Computing and Informatics, Vol. 34, 2015, 495–516

AN EXTENSION OF TWO CONJUGATE DIRECTION METHODS TO MARKOV CHAIN PROBLEMS

Chun Wen, Ting-Zhu Huang

University of Electronic Science and Technology of China
School of Mathematical Sciences
611731, Chengdu, Sichuan, P.R. China
e-mail: wchun17@163.com, tingzhuhuang@126.com

Tomohiro Sogabe

Aichi Prefectural University
Department of Information Science and Technology
Aichi 480-1198, Japan
e-mail: sogabe@ist.aichi-pu.ac.jp

Abstract. Motivated by the recent applications of the conjugate residual method to nonsymmetric linear systems by Sogabe, Sugihara and Zhang [An extension of the conjugate residual method to nonsymmetric linear systems. J. Comput. Appl. Math., Vol. 266, 2009, pp. 103–113], this paper describes two conjugate direction methods, BiCR and BiCG, and attempts to extend their applications to compute the stationary probability distribution for an irreducible Markov chain with the aim of finding an alternative basic solver. Numerical experiments show the feasibility of the BiCR and BiCG to some extent, with applications to several practical Markov chain problems.

Keywords: Krylov subspace methods, conjugate direction methods, Markov chains, stationary probability distribution

Mathematics Subject Classification 2010: 60J22, 65C40

1 INTRODUCTION

Krylov subspace methods are one of the most widespread and extensively accepted techniques for numerical solutions of today's large-scale linear systems of the form

$$Ax = b, \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$, $x, b \in \mathbb{R}^n$ with b known and x unknown. Most of those linear systems arise from various fields of computational science and engineering. Such examples are electromagnetic applications, digital image processing, discrete- and continuous-time Markov chains.

When the coefficient matrix A is symmetric, if A is positive definite, then the well-known Conjugate Gradient (CG) method [1] is the best choice for solving the linear systems (1), while if A is indefinite, then mathematically equivalent Minimum Residual (MINRES) method [2] and Conjugate Residual (CR) method [3] are applied.

On the other hand, when the coefficient matrix A is nonsymmetric, Krylov subspace methods based on minimizing the residual $r = b - Ax$ may be efficient, such as the Generalized Minimum Residual (GMRES) method [4] and the Generalized Conjugate Residual (GCR) method [5] have been proposed as the generalized revisions of the MINRES and CR methods. In addition, Krylov subspace methods based on bi-orthogonality are also used, such as the Bi-Conjugate Gradient (BiCG) method [6, 7] has been given plenty of attentions to develop and improve its performance. As the successful representations of the modified BiCG method, the Conjugate Gradients Squared (CGS) method, Biconjugate Gradient Stabilized (BiCGSTAB) method and Quasi-Minimal Residual (QMR) method have been studied by Sonneveld [8], van der Vorst [9], and Freund and Nachtigal [10], respectively. Note that the convergence behavior of all these methods is fairly well understood if the matrix A in (1) is nonsingular [11].

However, this paper is concerned with the computation of the linear systems (1) when the coefficient matrix A is singular. In recent years, many researches have devoted much effort to analyze the performance of Krylov subspace methods for singular systems. For instance, the interested readers can refer to the works for the CG method [12, 13, 14, 15], the QMR and its variant, transpose-free QMR (TFQMR) methods [16, 17, 18] and the GMRES-type methods [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32].

Based on the work of Sogabe, Sugihara and Zhang [33], the main attention of the present paper is to describe two conjugate direction methods and attempt to extend their applications to compute the singular systems which arise from Markov chains. One of the conjugate direction methods is the BiCG method [6, 7]. Another is the BiCR method, which has been proposed by Sogabe, Sugihara and Zhang [33] according to one of the simplest derivations of the BiCG method given by van der Vorst [34]. As a matter of fact, the works of comparing different conjugate direction methods are not new, e.g., Broyden has given a comparison of differ-

ent conjugate direction methods for the real and nonsymmetric nonsingular matrix in [35].

The remainder of the paper is organized as follows. In Section 2, Markov chain modellings are briefly introduced. In Section 3, based on the work of [33], we extend the BiCG and BiCR methods to solve Markov chain problems. In Section 4, numerical experiments on several Markov chains are made and comparative analyses between the BiCR and BiCG methods are given. Finally, conclusions and future work are presented in Section 5.

Throughout this paper, the following notations will be used:

- A^{T}: the transpose of the matrix $A \in \mathbb{R}^{n\times n}$;
- (x, y): the dot product given by $x^{\mathrm{T}}y$ with $x, y \in \mathbb{R}^n$;
- $\mathscr{R}(A)$: the range space of the matrix A;
- $\mathscr{N}(A)$: the null space of the matrix A;
- $\mathscr{K}_n(A, r)$: the Krylov subspace generated by a vector r, i.e.,

$$\mathscr{K}_n(A, r) = \mathrm{span}\{r, Ar, \ldots, A^{n-1}r\};$$

- $k = \mathrm{index}(A)$: the smallest nonnegative integer k with $\mathscr{R}(A^k) = \mathscr{R}(A^{k+1})$.

2 MARKOV CHAIN MODELLINGS

In this section, Markov chain modellings which represent a class of singular systems are briefly introduced and certain related properties are presented.

Markov chains are one of the most important kinds of models in simulation. Generally speaking, Markov chains are divided into discrete-time Markov chains (DTMCs) and continuous-time Markov chains (CTMCs) [36, 37, 38]. For a finite and irreducible Markov chain, there exists a unique stationary probability distribution π whose elements are strictly greater than zero. If the Markov chain is a DTMC, then it has

$$\pi P = \pi,\ \pi > 0,\ \pi e = 1, \tag{2}$$

where $P \in \mathbb{R}^{n\times n}$ is a transition probability matrix, and $e = (1, \ldots, 1)^{\mathrm{T}}$ is a $n \times 1$ vector. This problem can be considered as an eigenvalue problem. If the Markov chain is a CTMC, then we have

$$\pi Q = 0,\ \pi > 0,\ \pi e = 1, \tag{3}$$

where $Q \in \mathbb{R}^{n\times n}$ is an infinitesimal generator, and e is a column vector as given above.

Observe that both the Equations (2) and (3) may be put into the same form. Let $I \in \mathbb{R}^{n\times n}$ be an identity matrix, then the Equation (2) may be rewritten as

$\pi(P-I)=0$, which has the same form as the Equation (3). On the other hand, we may discrete a CTMC. From the Equation (3), we can have

$$\pi(Q\triangle T+I)=\pi,\ \pi>0,\ \pi e=1, \tag{4}$$

which has the same form as the Equation (2). In the discreted Markov chain, transitions take place at interval $\triangle T$, where $\triangle T$ is chosen sufficiently small such that the probability of two transitions taking place in time $\triangle T$ is negligible, e.g., one possibility is to take

$$\triangle T \leq \frac{1}{\max_i |q_{ii}|}.$$

In this case, $(Q\triangle T+I)$ is stochastic and the stationary probability distribution π of the CTMC, obtained from the Equation (3), is identical to that of the discreted Markov chain, obtained from the Equation (4). Hence, numerical methods designed to compute the stationary probability distribution of DTMCs can be used to compute the stationary probability distribution of CTMCs, and vice versa.

Here our concern is to compute the stationary probability distribution of CTMCs. The linear systems (3) may arise from information retrieval and web ranking [39, 40, 41, 42], queueing systems [43, 44, 45, 46], stochastic automata networks [47, 48, 49], manufacturing systems and inventory control [50] and communication systems [51, 52, 53]. In order to analyze their performance measures, it is required to find out the corresponding stationary probability distribution, i.e., the problem is changed to consider how to solve the linear systems (3) efficiently.

For simplicity, we rewrite the linear systems (3) to the form

$$Ax=0,\ \text{with}\ A=Q^{\mathrm{T}},\ x=\pi^{\mathrm{T}}. \tag{5}$$

Recently, there are large amounts of works that have been devoted to solving the linear systems (5). Such examples are the direct methods in [11, 36, 54] and the Krylov subspace methods in [12, 17, 18, 20, 23, 27, 28, 46, 48, 49, 54, 55]. In addition, the readers may find other efficient methods which have been studied to solve the linear systems (5) in [56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69] and references therein.

To some extent, it is probably fair to say that none of these methods mentioned before is completely satisfactory, since each method has its own advantages and disadvantages. For example, GMRES and GCR methods apply the minimal residual smoothing process so that they show smooth convergence behavior; however, GMRES and GCR methods may become impractical because the computational work and memory increase linearly with the number of iterations increasing; see, e.g., [11, 33]. Additionally, multilevel method can transfer a large linear system into a smaller one by some aggregation strategies so that the numerical solution can be computed in an efficient way; however, it may be expensive for unstructured problems since much time will be spent in constructing the restriction and prolongation operators and transforming in each level; see, e.g., [57, 58].

Motivated by the recent applications of the conjugate residual method to non-symmetric linear systems by Sogabe, Sugihara and Zhang [33], we attempt to extend the applications of the BiCR and BiCG methods to solve the stationary probability distribution of the singular linear systems (5) in order to make them become possible tools for the numerical solution of Markov chains in the next section.

Let us end this section with the following remarks:

Remark 1. The coefficient matrix A of (5) has zero column sum, positive diagonal entries and non-positive off-diagonal entries. Thus A is singular.

Remark 2. The coefficient matrix A of (5) is irreducible. Thus from the well-known Perron and Frobenius theory; see, e.g., [38], A has a one-dimensional null-space with a positive vector, i.e.,

$$\mathscr{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\} = \text{span}\{z\} \text{ for some positive vector } z. \tag{6}$$

Remark 3. The singular linear systems (5) are consistent. And all the Markov chain matrices have the property $\text{index}(A) = 1$ which is equivalent to $\mathscr{R}(A) \cap \mathscr{N}(A) = \varnothing$. This point has been discussed in papers [17, 24, 27, 28, 31, 37, 67, 68, 69] as the condition of studying the related convergence theory.

3 AN EXTENSION OF BICG AND BICR TO MARKOV CHAINS

In this section, we extend the proposed methods in [33] to solve the stationary probability distribution of the singular linear system (5) which arises from Markov chain problems.

It is well known that BiCG and BiCR methods are Krylov subspace methods for solving the large-scale linear system (1), where the coefficient matrix A is often non-singular and nonsymmetric. Particularly, as far as the BiCG method is concerned, there are several ways to derive it, one of the simplest derivations can be found in [33, 34]. However, both the BiCG and BiCR methods do not apply to solve the stationary probability distribution of an irreducible Markov chain extensively, since the coefficient matrix A in the linear system (5) which comes from some Markov modellings is singular. With this in mind, we extend the BiCG and BiCR methods to the linear system (5) as the following framework for the purpose of easily transforming them into computer code and rapidly comparing their complexities. The readers may be referred to see [11, 33, 34] for understanding their specific derivation processes which is not to be introduced here again.

Observe Algorithms 1 and 2, it is not difficult to find that an advantage of the BiCR and BiCG methods is that both of them possess short-term recurrences, which is the basis for their successful applications. Their computational cost can be found in Table 1 of [33]. More, some properties of the BiCG and BiCR methods have been discussed by Sogabe, Sugihara and Zhang, for instance, if breakdown does not occur in Algorithms 1 and 2, then there are

Algorithm 1: The BiCR method for Markov chains

Choose an initial guess x_0, with $x_0 \geq 0$ and $\|x_0\|_1 = 1$,
Compute $r_0 = -Ax_0$ and choose s_0 such that $(s_0, Ar_0) \neq 0$, e.g., $s_0 = r_0$,
Set $p_0 = r_0$, $q_0 = s_0$,
Compute $w_0 = Ap_0$, $v_0 = A^{\mathrm{T}}q_0$, $c_0 = v_0^{\mathrm{T}}w_0$, $z_0 = Ar_0$, $d_0 = s_0^{\mathrm{T}}z_0$,
For $j = 1, 2, \ldots,$ do
    If $c_{j-1} = 0$, then the method fails
    $\alpha_{j-1} = d_{j-1}/c_{j-1}$
    $x_j = x_{j-1} + \alpha_{j-1}p_{j-1}$
    $r_j = r_{j-1} - \alpha_{j-1}w_{j-1}$
    $s_j = s_{j-1} - \alpha_{j-1}v_{j-1}$
    Check convergence; continue if necessary
    $z_j = Ar_j$
    $d_j = s_j^{\mathrm{T}}z_j$
    $\beta_{j-1} = d_j/d_{j-1}$
    $p_j = r_j + \beta_{j-1}p_{j-1}$
    $q_j = s_j + \beta_{j-1}q_{j-1}$
    $w_j = z_j + \beta_{j-1}w_{j-1}$
    $v_j = A^{\mathrm{T}}q_j$
    $c_j = v_j^{\mathrm{T}}w_j$
End For

Algorithm 2: The BiCG method for Markov chains

Choose an initial guess x_0, with $x_0 \geq 0$ and $\|x_0\|_1 = 1$,
Compute $r_0 = -Ax_0$ and choose s_0 such that $(s_0, r_0) \neq 0$, e.g., $s_0 = r_0$,
Set $p_0 = r_0$, $q_0 = s_0$,
Compute $w_0 = Ap_0$, $v_0 = A^{\mathrm{T}}q_0$, $c_0 = q_0^{\mathrm{T}}w_0$, $d_0 = s_0^{\mathrm{T}}r_0$,
For $j = 1, 2, \ldots,$ do
    If $c_{j-1} = 0$, then the method fails
    $\alpha_{j-1} = d_{j-1}/c_{j-1}$
    $x_j = x_{j-1} + \alpha_{j-1}p_{j-1}$
    $r_j = r_{j-1} - \alpha_{j-1}w_{j-1}$
    $s_j = s_{j-1} - \alpha_{j-1}v_{j-1}$
    Check convergence; continue if necessary
    $d_j = s_j^{\mathrm{T}}r_j$
    $\beta_{j-1} = d_j/d_{j-1}$
    $p_j = r_j + \beta_{j-1}p_{j-1}$
    $q_j = s_j + \beta_{j-1}q_{j-1}$
    $w_j = Ap_j$
    $v_j = A^{\mathrm{T}}q_j$
    $c_j = q_j^{\mathrm{T}}w_j$
End For

Theorem 1. [33]: For any $i \neq j$, the following bi-orthogonality properties hold,

1. $(s_i, Ar_j) = 0$,
2. $(A^{\mathrm{T}}q_i, Ap_j) = 0$.

and

Theorem 2. [33]: Some further properties of the BiCR method are

1. $(s_i, Ap_j) = 0$, for $i > j$,
2. $(s_i, Ar_i) = (s_i, Ap_i)$,
3. $(A^{\mathrm{T}}s_i, Ap_i) = (A^{\mathrm{T}}q_i, Ap_i)$.

Recently, a large amount of excellent experts has paid their attentions to study the convergence analysis of Krylov subspace methods when applied to singular systems; see, e.g., [17, 24, 27, 28, 31, 37, 67, 68, 69]. In fact, as mentioned by Ipsen and Meyer [30], when the coefficient matrix A is singular, even if a solution exists, it may not lie in the Krylov space $\mathscr{K}_n(A, r)$. They decomposed the space into $\mathbb{R}^n = \mathscr{R}(A^k) \oplus \mathscr{N}(A^k)$ with $k = \mathrm{index}(A)$ for overcoming this problem. And then they restrained the right-hand vector b by means of the Jordan canonical form of A and kept b away from the nilpotent part of A such that the solution of a general square system $Ax = b$ may lie in a Krylov subspace.

From Remarks 1, 2 and 3 given in Section 2, the coefficient matrix A of (5) is found to be a singular and irreducible matrix with a one-dimensional null space spanned by a positive vector. In particular, the matrix A possesses the property $\mathrm{index}(A) = 1$. For the consistent systems (5), general conditions have been provided in [17] and [30] under which the Krylov subspace method is convergent. It is natural that the convergence analysis of the Krylov subspace method is also true for the BiCR and BiCG methods when applied to singular systems that arise from Markov chains, since both of them are Krylov subspace methods based on bi-orthogonality. Hence, suppose there is no breakdown, then as a direct consequence of the convergence theory given in [17] and [30], a simple result for the BiCR and BiCG methods applied to Markov chain problems is presented as follows.

Theorem 3. For the consistent systems (5), if the coefficient matrix A is a singular and irreducible matrix with $\mathrm{index}(A) = 1$, then the BiCR and BiCG methods determine a solution for the linear systems (5).

Theorem 3 indicates the convergence of Algorithms 1 and 2. Indeed, it is well known that the convergence rate of Krylov subspace methods have close relation with the distribution of the eigenvalues of the coefficient matrix [11]. The more the eigenvalues are clustered together, the faster these Krylov subspace methods will converge. Hence, a desired eigenvalue distribution is wished to be obtained by the applications of preconditioning techniques. Based on the unpreconditioned Algorithms 1 and 2, the preconditioned BiCG and BiCR methods are given in Algorithms 3 and 4 respectively, in which the matrix $M \in \mathbb{R}^{n\times n}$ denotes the preconditioner.

Algorithm 3: The preconditioned BiCR method for Markov chains

Choose an initial guess x_0, with $x_0 \geq 0$ and $\|x_0\|_1 = 1$,
Compute $r_0 = -Ax_0$ and choose s_0 such that $(s_0, Ar_0) \neq 0$, e.g., $s_0 = r_0$,
Compute $g_0 = M^{-1}r_0$, $h_0 = M^{-\mathrm{T}}s_0$,
Set $p_0 = g_0$, $q_0 = h_0$,
Compute $w_0 = Ap_0$, $v_0 = A^{\mathrm{T}}q_0$, $t_0 = M^{-\mathrm{T}}v_0$, $c_0 = t_0^{\mathrm{T}}w_0$, $z_0 = Ag_0$, $d_0 = h_0^{\mathrm{T}}z_0$,
For $j = 1, 2, \ldots,$ do
    If $c_{j-1} = 0$, then the method fails
    $\alpha_{j-1} = d_{j-1}/c_{j-1}$
    $x_j = x_{j-1} + \alpha_{j-1}p_{j-1}$
    $r_j = r_{j-1} - \alpha_{j-1}w_{j-1}$
    $s_j = s_{j-1} - \alpha_{j-1}v_{j-1}$
    Check convergence; continue if necessary
    $g_j = M^{-1}r_j$
    $h_j = h_{j-1} - \alpha_{j-1}t_{j-1}$
    $z_j = Ag_j$
    $d_j = h_j^{\mathrm{T}}z_j$
    $\beta_{j-1} = d_j/d_{j-1}$
    $p_j = g_j + \beta_{j-1}p_{j-1}$
    $q_j = h_j + \beta_{j-1}q_{j-1}$
    $w_j = z_j + \beta_{j-1}w_{j-1}$
    $v_j = A^{\mathrm{T}}q_j$
    $t_j = M^{-\mathrm{T}}v_j$
    $c_j = t_j^{\mathrm{T}}w_j$
End For

4 NUMERICAL RESULTS AND ANALYSIS

In this section, we try to make some numerical experiments to illustrate numerical behavior of the BiCG and BiCR methods, with applications to several practical Markov chain problems. For convenience, we let PBiCG and PBiCR denote the preconditioned BiCG and BiCR methods, respectively. These methods are tested with respect to the number of iterations (IT), computational time in seconds (CPU), and their convergence histories which are plotted in the following figures with IT (on the horizontal axis) versus Relres (on the vertical axis), where Relres is defined as $\log_{10}(\|r_n\|_2/\|r_0\|_2)$.

All experiments were run by using the Matlab 7.0.4 implementation on Microsoft Window XP with 2.27GHz 64-bit processor and 2GB memory. And in all cases the initial guess is generated by random sampling with a uniform (0,1) distribution and normalized to one in the one norm, since this choice seems to work well for almost all of the tests considered here. The stopping criteria used here were

$$\|r_n\|_2/\|r_0\|_2 < \mathrm{tol},$$

Algorithm 4: The preconditioned BiCG method for Markov chains

Choose an initial guess x_0, with $x_0 \geq 0$ and $\|x_0\|_1 = 1$,
Compute $r_0 = -Ax_0$ and choose s_0 such that $(s_0, r_0) \neq 0$, e.g., $s_0 = r_0$,
Compute $g_0 = M^{-1}r_0$, $h_0 = M^{-\mathrm{T}}s_0$,
Set $p_0 = g_0$, $q_0 = h_0$,
Compute $w_0 = Ap_0$, $v_0 = A^{\mathrm{T}}q_0$, $c_0 = q_0^{\mathrm{T}}w_0$, $d_0 = s_0^{\mathrm{T}}g_0$,
For $j = 1, 2, \ldots,$ do
    If $c_{j-1} = 0$, then the method fails
    $\alpha_{j-1} = d_{j-1}/c_{j-1}$
    $x_j = x_{j-1} + \alpha_{j-1}p_{j-1}$
    $r_j = r_{j-1} - \alpha_{j-1}w_{j-1}$
    $s_j = s_{j-1} - \alpha_{j-1}v_{j-1}$
    Check convergence; continue if necessary
    $g_j = M^{-1}r_j$
    $h_j = M^{-\mathrm{T}}s_j$
    $d_j = s_j^{\mathrm{T}}g_j$
    $\beta_{j-1} = d_j/d_{j-1}$
    $p_j = g_j + \beta_{j-1}p_{j-1}$
    $q_j = h_j + \beta_{j-1}q_{j-1}$
    $w_j = Ap_j$
    $v_j = A^{\mathrm{T}}q_j$
    $c_j = q_j^{\mathrm{T}}w_j$
End For

where r_n is the residual of the n^{th} iteration, and $\mathrm{tol} = 10^{-3}$. Note that the signs "N" and "$\mathrm{nnz}(A)$" denote the size and the number of nonzero of the coefficient matrix A in the singular linear systems (5), respectively.

4.1 Example 1: The M/M/1 Queue

This example is taken from Latouche and Ramaswami [56], which is a canonical example of homogeneous birth-and-death processes where the only allowed transitions are from the state n to the next higher state $n+1$ for all $n \geq 0$ and from n to $n-1$ for $n \geq 1$. It is a single server queueing system with an infinite waiting room. Customers arrive in the system at the renewal epochs of a Poisson process with parameter $\lambda = 1$. If there are other customers already present, then they enter a waiting room, otherwise they obtain service immediately. The service rate is exponential with parameter $\mu = 1$.

The preconditioner for this test problem is chosen to be $M = U + U^{\mathrm{T}}$, where $U \in \mathbb{R}^{n\times n}$ is the upper triangular matrix of A. To illustrate that the eigenvalue distributions of the preconditioned matrix $M^{-1}A$ are more clustered together than those of the original coefficient matrix A, Figure 1 has plotted the eigenvalue distributions of A and $M^{-1}A$ when the size of this example is $N = 1\,000$. It is clear that the eigenvalue distributions of $M^{-1}A$ are desirable.

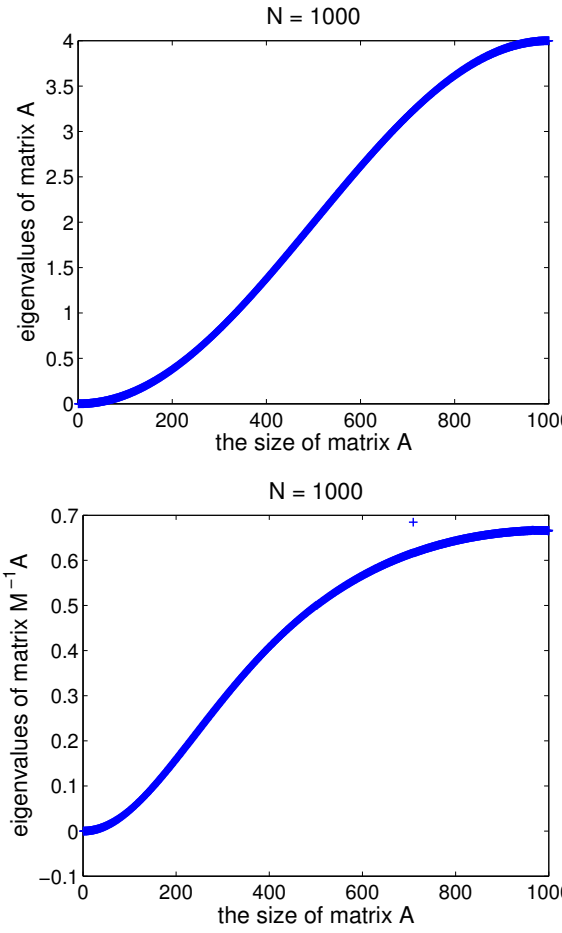

Figure 1. Comparison of the eigenvalue distributions between the original matrix A (top) and the preconditioned matrix $M^{-1}A$ (bottom) for Example 1 when $N = 1\,000$

N	nnz(A)	BiCG	BiCR	PBiCG	PBiCR
1 000	2 998	28 (0.013)	20 (0.010)	17 (0.008)	12 (0.005)
5 000	14 998	29 (0.032)	20 (0.022)	17 (0.020)	12 (0.017)
10 000	29 998	29 (0.061)	20 (0.045)	17 (0.039)	12 (0.035)
50 000	149 998	29 (0.356)	20 (0.261)	17 (0.211)	12 (0.172)
100 000	299 998	29 (0.731)	20 (0.546)	17 (0.518)	12 (0.449)
200 000	599 998	29 (1.466)	20 (1.203)	17 (1.171)	12 (0.921)
500 000	1 499 998	29 (2.749)	20 (2.512)	17 (2.135)	12 (1.647)

Table 1. IT and CPU (in brackets) of these methods for Example 1.

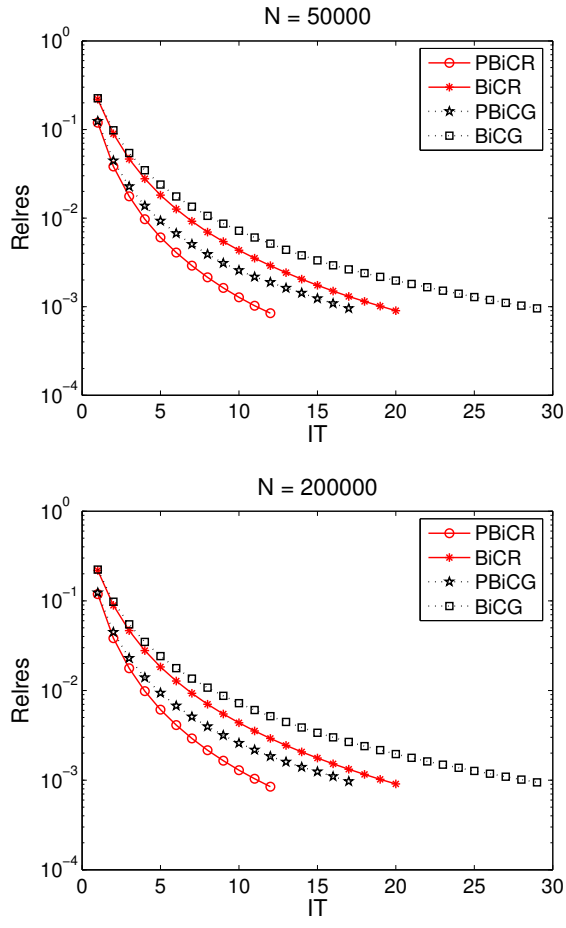

Figure 2. Convergence histories of these methods for Example 1

Table 1 lists the number of iterations and the computational time in seconds (in brackets) for the four algorithms on this M/M/1 queueing problem. We observe that both the preconditioned BiCG and BiCR methods run faster than the unpreconditioned BiCG and BiCR methods, while the PBiCR method is superior to the PBiCG method both in terms of IT and CPU time. Especially, it is not difficult to find that these methods are quite stable even though the size of this test problem has a dramatic increase.

Figure 2 has shown the convergence histories of these four algorithms for this test problem when the sizes of the matrix A are $N = 50\,000$ and $N = 200\,000$, respectively. One can see that the PBiCR and PBiCG methods use fewer iteration steps than the BiCR and BiCG methods, to reach the desired accuracy. It turns

out that these conjugate direction methods are extended to solve Markov chain modellings are feasible and efficient.

4.2 Example 2: Uniform 2D Lattice

The second test problem is a 2D lattice with uniform weights. As discussed in [58], it is similar to an isotropic elliptic PDE problem. Here we let the 2D lattice be square and use h to denote the number of nodes in every row or column, then the total size of the coefficient matrix A in the singular linear systems (5) is $N = h^2$.

In this experiments, let the preconditioner be $M = U + U^{\mathrm{T}}$ as given in the first test problem. To show the effectiveness of this preconditioner, Figure 3 depicts the eigenvalue distributions of the coefficient matrix A and those of the preconditioned matrix $M^{-1}A$ when $h = 50$, i.e., $N = 2\,500$. Observe that the application of the preconditioner M has improved the eigenvalue properties of the original matrix A, which explains why the preconditioned BiCR and BiCG algorithms are superior to the unpreconditioned BiCR and BiCG algorithms.

h	N	nnz(A)	BiCG	BiCR	PBiCG	PBiCR
50	2 500	12 300	17 (0.016)	16 (0.015)	10 (0.010)	9 (0.008)
100	10 000	49 600	18 (0.062)	16 (0.053)	11 (0.038)	9 (0.030)
200	40 000	199 200	18 (0.271)	16 (0.254)	11 (0.181)	9 (0.166)
300	90 000	448 800	19 (0.649)	16 (0.580)	11 (0.423)	9 (0.396)
400	160 000	798 400	19 (1.193)	16 (1.036)	11 (0.903)	9 (0.725)
500	250 000	1 248 000	19 (1.863)	16 (1.703)	11 (1.213)	9 (1.001)

Table 2. IT and CPU (in brackets) of these methods for Example 2.

Numerical results for this test problem are provided in Table 2. Again, it seems that both the preconditioned BiCG and BiCR methods run faster than the unpreconditioned BiCG and BiCR methods, while the PBiCR method outperforms the PBiCG method both in terms of iteration steps and computational time. Furthermore, these methods are quite stable for this uniform 2D lattice problem, even though there exists some small changes in the sense of iteration counts when the size of this problem is not large.

Convergence histories of these four algorithms for this example have been given in Figure 4 when $h = 300$ and $h = 500$, respectively. From an intuitive point, it is easy to find that the PBiCR and PBiCG methods converges faster than the BiCR and BiCG methods. Particularly, the PBiCR method performs the best.

4.3 Example 3: Two-Queue Overflow Networks

The last test problem is the two-queue overflow networks with the customer arrival rate and service rate of the servers being λ_i and μ_i $(i = 1, 2)$, respectively. Suppose the number of servers is s_i and the waiting space is $L_i - s_i - 1$ $(i = 1, 2)$. The

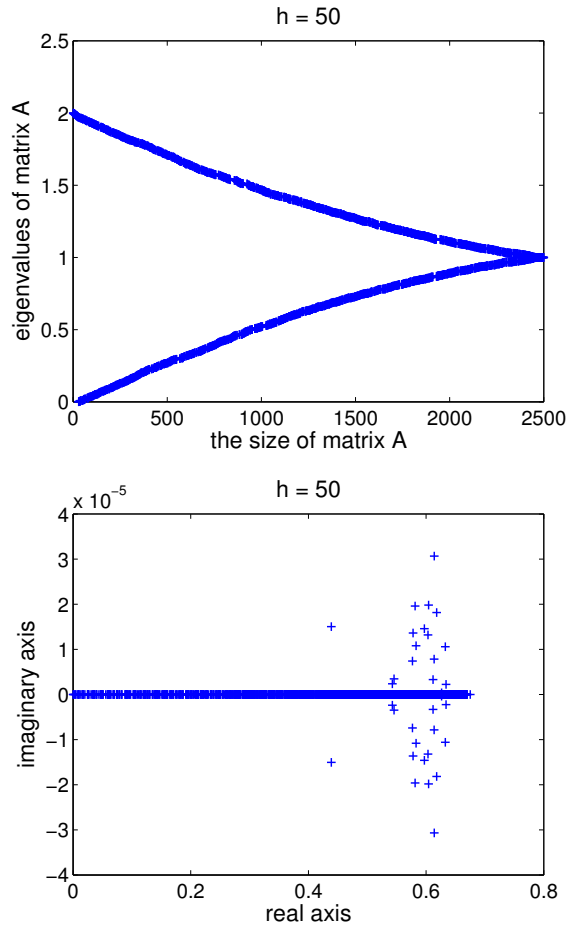

Figure 3. Comparison of the eigenvalue distributions between the original matrix A (top) and the preconditioned matrix $M^{-1}A$ (bottom) for Example 2 when $h = 50$

queueing discipline is first-come-first-served. Specifically, we allow the overflow of customers to occur from queue 2 to queue 1 when queue 2 is full and there is still a waiting space in queue 1. The graph of the two-queue overflow networks and the form of its generator matrix have been presented in a few papers; see, e.g., [43, 49].

For simplicity, in this test, we have set $s_1 = s_2 = 1$, $\lambda_1 = \lambda_2 = 1$ and $\mu_1 = \mu_2 = 1$. Let the incomplete LU factorization of the coefficient matrix A be

$$A = LU + R,$$

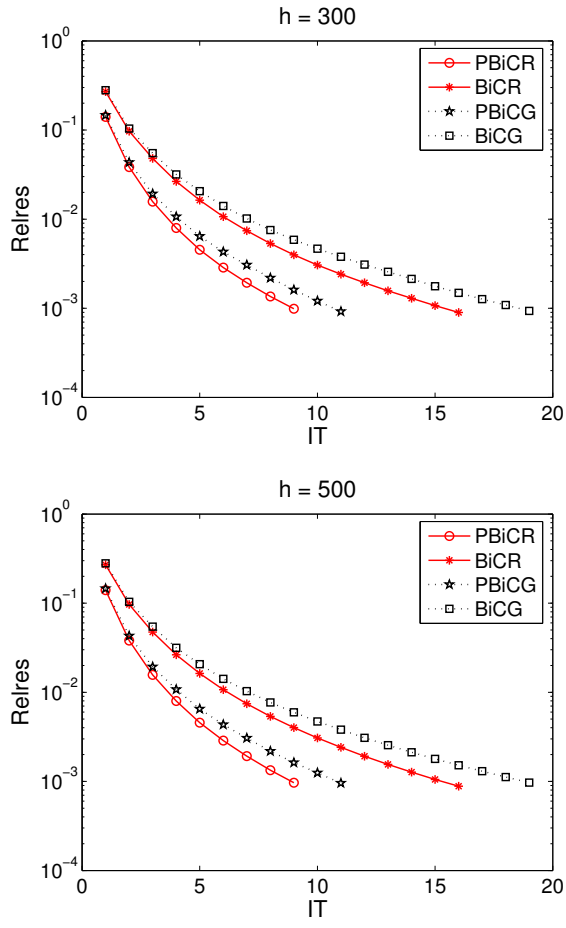

Figure 4. Convergence histories of these methods for Example 2

where $L \in \mathbb{R}^{n\times n}$ and $U \in \mathbb{R}^{n\times n}$ are the lower and upper triangular matrices, respectively. Then the preconditioner for this test problem is chosen to be $M = LU$ with the drop tolerance of the incomplete LU factorization is 0.01. In order to illustrate the efficiency of the preconditioner M, Figure 5 shows the eigenvalue distributions of the original matrix A and those of the preconditioned matrix $M^{-1}A$ when $L_1 = 32$ and $L_2 = 32$, i.e., $N = 1\,024$. It is clear that the eigenvalue distributions of the preconditioned matrix $M^{-1}A$ is desirable, most of them are closely clustered around 1.

Table 3 has supplied numerical results for this test problem. By a close look at the results in Table 3, we find that the iteration steps and computational time needed by the preconditioned BiCR and BiCG methods are less than those of the

L_1	L_2	N	nnz(A)	BiCG	BiCR	PBiCG	PBiCR
32	32	11 024	41 992	22 (0.025)	16 (0.011)	2 (0.010)	2 (0.009)
64	64	41 096	201 224	22 (0.125)	16 (0.111)	2 (0.104)	2 (0.100)
128	128	161 384	811 408	22 (0.181)	16 (0.154)	2 (0.102)	2 (0.101)
128	256	321 768	1631 327	22 (0.295)	16 (0.202)	2 (0.126)	2 (0.123)
256	256	651 536	3261 656	22 (0.533)	16 (0.431)	2 (0.355)	2 (0.349)
256	512	1311 072	6541 335	22 (1.093)	16 (0.856)	2 (0.657)	2 (0.648)

Table 3. IT and CPU (in brackets) of these methods for Example 3

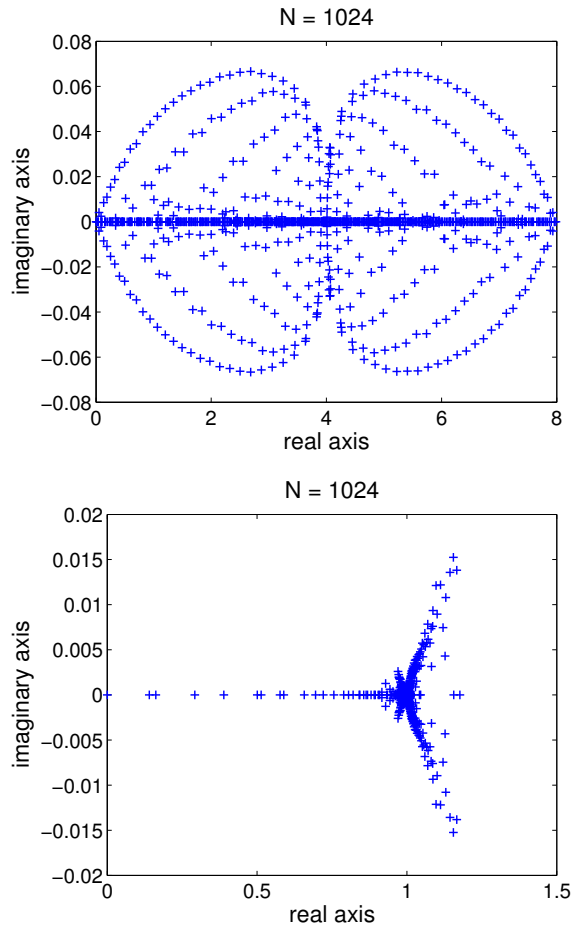

Figure 5. Comparison of the eigenvalue distributions between the original matrix A (top) and the preconditioned matrix $M^{-1}A$ (bottom) for Example 3 when $L_1 = 32$, $L_2 = 32$

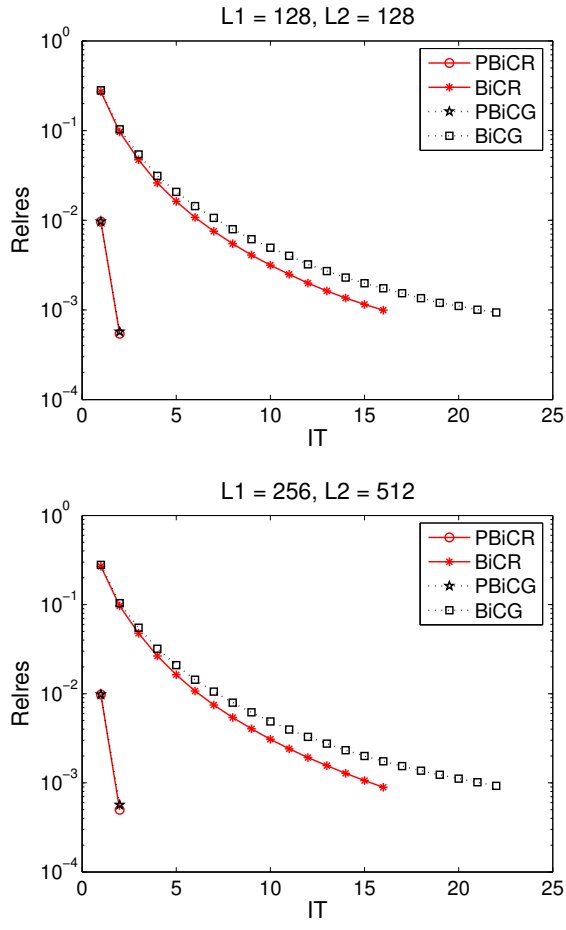

Figure 6. Convergence histories of these methods for Example 3

unpreconditioned BiCR and BiCG methods. Moreover, the data in Table 3 indicate the efficiency and feasibility of Algorithms 1–4 again, since the number of iterations required by them are almost unchanged.

Additionally, an intuitive understanding can be obtained from Figure 6, where the convergence histories of these four methods for this test problem are plotted when $L_1 = 128, L_2 = 128$ and $L_1 = 256, L_2 = 512$, respectively. It is clear that the BiCR and BiCG methods with one suitable preconditioner are not bad choices.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have described two conjugate direction methods, BiCR and BiCG, and extended their applications to solve the stationary probability distribution for Markov chain modellings. Some properties of the BiCR and BiCG methods are considered, and the corresponding preconditioned versions are provided in Algorithms 3 and 4 with the aim of easily transforming them into a computer code. Numerical experiments have been made on several practical Markov chain problems. Figures 1, 3 and 5 have indicated that eigenvalue properties of the original matrix A are able to be improved by using suitable preconditioners. Numerical results in Tables 1–3 have shown that the required iteration steps and computational time by the preconditioned BiCR and BiCG methods are less than those of unpreconditioned BiCR and BiCG methods. Furthermore, the convergence histories of these four algorithms have been compared in Figures 2, 4 and 6, which illustrate their efficiency and feasibility from an intuitive point.

Note that we do not consider other methods in this paper since our aim here is to indicate the feasibility of the BiCG and BiCR methods in solving singular linear system (5) which arise from Markov chain problems. Actually, from the previous analysis, we know that both the BiCG and BiCR methods possess the short-term recurrences, which may make them outperform certain other iterative methods like the classical Jacobi and Gauss-Seidel, and so on. We leave these numerical comparisons among different methods for the future research. Moreover, it would be interesting to extend these proposed methods to other sophisticated Markov chain problems [45, 46, 51, 52].

Acknowledgment

The authors would like to express their great thankfulness to the referees for their constructive comments and helpful suggestions in revising the paper.

REFERENCES

[1] Hestenes, M. R.—Stiefel, E. L.: Methods of Conjugate Gradients for Solving Linear Systems. J. Res. Nat. Bureau Standards, Vol. 49, 1952, pp. 409–436.

[2] Paige, C. C.—Saunders, M. A.: Solution of Sparse Indefinite Systems of Linear Equations. SIAM J. Numer. Anal., Vol. 12, 1975, pp. 617–629.

[3] Stiefel, E.: Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme. Comment. Math. Helv., Vol. 29, 1955, pp. 157–179.

[4] Saad, Y.—Schultz, M. H.: GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems. SIAM J. Sci. Stat. Comput., Vol. 7, 1986, pp. 856–869.

[5] Eisenstat, S. C.—Elman, H. C.—Schultz, M. H.: Variational Iterative Methods for Nonsymmetric Systems of Linear Equations. SIAM J. Sci. Numer Anal., Vol. 20, 1983, pp. 345–537.

[6] Fletcher, R.: Conjugate Gradient Methods for Indefinite Systems. Lecture Notes in Math., Springer-Verlag, Berlin, Heidelberg, New York, Vol. 506, 1976, pp. 73–89.

[7] Lanczos, C.: Solution of Systems of Linear Equations by Minimized Iterations. J. Res. Nat. Bureau Standards, Vol. 49, 1951, pp. 33–53.

[8] Sonneveld, P.: CGS, A Fast Lanczos-Type Solver for Nonsymmetric Linear Systems. SIAM J. Sci. Stat. Comput., Vol. 10, 1989, pp. 36–52.

[9] Van der Vorst, H. A.: Bi-CGSTAB: A Fast and Smoothly Converging Variant of Bi-CG for the Solution of Nonsymmetric Linear Systems. SIAM J. Sci. Stat. Comput., Vol. 13, 1992, pp. 631–644.

[10] Freund, R. W.—Nachtigal, N. M.: QMR: A Quasi-Minimal Residual Method for Non-Hermitian Linear Systems. Numer. Math., Vol. 60, 1991, pp. 315–339.

[11] Saad, Y.: Iterative Methods for Sparse Linear Systems. SIAM, Philadelphia, PA, 2003.

[12] Tanabe, K.: The Conjugate Gradient Method for Computing all the Extremal Stationary Probability Vectors of a Stochastic Matrix. Annals of the Institute of Statistical Mathematics, Vol. 37, 1985, pp. 173–187.

[13] Kammerer, W. J.—Nashed, M. Z.: On the Convergence of the Conjugate Gradient Method for Singular Linear Operator Equations. SIAM J. Numer. Anal., Vol. 9, 1972, pp. 165–181.

[14] Kaasschieter, E. F.: Preconditioned Conjugate Gradients for Solving Singular Systems. J. Comput. Appl. Math., Vol. 24, 1998, pp. 265–275.

[15] Fischer, B.—Hanke, M.—Hochbruck, M.: A Note on Conjugate-Gradient Type Methods for Indefinite and/or Inconsistent Linear Systems. Numer. Algo., Vol. 11, 1996, pp. 181–189.

[16] Freund, R. W.: A Transpose-Free Quasi-Minimal Residual Algorithm for Non-Hermitian Linear Systems. SIAM J. Sci. Comput., Vol. 14, 1993, pp. 425–448.

[17] Freund, R. W.—Hochbruck, M.: On the Use of Two QMR Algorithms for Solving Singular Systems and Applications in Markov Chain Modeling. Numer. Linear Algebra Appl., Vol. 1, 1994, pp. 403–420.

[18] Wei, Y.—Wu, H.: Convergence Properties of Krylov Subspace Methods for Singular Linear Systems with Arbitrary Index. J. Comput. Appl. Math., Vol. 114, 2000, pp. 305–318.

[19] Abe, K.—Ogata, H.—Sugihara, M.—Zhang, S. L.—Mitsui, T.: Convergence Theory of the CR Method for Linear Singular Systems. Transactions of the Japan Society for Industrial and Applied Mathematics, Vol. 9, 1999, pp. 1–13 (in Japanese).

[20] Hayami, K.: On the Behaviour of the Conjugate Residual Method for Singular Systems. In: Shi, Z.-C., Kawarada, H. (Eds.): Proceedings of Fifth China-Japan Seminar on Numerical Mathematics, Shanghai, 2000, Science Press, Beijing/New York, 2002, pp. 117–126.

[21] Hayami, K.: On the Convergence of the Conjugate Residual Method for Singular Systems. Transactions of the Japan Society for Industrial and Applied Mathematics, Vol. 13, 2003, pp. 1–33 (in Japanese).

[22] Zhang, S. L.—Oyanagi, Y.—Sugihara, M.: Necessary and Sufficient Conditions for the Convergence of Orthomin(k) on Inconsistent Linear Systems. Numer. Math., Vol. 87, 2000, pp. 391–405.

[23] Sidi, A.: DGMRES: A GMRES-Type Algorithm for Drazin-Inverse Solution of Singular Nonsymmetric Linear Systems. Linear Algebra Appl., Vol. 335, 2001, pp. 189–204.

[24] Hayami, K.—Sugihara, M.: A Geometric View of Krylov Subspace Methods on Singular Systems. Numer. Linear Algebra Appl., Vol. 18, 2011, pp. 449–469.

[25] Hayami, K.—Sugihara, M.: On the Convergence of the GCR(k) Method for Singular Systems. NII Technical Reports, NII-2004-009E, National Institute of Informatics, Tokyo, 2004, pp. 1–24.

[26] Schneider, O.: Krylov Subspace Methods and Their Generalizations for Solving Singular Operator Equations with Applications to Continuous Time Markov Chains. Doctoral Dissertation, Fakultät für Mathematik und Informatik der Technischen Universität Bergakademie Freiberg, November 2005.

[27] Brown, P.—Walker, H. F.: GMRES on (Nearly) Singular Systems. SIAM J. Matrix Anal. Appl., Vol. 18, 1997, pp. 37–51.

[28] Du, X. D.—Szyld, D. B.: Inexact GMRES for Singular Linear Systems. BIT Numerical Mathematics, Vol. 48, 2008, pp. 511–531.

[29] Smoch, L.: Some Results about GMRES in the Singular Case. Numer. Algo., Vol. 22, 1999, pp. 193–212.

[30] Ipsen, I. C. F.—Meyer, C. D.: The Idea Behind Krylov Methods. American Mathematical Monthly, Vol. 105, 1998, pp. 889–899.

[31] Reichel, L.—Ye, Q.: Breakdown-Free GMRES for Singular Systems. SIAM J. Matrix Anal. Appl., Vol. 26, 2005, pp. 1001–1021.

[32] Calvetti, D.—Lewis, B.—Reichel, L.: GMRES-Type Methods for Inconsistent Systems. Linear Algebra Appl., Vol. 316, 2000, pp. 157–169.

[33] Sogabe, T.—Sugihara, M.—Zhang, S. L.: An Extension of the Conjugate Residual Method to Nonsymmetric Linear Systems. J. Comput. Appl. Math., Vol. 266, 2009, pp. 103–113.

[34] Van der Vorst, H. A.: Iterative Krylov Methods for Large Linear Systems. Cambridge University Press, Cambridge, 2003, pp. 97–98.

[35] Broyden, C. G.: A Comparison of Three Basic Conjugate Direction Methods. Numer. Linear Algebra Appl., Vol. 3, 1996, pp. 473–489.

[36] Stewart, W. J.: Probability, Markov Chains, Queues, and Simulation. Princeton University Press, 2009.

[37] Berman, A.—Plemmons, R.: Nonnegative Matrices in Mathematical Sciences. SIAM, Philadelphia, 1994.

[38] Senta, E.: Non-Negative Matrices. John Wiley, New York, 1973.

[39] Page, L.—Brin, S.—Motwani, R.—Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999–0120, Computer Science Department, Stanford, 1999.

[40] Broder, A. Z.—Lempel, R.—Maghoul, F.—Pedersen, J.: Efficient PageRank Approximation via Graph Aggregation. Inf. Retrieval, Vol. 9, 2006, pp. 123–138.

[41] Kamvar, S. D.—Haveliwala, T. H.—Manning, C. D.—Golub, G. H.: Extrapolation Methods for Accelerating PageRank Computations. Proceedings of the Twelfth International World Wide Web Conference, 2003.

[42] Langville, A. N.—Meyer, C. D.: A Survey of Eigenvector Methods for Web Information Retrieval. SIAM Review, Vol. 47, 2005, pp. 135–161.

[43] Kaufman, L.: Matrix Methods for Queueing Problems. SIAM J. Sci. Stat. Comput., Vol. 4, 1982, pp. 525–552.

[44] Buchholz, P.: A Class of Hierarchical Queueing Networks and Their Analysis. Queueing Systems, Vol. 15, 1994, pp. 59–80.

[45] Gelenbe, E.: Product-Form Queueing Networks with Negative and Positive Customers. Journal of Applied Probability, Vol. 28, 1991, pp. 656–663.

[46] Ching, W. K.: Iterative Methods for Queuing Systems with Batch Arrivals and Negative Customers. BIT, Vol. 43, 2003, pp. 285–296.

[47] Plateau, B.—Atif, K.: Stochastic Automata Network for Modelling Parallell Systems. IEEE Transactions on Software Engineering, Vol. 17, 1991, pp. 1093–1108.

[48] Stewart, W. J.—Atif, K.—Plateau, B.: The Numerical Solution of Stochastic Automata Networks. Euro. J. Operat. Res., Vol. 86, 1995, pp. 503–525.

[49] Chan, R.—Ching, W. K.: Circulant Preconditioners for Stochastic Automata Networks. Numer. Math., Vol. 87, 2000, pp. 35–57.

[50] Buzacott, J.—Shanthikumar, J.: Stochastic Models of Manufacturing Systems. Prentice-Hall International Editions, New Jersey, 1993.

[51] Meier-Hellstern, K.: The Analysis of a Queue Arising in Overflow Models. IEEE Trans. Commun., Vol. 37, 1989, pp. 367–372.

[52] Heffes, H.—Lucantoni, D.: A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance. IEEE J. Select. Areas Commun., Vol. 4, 1986, pp. 856–868.

[53] Young, H.—Byung, B.—Chong, K.: Performance Analysis of Leaky-Bucket Bandwidth Enforcement Strategy for Bursty Traffics in an ATM Network. Comput. Net. ISDN Syst., Vol. 25, 1992, pp. 295–303.

[54] Philippe, B.—Saad, Y.—Stewart, W. J.: Numerical Methods in Markov Chain Modeling. Operations Research, Vol. 40, 1992, pp. 1156–1179.

[55] Saad, Y.: Preconditioned Krylov Subspace Methods for the Numerical Solution of Markov Chains. In: Stewart, W. J. (Ed.): Computations with Markov Chains, Kluwer Academic Publishers, 1995, pp. 49–64.

[56] Latouche, G.—Ramaswami, V.: Introduction to Matrix Analytic Methods in Stochastic Modeling. ASA-SIAM Series on Statistics and Applied Probability, 1999.

[57] Isensee, C.—Horton, G.: A Multi-Level Method for the Steady State Solution of Discrete-Time Markov Chains. Proceedings of the 2nd Balkan Conference in Informatics, Ohrid, Macedonia, November 2005, pp. 413–420.

[58] De Sterck, H.—Manteuffel, T. A.—McCormick, S. F.—Miller, K.—Pearson, J.—Ruge, J.—Sanders, G.: Smoothed Aggregation Multigrid for Markov Chains. SIAM J. Sci. Comput., Vol. 32, 2010, pp. 40–61.

[59] De Sterck, H.—Henson, V. E.—Sanders, G.: Multilevel Aggregation Methods for Small-World Graphs with Application to Random-Walk Ranking. Computing and Informatics, Vol. 30, 2011, pp. 225–246.

[60] Krieger, U. R.: On a Two-Level Multigrid Solution Method for Finite Markov Chains. Linear Algebra Appl., Vol. 223, 1995, pp. 415–438.

[61] Horak, V. C.—Berka, T.—Vajteršic, M.: Parallel Classification with Two-Stage Bagging Classifiers. Computing and Informatics, Vol. 32, 2013, pp. 661–677.

[62] Bini, D. A.—Latouche, G.—Meini, B.: Numerical Methods for Structured Markov Chains. Oxford University Press, London, 2005.

[63] Benzi, M.—Uçar, B.: Block Triangular Preconditioners for M-Matrices and Markov Chains. Electronic Trans. Numerical Analysis, Vol. 26, 2007, pp. 209–227.

[64] Stewart, W. J.—Cao, W. L.: Iterative Aggregation/Disaggregation Techniques for Nearly Uncoupled Markov Chains. J. ACM, Vol. 32, 1985, pp. 702–719.

[65] Schweitzer, P. J.—Kindle, K. W.: An Iterative Aggregation-Disaggregation Algorithm for Solving Linear Equations. Appl. Math. Comput., Vol. 18, 1986, pp. 313–354.

[66] Marek, I.—Mayer, P.: Convergence Analysis of an Iterative Aggregation/Disaggregation Method for Computing Stationary Probability Vector of Stochastic Matrices. Numer. Linear Algebra Appl., Vol. 5, 1998, pp. 253–274.

[67] Marek, I.—Szyld, D. B.: Algebraic Schwarz Methods for the Numerical Solution of Markov Chains. Linear Algebra Appl., Vol. 386, 2004, pp. 67–81.

[68] Bru, R.—Pedroche, F.—Szyld, D. B.: Additive Schwarz Iterations for Markov Chains. SIAM J. Matrix Anal. Appl., Vol. 27, 2005, pp. 445–458.

[69] Borovac, S.: A Graph Based Approach to the Convergence of One Level Schwarz Iterations for Singular M-Matrices and Markov Chains. SIAM J. Matrix Anal. Appl., Vol. 30, 2008, pp. 1371–1391.

Chun Wen received her Ph.D. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, Sichuan, China. She is currently Associate Professor in the School of Mathematical Sciences, UESTC. Her research interests include iterative methods of linear systems and saddle point problems, AMG methods with application to Markov chains, and preconditioning technologies.

Ting-Zhu Huang received his Ph.D. degree in computational mathematics from the Department of Mathematics, Xi'an Jiaotong University, Xi'an, China. He is currently Professor in the School of Mathematical Sciences, UESTC. He is currently an editor of The Scientific World Journal, Advances in Numerical Analysis, J. Appl. Math., J. Pure and Appl. Math.: Adv. and Appl., J. Electronic Sci. and Tech. of China, etc. His current research interests include numerical linear algebra and scientific computation with applications, iterative methods of linear systems, numerical algorithms for image processing, preconditioning technologies, and matrix analysis with applications, etc.

Tomohiro Sogabe received his Ph.D. degree in engineering from the University of Tokyo, Japan. He is currently Associate Professor in the Department of Information Science and Technology, Aichi Prefectural University. His area of research is numerical linear algebra and scientific computing. He is a member of Editorial Board, Transactions of JSIAM.