatrix}_{(l-1) \times (l-1)}. \tag{5}$$

The matrix B and the elements $b_{ii}, i = 1, 2, \dots, l - 1$ are called diagonal retention matrix and diagonal retention elements, the matrix C and the elements $c_{ii}, i = 1, 2, \dots, l - 1$ are called diagonal splitting matrix and diagonal splitter.

Second, we construct the following iterative methods:

$$I : \begin{cases} b_{11}x_1^{(k+1)} + c_{11}x_1^{(k)} + a_{12}x_2^{(k)} + a_{13}x_3^{(k)} + \dots + a_{1,l-1}x_{l-1}^{(k)} = b_1, \\ a_{21}x_1^{(k)} + b_{22}x_2^{(k+1)} + c_{22}x_2^{(k)} + a_{23}x_3^{(k)} + \dots + a_{2,l-1}x_{l-1}^{(k)} = b_2, \\ \vdots \\ a_{l-1,1}x_1^{(k)} + a_{l-1,2}x_2^{(k)} + a_{l-1,3}x_3^{(k)} + \dots + b_{l-1,l-1}x_{l-1}^{(k+1)} + c_{l-1,l-1}x_{l-1}^{(k)} = b_{l-1}, \end{cases} \tag{6}$$

which is called generalized Jacobi (G-Jacobi) iterative method. When $C=0$, namely, $c_{ii}=0, i=1, 2, \dots, l - 1$, method I is just the classical Jacobi method.

In Eq. (6), we can see that once the vector $x^{(k)}$ is given, the new vector $x^{(k+1)}$ can be computed. The G-Jacobi iteration method has the essential parallelism as the same as the Jacobi method. Eq. (6) can be written in element form to show exactly how to update the approximate solution vector, namely,

$$x_i^{(k+1)} = \frac{1}{b_{ii}} \left(b_i - \sum_{j \neq i} a_{ij}x_j^{(k)} - c_{ii}x_i^{(k)} \right), \quad i = 1, 2, \dots, l - 1.$$

Next, we can construct another new iterative method,

$$II : \begin{cases} b_{11}x_1^{(k+1)} + c_{11}x_1^{(k)} + a_{12}x_2^{(k)} + a_{13}x_3^{(k)} + \dots + a_{1,l-1}x_{l-1}^{(k)} = b_1, \\ a_{21}x_1^{(k+1)} + b_{22}x_2^{(k+1)} + c_{22}x_2^{(k)} + a_{23}x_3^{(k)} + \dots + a_{2,l-1}x_{l-1}^{(k)} = b_2, \\ \vdots \\ a_{l-1,1}x_1^{(k+1)} + a_{l-1,2}x_2^{(k+1)} + a_{l-1,3}x_3^{(k+1)} + \dots + b_{l-1,l-1}x_{l-1}^{(k+1)} + c_{l-1,l-1}x_{l-1}^{(k)} = b_{l-1}, \end{cases} \tag{7}$$

which is also called generalized Gauss-Seidel (G-GS) iterative method. Similarly, When $C=0$, namely, $c_{ii} = 0, i = 1, 2, \dots, l - 1$, the method II is the classical Gauss-Seidel method. The G-GS iterative method can be written in the form

$$x_i^{(k+1)} = \frac{1}{b_{ii}} \left(b_i - \sum_{j < i} a_{ij}x_j^{(k+1)} - \sum_{j > i} a_{ij}x_j^{(k)} - c_{ii}x_i^{(k)} \right), \quad i = 1, 2, \dots, l - 1.$$

It can be seen that the latest approximations to the components of x are used in the update of subsequent components. The old components $x^{(k)}$ are overwritten with those of $x^{(k+1)}$ as soon as they are computed.

The improvement of the G-GS method can be provided by introducing a relaxation parameter ω . We add a relaxation factor ω to construct the G-SOR iterative method as follows:

$$III : x^{(k+1)} = x^{(k)} + \omega \Delta x^{(k)},$$

here $\Delta x^{(k)} = x^{(k+1)} - x^{(k)}$, while $x^{(k+1)}$ is obtained by the Eq. (7).

The G-SOR method is defined by

$$x_i^{(k+1)} = \frac{\omega}{b_{ii}} \left(b_i - \sum_{j < i} a_{ij}x_j^{(k+1)} - \sum_{j > i} a_{ij}x_j^{(k)} - c_{ii}x_i^{(k)} \right) + (1 - \omega)x_i^{(k)}, \quad i = 1, 2, \dots, l - 1. \tag{8}$$

When $\omega=1$, the G-SOR method is equal to G-GS method; and when $c_{ii}=0, i=1, 2, \dots, l-1$, the G-SOR method is just the classical SOR method.

For above three generalized iterative methods, their iterative matrices are

$$\begin{cases} M_I = B^{-1}(L + U - C) = (D - C)^{-1}(L + U - C), \\ M_{II} = (B - L)^{-1}(U - C) = (D - C - L)^{-1}(U - C), \\ M_{III} = (B - \omega L)^{-1}((1 - \omega)B + \omega(U - C)). \end{cases}$$

Therefore, the classical Jacobi, GS and SOR iteration methods are the special cases of the presented generalized iteration methods. The theoretical analyses can also be provided.

2.2 Convergence analysis

Theorem 1^[2] Given the initial vector $x^{(0)}$, the G-Jacobi iterative method is convergent only if the spectral radius of the iterative matrix M_I satisfies $\rho(M_I) < 1$.

Lemma 1^[2] If the matrix A is strictly diagonal dominant or irreducible diagonal dominant, then the matrix A is nonsingular.

Definition 1^[2] When the matrix A is strictly diagonal dominant or irreducible diagonal dominant symmetric matrix, and its diagonal elements are all positive, which is called that A is positive definite matrix.

Theorem 2 If the coefficient matrix A in Eq. (2) is symmetric, and the diagonal retention matrix $B > 0$, then the G-Jacobi iterative method is convergent only if A and $2B - A$ are all positive definite.

Proof Because the matrix A is symmetric, and $B > 0$, we get

$$M_I = B^{-1}(L + U - C) = I - B^{-1}A = B^{1/2}(I - B^{-1/2}AB^{-1/2})B^{1/2}.$$

It shows that M_I is equivalent to $I - B^{-1/2}AB^{-1/2}$, and both of them have the same eigenvalues. Moreover, $I - B^{-1/2}AB^{-1/2}$ is symmetric, which has real eigenvalues.

(i) The necessity. If the iterative method is convergent, then $\rho(M_I) < 1$. We can obtain that all of the eigenvalues of the matrix $I - B^{-1/2}AB^{-1/2}$ are belong to $(-1, 1)$, which means

$$|\lambda_{I - B^{-1/2}AB^{-1/2}}| < 1,$$

then

$$0 < \lambda_{B^{-1/2}AB^{-1/2}} < 2.$$

So the matrix $B^{1/2}AB^{-1/2}$ is positive definite, which indicates A is positive definite. Furthermore, due to

$$B^{1/2}(2B - A)B^{-1/2} = 2I - B^{1/2}AB^{-1/2},$$

then the matrix $2B - A$ is positive definite.

(ii) The sufficiency. Since $B^{1/2}(I - M_I)B^{-1/2} = B^{1/2}AB^{-1/2}$, which means that the eigenvalues of $I - M_I$ are all greater than 0, then we have

$$\lambda_{M_I} < 1 \quad (9)$$

Otherwise, due to $A = B(I - M_I)$, it has $B^{-1/2}(2B - A)B^{-1/2} = B^{1/2}(I + M_I)B^{-1/2}$.

Since $2B - A$ is positive definite, $\lambda_{I + M_I} > 1$, we obtain

$$\lambda_{M_I} > -1. \quad (10)$$

From (9) and (10), $\rho(M_I) < 1$. So the G-Jacobi iterative method is convergent.

Theorem 3 If the coefficient matrix A in Eq. (2) is strictly diagonal dominant or irreducible diagonal dominant, and the diagonal retention matrix $B = \text{diag}(b_{ii})_{(l-1) \times (l-1)}$, $b_{ii} > 0, i = 1, 2, \dots, l - 1$, then the G-Jacobi iterative method is convergent.

Proof Due to $b_{ii} > 0, i = 1, 2, \dots, l - 1$, then the diagonal retention matrix B is reversible. Assume $\tilde{\lambda}$ is the eigenvalue of the iterative matrix M_I , which satisfy $|\tilde{\lambda}| \geq 1$. Then we can obtain that the matrix $\tilde{\lambda}B - L - U + C$ is also strictly diagonal dominant or irreducible diagonal dominant because of the strictly diagonal dominance or irreducible diagonal dominance of the matrix A . From the Lemma 1, we have

$$\det(\tilde{\lambda}I - M_I) = \det[\tilde{\lambda}I - B^{-1}(L + U - C)] = \det(B^{-1})\det(\tilde{\lambda}B - L - U + C) \neq 0.$$

which is contradictory. Therefore, the spectral radius $\rho(M_I) < 1$. By Theorem 1, the G-Jacobi iterative method is convergent.

Theorem 4 If the coefficient matrix A in Eq. (2) is a real symmetric positive definite matrix, and the diagonal retention matrix $B > 0$, then the G-SOR iterative method is convergent when

$$0 < \omega < \frac{2\|B\|_2}{\|B\|_2 + \|C\|_2}.$$

Here $\|\cdot\|_2$ means 2-norm, and ① If $\|B\|_2 \leq \|C\|_2$, then $0 < \omega < 1$. ② If $\|B\|_2 > \|C\|_2$, then $\frac{2\|B\|_2}{\|B\|_2 + \|C\|_2} > 1$, which means $\omega = 1$, the G-GS iterative method is convergent. ③ If $\|C\|_2 = 0$, the G-SOR method is just the SOR method, which is convergent when $0 < \omega < 2$.

Proof Assume λ is any eigenvalue of the iterative matrix M and v is the corresponding eigenvector, then we have

$$(B - \omega L)^{-1}[(1 - \omega)B + \omega(U - C)]v = \lambda v. \tag{11}$$

Let v^H be the conjugate transpose of v . Denote $v^H B v = \delta_1, v^H C v = \delta_2, v^H U v = \alpha + i\beta, i = \sqrt{-1}$. $\delta_1, \delta_2, \alpha, \beta$ are all real numbers. Since $L = U^T$, we get

$$v^H L v = \alpha - i\beta. \tag{12}$$

Taking the above relationships into (11), then

$$\lambda = \frac{\delta_1 - \omega(\delta_1 + \delta_2 - \alpha) + i\omega\beta}{\delta_1 - \omega\alpha + i\omega\beta}. \tag{13}$$

Computing the square of module in both sides for Eq. (13), we get

$$|\lambda|^2 = \frac{[\delta_1 - \omega(\delta_1 + \delta_2 - \alpha)]^2 + \omega^2\beta^2}{(\delta_1 - \omega\alpha)^2 + \omega^2\beta^2}.$$

Note that

$$d = [\delta_1 - \omega(\delta_1 + \delta_2 - \alpha)]^2 - (\delta_1 - \omega\alpha)^2 = \omega(2\delta_1 - \omega\delta_1 - \omega\delta_2)(2\alpha - \delta_1 - \delta_2),$$

due to A is real symmetric positive and $B > 0$, then

$$\delta_1 > 0, \delta_1 + \delta_2 > 0, v^H A v = \delta_1 + \delta_2 - 2\alpha > 0,$$

we can obtain that $d < 0$, namely, $|\lambda| < 1$ when

$$0 < \omega < \frac{2\delta_1}{\delta_1 + \delta_2}.$$

Theorem 5 If the coefficient matrix A in Eq. (2) is strictly diagonal dominant or irreducible diagonal dominant, and satisfies $|b_{ii}| > \sum_{i \neq j} |a_{ij}| + |c_{ii}|$, then the G-SOR iterative method is convergent when $\omega \in (0, 1]$.

Proof Assume the iterative matrix $M_{\square} = (B - \omega L)^{-1}[(1 - \omega)B + \omega(U - C)]$ has the eigenvalue $\tilde{\lambda}$ which is satisfied $|\tilde{\lambda}| \geq 1$, then

$$\begin{aligned} \det(\tilde{\lambda}I - M_{\square}) &= \det[(B - \omega L)^{-1}] \det[\tilde{\lambda}(B - \omega L) - (1 - \omega)B - \omega(U - C)] = \\ &= \det[(B - \omega L)^{-1}] \det[(\tilde{\lambda} - 1 + \omega)B - \omega(\tilde{\lambda}L + U - C)]. \end{aligned}$$

Since $\omega \in (0, 1]$ and B is strictly diagonal dominant or irreducible diagonal dominant, we have

$$|\tilde{\lambda} - 1 + \omega| |b_{ii}| > \omega |\tilde{\lambda}| \left[\sum_{i > j} |a_{ij}| + \sum_{i < j} |a_{ij}| + |c_{ii}| \right] > \omega \left[|\tilde{\lambda}| \sum_{i > j} |a_{ij}| + \sum_{i < j} |a_{ij}| + |c_{ii}| \right].$$

From Lemma 1, we have

$$\det[(\tilde{\lambda} - 1 + \omega)B - \omega(\tilde{\lambda}L + U - C)] \neq 0,$$

which is contradictory. Therefore, the spectral radius of G-SOR iterative matrix $\rho(M_{\square}) < 1$. So the G-SOR iterative method is convergent when $\omega \in (0, 1]$.

3 Tridiagonal linear equations

In this section, we take tridiagonal linear equations for example to provide the eigenvalues, eigenvectors, the range of the diagonal splitter, and the optimal relaxation factors of the presented new iterative methods.

Consider the linear equations (2) with the coefficient matrix

$$\bar{A} = \begin{bmatrix} d & \beta & & & \\ \beta & d & \beta & & \\ & \ddots & \ddots & \ddots & \\ & & & \beta & d \end{bmatrix}_{(l-1) \times (l-1)}, \quad (14)$$

we split the diagonal matrix of A into two parts, namely, $\text{diag}(A) = B + C$, here

$$\bar{B} = \begin{bmatrix} d - \gamma & & & & \\ & d - \gamma & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & d - \gamma \end{bmatrix}_{(l-1) \times (l-1)}, \quad \bar{C} = \begin{bmatrix} \gamma & & & & \\ & \gamma & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \gamma \end{bmatrix}_{(l-1) \times (l-1)}, \quad (15)$$

where γ is defined as the diagonal splitter.

Theorem 6 Let $\bar{A}_{(l-1) \times (l-1)}$ in (14) be the coefficient matrix for linear equations $\bar{A}\mathbf{x} = \bar{\mathbf{b}}$, the diagonal retention matrix \bar{B} and the diagonal splitting matrix \bar{C} are defined by (15), then the eigenvalues of the G-Jacobi iteration matrix are

$$\lambda_k = \frac{\gamma}{d - \gamma} + \frac{2\beta}{d - \gamma} \cos \frac{k\pi}{l}, \quad k = 1, 2, \dots, l-1. \quad (16)$$

Theorem 7 Let $\bar{A}_{(l-1) \times (l-1)}$ in (14) be the coefficient matrix for linear equations $\bar{A}\mathbf{x} = \bar{\mathbf{b}}$, the diagonal retention matrix \bar{B} and the diagonal splitting matrix \bar{C} are defined by (15), then the eigenvalues of the G-SOR iteration matrix are

$$\lambda_k = \left(\frac{\omega\beta}{d - \gamma} \cos \frac{k\pi}{l} + \sqrt{\frac{\omega^2\beta^2}{(d - \gamma)^2} \cos^2 \frac{k\pi}{l} - \left(\frac{\omega d}{d - \gamma} - 1\right)} \right)^2, \quad k = 1, 2, \dots, l-1. \quad (17)$$

The proofs of the Theorem 6 and Theorem 7 can be obtained by references [6]. From the Theorem 6 and Theorem 7, we can get the convergence domains about diagonal splitter γ of G-Jacobi and G-GS methods with the coefficient matrix in (14).

Theorem 8 Let $\bar{A}_{(l-1) \times (l-1)}$ be a coefficient matrix as in (14) for linear equations $\bar{A}\mathbf{x} = \bar{\mathbf{b}}$, which is diagonally dominant and $d > 0$. $\text{diag}(\bar{A}) = \bar{B} + \bar{C}$, $\bar{B} > 0$ means the diagonal splitter $\gamma < d$ as in (15). Then the G-Jacobi method is convergent when

$$\gamma < \frac{d}{2} (1 + \mu_{\min}), \quad (18)$$

here μ_{\min} is the minimum of the eigenvalues for the classical Jacobi iteration matrix.

Proof Note λ and μ are the eigenvalues of G-Jacobi and Jacobi iteration matrices. From the Theorem 6, we have

$$\lambda = \frac{\mu d - \gamma}{d - \gamma}, \quad (19)$$

when $|\lambda| < 1$, namely,

$$-1 < \frac{\mu d - \gamma}{d - \gamma} < 1 \quad (20)$$

is satisfied, the G-Jacobi method is convergent. So the convergence domain is

$$\gamma < \frac{d}{2} (1 + \mu_{\min}).$$

Similarly, we provide the following theorem.

Theorem 9 Let $\bar{A}_{(l-1) \times (l-1)}$ be an coefficient matrix as in (14) for linear equations $\bar{A}x = \bar{b}$ which is diagonally dominant and $d > 0$. Then the G-GS method is convergent when

$$\gamma < \frac{d}{2}(2 - \rho(J)), \quad (21)$$

here $\rho(J)$ is the spectral radius of Jacobi iteration matrix correspondingly.

Proof Note μ is the eigenvalue of Jacobi iterative matrix. From Eq. (17), to ensure the convergence of the G-GS method, it need that

$$\left| \frac{\mu d}{2(d - \gamma)} + \sqrt{\frac{\mu^2 d^2}{4(d - \gamma)^2} - \frac{\gamma}{d - \gamma}} \right| < 1,$$

namely,

$$\frac{\mu d}{2(d - \gamma)} + \sqrt{\frac{\mu^2 d^2}{4(d - \gamma)^2} - \frac{\gamma}{d - \gamma}} < 1, \quad (22)$$

and

$$\frac{\mu d}{2(d - \gamma)} + \sqrt{\frac{\mu^2 d^2}{4(d - \gamma)^2} - \frac{\gamma}{d - \gamma}} > -1. \quad (23)$$

For (22), we can get

$$\gamma < \frac{d}{2}(2 - \mu).$$

For (23), we get

$$\gamma < \frac{d}{2}(2 + \mu).$$

Therefore, when

$$\gamma < \frac{d}{2}(2 - \rho(J)),$$

the G-GS iteration method is convergent.

Theorem 10 Let $\bar{A}_{(l-1) \times (l-1)}$ in (14) be the coefficient matrix for linear equations $\bar{A}x = \bar{b}$. The diagonal retention matrix $\bar{B} > 0$ and the diagonal splitting matrix \bar{C} are defined by (15), then the G-SOR iterative method is convergent when

$$0 < \omega < \frac{2(d - \gamma)}{d}.$$

The theorem can be obtained by Theorem 4.

Theorem 11 Let $A_{(l-1) \times (l-1)}$ in (14) be the coefficient matrix for linear equations $\bar{A}x = \bar{b}$, the diagonal retention matrix $\bar{B} > 0$ and the diagonal splitting matrix \bar{C} are defined by (15), the optimal relaxation factor of the G-SOR method is

$$\omega_{\text{opt}} = \frac{2(d - \gamma)}{d + \sqrt{d^2 - (\gamma + d\rho(\mathbf{M}_I) - \gamma\rho(\mathbf{M}_I))^2}}, \quad (24)$$

and

$$\rho(\mathbf{M}_{\text{III}}) = \frac{1 - \sqrt{1 - \left(\frac{\gamma + (d - \gamma)\rho(\mathbf{M}_I)}{d}\right)^2}}{1 + \sqrt{1 - \left(\frac{\gamma + (d - \gamma)\rho(\mathbf{M}_I)}{d}\right)^2}},$$

here $\rho(\mathbf{M}_I)$ is the spectral radius of G-Jacobi iteration matrix.

4 Applications

In this section, we use the presented G-Jacobi, G-GS and G-SOR iteration methods to solve linear algebraic

equations and 2D and 3D diffusion problems to examine their superiority and accuracy. For time dependent problems, we can choose more appropriate splittings of diagonal matrices to establish the algorithms. Due to the strictly diagonal dominance of coefficient matrix obtained by the fully implicit discretization for parabolic problems, the choice of splitting is more intuitive. Denote $\|E\|_{\infty} = \max |x^{(s)} - x^*|$ as the absolute errors, while $x^{(s)}$ and x^* are the s th iterative solution and exact solution. A laptop with a 2.8 GHz Intel Core i7 processor and MATLAB running in Windows 10 is used for computations.

4.1 Linear equations

(I) Consider the linear equations $Ax=b$ with

$$A = \begin{bmatrix} 3 & -1 & & & \\ -1 & 3 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & & -1 & 3 \end{bmatrix}_{(l-1) \times (l-1)}, \quad b = \begin{bmatrix} 2 \\ 1 \\ \vdots \\ 2 \end{bmatrix}_{(l-1)}.$$

The exact solution is $x^* = [1, 1, \dots, 1]^T$. Taking the linear equations with the order of $l = 101$ for example, Fig. 1 provides the iteration numbers of the G-Jacobi and G-GS iteration methods with the different values of diagonal splitter γ . It is shown that the G-Jacobi iteration method is convergent when $\gamma < 0.5$, which is matched with the results obtained by the theory analysis

$$\gamma < \frac{3}{2} (1 + \mu_{\min}) \approx 0.5005$$

in Theorem 8; The G-GS iteration method is convergent when $\gamma < 2$ nearly, which is matched with the results obtained by the theory analysis

$$\gamma < \frac{3}{2} (2 - \rho(J)) \approx 2.0005$$

in Theorem 9. Table 1 shows obviously the iteration numbers of the G-Jacobi and G-GS iteration methods when $l = 101$, $\|E\|_{\infty} \leq 1 \times 10^{-4}$, which are matched with the theoretical analysis. Moreover, when $\gamma=0$, the iteration numbers 34, 14 are corresponding to the classical Jacobi and GS methods, which are not the smallest. We find that the iteration numbers of the G-jacobi method reach to 22 nearly at $\gamma=-0.4/0.3$. The G-GS method leads to 13 when γ is about 0.1/0.2. In fact, the presented G-GS method has already achieve the same iterative numbers as the classical SOR method.

$$l = 101, \|E\|_{\infty} \leq 1 \times 10^{-4}.$$

(II) Consider the linear equations $Ax=b$ with

$$A = \begin{bmatrix} 3 & -1 & & & \\ -1 & 3 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & & -1 & 3 \end{bmatrix}_{(l-1) \times (l-1)}, \quad b = \begin{bmatrix} 2 \\ 1 \\ \vdots \\ 2 \end{bmatrix}_{(l-1)}.$$

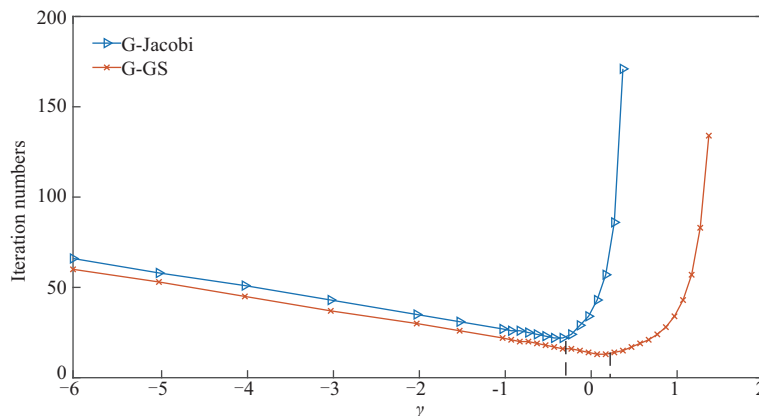


Fig. 1 The iteration numbers of the G-Jacobi and G-GS methods with different γ when $l = 101, \|E\|_{\infty} \leq 1 \times 10^{-4}$

Table 1 The iteration numbers of the presented methods when $l = 101, \|E\|_{\infty} \leq 1 \times 10^{-4}$

γ	G-Jacobi	G-GS
-0.6	24	19
-0.5	23	18
-0.4	22	17
-0.3	22	16
-0.2	24	16
-0.1	29	15
0	34	14
0.1	43	13
0.2	57	13
0.3	86	14

The exact solution is $x^* = [1, 1, \dots, 1]^T$. Taking the linear equations with the order of $l=101$ for example, Fig. 2 provides the iteration numbers of the G-Jacobi and G-GS iteration methods with the different values of diagonal splitter γ . The eigenvalues of the coefficient matrix in problem (2) are same as that of the matrix in problem (1), so they have the same range of the diagonal splitter for the convergence.

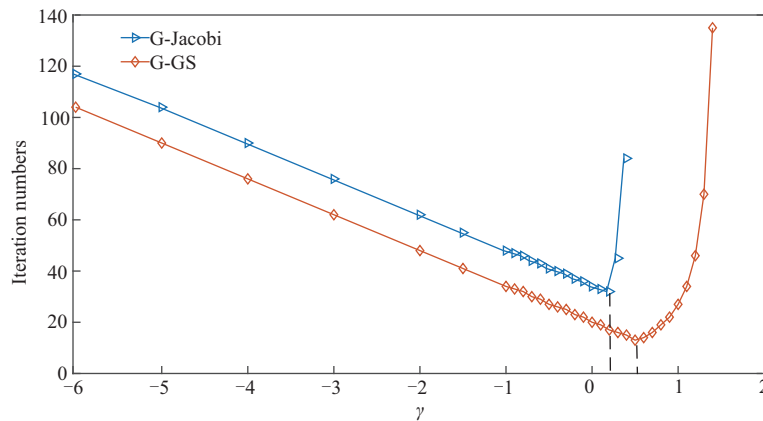


Fig. 2 The iteration numbers of the G-Jacobi and G-GS methods with different γ when $l = 101, \|E\|_{\infty} \leq 1 \times 10^{-4}$

Table 2 shows the iteration numbers of the G-Jacobi and G-GS iteration methods when $l = 101, \|E\|_{\infty} \leq 1 \times 10^{-4}$, which are matched with the theoretical analysis. When γ , the iteration numbers 34, 20 are corresponding to the classical Jacobi and GS methods, which are not the smallest. We find that the iteration numbers of the G-Jacobi method reach to 32 nearly at $\gamma=0.2$. The G-GS method leads to 13 when γ is about 0.5. These results indicate that the classical Jacobi and GS methods of matrix splitting are not optimal compared to the diagonal element splitting for G-Jacobi and G-GS methods, as the new methods use more initial information for approximation.

4.2 2D diffusion problem

Consider the following diffusion problem

$$C_v \frac{\partial T}{\partial t} - K \Delta T(x, y) = f(x), (x, y) \in \Omega = (0, L)^2, t > 0, \tag{25}$$

with the initial and boundary conditions

$$\begin{cases} T(x, y, 0) = \sin(x + y), \\ T(0, y, t) = \sin(y + t), \\ T(1, y, t) = \sin(1 + y + t), \\ T(x, 0, t) = \sin(x + t), \\ T(x, 1, t) = \sin(x + 1 + t), \end{cases}$$

Table 2 The iteration numbers of the presented methods when $l = 101, \|E\|_{\infty} \leq 1 \times 10^{-4}$

γ	G-Jacobi	G-GS
-0.6	43	29
-0.5	41	27
-0.4	40	26
-0.3	39	25
-0.2	37	23
-0.1	36	22
0	34	20
0.1	33	19
0.2	32	17
0.3	45	16
0.4	84	15

where $C_v = 1$ is the specific heat, $K = 1, f(x) = 2\sin(x + y + t) + \cos(x + y + t)$. The exact solution is $T = \sin(x + y + t)$. We divide the domain Ω with the space step $h_x = L/l$ in x direction and $h_y = L/m$ in y direction (for simplicity, we only consider the case of $h = h_x = h_y$), and time step τ , while l, m and J_1 are positive integers and $x_i = ih, y_j = jh, i, j = 0, 1, \dots, l; t_n = n\tau, n = 0, 1, 2, \dots, J_1$. Denote (x_i, y_j, t_n) as $(ih, jh, n\tau)$ and $T_{i,j}^{n,(s)}$ as numerical solution on the n th time level with s th iteration at the node $(x_i, y_j), s = 0, 1, \dots, J_2$. We use the 2D fully implicit (2D F-I) scheme to discrete Eq. (25), due to its unconditionally stability. The discreted linear equations can be written as $A_2 T^{n+1} = b_2^n$ while A_2, b_2^n are the coefficient matrix and right term^[21]. For the presented generalized iteration methods, we choose the optimal diagonal splitter $\gamma=2rK$ and optimal relaxation factors in Eq. (24) to examine the superiority and accuracy of the presented generalized iteration methods. Denote E^n as the absolute errors,

$$\|E^n\|_{\infty} = \max_i |T(x_i, y_j, t_n) - T_{i,j}^n|.$$

Table 3 provides the absolute errors of the 2D Jacobi, 2D G-Jacobi, 2D GS, 2D G-GS, 2D SOR, and 2D G-SOR methods when $\tau = 0.001, J_1 = 10, J_2 = 10$, which shows the errors of 2D G-SOR method is smallest. The accuracy of 2D G-GS method even exceeds 2D SOR method. Fig. 3 gives the iteration numbers of 2D Jacobi, 2D G-Jacobi, 2D GS, 2D G-GS, 2D SOR and 2D G-SOR iteration methods under the errors control $\|E\|_{\infty} \leq 1 \times 10^{-4}$ within 200 times steps. From Fig. 3, we observe that the iteration numbers of 2D G-SOR method is least, and 2D G-GS method is also efficient even more than 2D SOR method.

Table 3 The comparison of the errors when $\tau = 0.001, J_1 = 10, J_2 = 10$

r	h	2D Jacobi	2D G-Jacobi	2D GS	2D G-GS	2D SOR($\bar{\omega}_{opt}$)	2D G-SOR(ω_{opt})
1.024 0	1/64	9.0025×10^{-4}	6.7991×10^{-4}	1.3871×10^{-4}	8.9888×10^{-6}	9.0636×10^{-6}	4.0543×10^{-6}
4.096 0	1/128	4.9261×10^{-3}	4.8286×10^{-3}	2.6395×10^{-3}	7.6428×10^{-5}	1.3540×10^{-3}	3.6689×10^{-5}
16.384 0	1/256	8.2177×10^{-3}	8.2060×10^{-3}	8.2060×10^{-3}	2.7491×10^{-4}	5.7198×10^{-3}	2.0946×10^{-4}
65.536 0	1/512	9.4802×10^{-3}	9.4793×10^{-3}	8.9980×10^{-3}	3.9467×10^{-4}	8.5757×10^{-3}	3.4420×10^{-4}

4.3 3D diffusion problem

Consider the following diffusion problem

$$C_v \frac{\partial T}{\partial t} - K\Delta T(x, y, z) = f(x), (x, y, z) \in \Omega = (0, L)^3, t > 0, \tag{26}$$

with the initial and boundary conditions

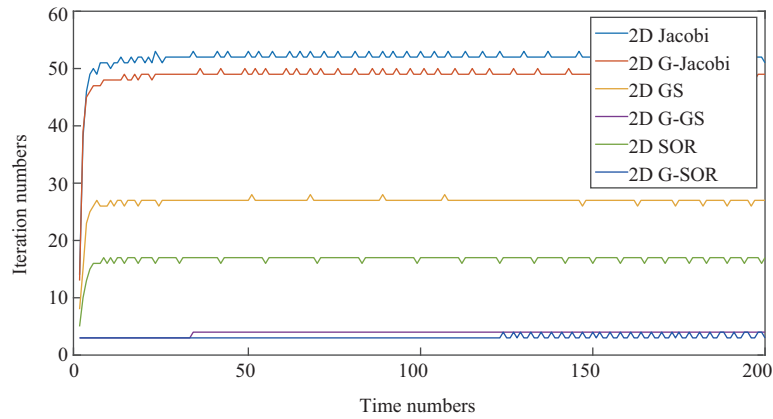


Fig. 3 The iteration numbers of the 2D Jacobi, 2D G-Jacobi, 2D GS, 2D G-GS, 2D SOR, and 2D G-SOR methods when $h = 1/64$, $\tau = 0.0005$, $J_1 = 200$, $\|E\|_{\infty} \leq 1 \times 10^{-4}$

$$\begin{cases} T(x, y, z, 0) = \sin(x + y + z), \\ T(0, y, z, t) = \sin(y + z + t), \\ T(1, y, z, t) = \sin(1 + y + z + t), \\ T(x, 0, z, t) = \sin(x + z + t), \\ T(x, 1, z, t) = \sin(x + 1 + z + t), \\ T(x, y, 0, t) = \sin(x + y + t), \\ T(x, y, 1, t) = \sin(x + y + 1 + t), \end{cases}$$

where $C_v = 1$ is the specific heat, $K = 1$, $f(x) = 3\sin(x + y + z + t) + \cos(x + y + z + t)$. The exact solution is $T = \sin(x + y + z + t)$. We divide the domain Ω with the space step $h_x = L/l$ in x direction, $h_y = L/m$ in y direction and $h_z = L/q$ in z direction (for simplicity, we only consider the case of $h = h_x = h_y = h_z$), and time step τ , while l, m, q and J_1 are positive integers and $x_i = ih, i = 0, 1, \dots, l; y_j = jh, j = 1, \dots, m; z_k = kh, k = 0, 1, \dots, q; t_n = n\tau, n = 0, 1, 2, \dots, J_1$. Denote (x_i, y_j, z_k, t_n) as $(ih, jh, kh, n\tau)$, and $T_{i,j,k}^n$ as numerical solution on the n th time level at the node (x_i, y_j, z_k) . We use the 3D fully implicit (3D F-I) scheme to discrete Eq. (26) since it is stable unconditionally. The discretization can be written as $A_3 T^{n+1} = b_3^n$, while A_3, \tilde{b}_3^n are the coefficient matrix and right term^[21]. For the presented generalized iteration methods, we choose the optimal diagonal splitter $\gamma=3rK$ and optimal relaxation factors in Eq. (24) to examine the superiority and accuracy of the presented generalized iteration methods.

Table 4 gives the absolute errors of the 3D Jacobi, 3D G-Jacobi, 3D GS, 3D G-GS, 3D SOR, and 3D G-SOR methods when $\tau = 0.001, J_1 = 10, J_2 = 5$, which shows the errors of 3D G-SOR method is the smallest of these iteration methods. The accuracy of 3D G-GS and 3D G-SOR methods even exceed 3D SOR method on accuracy. Fig. 4 provides the iteration numbers of 3D Jacobi, 3D G-Jacobi, 3D GS, 3D G-GS, 3D SOR, and 3D G-SOR methods within 200 times levels when $\tau = 0.0005$. The results show that the presented generalized iteration methods are performed very efficient because the smaller iteration numbers can save the computational cost well and speed up the calculation of the solver on each iteration.

Table 4 The comparison of the errors when $\tau = 0.001, J_1 = 10, J_2 = 5$

r	h	3D Jacobi	3D G-Jacobi	3D GS	3D G-GS	3D SOR($\bar{\omega}_{opt}$)	3D G-SOR(ω_{opt})
0.225 0	1/32	$3.560 8 \times 10^{-4}$	$1.831 3 \times 10^{-4}$	$8.445 7 \times 10^{-5}$	$6.113 9 \times 10^{-6}$	$6.186 4 \times 10^{-5}$	$5.771 1 \times 10^{-6}$
1.024 0	1/64	$3.571 9 \times 10^{-3}$	$3.313 3 \times 10^{-3}$	$1.623 8 \times 10^{-3}$	$3.707 0 \times 10^{-4}$	$4.837 4 \times 10^{-4}$	$1.596 1 \times 10^{-4}$
4.096 0	1/128	$7.301 0 \times 10^{-3}$	$7.261 1 \times 10^{-3}$	$5.543 1 \times 10^{-3}$	$1.205 1 \times 10^{-3}$	$4.044 6 \times 10^{-3}$	$8.202 8 \times 10^{-4}$
16.384 0	1/256	$9.132 1 \times 10^{-3}$	$9.128 1 \times 10^{-3}$	$8.380 1 \times 10^{-3}$	$1.378 6 \times 10^{-3}$	$7.655 3 \times 10^{-3}$	$1.229 7 \times 10^{-3}$

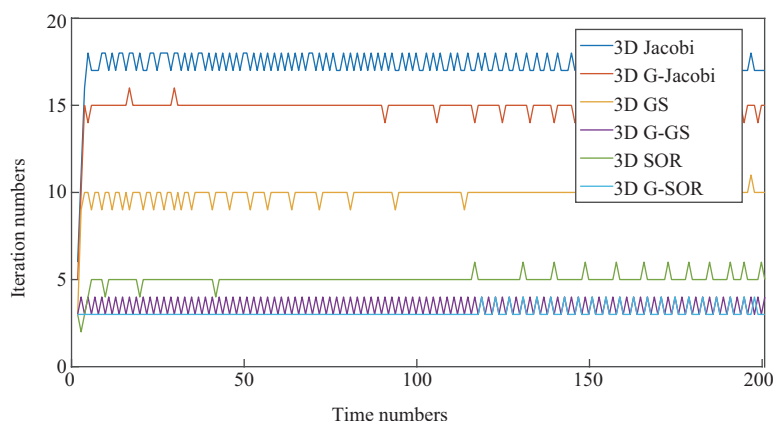


Fig. 4 The iteration numbers of the 3D Jacobi, 3D G-Jacobi, 3D GS, 3D G-GS, 3D SOR, and 3D G-SOR methods when $h = 1/32$, $\tau = 0.0005$, $J_1 = 200$, $\|E\|_{\infty} \leq 1 \times 10^{-4}$

5 Conclusions

A class of diagonal matrix splitting iteration methods is provided, and the classical Jacobi Gauss-Seidel and SOR methods are the special cases of these new methods. By selecting appropriate diagonal element splitting, better iteration methods can be constructed. We give the eigenvalues, convergence domains and the optimal relaxation factors of the presented methods for tridiagonal coefficient matrix. This kind of generalized methods are applied to solve the algebraic linear equations and 2D and 3D diffusion problems with fully implicit discretization. It is found that fewer iteration numbers are used to improve the calculation efficiency and save the calculation cost compared to the classical Jacobi, Gauss-Seidel and SOR methods. The new methods can also be used to solve ill conditioned equations, such as saddle point problem and nonlinear equations.

References:

- [1] VARGA R S. Matrix iterative analysis [M]. 2nd ed. Springer Series in Computational Mathematics, Vol. 127. Berlin: Springer-Verlag, 2000.
- [2] YOUNG D M. Iterative solution of large linear systems [M]. New York: Academic Press, 1971.
- [3] GAUSS C F. Letter to Gerling, December 26, 1823[J]. Werke, 1903, 9: 278-281. English translation by FORSYTHE G E. Mathematical Tables and Other Aids to Computation, 1951, 5(36): 255-258.
- [4] HAGEMAN L A, YOUNG D M. Applied iterative methods[M]. New York: Academic Press, 1981.
- [5] AXELSSON O. Iterative solution methods [M]. Cambridge: Cambridge University Press, 1994.
- [6] SAAD Y. Iterative methods for sparse linear systems [M]. 2nd ed. Philadelphia: SIAM, 2003.
- [7] BAI Zhongzhi, GOLUB G H, NG M K. Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems[J]. SIAM Journal on Matrix Analysis and Applications, 2003, 24(3):603-626.
- [8] BAI Zhongzhi, GOLUB G H. Accelerated Hermitian and skew-Hermitian splitting iteration methods for saddle-point problems[J]. IMA Journal of Numerical Analysis, 2007, 27(1): 1-23. DOI:10.1093/imanum/drl017.
- [9] BENZI M. Preconditioning techniques for large linear systems: A survey [J]. Journal of Computational Physics, 2002, 182(2):418-477.
- [10] BAI Zhongzhi. On Hermitian and skew-Hermitian splitting iteration methods for continuous Sylvester equations [J]. Journal of Computational Mathematics, 2010, 28(1): 39-50.
- [11] LI Wen, SUN Weiwei. Modified Gauss-Seidel type methods and Jacobi type methods for Z -matrices [J]. Linear Algebra and its Applications, 2000, 317(1/2/3): 227-240.
- [12] BERTACCINI D, DURASTANTE F. Interpolating preconditioners for the solution of sequence of linear systems [J]. Computers & Mathematics with Applications, 2016, 72(4): 1118-1130.
- [13] SALKUYEH D K, HEZARI D, EDALATPOUR V. Generalized successive overrelaxation iterative method for a class of complex symmetric linear system of equations [J]. International Journal of Computer Mathematics, 2015, 92(4): 802-815.
- [14] ZHAO Wanchen, SHAO Xinhui. New matrix splitting iteration method for generalized absolute value equations [J]. AIMS Mathematics, 2023, 8(5): 10558-10578.

- [15] LI Tianyi, CHEN Fang, FANG Zhiwei, et al. Two-parameter modified matrix splitting iteration method for Helmholtz equation [J]. *International Journal of Computer Mathematics*, 2024, 101(9/10): 1205-1218.
- [16] BAI Zhongzhi. A two-step matrix splitting iteration paradigm based on one single splitting for solving systems of linear equations [J]. *Numerical Linear Algebra with Applications*, 2024, 31(3): e2510. DOI: 10.1002/nla.2510.
- [17] HADJIDIMOS A. Successive overrelaxation (SOR) and related methods [J]. *Journal of Computational and Applied Mathematics*, 1991, 20: 75-89.
- [18] WOZNICKI Z I, JEDRZEJEC H A. A new class of modified line-SOR algorithms [J]. *Journal of Computational and Applied Mathematics*, 2001, 131(1/2): 89-142.
- [19] YOUSSEF I K, FARID M M. On the accelerated overrelaxation method [J]. *Pure and Applied Mathematics Journal*, 2015, 4(1): 26-31.
- [20] PAN Yunming, XU Qiuyan, LIU Zhiyong, et al. A class of new successive permutation iterative algorithms with the relaxation factor for radiation diffusion equation [J]. *Computational and Applied Mathematics*, 2023, 42(5): Article 203. DOI: 10.1007/s40314-023-02341-7.
- [21] XU Qiuyan, LIU Zhiyong. A class of new successive permutation iterative algorithms for diffusion equation [J]. *Journal of Difference Equations and Applications*, 2021, 27(9): 1355-1372.

一种基于对角矩阵分裂的快速迭代方法及其应用

许秋燕

(宁夏大学 数学统计学院, 宁夏 银川 750021)

摘要:线性方程组的快速求解一直是科学计算领域的研究热点之一。文中提出了一种对角矩阵分裂迭代方法,其不同于经典的矩阵分裂。以系数矩阵对角元的分解为核心,构建了若干新的预条件子。以三对角系数矩阵为例,从理论上分析了新方法的收敛域与最佳松弛因子。将所提新迭代方法应用于线性代数方程组,以及全隐离散下的二维、三维扩散问题的求解。数值实验结果与理论分析一致,同时显示了新方法大幅减少了迭代次数。所提出的迭代方法显著优于部分经典迭代方法。

关键词:迭代;矩阵分裂;扩散方程;收敛性;最佳松弛因子

(责任编辑 张 娣)