COMMENTS ON A SHIFTED CYCLIC REDUCTION ALGORITHM FOR QUASI-BIRTH-DEATH PROBLEMS*

CHUN-HUA GUO†

Abstract. A shifted cyclic reduction algorithm has been proposed by He, Meini, and Rhee [*SIAM J. Matrix Anal. Appl.*, 23 (2001), pp. 673–691] for finding the stochastic matrix G associated with discrete-time quasi-birth-death (QBD) processes. We point out that the algorithm has quadratic convergence even for null recurrent QBDs. We also note that the approximations (to the matrix G) obtained by their algorithm are always stochastic when they are nonnegative.

Key words. matrix equations, minimal nonnegative solution, Markov chains, cyclic reduction, iterative methods, convergence rate

AMS subject classifications. 15A24, 15A51, 60J10, 60K25, 65H05

1. Introduction. A discrete-time quasi-birth-death process (QBD) is a Markov chain with state space $\{(i,j)\,|\,i \geq 0, 1 \leq j \leq m\}$, which has a transition probability matrix of the form

$$P = \begin{pmatrix} C_0 & C_1 & 0 & 0 & \cdots \\ A_0 & A_1 & A_2 & 0 & \cdots \\ 0 & A_0 & A_1 & A_2 & \cdots \\ 0 & 0 & A_0 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where C_0, C_1, A_0, A_1, and A_2 are $m \times m$ nonnegative matrices such that P is stochastic. In particular, $(A_0 + A_1 + A_2)e = e$, where e is the column vector with all components equal to one. The matrix P is also assumed to be irreducible.

We assume that $A = A_0 + A_1 + A_2$ is irreducible. Thus, there exists a unique vector $\alpha > 0$ with $\alpha^T e = 1$ and $\alpha^T A = \alpha^T$. The vector α is called the stationary probability vector of A. The QBD is positive recurrent if $\alpha^T A_0 e > \alpha^T A_2 e$, and null recurrent if $\alpha^T A_0 e = \alpha^T A_2 e$.

The minimal nonnegative solution G of the matrix equation

$$G = A_0 + A_1 G + A_2 G^2 \tag{1.1}$$

plays an important role in the study of the QBD (see [8]). We will also need the equation

$$F = A_2 + A_1 F + A_0 F^2, \tag{1.2}$$

and let F be its minimal nonnegative solution. It is well known (see [8], for example) that if the QBD is positive recurrent, then G is stochastic and F is substochastic with spectral radius $\rho(F) < 1$; if the QBD is null recurrent, then G and F are both stochastic.

Recently, a shift technique has been introduced in [6] to a cyclic reduction (CR) algorithm (see [3]) for finding the matrix G in the positive recurrent case, assuming that the only eigenvalue of G on the unit circle is the simple eigenvalue 1. In this note we will make some comments on that interesting paper.

*This work was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

†Department of Mathematics and Statistics, University of Regina, Regina, SK S4S 0A2, Canada (chguo@math.uregina.ca).

2. Comments. The shift technique introduced in [6] is $H = G - eu^T$, where $u > 0$ and $u^T e = 1$. Then the eigenvalues of H are those of G except that in H the eigenvalue 1 of G is replaced by 0, and H is a solution of the new equation

$$H = B_0 + B_1 H + B_2 H^2, \tag{2.1}$$

where

$$B_0 = A_0(I - eu^T), \quad B_1 = A_1 + A_2 eu^T, \quad B_2 = A_2. \tag{2.2}$$

The shifted CR algorithm is obtained in [6] by applying the CR algorithm to (2.1). For positive recurrent QBDs, it is shown in [6] that the convergence of the shifted CR algorithm is quadratic and faster than that of the CR algorithm, provided that no breakdown occurs. Here we point out that the same is true for null recurrent QBDs. This is a very important feature of the shift technique. Without using the shift technique, all previous methods for finding the matrix G have only linear or sublinear convergence for null recurrent QBDs. For example, the convergence of the Latouche–Ramaswami (LR) algorithm [7] is linear with rate 1/2 for null recurrent QBDs (see [5]). Since the CR algorithm and the LR algorithm are closely related (see [2]), the convergence of the CR algorithm is also linear with rate 1/2 for null recurrent QBDs. Once we have shown that the shift technique recovers quadratic convergence for the CR algorithm in the null recurrent case, the same is true for the LR algorithm.

Some work is needed to justify our claim about the shifted CR algorithm for null recurrent QBDs.

Let

$$A(\lambda) = -A_0 + (I - A_1)\lambda - A_2\lambda^2$$

be the matrix polynomial corresponding to (1.1), and

$$B(\lambda) = -B_0 + (I - B_1)\lambda - B_2\lambda^2$$

be the matrix polynomial associated with (2.1). We first point out that there is a simple proof for the following generalization of Theorem 3.1 in [6].

LEMMA 2.1. *The zeros of* $\det(B(\lambda))$ *are obtained from the zeros of* $\det(A(\lambda))$ *by replacing one zero* 1 *by* 0.

Proof. Since

$$A(\lambda) = (I - A_1 - A_2 G - \lambda A_2)(\lambda I - G),$$

$$B(\lambda) = (I - B_1 - B_2 H - \lambda B_2)(\lambda I - H),$$

and

$$\begin{aligned} I - B_1 - B_2 H - \lambda B_2 &= I - (A_1 + A_2 eu^T) - A_2(G - eu^T) - \lambda A_2 \\ &= I - A_1 - A_2 G - \lambda A_2, \end{aligned}$$

the assertion follows immediately. ∎

Note that $\det(A(\lambda))$ has two zeros 1 for null recurrent QBDs, as seen from the following special case of Theorem 4 in [4].

LEMMA 2.2. *Assume that* $\det(A(\lambda)) \neq 0$ *if* $|\lambda| = 1, \lambda \neq 1$. *Then*

(1) *If the QBD is positive recurrent, then* $\det(A(\lambda))$ *has* $m-1$ *zeros inside the unit circle, one zero* 1*, and* m *zeros outside the unit circle (zeros at infinity are added, if the degree of* $\det(A(\lambda))$ *is less than* $2m$*).*

(2) *If the QBD is null recurrent, then* $\det(A(\lambda))$ *has* $m-1$ *zeros inside the unit circle, two zeros* 1*, and* $m-1$ *zeros outside the unit circle (zeros at infinity are added, if the degree of* $\det(A(\lambda))$ *is less than* $2m$*).*

We note that the assumption in Lemma 2.2 is equivalent to our earlier assumption that the only eigenvalue of G on the unit circle is the simple eigenvalue 1 (see [4]).

COROLLARY 2.3. *If the QBD is positive recurrent then* $\det(B(\lambda))$ *has* m *zeros inside the unit circle and no zeros on the unit circle; if the QBD is null recurrent then* $\det(B(\lambda))$ *has* m *zeros inside the unit circle, one (simple) zero* 1 *on the unit circle, and* $m-1$ *zeros outside the unit circle.*

When the QBD is positive recurrent, $u^TFe<1$ and $I-eu^TF$ is nonsingular (see [6]). The following result plays a crucial role in [6] for the convergence analysis of the shifted CR algorithm.

LEMMA 2.4. [6] *When the QBD is positive recurrent,*

$$K=(I-eu^TF)F(I-eu^TF)^{-1} \tag{2.3}$$

is a solution of

$$K=B_2+B_1K+B_0K^2. \tag{2.4}$$

When the QBD is null recurrent, we have $Fe=e$. Thus, $(I-eu^TF)e=0$ and $I-eu^TF$ is singular. The question then arises as to whether the norm of the matrix K in (2.3) will get arbitrarily large when the QBD becomes nearly null recurrent. As noted in [6], there is a K-dependent operator norm $\|\cdot\|_K$ such that $\|K\|_K=\rho(K)=\rho(F)<1$. However, the norm $\|\cdot\|_K$ would be drastically different from practically useful norms like $\|\cdot\|_\infty$ as the QBD becomes nearly null recurrent, if $\|K\|_\infty$ couldn't be bounded independent of the nearness to null recurrence. We have the following positive result in this regard. The result will also be the basis for proving quadratic convergence of the shifted CR algorithm in the null recurrent case.

LEMMA 2.5. *If the QBD is positive recurrent, then for the matrix* K *in* (2.3)

$$\|K\|_\infty<3+\frac{2}{\min_{1\le i\le m}u_i},$$

where u_i *is the ith component of* u*. In particular,* $\|K\|_\infty<3+2m$ *if* $u=\frac{1}{m}e$*.*

Proof. By the Sherman–Morrison–Woodbury formula,

$$(I-eu^TF)^{-1}=I+\frac{1}{1-u^TFe}eu^TF=I+\frac{1}{u^T(e-Fe)}eu^TF.$$

Thus,

$$K=(I-eu^TF)F+\frac{1}{u^T(e-Fe)}(I-eu^TF)Feu^TF.$$

Note that

$$\begin{aligned}(I-eu^TF)Feu^TF&=(I-eu^TF)eu^TF-(I-eu^TF)(e-Fe)u^TF\\&=eu^T(e-Fe)u^TF-(I-eu^TF)(e-Fe)u^TF.\end{aligned}$$

Therefore,

$$K = (I - eu^T F)F + eu^T F - \frac{1}{u^T(e - Fe)}(I - eu^T F)(e - Fe)u^T F.$$

It follows that

$$\|K\|_\infty \le \|I - eu^T F\|_\infty \|F\|_\infty + \|eu^T F\|_\infty + \|I - eu^T F\|_\infty \|u^T F\|_\infty \frac{\|e - Fe\|_\infty}{u^T(e - Fe)}.$$

Since $Fe \le e$, $u^T Fe < 1$ and $eu^T Fe < e$, we have $\|F\|_\infty \le 1$, $\|u^T F\|_\infty < 1$ and $\|eu^T F\|_\infty < 1$. Thus, $\|I - eu^T F\|_\infty < 2$ and

$$\|K\|_\infty < 3 + \frac{2}{\min_{1\le i\le m} u_i}.$$

This completes the proof. ∎

For the null recurrent case, the role of Lemma 2.4 will be assumed by the following result.

THEOREM 2.6. *If the QBD is null recurrent, then* (2.4) *has a solution K having one eigenvalue* 1 *and $m-1$ eigenvalues inside the unit circle.*

Proof. Since the QBD is irreducible, $A_2 \ne 0$. Suppose that $A_2(i,j)$, the (i,j) element of A_2, is positive. For any ϵ with $0 < \epsilon < A_2(i,j)$, define

$$A_0(\epsilon) = A_0, \quad A_1(\epsilon) = A_1 + \epsilon E_{ij}, \quad A_2(\epsilon) = A_2 - \epsilon E_{ij},$$

where E_{ij} is the matrix with one in the (i,j) position and zeros elsewhere. Since $\alpha^T A_0(\epsilon)e > \alpha^T A_2(\epsilon)e$, where α is the stationary probability vector of $A = A_0 + A_1 + A_2 = A_0(\epsilon) + A_1(\epsilon) + A_2(\epsilon)$, the QBD corresponding to $(A_0(\epsilon), A_1(\epsilon), A_2(\epsilon))$ is positive recurrent. We now define

$$B_0(\epsilon) = A_0(\epsilon)(I - eu^T), \quad B_1(\epsilon) = A_1(\epsilon) + A_2(\epsilon)eu^T, \quad B_2(\epsilon) = A_2(\epsilon)$$

and let F_ϵ be the minimal nonnegative solution of

$$F = A_2(\epsilon) + A_1(\epsilon)F + A_0(\epsilon)F^2.$$

Thus, $\rho(F_\epsilon) < 1$. Moreover, $K_\epsilon = (I - eu^T F_\epsilon)F_\epsilon(I - eu^T F_\epsilon)^{-1}$ is a solution of

$$K = B_2(\epsilon) + B_1(\epsilon)K + B_0(\epsilon)K^2$$

by Lemma 2.4. Let the sequence $\{\epsilon_n\}$ be such that $0 < \epsilon_n < A_2(i,j)$ and $\lim \epsilon_n = 0$. Since the sequence $\{K_{\epsilon_n}\}$ is bounded by Lemma 2.5, it has a limit point K. It is clear that this matrix K is a solution of (2.4). Since $\rho(K_{\epsilon_n}) < 1$, we have $\rho(K) \le 1$. Since the zeros of $\det(\hat{B}(\lambda))$, where

$$\hat{B}(\lambda) = -B_2 + (I - B_1)\lambda - B_0\lambda^2,$$

are the reciprocals of the zeros of $\det(B(\lambda))$ and the eigenvalues of K are part of the zeros of $\det(\hat{B}(\lambda))$, we know from Corollary 2.3 that K has $m-1$ eigenvalues inside the unit circle and one eigenvalue 1. ∎

The shifted CR algorithm generates a sequence $\widehat{B}_1^{(n)}$ (if no breakdown occurs) and approximations $\tilde{H}_n$ to the matrix H is obtained by $\tilde{H}_n = (I - \widehat{B}_1^{(n)})^{-1}B_0$ (see

[6]). Approximations $\tilde{G}_n$ to the matrix G can be obtained using $\tilde{G}_n = \tilde{H}_n + eu^T$. It is noted in [6] that we also have $\tilde{G}_n = (I - \widehat{B}_1^{(n)})^{-1} A_0$.

For the null recurrent case, the spectral properties of the matrix K in Theorem 2.6 are crucial to show the quadratic convergence of the sequence $\{\tilde{G}_n\}$. Once Theorem 2.6 is proved, quadratic convergence follows from known results.

Let K be the solution of (2.4) given by Lemma 2.4 for the positive recurrent case and given by Theorem 2.6 for the null recurrent case. We have from the discussions in [6] or from Theorem 16 and Remark 17 of [1] that

$$\limsup_{n\to\infty} \sqrt[2^n]{\|\tilde{G}_n - G\|_\infty} \le \rho(K)\rho(H) = \rho(F)\rho(H) < 1. \tag{2.5}$$

In particular, $\tilde{G}_n$ converges to G quadratically for both positive recurrent and null recurrent QBDs. If we apply the CR algorithm directly to (1.1), the approximations G_n for G are such that

$$\limsup_{n\to\infty} \sqrt[2^n]{\|G_n - G\|_\infty} \le \rho(F)\rho(G) \le 1. \tag{2.6}$$

Thus, the convergence of $\{G_n\}$ is slower than that of $\{\tilde{G}_n\}$. One good thing about the sequence $\{G_n\}$ is that it is monotonically increasing to G (see [3]). Thus, $\|G_n e - e\|_\infty = \|(G_n - G)e\|_\infty = \|G_n - G\|_\infty$. So, the actual error $\|G_n - G\|_\infty$ can be obtained easily even though G is not known. For the sequence $\{\tilde{G}_n\}$, the actual error $\|\tilde{G}_n - G\|_\infty$ cannot be obtained in this way. In fact, since $B_0 e = A_0(I - eu^T)e = 0$, we have $\tilde{H}_n e = 0$ and $\tilde{G}_n e = e$ for each $n \ge 0$. Therefore, the matrices $\tilde{G}_n$ are stochastic when they are nonnegative, and we always have $\|\tilde{G}_n e - e\|_\infty = 0$ (in exact arithmetic) no matter how large $\|\tilde{G}_n - G\|_\infty$ is. Nevertheless, computing the values $\|\tilde{G}_n e - e\|_\infty$ in the presence of rounding errors is still of interest. If these values are close to the machine epsilon, we could reasonably assume that the effect of rounding errors on the algorithm is minor. On the other hand, we would have to use the residual error to measure the accuracy of the approximation $\tilde{G}_n$.

We define functions $\mathcal{F}_A,\ \mathcal{F}_B : \mathbb{R}^{m\times m} \to \mathbb{R}^{m\times m}$ by

$$\mathcal{F}_A(X) = X - A_0 - A_1 X - A_2 X^2, \quad \mathcal{F}_B(X) = X - B_0 - B_1 X - B_2 X^2.$$

In [6], the accuracy of the approximations $\tilde{G}_n$ and G_n is compared using the residual errors $\|\mathcal{F}_A(\tilde{G}_n)\|_\infty$ and $\|\mathcal{F}_A(G_n)\|_\infty$. The reported values for $\|\mathcal{F}_A(\tilde{G}_n)\|_\infty$ and $\|\mathcal{F}_A(G_n)\|_\infty$ are roughly of the same magnitude. We note that this does not mean that $\tilde{G}_n$ and G_n have roughly the same accuracy. In fact, $\tilde{G}_n$ is typically much more accurate than G_n when the QBD is null recurrent or nearly null recurrent. The reason for this is the following. When the QBD is null recurrent, the Fréchet derivative of $\mathcal{F}_A$ at the solution G is a singular map. Thus, in general, $\|G_n - G\|_\infty$ is not of the order of $\|\mathcal{F}_A(G_n)\|_\infty = \|\mathcal{F}_A(G_n) - \mathcal{F}_A(G)\|_\infty$. On the other hand, the Fréchet derivative of $\mathcal{F}_B$ at the solution H is a nonsingular map. Using $\tilde{H}_n e = 0$, it is easy to show that $\mathcal{F}_A(\tilde{G}_n) = \mathcal{F}_B(\tilde{H}_n)$. Thus,

$$\|\tilde{G}_n - G\|_\infty = \|\tilde{H}_n - H\|_\infty = O(\|\mathcal{F}_B(\tilde{H}_n) - \mathcal{F}_B(H)\|_\infty) = O(\|\mathcal{F}_A(\tilde{G}_n)\|_\infty).$$

Acknowledgment. The author thanks the referees for their helpful comments on this note.

REFERENCES

[1] D. A. Bini, L. Gemignani, and B. Meini, *Computations with infinite Toeplitz matrices and polynomials*, Linear Algebra Appl., 343–344 (2002), pp. 21–61.

[2] D. A. Bini, G. Latouche, and B. Meini, *Solving matrix polynomial equations arising in queueing problems*, Linear Algebra Appl., 340 (2002), pp. 225–244.

[3] D. Bini and B. Meini, *On the solution of a nonlinear matrix equation arising in queueing problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 906–926.

[4] H. R. Gail, S. L. Hantler, and B. A. Taylor, *Spectral analysis of $M/G/1$ and $G/M/1$ type Markov chains*, Adv. in Appl. Probab., 28 (1996), pp. 114–165.

[5] C.-H. Guo, *Convergence analysis of the Latouche–Ramaswami algorithm for null recurrent quasi-birth-death processes*, SIAM J. Matrix Anal. Appl., 23 (2002), pp. 744–760.

[6] C. He, B. Meini, and N. H. Rhee, *A shifted cyclic reduction algorithm for quasi-birth-death problems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 673–691.

[7] G. Latouche and V. Ramaswami, *A logarithmic reduction algorithm for quasi-birth-death processes*, J. Appl. Probab., 30 (1993), pp. 650–674.

[8] G. Latouche and V. Ramaswami, *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM, Philadelphia, PA, 1999.