Calcolo manuscript No.
(will be inserted by the editor)

A probabilistic interpretation of Cyclic Reduction and its relationships with Logarithmic Reduction

Dario A. Bini[1], **Beatrice Meini**[1], **V. Ramaswami**[2]

[1] Dipartimento di Matematica, Università di Pisa, Pisa, Italy
[2] AT&T Labs-Research, Florham Park, NJ 07932, USA

Abstract. We provide a simple convergence proof for the cyclic reduction algorithm for $M/G/1$ type Markov chains together with a probabilistic interpretation which helps to better understand the relationships between Logarithmic Reduction and Cyclic Reduction.

1. Introduction

A Markov chain of M/G/1 type [11], [10], [3] has a two-dimensional state space $\{(i,j) : i \geq 0,\ 1 \leq j \leq m\}$, where m is a positive integer, is skip-free downward in its first coordinate, and except possibly for the set of boundary states $\{(0,j) : 1 \leq j \leq m\}$ is spatially homogeneous with respect to the first coordinate. Specifically, calling the set of states $\{(i,j) : 1 \leq j \leq m\}$ as level i and partitioning the state space according to levels $0, 1, \cdots$, we can write the transition matrix of the Markov chain in the form

$$P = \begin{bmatrix} B_0 & B_1 & B_2 & B_3 & \cdots \\ A_{-1} & A_0 & A_1 & A_2 & \cdots \\ 0 & A_{-1} & A_0 & A_1 & \ddots \\ 0 & 0 & A_{-1} & A_0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

where B_n and A_{n-1}, for $n \geq 0$, are of order $m \times m$ and such that $\sum_{n=-1}^{\infty} A_n$, $\sum_{n=0}^{\infty} B_n$ are row-stochastic.

An important quantity in the algorithmic analysis of a Markov chain of the $M/G/1$ type is the matrix G whose element g_{ij} is the probability that, starting in the state $(1, i)$, the Markov chain makes an eventual first passage into level 0 and does so by visiting the specific state $(0, j)$. It is well-known [11] that G is the minimal non-negative solution of the matrix equation

$$G = \sum_{n=-1}^{\infty} A_n G^{n+1}. \quad (1)$$

The special case of Markov chains of the above form with $A_n = B_n = 0$ for $n \geq 2$ goes by the name Quasi-Birth-and-Death process (QBD). For QBDs, an iterative, quadratically convergent algorithm for G, briefly denoted by LR or Logarithmic Reduction and based on censoring, was derived by Latouche and Ramaswami [9] which involves viewing the Markov chain successively on levels of the form $\{0, 2^k, 2 \cdot 2^k, 3 \cdot 2^k, \cdots\}$, $k \geq 0$. The k-th iterate in this algorithm computes an approximation for G by taking into account all paths that go up to but not above the level 2^k. That probabilistic interpretation immediately yields the monotonicity of the iterates as well as their quadratic convergence to the matrix G, in the nonnull recurrent case, quite easily.

An algorithm similar to LR was derived by Bini and Meini [1], [2] for the general $M/G/1$ type Markov chains. They called this algorithm Cyclic Reduction (CR) due to the similarity of the method with an odd-even reduction algorithm for linear systems due to Buzbee, Golub and Nielson [4]. The algorithm is based on methods of linear algebra such as Schur complementation as they apply to Toeplitz systems, and the approach taken does not rely on probabilistic arguments. In fact, CR can be described either in an algebraic way or in an analytic way by means of a functional representation [3].

Bini, Latouche and Meini did demonstrate the closeness of their method to LR in the case of QBDs in [3]. Here, we further analyze this closeness by means of a probabilistic interpretation, where the CR algorithm is viewed as applying recursive censoring to a suitable M/G/1 Markov chain. This interpretation enables us to provide a convergence proof for the algorithm in a much easier way than previously possible.

The idea is to transform a positive recurrent Markov chain into a transient one for which the convergence proof of CR is almost immediate. It is interesting to observe that this reduction, reinterpreted in the language of linear algebra as based on a diagonal rescaling, provides a convergence proof much simpler and more elegant than the one previously known, which is also independent of the probabilistic interpretation.

A probabilistic interpretation of CR by means of successive censorings was previously obtained by Hunt in [7], [8], where at the kth censoring step the levels $1 + n \cdot 2^k$, $n = 0, 1, 2, \ldots$, are considered. Unlike in [7], in our approach the levels involved at the kth step are $0, 1 + n \cdot 2^k$, for $n = 0, 1, 2, \ldots$. The advantage of including level 0 is that there is a one-to-one correspondence between the matrix sequences generated by CR and the stochastic complements obtained by means of the successive censorings. The interpretation in terms of the simple odd-even censoring of [7] does not give this correspondence immediately.

The paper is organized as follows. In section 2 we briefly recall CR and the proof of its convergence in the transient case; then we show that the positive recurrent case can be reduced to the transient one, and provide a simple convergence proof. In Section 3 we give a probabilistic interpretation of CR which relates this algorithm to LR.

2. A simple convergence proof of CR

The method of CR can be synthesized as follows. Equation (1) can be rewritten in terms of the linear system

$$\begin{bmatrix} I - A_0 & -A_1 & -A_2 & \ldots \\ -A_{-1} & I - A_0 & -A_1 & \ddots \\ & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} G \\ G^2 \\ G^3 \\ \vdots \end{bmatrix} = \begin{bmatrix} A_{-1} \\ 0 \\ \vdots \end{bmatrix}. \tag{2}$$

By rearranging the equations and the unknowns according to the even/odd block permutation yields

$$\left[\begin{array}{c|c} U_{1,1} & U_{1,2} \\ \hline U_{2,1} & U_{2,2} \end{array}\right] \left[\begin{array}{c} G^2 \\ G^4 \\ \vdots \\ \hline G \\ G^3 \\ \vdots \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ \vdots \\ \hline A_{-1} \\ 0 \\ \vdots \end{array}\right]$$

where

$$U_{1,1} = \begin{bmatrix} I - A_0 & -A_2 & -A_4 & \dots \\ & I - A_0 & -A_2 & \ddots \\ & & I - A_0 & \ddots \\ 0 & & & \ddots \end{bmatrix}, \quad U_{1,2} = \begin{bmatrix} -A_{-1} & -A_1 & -A_3 & \dots \\ & -A_{-1} & -A_1 & \ddots \\ & & -A_{-1} & \ddots \\ 0 & & & \ddots \end{bmatrix},$$

$U_{2,2} = U_{1,1}$ and $U_{2,1}$ is obtained by removing the first block column of $U_{1,2}$.

Eliminating the even powers of G by means of a Schur complementation yields the new system

$$S^{(1)} \begin{bmatrix} G \\ G^3 \\ G^5 \\ \vdots \end{bmatrix} = \begin{bmatrix} A_{-1} \\ 0 \\ \vdots \end{bmatrix}, \tag{3}$$

where

$$S^{(1)} = U_{2,2} - U_{2,1}U_{1,1}^{-1}U_{1,2} = \begin{bmatrix} I - \widehat{A}_0^{(1)} & -\widehat{A}_1^{(1)} & -\widehat{A}_2^{(1)} & \dots \\ -A_{-1}^{(1)} & I - A_0^{(1)} & -A_1^{(1)} & \dots \\ & -A_{-1}^{(1)} & I - A_0^{(1)} & \dots \\ 0 & & \ddots & \ddots \end{bmatrix} \tag{4}$$

Observe that the system (3) with matrix (4) has almost the same form of (2) so that we can apply once again the same technique.

The recursive application of this scheme, known as Cyclic Reduction, provides the following sequence of systems

$$\begin{bmatrix} I - \widehat{A}_0^{(k)} & -\widehat{A}_1^{(k)} & -\widehat{A}_2^{(k)} & \dots \\ -A_{-1}^{(k)} & I - A_0^{(k)} & -A_1^{(k)} & \dots \\ & -A_{-1}^{(k)} & I - A_0^{(k)} & \dots \\ 0 & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} G \\ G^{2^k+1} \\ G^{2 \cdot 2^k+1} \\ \vdots \end{bmatrix} = \begin{bmatrix} A_{-1} \\ 0 \\ \vdots \end{bmatrix}. \tag{5}$$

The first block equation yields the following expression for the matrix G:

$$\begin{aligned} G &= G_k + (I - \widehat{A}_0^{(k)})^{-1} \textstyle\sum_{j=1}^{+\infty} \widehat{A}_j^{(k)} G^{j \cdot 2^k+1}, \\ G_k &= (I - \widehat{A}_0^{(k)})^{-1} A_{-1}. \end{aligned} \tag{6}$$

Equation (4) can be similarly rewritten at the general step k and provides a means for relating the matrices $A_i^{(k+1)}$, $\widehat{A}_i^{(k+1)}$ to $A_i^{(k)}$, $\widehat{A}_i^{(k)}$, respectively, at two consecutive steps of CR. A more compact and computationally efficient way for describing the general CR step relies on a functional representation of this algorithm; the reader can find more details on this topic in [3].

The matrices $A_{i-1}^{(k)}$ and $\widehat{A}_i^{(k)}$, $i \geq 0$, are nonnegative and such that $\sum_{i=-1}^{+\infty} A_i^{(k)}$ and $A_{-1} + \sum_{i=0}^{+\infty} \widehat{A}_i^{(k)}$ are row stochastic. This implies that $\|A_i^{(k)}\|$ and $\|\widehat{A}_i^{(k)}\|$ are uniformly bounded for any matrix norm $\|\cdot\|$. It is a simple matter to prove that $\|(I - \widehat{A}_0^{(k)})^{-1}\|$ is bounded as well [3].

The convergence of G_k to G is equivalent to proving that

$$(I - \widehat{A}_0^{(k)})^{-1} \sum_{j=1}^{+\infty} \widehat{A}_j^{(k)} G^{j \cdot 2^k + 1}$$

converges to zero as $k \to \infty$. In the book [3], this property is proved in the positive recurrent case by relying on analytic tools on functions of a complex variable, and in the transient case by means of almost immediate algebraic manipulations. The latter proof is recalled in Section 2.1.

We provide a new argument in Seciton 2.2 to show that the positive recurrent case is solved by means of a simple reduction to the transient case. 2.2.

2.1. The transient case

For a transient Markov chain we have $\rho(G) = \eta < 1$, where ρ denotes the spectral radius of a matrix (see [5]); this implies that for any $\epsilon > 0$ and for any matrix norm $\|\cdot\|$ there exists $\theta > 0$ such that

$$\|G^n\| \leq \theta(\eta + \epsilon)^n \quad \text{for any} \quad n > 0. \tag{7}$$

In the case where all the eigenvalues of G of modulus η are simple we may choose $\epsilon = 0$. This is the case that we consider in the next section.

From (6) we deduce that

$$\|G - G_k\| \leq \|(I - \widehat{A}_0^{(k)})^{-1}\| \sum_{j=1}^{+\infty} \|\widehat{A}_j^{(k)}\| \cdot \|G^{j \cdot 2^k + 1}\| \leq \gamma(\eta + \epsilon)^{2^k}$$

for a suitable constant γ, since $\|\widehat{A}_j^{(k)}\|$ and $\|(I - \widehat{A}_0^{(k)})^{-1}\|$ are uniformly bounded. This proves quadratic convergence with rate $\eta + \epsilon$.

2.2. Reducing a positive recurrent to a transient Markov chain

Suppose that the Markov chain whose homogeneous part is defined by A_i, $i = -1, 0, 1, \ldots$, is positive recurrent so that $\rho(G) = 1$ (see [5]). Let us assume that the matrix function $A(z) = \sum_{i=-1}^{+\infty} A_i z^{i+1}$ is analytic for $|z| < R$, for a suitable $R > 1$, and that there exists a zero of $\det(zI - A(z))$ of modulus greater than 1. Under these assumptions it has been proved by Gail, Hantler and Taylor [6] that there exist a real ξ and a positive vector $\boldsymbol{u} = (u_i)$ such that $1 < \xi < R$ and $A(\xi)\boldsymbol{u} = \xi\boldsymbol{u}$. Moreover, ξ is a simple zero of $\det(zI - A(z))$, and it is such that every other zero of $\det(zI - A(z))$ outside the unit circle is strictly greater in modulus than ξ. The above assumptions are the same as the ones stated in [3, Chapter 7] and are satisfied in general if the matrix function $A(z)$ is entire or rational [6].

Let $D = \mathrm{Diag}(\boldsymbol{u})$ be the diagonal matrix having $u_1, \ldots, u_m$ as diagonal entries, and define

$$\begin{aligned} C_i &= \xi^i D^{-1} A_i D, \quad i = -1, 0, 1, \ldots, \\ H &= \xi^{-1} D^{-1} G D. \end{aligned} \tag{8}$$

It is readily verified that $C_i \geq 0$ and the matrix $\sum_{i=-1}^{+\infty} C_i$ is row stochastic. Therefore, the matrices C_i, $i = -1, 0, 1, \ldots$, define the homogeneous part of an M/G/1-type Markov chain. Moreover, it is a simple matter to show that G is the nonnegative minimal solution of the equation (1) if and only if H is the nonnegative minimal solution of

$$H = \sum_{i=-1}^{\infty} C_i H^{i+1}. \tag{9}$$

Since $\rho(H) = \xi^{-1}\rho(G)$ and $\rho(G) = 1$, the matrix H has m eigenvalues of modulus strictly less than 1 and, therefore the Markov chain associated with (9) is transient (see [5]).

From equation (4) one finds that applying the CR algorithm to (9) generates matrices $\{C_{i-1}^{(k)}\}_{i\geq 0}$ and $\{\widehat{C}_i^{(k)}\}_{i\geq 0}$ which are related to the matrices $\{A_{i-1}^{(k)}\}_{i\geq 0}$ and $\{\widehat{A}_i^{(k)}\}_{i\geq 0}$ by the following equations:

$$\begin{aligned} C_i^{(k)} &= \xi^{i\cdot 2^k} D^{-1} A_i^{(k)} D, \quad i = -1, 0, 1, \ldots, \\ \widehat{C}_i^{(k)} &= \xi^{i\cdot 2^k} D^{-1} \widehat{A}_i^{(k)} D, \quad i = 0, 1, 2, \ldots. \end{aligned}$$

Moreover, CR generates approximations $\{H_k\}$ to the matrix H which are related to $\{G_k\}$ by the simple equation

$$G_k = \xi D H_k D^{-1}, \quad k \geq 0.$$

Thus, convergence of CR for the positive recurrent case follows from convergence in the transient case. More specifically, since the eigenvalues of G with modulus 1 are all simple (see [5]) the eigenvalues of H of modulus ξ^{-1} are simple and we may choose $\epsilon = 0$ in the inequality (7). Thus, we have

$$\|G_k - G\| = \xi\|D(H_k - H)D^{-1}\| \le \xi\|D\| \cdot \|D^{-1}\| \cdot \|H_k - H\| \le \gamma'\xi^{-2^k},$$

for a suitable constant $\gamma' > 0$; that is, a quadratic convergence obtains with convergence rate ξ^{-1}.

Concerning the convergence of the matrix sequences $\{A_{i-1}^{(k)}\}_{k\ge0}$ and $\{\widehat{A}_i^{(k)}\}_{k\ge0}$ as $k \to \infty$, from the relation $A_i^{(k)} = \xi^{-i\cdot2^k} DC_i^{(k)} D^{-1}$ and from the property $\|C_i^{(k)}\|_1 \le 1$ for any $i \ge -1$ and $k \ge 0$, we deduce that

$$\|A_i^{(k)}\|_1 \le \sigma\xi^{-i\cdot2^k}, \quad i = 0, 1, 2, \dots, \quad k = 0, 1, 2, \dots,$$

for a suitable constant σ. Therefore the sequence $\{A_i^{(k)}\}_k$, for $i \ge 1$, converges quadratically to zero as $k \to \infty$ with convergence rate ξ^{-i}. Similarly, we find that

$$\|\widehat{A}_i^{(k)}\|_1 \le \hat{\sigma}\xi^{-i\cdot2^k}, \quad i = 0, 1, 2, \dots, \quad k = 0, 1, 2, \dots,$$

for a suitable constant $\hat{\sigma}$, i.e., the sequence $\{\widehat{A}_i^{(k)}\}_k$, for $i \ge 1$, converges quadratically to zero as $k \to \infty$ with convergence rate ξ^{-i}.

3. Probabilistic interpretation: CR and LR

The computation of G for an $M/G/1$ type chain is indeed the computation of the G matrix for the chain on levels $\{0, 1, 2, \cdots\}$ with block partitioned transition matrix

$$P_0 = \begin{bmatrix} I & 0 & 0 & 0 & \cdots \\ A_{-1} & A_0 & A_1 & A_2 & \cdots \\ 0 & A_{-1} & A_0 & A_1 & \ddots \\ 0 & 0 & A_{-1} & A_0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \qquad (10)$$

since the boundary matrices $\{B_n\}$ play no role with respect to G.

Now consider the embedded Markov chain on $\{0,1,3,5,7,\cdots\}$. To this purpose apply the permutation $\{0,1,3,5,\ldots,2,4,6,\ldots\}$ and partition the permuted matrix as

$$\begin{bmatrix} I & 0 & 0 \\ A_{-1} & \hat{A}^{(0)}_{\text{even}} & \hat{A}^{(0)}_{\text{odd}} \\ 0 & W^{(0)} & U^{(0)}_{2,1} \\ 0 & Y^{(0)} & U^{(0)}_{1,1} \end{bmatrix},$$

where $W^{(0)} = \left[O \; U^{(0)}_{2,2} \right]$ with O being the infinite block column with null entries, $Y^{(0)} = \left[V^{(0)} \; U^{(0)}_{1,2} \right]$ with $V^{(0)} = [-A^T_{-1}, 0, \ldots]^T$ and the U-matrices are the matrices $U_{i,j}$ in Section 2.

When we construct the embedded Markov chain P_1, what we need to do is to delete the last block of columns and rows corresponding to the set $\{2,4,6,\cdots\}$ and also to make the following replacements in the other columns:

$$\begin{aligned} \hat{A}_{\text{even}} &\leftarrow \hat{A}_{\text{even}} + \hat{A}_{\text{odd}}[I - U_{1,1}]^{-1} Y \\ Y &\leftarrow Y + U_{2,1}\,[I - U_{1,1}]^{-1}\, W \end{aligned}$$

This step is also called censoring the Markov chain on the levels $\{0,1,3,5,7,\cdots\}$ and amounts to looking at the Markov chain only when it is on this set, i.e., eliminating periods when levels that are an odd distance away to the right from level 1 are visited. That gives a Markov chain on levels $\{0,1,3,5,7,...\}$ whose transition matrix is given in partitioned form by

$$P_1 = \begin{bmatrix} I & 0 & 0 & 0 & \cdots \\ A_{-1} & \hat{A}^{(1)}_0 & \hat{A}^{(1)}_1 & \hat{A}^{(1)}_2 & \cdots \\ 0 & A^{(1)}_{-1} & A^{(1)}_0 & A^{(1)}_1 & \cdots \\ 0 & 0 & A^{(1)}_{-1} & A^{(1)}_0 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{11}$$

Observe that except for the first and the second block row, the remaining block rows have constant blocks along the diagonals. Observe also that this step corresponds to computing the Schur complement $S^{(1)}$ of (4). It is also trivial to note that the G-matrices associated with P_0 and P_1 are identical, and computing G using these Markov chains simply amounts to nothing more than a different way of book keeping of the various levels visited in the first passage time to level 0.

More specifically, the matrix G solves the equation

$$G = A_{-1} + \sum_{i=0}^{+\infty} \widehat{A}_i^{(1)} G^{i+1},$$

while the matrix $G_1 = G^2$ solves the equation

$$G_1 = \sum_{i=-1}^{+\infty} A_i^{(1)} G_1^{i+1}.$$

Simple path-based probabilistic arguments establish that the matrices $A_k^{(1)}$, $\widehat{A}_k^{(1)}$ are the same as the ones obtained in equation (3) by means of CR. In the linear algebra terminology, the above process amounts to eliminating the last set of linear equations through the creation of a Schur complement whereas in the probabilistic approach this amounts to visits to the last set of levels being accounted through the matrix $[I - U_{1,1}]^{-1}$ which can also be written as the sum $\sum_{n=0}^{\infty} [U_{1,1}]^n$ of the probabilities that n successive transitions occur within the last set of states before the set of interest is entered into.

We can now start with the chain given by P_1, split the state space again as $\{0\}$, $\{1, 5, 9, \cdots\}$, $\{3, 7, 11, \cdots\}$ and repeat the procedure. In general, we can do this k times and get a Markov chain on the levels $\{0, 1, 1 + 2^k, 1 + 2 \cdot 2^k, 1 + 3 \cdot 2^k, \cdots\}$ whose transition matrix is easily seen to be

$$P_k = \begin{bmatrix} I & 0 & 0 & 0 & \cdots \\ A_{-1} & \hat{A}_0^{(k)} & \hat{A}_1^{(k)} & \hat{A}_2^{(k)} & \cdots \\ 0 & A_{-1}^{(k)} & A_0^{(k)} & A_1^{(k)} & \cdots \\ 0 & 0 & A_{-1}^{(k)} & A_0^{(k)} & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{12}$$

A point to note is that the matrix G is the same for all these Markov chains, i.e.,

$$G = A_{-1} + \sum_{i=0}^{+\infty} \widehat{A}_i^{(k)} G^{i+1},$$

and all that we have done is to calculate it differently each time through different embedded Markov chains. Now in the k-th chain, we can form the approximation

$$G \approx [I - \hat{A}_0^{(k)}]^{-1} A_{-1}$$

by omitting visits to levels higher than level 1 in this embedded chain. Due to the skip-free downward structure of the Markov chain we started from, it is clear that this process amounts to deleting all paths that go to levels $1 + 2^k$ and higher in the original chain we started with.

In the language of linear algebra these properties correspond to the fact that the block vector in the left-hand side of (5) has G as first component and, consequently that G satisfies equation (6).

References

1. D.A. Bini and B. Meini. On cyclic reduction applied to a class of Toeplitz-like matrices arising in queueing problems, in *Computations with Markov Chains*, W.J. Stewart (editor), Kluwer Academic Publishers, Boston, 1995, pages 21–38.
2. D. A. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems, *SIAM J. Matrix Anal. Appl.*,, **17(4)**, 1996, 906–926.
3. D.A. Bini, G. Latouche and B. Meini. *Numerical Methods for Structured Markov Chains*, Oxford Science Publications, London, 2005.
4. B.L. Buzbee, G.H. Golub and C.W. Nielson. On direct methods for solving Poisson's equations, *SIAM J. Numer. Anal.*, **7**, 1970, 627–656.
5. 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(1)**, 1996, 114–165.
6. H.R. Gail, S.L. Hantler and B.A. Taylor. Matrix-geometric invariant measures for $G/M/1$ type Markov chains. *Comm. Statist. Stochastic Models*, **14(3)**, 1998, 537–569.
7. E. Hunt. A probabilistic algorithm for determining the fundamental matrix of a block $M/G/1$ Markov chain. *Math. Comput. Modelling*, **38(11-13)**, 2003, 1203–1209.
8. E. Hunt. Determining the fundamental matrix of a block $M/G/1$ Markov chain. In *Engineering, Technology and Management*, Eds. R.J. Wilson, S. Osaki and M.J. Faddy, University of Queensland, 1999, 446-455.
9. G. Latouche and V. Ramaswami. A logarithmic reduction algorithm for quaqsi-birth-and-death processes, *J. Appl. Prob.*, **30(3)**, 1993, 650–674.
10. G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*, SIAM & ASA, Philadelphia, PA, 1999.
11. M.F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and their Applications,* Marcel Dekker, Inc., NY, 1989.