Journal of Computational Mathematics, Vol.24, No.4, 2006, 539–552.

A SHIFT-SPLITTING PRECONDITIONER FOR NON-HERMITIAN POSITIVE DEFINITE MATRICES *[1]

Zhong-zhi Bai Jun-feng Yin
(*LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China*)

Yang-feng Su
(*Department of Mathematics, Fudan University, Shanghai 200433, China*)

Abstract

A shift splitting concept is introduced and, correspondingly, a shift-splitting iteration scheme and a shift-splitting preconditioner are presented, for solving the large sparse system of linear equations of which the coefficient matrix is an ill-conditioned non-Hermitian positive definite matrix. The convergence property of the shift-splitting iteration method and the eigenvalue distribution of the shift-splitting preconditioned matrix are discussed in depth, and the best possible choice of the shift is investigated in detail. Numerical computations show that the shift-splitting preconditioner can induce accurate, robust and effective preconditioned Krylov subspace iteration methods for solving the large sparse non-Hermitian positive definite systems of linear equations.

Mathematics subject classification: 65F10, 65F50.
Key words: Non-Hermitian positive definite matrix, Matrix splitting, Preconditioning, Krylov subspace method, Convergence.

1. Introduction

Let $\mathbb{C}^n$ represent the complex n-dimensional vector space, and $\mathbb{C}^{n\times n}$ the complex $n\times n$ matrix space. In this paper, we will consider preconditioning the large sparse system of linear equations

$$Ax=b, \qquad A\in\mathbb{C}^{n\times n} \quad \text{and} \quad x,b\in\mathbb{C}^n, \tag{1}$$

where A is a large sparse *non-Hermitian positive definite* matrix (i.e., its Hermitian part $H=\frac{1}{2}(A+A^*)$ is positive definite), and b and x are, respectively, the known and the unknown vectors. Here, we have used A^* to denote the conjugate transpose of the matrix A.

The system of linear equations (1) with a non-Hermitian positive definite coefficient matrix A arises in many problems in scientific and engineering computing, see [24, 2, 23]. When a Krylov subspace iteration method is employed to compute an approximation for its solution $x_*=A^{-1}b$, an economical and effective preconditioner is often demanded in order to improve the computational efficiency, the approximate accuracy and the numerical stability of the referred Krylov subspace iteration method, see [2, 23]. There have been many elegant preconditioners presented and studied in the literature in recent years [24, 1, 2, 23, 3, 4, 21], which are cheaply applicable and practically efficient for matrices of specific structures and properties. These preconditioners can be roughly categorized into the incomplete factorizations [18, 17, 5, 2] and the splitting iterations [24, 3, 4, 21]. Essentially, a preconditioner aims to transform the

* Received May 8, 2005; Final revised January 23, 2006.
[1] Research supported by The China NNSF Outstanding Young Scientist Foundation (No. 10525102), The National Natural Science Foundation (No. 10471146), and The National Basic Research Program (No. 2005CB321702), P.R. China.

original linear system (1) by a suitable linear transformation such that the spectral property of the coefficient matrix $A \in \mathbb{C}^{n\times n}$ is largely improved, and therefore, the convergence speed of the referred Krylov subspace iteration method is considerably accelerated. However, both incomplete factorization and splitting iteration are only applicable and efficient for special classes of matrices, e.g., a diagonally dominant or an irreducibly weakly diagonally dominant matrix. Even for a Hermitian positive definite matrix, its incomplete Cholesky factorization may break down [17]; and for a non-Hermitian positive definite matrix of strong skew-Hermitian part, the splitting iteration may diverge [9].

For the Hermitian positive definite system of linear equations, considering that the conjugate gradient method is quite efficient when its coefficient matrix has tightly clustered spectrum [2, 12, 23], Bai and Zhang [8] recently presented a class of *regularized conjugate gradient* (RCG) method by first shifting and contracting the spectrum of the coefficient matrix, and then approximating the iterates of the regularized iteration sequence by the *conjugate gradient* (CG) iteration [15, 10, 16]. Therefore, the RCG method is actually an inner/outer iteration method [19, 20, 13, 11, 14] with a standard splitting iteration as its outer iteration, and the CG iteration as its inner iteration. The shifted and contracted matrix leads to a linear polynomial preconditioner and the RCG iteration leads to a nonstationary iteration preconditioner for the Hermitian positive definite linear system.

For the non-Hermitian positive definite system of linear equations (1), in this paper we first present a shift splitting for the coefficient matrix A, and then construct a shift-splitting iteration scheme for the linear system (1). Theoretically, this scheme is proved to be convergent unconditionally to the exact solution of the system of linear equations (1). The shift splitting also naturally induces a simple but effective preconditioner for the coefficient matrix A. Moreover, the shift-splitting preconditioning matrix itself can be again approximated by employing an incomplete factorization or a splitting iteration. This leads to a so-called two-level preconditioner for the system of linear equations (1). In actual applications, we can suitably choose the shift in such a way that the induced splitting matrix has reasonably good diagonally dominant property such that its incomplete factorization or splitting iteration is existent, stable, and accurate. Hence, the two-level preconditioner can lead to a highly efficient Krylov subspace iteration method for solving the system of linear equations (1). These results extend and develop those for Hermitian positive definite linear system studied in [8] to non-Hermitian positive definite one.

The organization of the paper is as follows. In Section 2 we introduce the shift splitting concept and the shift-splitting iteration scheme, and discuss their convergent and preconditioning properties. In Section 3 we describe and analyze the two-level preconditioning technique which is defined by adopting a further approximation of the shift-splitting preconditioning matrix. Numerical results are given in Section 4 to show the feasibility and the effectiveness of the shift-splitting and the corresponding two-level preconditioners when they are employed to accelerate the Krylov subspace iteration methods. Finally, in Section 5 we use brief conclusions to end this paper.

2. The Shift-splitting Preconditioner

For a non-Hermitian positive definite matrix $A \in \mathbb{C}^{n\times n}$, we use $\lambda(A)$ to represent its eigenvalue and $\sigma(A)$ its spectrum set, $\beta_l(A)$ and $\beta_u(A)$ the lower and the upper bounds of the real parts of its eigenvalues, and $\gamma_l(A)$ and $\gamma_u(A)$ the lower and the upper bounds of the imaginary parts of its eigenvalues, respectively. That is to say, we have

$$\beta_l(A) \leq \Re(\lambda(A)) \leq \beta_u(A) \quad \text{and} \quad \gamma_l(A) \leq \Im(\lambda(A)) \leq \gamma_u(A),$$

where $\Re(\lambda)$ and $\Im(\lambda)$ represent the real and the imaginary parts of the complex λ. Without causing confusion, sometimes we may neglect the matrix A and simply write these bounds as

β_l, β_u and γ_l, γ_u. The condition number $\kappa_2(A)$ of the matrix $A \in \mathbb{C}^{n\times n}$ with respect to the Euclidean norm is given by $\kappa_2(A) = \|A\|_2\|A^{-1}\|_2$. We use $\|A\|$ to denote any consistent matrix norm defined in $\mathbb{C}^{n\times n}$, and I the identity matrix of suitable dimension.

Given a non-Hermitian positive definite matrix $A \in \mathbb{C}^{n\times n}$ and a positive parameter α, we can construct the shift splitting of the matrix A as follows:

$$\begin{aligned} A &\equiv M(\alpha) - N(\alpha) \\ &= \frac{1}{2}(\alpha I + A) - \frac{1}{2}(\alpha I - A). \end{aligned}$$

This splitting naturally leads to the shift-splitting iteration scheme

$$x^{(k+1)} = (\alpha I + A)^{-1}(\alpha I - A)x^{(k)} + 2(\alpha I + A)^{-1}b, \qquad k = 0, 1, 2, \ldots, \tag{2}$$

for solving the system of linear equations (1).

Because the shift-splitting iteration scheme (2) is only a single-step method, in actual applications it may have considerably less computing workloads than the two-step iteration methods such as the *Hermitian and skew-Hermitian splitting* (HSS) iteration [7] and the *positive-definite and skew-Hermitian splitting* (PSS) iteration [6]. See also [21].

Evidently, the shift-splitting iteration scheme (2) is convergent for all initial vector $x^{(0)} \in \mathbb{C}^n$ if and only if the spectral radius of its iteration matrix

$$T(\alpha) = (\alpha I + A)^{-1}(\alpha I - A)$$

is less than one, i.e., $\rho(T(\alpha)) < 1$. The following theorem precisely describes the unconditional convergence property of the shift-splitting iteration method.

Theorem 2.1. *Let $A \in \mathbb{C}^{n\times n}$ be a non-Hermitian positive definite matrix and α a positive constant. Then the spectral radius $\rho(T(\alpha))$ of the iteration matrix $T(\alpha)$ of the shift-splitting iteration is bounded by $\varrho(\alpha) = \|(\alpha I + A)^{-1}(\alpha I - A)\|$. Consequently, we have*

$$\rho(T(\alpha)) \le \varrho(\alpha) < 1, \qquad \forall \alpha > 0,$$

i.e., the shift-splitting iteration (2) is convergent unconditionally to the exact solution $x_* \in \mathbb{C}^n$ of the system of linear equations (1).

Proof. It is obvious that $\rho(T(\alpha)) \le \varrho(\alpha)$ holds for all $\alpha > 0$. From Lemma 2.1 in [6], we further know that $\varrho(\alpha) < 1$, $\forall \alpha > 0$.

Recall that when A is skew-Hermitian $T(\alpha)$ is the Cayley transform which maps all the purely imaginary eigenvalues of the matrix A onto the unit circle centered at origin. Therefore, in general, we will call $T(\alpha)$ the generalized Cayley transform which maps all eigenvalues of the positive definite matrix A into the interior of the unit disk centered at origin.

When A is Hermitian positive definite we easily see that $\rho(T(\alpha)) = \varrho(\alpha)$,

$$\varrho(\alpha) = \max_{\lambda\in\sigma(A)} \left|\frac{\alpha - \lambda}{\alpha + \lambda}\right| < 1, \qquad \forall \alpha > 0,$$

and $\varrho(\alpha)$ attains its minimum $\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}$ at $\alpha = \sqrt{\beta_l(A)\beta_u(A)}$. Now, the convergence rate of the shift-splitting iteration (2) is the same as those of the conjugate gradient method, the HSS method, and the PSS method. Actually, the shift-splitting iteration, the HSS iteration, and the PSS iteration all reduce to the same iteration scheme for the system of linear equations (1). Moreover, in [8] the regularized splitting for the matrix A is defined as

$$\begin{aligned} A &\equiv \widetilde{M}(\alpha) - \widetilde{N}(\alpha) \\ &= (\alpha I + A) - \alpha I \end{aligned}$$

and, correspondingly, its iteration matrix is given by

$$\widetilde{T}(\alpha) = \alpha(\alpha I + A)^{-1},$$

the corresponding convergence factor is given by

$$\widetilde{\varrho}(\alpha) := \rho(\widetilde{T}(\alpha)) = \max_{\lambda \in \sigma(A)} \frac{\alpha}{\alpha + \lambda},$$

and $\widetilde{\varrho}(\alpha)$ attains its minimum 0 at $\alpha = 0$. See also [17]. Evidently, the splitting matrices $M(\alpha)$ and $\widetilde{M}(\alpha)$ are only different by a factor of $\frac{1}{2}$. However, this little difference seems intrinsic and crucial as it may lead to drastic difference between the properties of the functions $\rho(\alpha)$ and $\widetilde{\varrho}(\alpha)$ with respect to $\alpha \in (0, +\infty)$.

When A is non-Hermitian positive definite, the convergence rate of the shift-splitting iteration (2) is about equal to that of the PSS iteration.

In addition, we easily see that $\varrho(\alpha)$ is the contraction factor of the shift-splitting iteration (2), as it holds that

$$\|x^{(k+1)} - x_*\| \le \varrho(\alpha)\|x^{(k)} - x_*\|, \qquad k = 0, 1, 2, \ldots.$$

When the lower eigenvalue bound of the matrix $H = \frac{1}{2}(A + A^*)$ and the upper eigenvalue bound of the matrix (AA^*) (i.e., $\|A\|_2$) are available, we can further derive the following estimate about $\varrho(\alpha)$.

Theorem 2.2. (The convergence theorem).
Let $A \in \mathbb{C}^{n\times n}$ be a non-Hermitian positive definite matrix, $H = \frac{1}{2}(A + A^*)$ be its Hermitian part, and α a positive constant. Denote by $\beta_l(H)$ the lower bound of the eigenvalues of the matrix H. Then the contraction factor $\varrho(\alpha)$ of the shift-splitting iteration (2) can be bounded by

$$\varrho_u(\alpha) = \sqrt{\frac{\alpha^2 - 2\beta_l(H) + \|A\|_2^2}{\alpha^2 + 2\beta_l(H) + \|A\|_2^2}}, \quad \forall \alpha > 0.$$

It then follows straightforwardly that $\varrho_u(\alpha)$ attains its minimum $\sqrt{\frac{\|A\|_2 - \beta_l(H)}{\|A\|_2 + \beta_l(H)}}$ when $\alpha = \|A\|_2$.

Proof. By direct computations we have

$$\begin{aligned} \varrho(\alpha)^2 &= \rho((\alpha I + A)^{-1}(\alpha I - A)(\alpha I - A)^*(\alpha I + A)^{-*}) \\ &= \max_{x \ne 0} \frac{x^*(\alpha I - A)(\alpha I - A)^* x}{x^*(\alpha I + A)(\alpha I + A)^* x} \\ &= \max_{x \ne 0} \frac{\alpha^2 x^* x - 2\alpha x^* H x + x^* A A^* x}{\alpha^2 x^* x + 2\alpha x^* H x + x^* A A^* x} \\ &\le \varrho_u(\alpha)^2. \end{aligned}$$

Hence, it holds that $\varrho(\alpha) \le \varrho_u(\alpha)$.

As a matter of fact, any matrix splitting not only can automatically lead to a splitting iteration method, but also can naturally induce a splitting preconditioner for the Krylov subspace methods. The splitting preconditioner corresponds to the shift-splitting iteration (2) is given by

$$M(\alpha) = \frac{1}{2}(\alpha I + A). \tag{3}$$

We call this preconditioner the *shift-splitting preconditioner* for the matrix A.

The shift-splitting preconditioner is simple but quite effective in actual applications. In fact, because A is positive definite, its diagonal entries have positive real parts. It then follows that $(\alpha I + A)$ is more diagonally dominant than A itself. Therefore, instead of the original matrix A, we can obtain a more accurate and stable incomplete factorization or splitting approximation to the shifted matrix $(\alpha I + A)$.

The following theorem describes the eigenvalue distribution of the preconditioned matrix $M(\alpha)^{-1}A$ with respect to the shift splitting.

Theorem 2.3. (The eigen-distribution theorem).
Let $A \in \mathbb{C}^{n\times n}$ be a non-Hermitian positive definite matrix and α a positive constant. Then $M(\alpha)$ and $M(\alpha)^{-1}A$ have the same eigenvector sets as A.

Let the eigenvalues of A satisfy

$$\beta_l(A) \leq \Re(\lambda(A)) \leq \beta_u(A) \quad \textit{and} \quad |\Im(\lambda(A))| \leq \gamma(A),$$

with $\beta_l(A) > 0$. Then the eigenvalues of the preconditioned matrix $M(\alpha)^{-1}A$ with respect to the shift-splitting preconditioner $M(\alpha)$ of A satisfy

$$\beta_l(M(\alpha)^{-1}A) \leq \Re(\lambda(M(\alpha)^{-1}A)) \leq \beta_u(M(\alpha)^{-1}A) \quad \textit{and} \quad |\Im(\lambda(M(\alpha)^{-1}A))| \leq \gamma(M(\alpha)^{-1}A),$$

where

$$\left\{\begin{array}{lcl} \beta_l(M(\alpha)^{-1}A) & = & \frac{2\beta_l(A)}{\alpha+\beta_l(A)}, \\ \beta_u(M(\alpha)^{-1}A) & = & \max\left\{\frac{2[(\alpha+\beta_l(A))\beta_l(A)+\gamma(A)^2]}{(\alpha+\beta_l(A))^2+\gamma(A)^2}, \quad \frac{2[(\alpha+\beta_u(A))\beta_u(A)+\gamma(A)^2]}{(\alpha+\beta_u(A))^2+\gamma(A)^2}\right\}, \\ \gamma(M(\alpha)^{-1}A) & = & \frac{2\alpha\gamma(A)}{(\alpha+\beta_l(A))^2}. \end{array}\right.$$

Proof. It is obvious that the matrices A, $M(\alpha)$ and $M(\alpha)^{-1}A$ have the same eigenvector sets.

Let λ_α be an eigenvalue of the matrix $M(\alpha)^{-1}A$ and x_α be a corresponding eigenvector, i.e., $M(\alpha)^{-1}Ax_\alpha = \lambda_\alpha x_\alpha$. Then we have

$$Ax_\alpha = \lambda_\alpha M(\alpha)x_\alpha,$$

and hence,

$$\lambda_\alpha = \frac{x_\alpha^* Ax_\alpha}{x_\alpha^* M(\alpha)x_\alpha} = \frac{2x_\alpha^* Ax_\alpha}{\alpha x_\alpha^* x_\alpha + x_\alpha^* Ax_\alpha}.$$

Denote by $\xi + \imath\eta = \frac{x_\alpha^* Ax_\alpha}{x_\alpha^* x_\alpha}$, where $\imath$ is the imaginary unit. Then $\xi > 0$ and

$$\lambda_\alpha = \frac{2(\xi+\imath\eta)}{\alpha+(\xi+\imath\eta)} = \frac{2[(\alpha+\xi)\xi+\eta^2+\imath\alpha\eta]}{(\alpha+\xi)^2+\eta^2}.$$

As

$$\Re(\lambda_\alpha) = \frac{2[(\alpha+\xi)\xi+\eta^2]}{(\alpha+\xi)^2+\eta^2},$$

we easily obtain

$$\Re(\lambda_\alpha) \geq 2\cdot\min\left\{\frac{\xi}{\alpha+\xi}, \quad 1\right\} = \frac{2\xi}{\alpha+\xi} \geq \frac{2\beta_l(A)}{\alpha+\beta_l(A)} = \beta_l(M(\alpha)^{-1}A)$$

and

$$\begin{array}{rcl} \Re(\lambda_\alpha) & \leq & \dfrac{2[(\alpha+\xi)\xi+\gamma(A)^2]}{(\alpha+\xi)^2+\gamma(A)^2} \\ & \leq & \max\left\{\dfrac{2[(\alpha+\beta_l(A))\beta_l(A)+\gamma(A)^2]}{(\alpha+\beta_l(A))^2+\gamma(A)^2}, \quad \dfrac{2[(\alpha+\beta_u(A))\beta_u(A)+\gamma(A)^2]}{(\alpha+\beta_u(A))^2+\gamma(A)^2}\right\} \\ & = & \beta_u(M(\alpha)^{-1}A), \end{array}$$

where we have applied the fact that $\Re(\lambda_\alpha)$ is a monotonically increasing function with respect to the variable $t := \eta^2$. As

$$\Im(\lambda_\alpha) = \frac{2\alpha\eta}{(\alpha+\xi)^2+\eta^2},$$

we easily obtain

$$|\Im(\lambda_\alpha)| = \frac{2\alpha|\eta|}{(\alpha+\xi)^2+\eta^2} \le \frac{2\alpha|\eta|}{(\alpha+\xi)^2} \le \frac{2\alpha\gamma(A)}{(\alpha+\beta_l(A))^2} = \gamma(M(\alpha)^{-1}A).$$

Essentially, we may choose a parameter α such that the eigenvalues of the preconditioned matrix $M(\alpha)^{-1}A$ are more tightly clustered than the original matrix A, and the computing cost of the induced preconditioned Krylov subspace method is, at least, not increased. It follows that there is a trade-off between these two requirements and, hence, finding an optimal parameter α is a very difficult problem in actual applications. If only the first requirement is considered, then we may minimize the function

$$\mathcal{A}(\alpha) := 2[\beta_u(M(\alpha)^{-1}A) - \beta_l(M(\alpha)^{-1}A)]\gamma(M(\alpha)^{-1}A)$$

and obtain an α. But this α is certainly not the required optimal parameter. Note that $\mathcal{A}(\alpha)$ is an estimate of the area of the rectangle including all eigenvalues of the preconditioned matrix $M(\alpha)^{-1}A$. Hence, minimizing $\mathcal{A}(\alpha)$ implies clustering the eigenvalues of $M(\alpha)^{-1}A$. Alternatively, we may choose a parameter α such that $\mathcal{A}(\alpha) \ll 2[\beta_u(A) - \beta_l(A)]\gamma(A)$ is at least satisfied.

Based on Theorem 2.3 we can estimate the asymptotic convergence rate of the Krylov subspace iteration methods, accelerated by the shift-splitting preconditioner $M(\alpha)$, for solving the system of linear equations (1). As an example, in the following we give an analysis for the convergence property of the preconditioned GMRES method [23].

To simplify the notations, we denote by

$$\mathbf{A}_\alpha = M(\alpha)^{-1}A \quad \text{and} \quad \mathbf{b}_\alpha = M(\alpha)^{-1}b,$$

and consider the preconditioned linear system

$$\mathbf{A}_\alpha\mathbf{x}_\alpha = \mathbf{b}_\alpha, \quad \text{with} \quad \mathbf{x}_\alpha \equiv x, \tag{4}$$

associated with the shift-splitting preconditioner $M(\alpha)$. As is known, GMRES is a polynomial iteration method in which the k-th residual is written as

$$\mathbf{r}_\alpha^{(k)} = \mathbf{b}_\alpha - \mathbf{A}_\alpha\mathbf{x}_\alpha^{(k)} = \mathcal{P}_k(\mathbf{A}_\alpha)\mathbf{r}_\alpha^{(0)}, \quad \mathcal{P}_k \in \Pi_k, \quad \mathcal{P}_k(0) = 1,$$

where Π_k is the set of polynomials of degree not greater than k. At each iteration step, the GMRES iterate is computed such that

$$\|\mathbf{r}_\alpha^{(k)}\|_2 = \min_{\mathcal{P}_k\in\Pi_k,\mathcal{P}_k(0)=1}\{\|\mathcal{P}_k(\mathbf{A}_\alpha)\mathbf{r}_\alpha^{(0)}\|_2\}.$$

It is well known that if $\mathbf{A}_\alpha$ is diagonalizable with the eigenvector matrix $\mathbf{W}_\alpha$, then the 2-norm of the residual is bounded from above and has the estimate

$$\frac{\|\mathbf{r}_\alpha^{(k)}\|_2}{\|\mathbf{r}_\alpha^{(0)}\|_2} \le \kappa_2(\mathbf{W}_\alpha)\cdot \min_{\mathcal{P}_k\in\Pi_k,\mathcal{P}_k(0)=1}\ \max_{\lambda\in\Upsilon(\mathbf{A}_\alpha)} |\mathcal{P}_k(\lambda)|,$$

where $\kappa_2(\mathbf{W}_\alpha)$ denotes the Euclidean condition number and $\Upsilon(\mathbf{A}_\alpha)$ a set which contains the spectrum of the matrix $\mathbf{A}_\alpha$. The convergence of GMRES is therefore essentially bounded by the quantity

$$\rho_k(\Upsilon(\mathbf{A}_\alpha)) = \min_{\mathcal{P}_k\in\Pi_k,\mathcal{P}_k(0)=1}\ \max_{\lambda\in\Upsilon(\mathbf{A}_\alpha)} |\mathcal{P}_k(\lambda)|.$$

The corresponding asymptotic convergence factor (see [24]) is defined by

$$\rho(\Upsilon(\mathbf{A}_\alpha)) = \lim_{k\to\infty} \rho_k(\Upsilon(\mathbf{A}_\alpha))^{\frac{1}{k}}. \tag{5}$$

To estimate (5), we will restrict our attention to the case where $\Upsilon(\mathbf{A}_\alpha)$ is an ellipse enclosing the eigenvalues of the matrix $\mathbf{A}_\alpha$ except for the origin[23]. Since the spectrum of $\mathbf{A}_\alpha$ is symmetric with respect to the real axis, we have only to consider the ellipse which is aligned with axis. Let $\mathcal{E}(\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d})$ denote the ellipse with center $\mathtt{d}$, foci $\mathtt{d}\pm\mathtt{c}$ and semi-axes $\mathtt{a}$ and $\mathtt{b}$, where $\mathtt{c}^2 = \mathtt{a}^2 - \mathtt{b}^2$. We note that the ellipse $\mathcal{E}(\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d})$ has either real or complex conjugate foci depending on the sign of $\mathtt{a}-\mathtt{b}$. The asymptotic convergence factor on these ellipses can be expressed as

$$\rho(\Upsilon(\mathbf{A}_\alpha)) = \frac{\mathtt{a}+\mathtt{b}}{\mathtt{d}+\sqrt{\mathtt{d}^2-\mathtt{c}^2}}. \tag{6}$$

We refer the readers to [23] for details.

According to our preconditioned matrix $\mathbf{A}_\alpha$, we have shown in Theorem 2.3 that its eigenvalues are enclosed in the rectangle

$$[\beta_l(\mathbf{A}_\alpha), \beta_u(\mathbf{A}_\alpha)] \times [-\gamma(\mathbf{A}_\alpha), \gamma(\mathbf{A}_\alpha)].$$

To estimate the asymptotic convergence rate of the GMRES, we compute an ellipse $\mathcal{E}(\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d})$ of smallest area containing this rectangle. Because the center of the rectangle is $(\tau, 0)$, where

$$\tau = \frac{\beta_u(\mathbf{A}_\alpha) - \beta_l(\mathbf{A}_\alpha)}{2}$$

and the length of the sides of the rectangle are

$$\chi = \beta_u(\mathbf{A}_\alpha) - \beta_l(\mathbf{A}_\alpha) \quad \text{and} \quad \omega = 2\gamma(\mathbf{A}_\alpha),$$

the ellipse $\mathcal{E}(\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d})$ which has the smallest area of all ellipses and encloses the rectangle is given by

$$\mathtt{a} = \frac{\sqrt{2}}{2}\chi, \quad \mathtt{b} = \frac{\sqrt{2}}{2}\omega, \quad \mathtt{c} = \frac{\sqrt{2}}{2}\sqrt{|\chi^2-\omega^2|}, \quad \mathtt{d} = \tau, \tag{7}$$

see [22]. By combining (6) and (7), we know that the asymptotic convergence rate of the GMRES method for solving the system of linear equations (4) is

$$\rho(\Upsilon(\mathbf{A}_\alpha)) = \frac{\chi+\omega}{\sqrt{2}\tau + \sqrt{2\tau^2 - |\chi^2-\omega^2|}}.$$

Consequently, if we choose a parameter α such that the function $\rho(\Upsilon(\mathbf{A}_\alpha))$ is minimized, then the eigenvalues of the preconditioned matrix $M(\alpha)^{-1}A$ may be clustered tightly.

3. The Two-level Preconditioners

In actual applications of the shift-splitting preconditioner $M(\alpha)$ (see (3)) to some Krylov subspace iteration methods such as GMRES, we need to solve the generalized residual equation

$$M(\alpha)z = r$$

at each of the iteration steps, where r is the current residual and z the generalized residual. This may be still costly and complicated, in particular, when A is large sparse and very ill-conditioned, even the matrix $(\alpha I + A)$ preserves well the sparse structure and is more diagonally dominant than the original matrix A. Therefore, we need to further approximate $M(\alpha)$ by another matrix, say $P(\alpha)$, which may be produced by some efficient and practical approximation process such as the incomplete triangular factorization (ILU) [18, 2], the incomplete orthogonal-triangular factorization (IQR) [23, 5], or the unsymmetric Gauss-Seidel iteration (UGS) [24, 2, 1, 3, 4]. This then yields the so-called two-level preconditioning technique that approximates the original matrix A by $P(\alpha)$ through two steps: first approximating A by $M(\alpha)$ and then approximating $M(\alpha)$ by $P(\alpha)$.

Because $(\alpha I + A)$ is as sparse as A and it is more diagonally dominant than A, we may choose a suitable shift α such that the incomplete factorizations or the relaxation iterations associated with $M(\alpha)$ are more robust and efficient than those associated with A itself.

Moreover, by Theorem 2.3 we know that the matrix $M(\alpha)^{-1}A$ has the same eigenvector set as the matrix A and its eigenvalues are more clustered. Hence, when $P(\alpha)$ is a good approximation to $M(\alpha)$, it may be also a good approximation to A itself from aspects of both eigenvalue clustering and eigenvector coinciding, as

$$P(\alpha)^{-1}A = (P(\alpha)^{-1}M(\alpha))(M(\alpha)^{-1}A).$$

4. Numerical Results

Consider the convection-diffusion equation

$$-\Delta \mathtt{u} + \beta(\mathtt{x},\mathtt{y})(\mathtt{x}\frac{\partial \mathtt{u}}{\partial \mathtt{x}} + \mathtt{y}\frac{\partial \mathtt{u}}{\partial \mathtt{y}}) + \gamma(\mathtt{x},\mathtt{y})\mathtt{u} = f(\mathtt{x},\mathtt{y}), \tag{8}$$

where $\beta(\mathtt{x},\mathtt{y})$ and $\gamma(\mathtt{x},\mathtt{y})$ are two smooth functions.

We use the finite difference scheme to discretize the convection-diffusion equation (8). That is to say, the convective term is discretized by the upwind difference scheme and the diffusive term is discretized by the centered difference scheme, on a uniform $N \times N$ grid. This leads to a system of linear equations (1) whose coefficient matrix A is of order $n = N^2$. In the experiments, we choose

$$\beta(\mathtt{x},\mathtt{y}) = qe^{\mathtt{x}+\mathtt{y}} \quad \text{and} \quad \gamma(\mathtt{x},\mathtt{y}) = 100(e^{1/\mathtt{x}}\cos(\mathtt{y}) + e^{1/\mathtt{y}}\cos(\mathtt{x})),$$

with q a positive constant used to control the magnitude of the convective term. In addition, we take the right-hand-side vector b such that the exact solution of the resulted system of linear equations (1) is $x_* = (1, 1, \ldots, 1)^T$.

In this section, we are going to test the numerical behaviour of the shift-splitting preconditioner $M(\alpha)$ by iteratively solving the above-described system of linear equations with the preconditioned GMRES method[23], for various choices of N and q. The shift-splitting preconditioner $M(\alpha)$ or its two-level alternative $P(\alpha)$ is employed to accelerate the GMRES iteration method. More specifically, the testing preconditioners include the standard ILU[18] and IGO[5] which are obtained from the incomplete LU and the incomplete Givens-orthogonal factorizations of the coefficient matrix A, and the new two-level shift-splitting ILU and two-level shift-splitting IGO (denoted respectively as TSILU and TSIGO in short) which are obtained from the incomplete LU and the incomplete Givens-orthogonal factorizations of the shift-splitting preconditioner $M(\alpha)$.

Each iteration process is started from an initial vector having all entries equal to zero, and terminated once either the iteration number is over 200 or the current iteration residual $r^{(k)} = b - Ax^{(k)}$ satisfies $\|r^{(k)}\|_2/\|r^{(0)}\|_2 \leq 10^{-6}$, where $r^{(0)} = b - Ax^{(0)}$ is the initial residual. The performance of the testing preconditioners is compared from aspects of number of iteration steps (denoted by "IT") and CPU times (denoted by "CPU"). All numerical results are implemented on Origin 3800 using C++ with double precision.

In Tables 1 and 2, we list the numbers of iteration steps and the total CPU times of the preconditioned GMRES methods. The α adopted in TSILU and TSIGO are the experimentally optimal shifts α_{exp} determined in the sense that the preconditioned GMRES with the preconditioner $M(\alpha)$ attains the least number of iteration steps among all experimental samples of the α.

From these two tables, we see that TSILU outperforms ILU and TSIGO outperforms IGO considerably in aspects of both iteration steps and CPU times. Roughly speaking, ILU shows better preconditioning effect than IGO, TSIGO, however, has much better preconditioning effect than IGO, in particular, when q becomes large.

To further illustrate the effect of the shift α on the preconditioning quality of the shift-splitting preconditioner, in Figures 1- 4 we plot the curves of IT and CPU versus α, respectively, when $N = 32$ and $q = 2000$ as well as when $N = 64$ and $q = 3000$, for the preconditioned GMRES methods. From these figures we can see that the iteration steps and the CPU times vary very irregularly when α is increasing. However, the functional relationships of the curves IT-α and CPU-α are intuitively comparable, especially when α is close to the experimentally optimal shift α_{exp}. In fact, the curves IT-α and CPU-α share the same minimum point α_{exp}.

Figures 5 and 6 depict the curves of IT and CPU versus q for the preconditioned GMRES methods. We see that the curves IT-q and CPU-q have almost the same shapes for each preconditioner, and TSILU and TSIGO have better numerical behaviours than ILU and IGO, respectively, for all q.

Figure 7 shows the functional relationship of the experimentally optimal shift α_{exp} with respect to q when $N = 32$. Clearly, both curves with respect to TSILU and TSIGO are much similar, except for $q \geq 4000$ when the behaviours of the two curves become drastically different.

Table 1: IT and CPU for the ILU- and the TSILU-preconditioned GMRES methods

N	Precond	q	1000	2000	3000	4000	5000
32	ILU	IT	14	34	44	48	41
		CPU	0.10	0.41	0.67	0.78	0.59
	TSILU	α_{exp}	31.8	177.0	166.7	77.8	25.1
		IT	4	10	21	21	28
		CPU	0.01	0.05	0.17	0.17	0.28
64	ILU	IT	5	22	40	59	65
		CPU	0.12	0.81	2.14	4.37	5.24
	TSILU	α_{exp}	0.5	131.0	134.4	37.3	149.2
		IT	5	4	12	19	18
		CPU	0.07	0.06	0.27	0.58	0.54

Table 2: IT and CPU for the IGO- and the TSIGO-preconditioned GMRES methods

N	Precond	q	1000	2000	3000	4000	5000
32	IGO	IT	16	26	45	81	75
		CPU	0.17	0.31	0.77	2.21	1.91
	TSIGO	α_{exp}	79.4	167.1	53.3	93.9	4.2
		IT	2	12	15	9	21
		CPU	0.01	0.09	0.13	0.07	0.22
64	IGO	IT	5	17	52	68	51
		CPU	0.18	0.67	3.87	6.24	3.74
	TSIGO	α_{exp}	0.1	39.6	177.2	139.3	50.6
		IT	5	4	3	9	12
		CPU	0.13	0.11	0.09	0.26	0.38

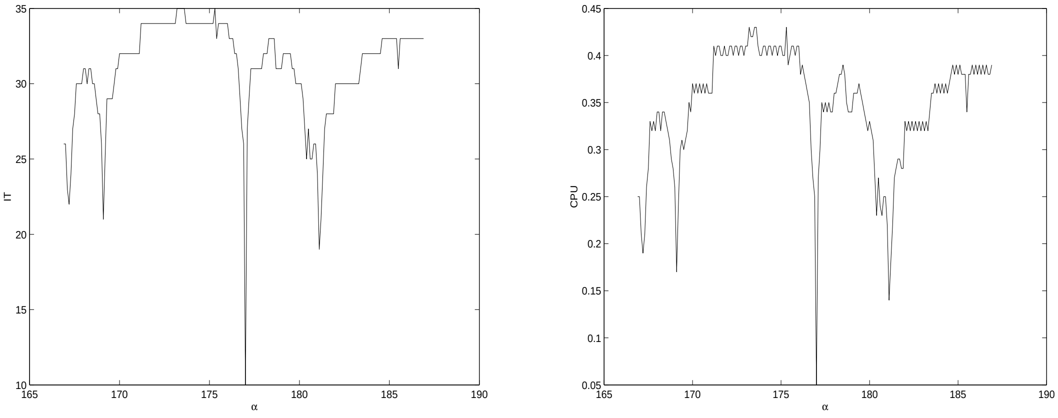

Figure 1: Curves of IT (left) and CPU (right) versus α for TSILU-preconditioned GMRES methods. ($N = 32$ and $q = 2000$)

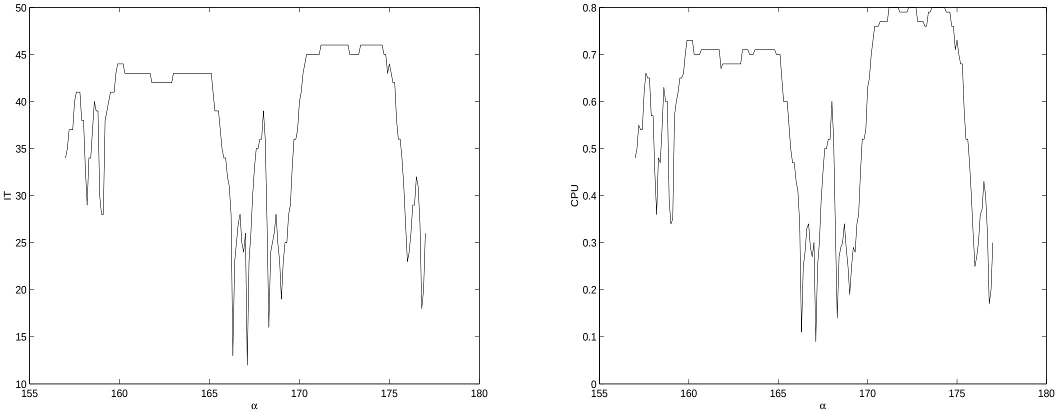

Figure 2: Curves of IT (left) and CPU (right) versus α for TSIGO-preconditioned GMRES methods. ($N = 32$ and $q = 2000$)

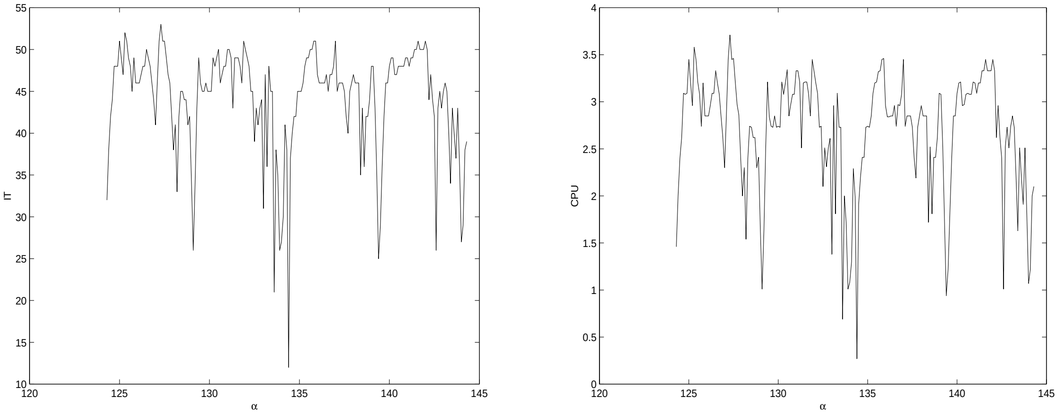

Figure 3: Curves of IT (left) and CPU (right) versus α for TSILU-preconditioned GMRES methods. ($N = 64$ and $q = 3000$)

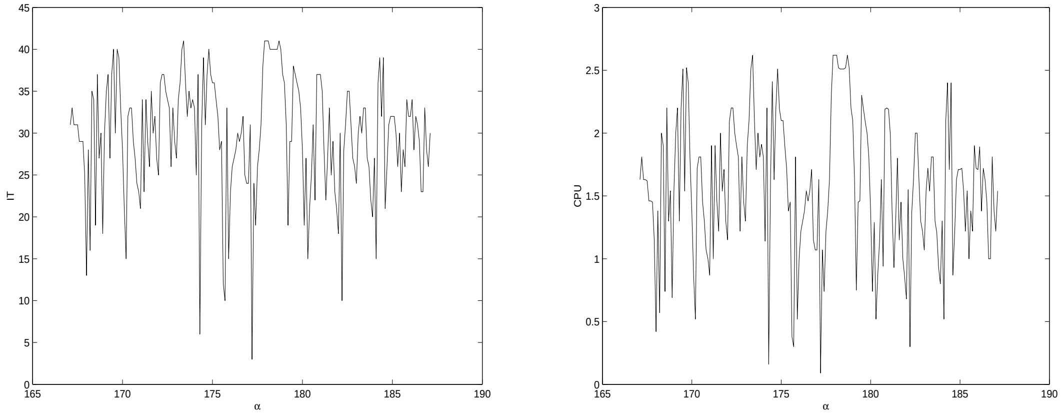

Figure 4: Curves of IT (left) and CPU (right) versus α for TSIGO-preconditioned GMRES methods. ($N = 64$ and $q = 3000$)

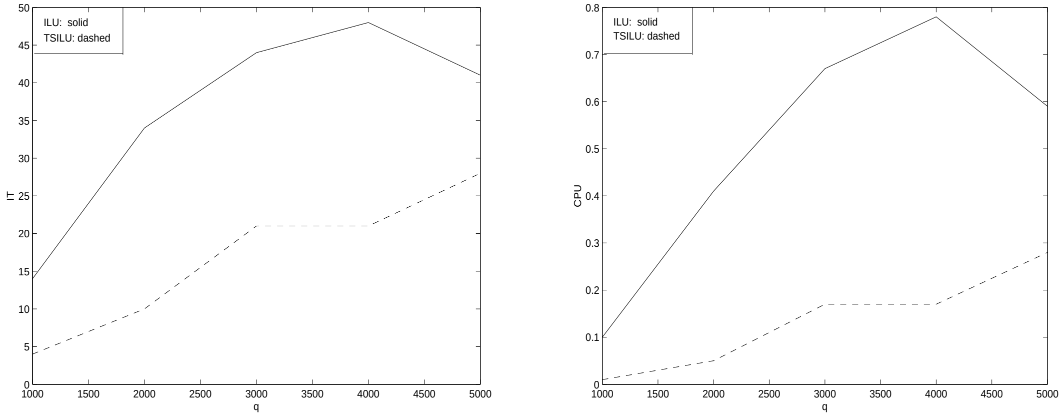

Figure 5: Curves of IT (left) and CPU (right) versus q. ($N = 32$)

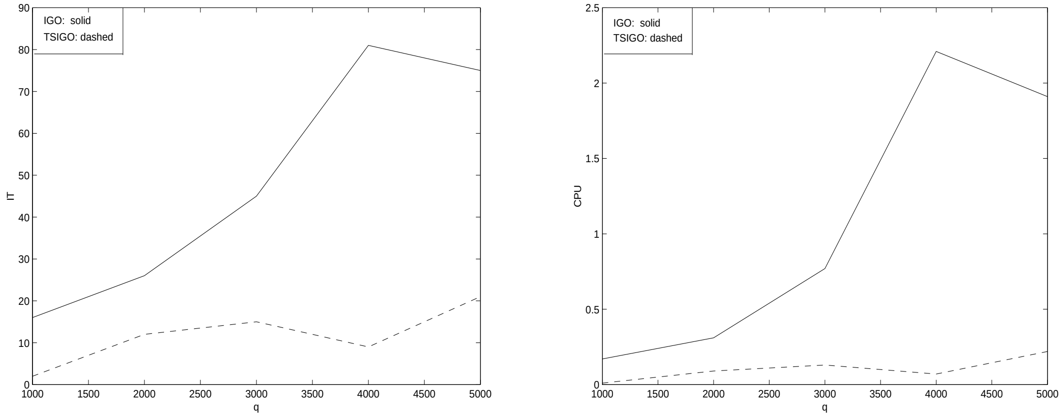

Figure 6: Curves of IT (left) and CPU (right) versus q. ($N = 64$)

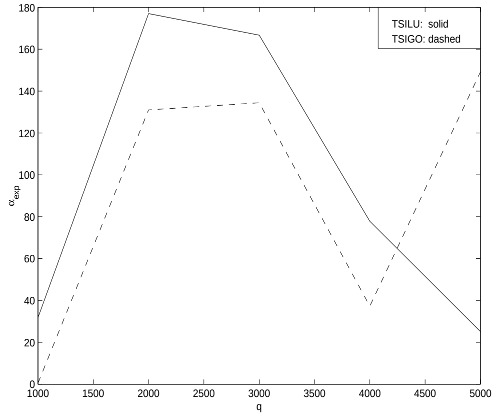

Figure 7: Curves of α_{exp} versus q. ($N = 32$)

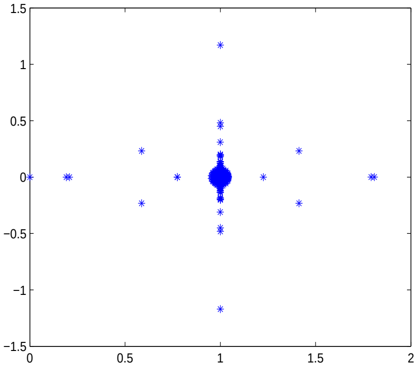

Figure 8: Distribution of the eigenvalues for the scaled matrix $\widehat{A} = \text{diag}(A)^{-1/2}\ A\ \text{diag}(A)^{-1/2}$

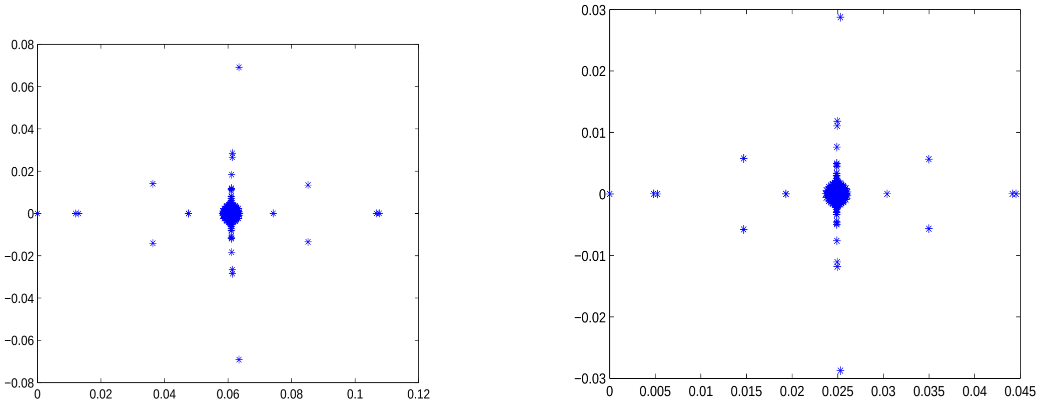

Figure 9: Distributions of the eigenvalues for the preconditioned matrices $\widehat{P}(\alpha)^{-1}\widehat{A}$, with $\widehat{P}(\alpha)$ the TSILU (left, $\alpha_{\text{exp}} = 31.8$) and the TSIGO (right, $\alpha_{\text{exp}} = 79.4$) factorizations of $\widehat{M}(\alpha) = \frac{1}{2}(\alpha I + \widehat{A})$. ($N = 32$ and $q = 1000$)

To further investigate the preconditioning property of the shift-splitting preconditioner $M(\alpha)$, in Figures 8 and 9 we plot the eigenvalue distributions of the scaled coefficient matrix $\widehat{A} := \text{diag}(A)^{-1/2}\ A\ \text{diag}(A)^{-1/2}$ and the correspondingly preconditioned matrix $\widehat{P}(\alpha)^{-1}\widehat{A}$ when $N = 32$ and $q = 1000$, where $\widehat{P}(\alpha)$ is either the ILU or the IGO factorization of the shift-splitting preconditioner $\widehat{M}(\alpha) = \frac{1}{2}(\alpha I + \widehat{A})$. Here, the symmetrically diagonal scaling is adopted to normalize the eigenvalues of the matrix A. Evidently, from these figures we observe that the eigenvalues of the preconditioned matrices are more tightly clustered than the original coefficient matrix and, therefore, the correspondingly induced preconditioned GMRES methods may converge quickly to the solution of the system of linear equations. This fact has been already confirmed by the numerical results shown in Tables 1 and 2.

5. Conclusions

We have presented and analyzed a class of shift-splitting preconditioners for non-Hermitian positive definite matrices, and used numerical examples to show that the new preconditioning strategy may potentially yield efficient preconditioners for Krylov subspace methods such as GMRES for solving large sparse positive definite systems of linear equations, provided a good estimate of the shift α can be prescribed. The choice of such an α is usually problem-dependent, and is also closely related to the Krylov subspace method used. Hence, how to determine the best shift α such that the induced preconditioned Krylov subspace method possesses fast convergence speed and low computation cost needs to be further studied in depth.

Acknowledgements. Part of this work was done during the first author's visit to Shanghai Key Laboratory of Contemporary Applied Mathematics of Fudan University during October-November, 2004. So, he is very much indebted to Prof. Yang-Feng Su for his kind invitation of this visit.

References

[1] O. Axelsson, A generalized SSOR method, *BIT Numerical Mathematics,* **12** (1972), 443-467.

[2] O. Axelsson, Iterative Solution Methods, *Cambridge University Press,* Cambridge, 1994.

[3] Z.-Z. Bai, A class of modified block SSOR preconditioners for symmetric positive definite systems of linear equations, *Advances in Computational Mathematics,* **10** (1999), 169-186.

[4] Z.-Z. Bai, Modified block SSOR preconditioners for symmetric positive definite linear systems, *Annals of Operations Research*, **103** (2001), 263-282.

[5] Z.-Z. Bai, I.S. Duff and A.J. Wathen, A class of incomplete orthogonal factorization methods. I: Methods and theories, *BIT Numerical Mathematics*, **41**:1 (2001), 53-70.

[6] Z.-Z. Bai, G.H. Golub, L.-Z. Lu and J.-F. Yin, Block triangular and skew-Hermitian splitting methods for positive-definite linear systems, *SIAM Journal on Scientific Computing*, **26**:3 (2005), 844-863.

[7] Z.-Z. Bai, G.H. Golub and M.K. Ng, Hermitian and skew-Hermitian splitting methods for non-Hermitian positive definite linear systems, *SIAM Journal on Matrix Analysis and Applications*, **24**:3 (2003), 603-626.

[8] Z.-Z. Bai and S.-L. Zhang, A regularized conjugate gradient method for symmetric positive definite system of linear equations, *Journal of Computational Mathematics*, **20**:4 (2002), 437-448.

[9] J. Bey and A. Reusken, On the convergence of basic iterative methods for convection-diffusion equations, *Numerical Linear Algebra with Applications*, **6** (1999), 329-352.

[10] J.W. Daniel, The conjugate gradient method for linear and nonlinear operator equations, *SIAM Journal on Numerical Analysis,* **4** (1967), 10-26.

[11] R.S. Dembo, S.C. Eisenstat and T. Steihaug, Inexact Newton methods, *SIAM Journal on Numerical Analysis,* **19** (1982), 400-408.

[12] G.H. Golub and C.F. Van Loan, Matrix Computations, 3rd Edition, *The Johns Hopkins University Press,* Baltimore and London, 1996.

[13] G.H. Golub and M.L. Overton, Convergence of a two-stage Richardson iterative procedure for solving systems of linear equations, In: Numerical Analysis (Proceedings of the Ninth Biennial Conference, Dundee, Scotland, 1981)(G.A. Watson, ed.), Lecture Notes Math. 912, *Springer, New York/Heidelberg/Berlin,* pp. 128-139.

[14] G.H. Golub and M.L. Overton, The convergence of inexact Chebyshev and Richardson iterative methods for solving linear systems, *Numerische Mathematik,* **53** (1988), 571-593.

[15] M.R. Hestenes and E.L. Stiefel, Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau Standards, Section B,* **49** (1952), 409-436.

[16] C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations, *SIAM,* Philadelphia, 1995.

[17] T.A. Manteuffel, An incomplete factorization technique for positive definite linear systems, *Mathematics of Computation,* **34** (1980), 473-497.

[18] J.A. Meijerink and H.A. van der Vorst, An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix, *Mathematics of Computation,* **31** (1977), 148-162.

[19] N.K. Nichols, On the convergence of two-stage iterative processes for solving linear equations, *SIAM Journal on Numerical Analysis,* **10** (1973), 460-469.

[20] N.K. Nichols, On the local convergence of certain two step iterative procedures, *Numerische Mathematik,* **24** (1975), 95-101.

[21] J.-Y. Pan, M.K. Ng and Z.-Z. Bai, New preconditioners for saddle point problems, *Applied Mathematics and Computation,* **172** (2006), 762-771.

[22] Y. Saad, Numerical Methods for Large Eigenvalue Problems: Theory and Algorithms, Manchester University Press, Manchester, 1992.

[23] Y. Saad, Iterative Methods for Sparse Linear Systems, PWS Publishing Company, Boston, 1996.

[24] R.S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1962.