

ON THE DUAL RELATIONSHIP BETWEEN MARKOV CHAINS OF GI/M/1 AND M/G/1 TYPE

P. G. TAYLOR,* *University of Melbourne*

B. VAN HOUDT,** *University of Antwerp*

Abstract

In 1990, Ramaswami proved that, given a Markov renewal process of M/G/1 type, it is possible to construct a Markov renewal process of GI/M/1 type such that the matrix transforms $\mathbf{G}(z, s)$ for the M/G/1-type process and $\mathbf{R}(z, s)$ for the GI/M/1-type process satisfy a duality relationship. In his 1996 PhD thesis, Bright used time reversal arguments to show that it is possible to define a different dual for positive-recurrent and transient processes of M/G/1 type and GI/M/1 type. Here we compare the properties of the Ramaswami and Bright dual processes and show that the Bright dual has desirable properties that can be exploited in the design of algorithms for the analysis of Markov chains of GI/M/1 type and M/G/1 type.

Keywords: Matrix-analytic models; quasi-birth-and-death process; processes of GI/M/1 type; processes of M/G/1 type

2000 Mathematics Subject Classification: Primary 60J10
Secondary 90B22

1. Introduction

Markov chains of GI/M/1 type and M/G/1 type were first defined and studied by Neuts [18]–[20]. Since that time, they have proved to be versatile classes of models finding wide application to the solution of problems in queueing theory and applied probability more generally.

In 1990, Ramaswami [21] showed that, given an M/G/1-type process, it is possible to construct a GI/M/1-type process such that the matrix \mathbf{G} for the M/G/1-type process and the matrix \mathbf{R} for the GI/M/1-type process satisfy a duality relationship. He called the GI/M/1-type process so constructed the *dual process*. A similar dual can be defined when the original chain is of GI/M/1 type. Ramaswami's duality relationship has proved to be useful in that it has enabled results for one of the classes to be extended to the other. This relationship has, for example, been implemented in the SMCSolver tool [7] in such a way that GI/M/1-type Markov chains are analyzed by converting them into dual M/G/1-type chains and performing the analysis within that framework.

Asmussen and Ramaswami [3] quickly followed Ramaswami's original paper with an alternative interpretation of the dual process that appealed to time reversal arguments. This gave insight into the physical meaning of the matrices involved. In his 1996 PhD thesis [8], Bright provided a different time-reversed interpretation of Ramaswami's dual process and

Received 18 November 2008; revision received 24 November 2009.

* Postal address: Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia.

Email address: p.taylor@ms.unimelb.edu.au

** Postal address: Performance Analysis of Telecommunications Systems Research Group, Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, Antwerp, B 2020, Belgium.

Email address: benny.vanhoudt@ua.ac.be

showed that, for positive-recurrent GI/M/1-type processes and transient M/G/1-type processes, it is possible to define a second dual process. It is also possible to define a second dual for a large class of transient GI/M/1-type processes and positive-recurrent M/G/1-type processes.

This second dual, which has not previously been discussed in the open literature, is of intrinsic interest in its own right. More recently, we have observed that, if we use the Bright dual instead of the Ramaswami dual to convert a positive-recurrent GI/M/1-type Markov chain into an M/G/1-type Markov chain, there is a significant improvement in the efficiency of the implementation of the cyclic reduction algorithm that is used to calculate G in the SMCSolver tool [7]. Furthermore, by combining both duals, we can transform a transient M/G/1-type Markov chain into a positive-recurrent M/G/1-type Markov chain, again allowing a more efficient computation of the G matrix. This has motivated us to write this paper, in which we present and extend the duality results of [8] and then explain why they can lead to efficiency gains in algorithmic analysis.

Following the example of Ramaswami, Bright defined his dual in the Markov renewal process context. He also considered level-dependent processes. For ease of presentation, we have decided to work within the context of level-independent continuous-time Markov chains, recognizing that the analysis can easily be reformulated in a discrete-time setting, and extended in a relatively straightforward manner to level-dependent Markov renewal processes.

This paper is organised as follows. In Section 2, we start with definitions of GI/M/1-type and M/G/1-type Markov chains, give a brief summary of the role of the matrices R and G in their analysis, and finish with a discussion of the Ramaswami dual, together with its properties and probabilistic interpretation in terms of time-reversed processes. Included is a new interpretation in terms of the time reversal of a doubly infinite extension of the original chain. This leads naturally, in Section 3, to the definition of the Bright dual process, which has different properties. In Section 4 we discuss the role of both duals in the implementation of algorithms for the derivation of R and G , while some numerical examples are given in Section 5.

2. GI/M/1-type and M/G/1-type Markov chains

Level-independent Markov chains of GI/M/1 and M/G/1 type have two-dimensional state spaces of the form $\{0, 1, \dots\} \times \{1, \dots, m\}$, with m considered here to be finite, and block lower-Hessenberg and block upper-Hessenberg generator matrices of the form

$$P_{GI} = \begin{bmatrix} B_0 & A_0 & \mathbf{0} & \mathbf{0} & \cdots \\ B_1 & A_1 & A_0 & \mathbf{0} & \cdots \\ B_2 & A_2 & A_1 & A_0 & \cdots \\ B_3 & A_3 & A_2 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{1}$$

and

$$P_M = \begin{bmatrix} B_0 & B_1 & B_2 & B_3 & \cdots \\ A_0 & A_1 & A_2 & A_3 & \cdots \\ \mathbf{0} & A_0 & A_1 & A_2 & \cdots \\ \mathbf{0} & \mathbf{0} & A_0 & A_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{2}$$

respectively. The class of level-independent *quasi-birth-and-death processes* (QBDPs) can be considered either as processes of M/G/1 type or as processes of GI/M/1 type. They have

generator matrices of the form (1) and (2), but the only nonzero blocks are on the diagonal and upper and lower diagonals.

In Neuts [19, p. 9], [20, p. 88] it was shown that the matrices \mathbf{R} and \mathbf{G} , which are the minimal nonnegative solutions to the equations

$$\mathbf{0} = \sum_{\ell=0}^{\infty} \mathbf{R}^{\ell} \mathbf{A}_{\ell} \quad (3)$$

and

$$\mathbf{0} = \sum_{\ell=0}^{\infty} \mathbf{A}_{\ell} \mathbf{G}^{\ell}, \quad (4)$$

play a crucial role in the derivation of performance measures for Markov chains with generator matrices of the form (1) and (2), respectively.

For a GI/M/1-type Markov chain, the (i, j) th entry of the matrix \mathbf{R} is the ratio of the expected total sojourn time in phase j of level $k + 1$ before first return to level k , conditional on the process starting in state (k, i) , to the expected sojourn time in state (k, i) . The matrix \mathbf{R} is nonnegative, with spectral radius less than or equal to 1, and the GI/M/1-type Markov chain is positive recurrent if and only if the spectral radius of \mathbf{R} is strictly less than 1. In this case, the stationary distribution $\mathbf{x} = (x_0, x_1, x_2, \dots)$ of the chain has the well-known matrix-geometric form, given by

$$\mathbf{x}_k = \mathbf{x}_0 \mathbf{R}^k.$$

The (i, j) th entry of the matrix \mathbf{G} is the probability that the M/G/1-type Markov chain will first enter level k in phase j given that it starts in phase i of level $k + 1$. The matrix \mathbf{G} is stochastic if and only if the chain is recurrent. Otherwise, \mathbf{G} is substochastic, with spectral radius strictly less than 1.

Because it is very rarely possible to evaluate either \mathbf{R} or \mathbf{G} analytically, a great deal of attention has been paid to iterative algorithms that can be shown to converge to the desired matrix; see, for example, [1], [4], [5, Chapters 6–9], [6], [14]–[16], [17, Chapter 8], [18], [19, pp. 36–40], and [20, p. 95]. Notable amongst these algorithms are the logarithmic reduction algorithm of Latouche and Ramaswami [16], which was designed for QBDPs, and the cyclic reduction algorithm of Bini and Meini [4], which was designed for processes of M/G/1 type. These algorithms are quadratically convergent and represent the current state of the art for efficiently analysing these processes. Together with various advanced features, which we shall explain in Section 4, these algorithms have been implemented in the SMCSolver tool [7], to which the second author of the current paper has been a major contributor. One feature of this tool is that it analyses GI/M/1-type Markov chains by using Ramaswami's duality result to convert them into M/G/1-type Markov chains and then employing the cyclic reduction algorithm to calculate \mathbf{G} .

As the \mathbf{B}_{ℓ} matrices in (1) and (2) do not affect the \mathbf{R} and \mathbf{G} matrices, respectively, we will refer to the GI/M/1-type Markov chain of (1) as $\mathcal{G}\mathcal{I}(A_{\ell})$ and use $\mathcal{M}(A_{\ell})$ to denote the M/G/1-type Markov chain characterized by (2).

Theorem 1, below, includes the special case of the duality result obtained by Ramaswami [21] for level-independent M/G/1-type continuous-time Markov chains, as well as its analogous result for GI/M/1-type Markov chains. This special case is equivalent to Corollary 5.1 of Asmussen [2], which was derived using Wiener–Hopf factorization methods. Here and elsewhere, we use \mathbf{X}^{\top} to denote the transpose of a matrix \mathbf{X} .

Theorem 1. Assume that $A = \sum_{\ell=0}^{\infty} A_{\ell}$ is an irreducible generator matrix. Let \mathbf{a} be the invariant probability vector of A , $\Delta^{(r)} = \text{diag}(\mathbf{a})$, \mathbf{R} be the minimal nonnegative solution to (3), and \mathbf{G} be the minimal nonnegative solution to (4).

Then, with $A_{\ell}^{(r)}$ defined by

$$A_{\ell}^{(r)} = (\Delta^{(r)})^{-1} A_{\ell}^{\top} (\Delta^{(r)}), \quad \ell \geq 0, \tag{5}$$

the matrix $A^{(r)} = \sum_{\ell=0}^{\infty} A_{\ell}^{(r)}$ is an irreducible generator matrix, and the \mathbf{R} and \mathbf{G} matrices of $\mathcal{G}\mathcal{I}(A_{\ell}^{(r)})$ and $\mathcal{M}(A_{\ell}^{(r)})$, denoted as $\mathbf{R}^{(r)}$ and $\mathbf{G}^{(r)}$, are given by

$$\mathbf{R}^{(r)} = (\Delta^{(r)})^{-1} \mathbf{G}^{\top} (\Delta^{(r)}) \tag{6}$$

and

$$\mathbf{G}^{(r)} = (\Delta^{(r)})^{-1} \mathbf{R}^{\top} (\Delta^{(r)}), \tag{7}$$

respectively.

The Ramaswami dual of the M/G/1-type process $\mathcal{M}(A_{\ell})$ is the GI/M/1-type process $\mathcal{G}\mathcal{I}(A_{\ell}^{(r)})$ with $A_{\ell}^{(r)}$ given by (5). Similarly, the dual of $\mathcal{G}\mathcal{I}(A_{\ell})$ is $\mathcal{M}(A_{\ell}^{(r)})$. It is straightforward to show that taking the dual of the dual results in the original process.

It follows from Theorem 1 that the matrices $\mathbf{R}^{(r)}$ and \mathbf{G} have the same eigenvalues. Thus, the spectral radius of \mathbf{G} is equal to 1 if and only if the spectral radius of $\mathbf{R}^{(r)}$ for the dual process is equal to 1, which tells us that the dual process of a recurrent level-independent M/G/1-type process is either transient or null recurrent and that the dual process of a transient level-independent M/G/1-type process is positive recurrent. Similarly, it follows that the matrices $\mathbf{G}^{(r)}$ and \mathbf{R} have the same eigenvalues, and so the dual process of a positive-recurrent level-independent GI/M/1-type process is transient and the dual process of a transient or null recurrent level-independent GI/M/1-type process is recurrent. Furthermore, we can also prove that the dual is null recurrent if and only if the original process is.

Ramaswami originally established his duality result via algebraic means. In [3] Asmussen and Ramaswami gave an alternative proof in terms of time reversal of sample paths. Below we give a brief outline of their arguments applied to the dual of an M/G/1-type process. The discussion can be easily extended to the dual of a GI/M/1-type process.

Consider a continuous-time Markov chain $\tilde{X}_M(t) = (L_M(t), J_M(t))$ on the state space $\{(\ell, j) : \ell = \pm 1, \pm 2, \dots, 1 \leq j \leq m\}$ with generator matrix of the form

$$Q_M = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \\ \dots & A_1 & A_2 & A_3 & A_4 & \dots \\ \dots & A_0 & A_1 & A_2 & A_3 & \dots \\ \dots & \mathbf{0} & A_0 & A_1 & A_2 & \dots \\ \dots & \mathbf{0} & \mathbf{0} & A_0 & A_1 & \dots \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{8}$$

where the matrix $A = \sum_{\ell=0}^{\infty} A_{\ell}$ is irreducible and aperiodic. Because of the structure of the state space, we refer to the process with generator matrix (8) as a ‘doubly infinite’ M/G/1-type process.

Asmussen and Ramaswami defined a second doubly infinite process, $\tilde{X}_{GI}(t) = (L_{GI}(t), J_{GI}(t))$ on the same state space as that above, with generator matrix given by

$$Q_{GI} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \dots \\ \dots & A_1^{(r)} & A_0^{(r)} & \mathbf{0} & \mathbf{0} & \dots \\ \dots & A_2^{(r)} & A_1^{(r)} & A_0^{(r)} & \mathbf{0} & \dots \\ \dots & A_3^{(r)} & A_2^{(r)} & A_1^{(r)} & A_0^{(r)} & \dots \\ \dots & A_4^{(r)} & A_3^{(r)} & A_2^{(r)} & A_1^{(r)} & \dots \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where $A_\ell^{(r)}$ is given by (5). This process has the skip-free upward property and, hence, we refer to it as a doubly infinite GI/M/1-type process.

In order to obtain a relationship that could be used to derive Ramaswami’s duality result, Asmussen and Ramaswami considered the discrete-time processes $K_n^M, n \geq 0$, and $K_n^{GI}, n \geq 0$, defined to be the processes of changes in levels at each state transition for $X_M(t)$ and $X_{GI}(t)$, respectively. The processes K_n^M and K_n^{GI} are such that $K_n^M \geq -1, n \geq 0$, and $K_n^{GI} \leq 1, n \geq 0$, and so reside in singly infinite state spaces. Using this property, Asmussen and Ramaswami derived a time reversal result; in the transformation from $\tilde{X}_M(t)$ to $\tilde{X}_{GI}(t)$ we reverse the original order of the phases, and we reverse the order and the direction of the jump sizes between levels.

The need for Asmussen and Ramaswami to consider the processes K_n^M and K_n^{GI} arose essentially because the doubly infinite Markov chains $\tilde{X}_M(t)$ and $\tilde{X}_{GI}(t)$ are transient, and Asmussen and Ramaswami did not use the concept of a time reversal for such processes. However, as Bright explained in his thesis [8], it is possible to construct and interpret a time-reversed process for transient continuous-time Markov chains. When we do this, a simpler time-reversed interpretation of the dual emerges.

Suppose that $\mathbf{m} = (\dots, m_{-1}, m_0, m_1, \dots)$ is a positive vector satisfying

$$\sum_{\ell=0}^{\infty} m_{v-\ell} A_\ell = \mathbf{0}, \quad v \in \mathbb{Z}. \tag{9}$$

Then, from an adaptation of the theorem in Section 3 of [13], \mathbf{m} is an invariant measure for $\tilde{X}_M(t)$ if the Markov process with generator matrix given by

$$q_M^{(r)}((k, j), (\ell, m)) = \frac{[m_\ell]_m}{[m_k]_j} q_M((\ell, m), (k, j))$$

is regular, that is, nonexplosive. It was shown in [8, p. 153] that this is always the case.

Now, setting $\mathbf{m} = (\dots, \mathbf{a}, \mathbf{a}, \mathbf{a}, \dots)$, we see that

$$\sum_{\ell=0}^{\infty} m_{v-\ell} A_\ell = \sum_{\ell=0}^{\infty} \mathbf{a} A_\ell = \mathbf{a} \mathbf{A} = \mathbf{0},$$

and, hence, $\mathbf{m} = (\dots, \mathbf{a}, \mathbf{a}, \mathbf{a}, \dots)$ is an invariant measure for the doubly infinite M/G/1-type process $X_M(t)$. The time reversal of $\tilde{X}_M(t)$ with respect to this invariant measure is $\tilde{X}_{GI}(t)$. Hence, we see that Ramaswami’s dual of $\tilde{X}_M(t)$ is the time reversal of $\tilde{X}_M(t)$ with respect to the invariant measure $\mathbf{m} = (\dots, \mathbf{a}, \mathbf{a}, \mathbf{a}, \dots)$.

For the doubly infinite M/G/1-type process $\tilde{X}_M(t)$, we can define the matrix \mathbf{G} in the same way as we did for the singly infinite process with generator \mathbf{P}_M . This matrix will still be the minimal nonnegative solution to (4). Similarly, for the doubly infinite GI/M/1-type process $\tilde{X}_{GI}(t)$, we can define the matrix $\mathbf{R}^{(r)}$ in the same way as we did for the singly infinite process with generator \mathbf{P}_{GI} and $\mathbf{R}^{(r)}$ will still be the minimum nonnegative solution to (3). Having defined \mathbf{G} and $\mathbf{R}^{(r)}$, the same algebraic arguments as those used in the proof of Theorem 1 give us the fact that

$$\mathbf{R}^{(r)} = (\mathbf{\Delta}^{(r)})^{-1} \mathbf{G}^\top (\mathbf{\Delta}^{(r)}).$$

3. An alternative dual process

In the previous section we saw that the Ramaswami dual of a doubly infinite M/G/1-type process is the time reversal with respect to a particular invariant measure. This observation suggests that if we take the time reversal with respect to a different invariant measure then we may obtain a different dual process. In this section we summarise the results of [8] to derive an alternative dual process in this fashion.

For the matrix sequence $\{A_\ell\}$ that characterizes the M/G/1-type process $\mathcal{M}(A_\ell)$ and the GI/M/1-type process $\mathcal{G}\mathcal{I}(A_\ell)$, assume that the matrix $\mathbf{A} = \sum_{\ell=0}^\infty A_\ell z^\ell$ is an irreducible $m \times m$ generator matrix. Let $A(z) = \sum_{\ell=0}^\infty A_\ell z^\ell$, $a(z) = \det(A(z))$, and

$$\boldsymbol{\mu} = \mathbf{a} \sum_{\ell=0}^\infty (\ell - 1) A_\ell \mathbf{e}, \tag{10}$$

where \mathbf{a} is defined in Theorem 1 and \mathbf{e} is a column vector of 1s. Neuts [19, p. 8] showed that $\mathcal{M}(A_\ell)$ is positive recurrent, null recurrent, and transient if and only if $\mu < 0$, $\mu = 0$, and $\mu > 0$, respectively, while $\mathcal{G}\mathcal{I}(A_\ell)$ is positive recurrent, null recurrent, and transient if and only if $\mu > 0$, $\mu = 0$, and $\mu < 0$, respectively.

3.1. Case 1: $\mu > 0$

As is shown in [11], if $\mu > 0$, $a(z)$ has exactly m zeros in the open unit disk and we can define

$$\eta = \max\{|z| : |z| < 1, a(z) = 0\}. \tag{11}$$

Furthermore, these m zeros were shown to be the m eigenvalues of the \mathbf{R} and \mathbf{G} matrices of $\mathcal{G}\mathcal{I}(A_\ell)$ and $\mathcal{M}(A_\ell)$, respectively [10]. It follows that η is the dominant eigenvalue of both \mathbf{R} and \mathbf{G} , implying that η is the asymptotic decay rate of the stationary vector $\boldsymbol{\pi}$ of the positive recurrent $\mathcal{G}\mathcal{I}(A_\ell)$, as well as the spectral radius of the \mathbf{G} matrix of the transient $\mathcal{M}(A_\ell)$. Finally, Neuts also showed that $A(\eta)$ has spectral radius equal to 0 [20, p. 91]. Hence, there exists a positive vector \mathbf{w}_η , normalised so that $\mathbf{w}_\eta \mathbf{e} = 1$, such that

$$\mathbf{w}_\eta \left(\sum_{\ell=0}^\infty A_\ell \eta^\ell \right) = \mathbf{0}. \tag{12}$$

Now, consider a doubly infinite level-independent M/G/1-type Markov chain. We observed in the previous section that a positive vector $\mathbf{m} = (\dots, m_{-1}, m_0, m_1, \dots)$ is an invariant measure for this process if it is a solution to (9).

If we set $m_k = \mathbf{w}_\eta \eta^{-k}$, $k \in \mathbb{Z}$, where \mathbf{w}_η is the solution to (12) such that $\mathbf{w}_\eta \mathbf{e} = 1$ then

$$\sum_{\ell=0}^\infty m_{v-\ell} A_\ell = \sum_{\ell=0}^\infty \mathbf{w}_\eta \eta^{\ell-v} A_\ell = \eta^{-v} \mathbf{w}_\eta \left(\sum_{\ell=0}^\infty A_\ell \eta^\ell \right) = \mathbf{0},$$

where the last equality follows from (12). Hence, $\mathbf{m}_k = \mathbf{w}_\eta \eta^{-k}$, $k \in \mathbb{Z}$, is an invariant measure for the doubly infinite M/G/1-type process. By an analogous argument, the vector $\mathbf{m}_k = \mathbf{w}_\eta \eta^k$, $k \in \mathbb{Z}$, is an invariant measure for the doubly infinite GI/M/1-type process.

The time reversal with respect to the invariant measure \mathbf{m}_k of the doubly infinite M/G/1-type process and doubly infinite GI/M/1-type process characterized by \mathbf{A}_ℓ is given by

$$\mathbf{A}_\ell^{(b)} = \eta^{\ell-1} (\mathbf{\Delta}^{(b)})^{-1} \mathbf{A}_\ell^\top (\mathbf{\Delta}^{(b)}), \quad \ell \geq 0,$$

with $\mathbf{\Delta}^{(b)} = \text{diag}(\mathbf{w}_\eta)$. This discussion suggests that the matrices $\mathbf{R}^{(b)}$ and $\mathbf{G}^{(b)}$, derived for the level-independent GI/M/1-type and M/G/1-type processes characterized by $\mathbf{A}_\ell^{(b)}$ and the matrices \mathbf{G} and \mathbf{R} for the original M/G/1-type and GI/M/1-type processes may satisfy a duality relationship similar to (6) and (7). In the following proposition we see that such a relationship does hold.

Proposition 1. *Let $\{\mathbf{A}_\ell\}$ be such that $\mathbf{A} = \sum_{\ell=0}^\infty \mathbf{A}_\ell$ is an irreducible generator matrix with $\mu > 0$ as defined by (10). Then, the matrices $\mathbf{R}^{(b)}$ and $\mathbf{G}^{(b)}$ derived for the processes $\mathcal{G}\mathcal{I}(\mathbf{A}_\ell^{(b)})$ and $\mathcal{M}(\mathbf{A}_\ell^{(b)})$, respectively, are given by*

$$\mathbf{R}^{(b)} = \eta^{-1} (\mathbf{\Delta}^{(b)})^{-1} \mathbf{G}^\top (\mathbf{\Delta}^{(b)}) \tag{13}$$

and

$$\mathbf{G}^{(b)} = \eta^{-1} (\mathbf{\Delta}^{(b)})^{-1} \mathbf{R}^\top (\mathbf{\Delta}^{(b)}). \tag{14}$$

Proof. The matrix $\mathbf{R}^{(b)}$ of $\mathcal{G}\mathcal{I}(\mathbf{A}_\ell^{(b)})$ satisfies the equation

$$\mathbf{0} = \sum_{\ell=0}^\infty (\mathbf{R}^{(b)})^\ell \eta^{\ell-1} (\mathbf{\Delta}^{(b)})^{-1} \mathbf{A}_\ell^\top (\mathbf{\Delta}^{(b)}). \tag{15}$$

Pre-multiplying (15) by $\eta \mathbf{\Delta}^{(b)}$, post-multiplying by $(\mathbf{\Delta}^{(b)})^{-1}$, and taking transposes of both sides, we see that

$$\mathbf{0} = \sum_{\ell=0}^\infty \mathbf{A}_\ell (\eta (\mathbf{\Delta}^{(b)})^{-1} (\mathbf{R}^{(b)})^\top (\mathbf{\Delta}^{(b)}))^\ell.$$

Thus, $\eta (\mathbf{\Delta}^{(b)})^{-1} (\mathbf{R}^{(b)})^\top (\mathbf{\Delta}^{(b)})$ is a nonnegative solution to (4) and, since \mathbf{G} is the minimal nonnegative solution to (4), it must be the case that

$$\eta (\mathbf{\Delta}^{(b)})^{-1} (\mathbf{R}^{(b)})^\top (\mathbf{\Delta}^{(b)}) \geq \mathbf{G},$$

which, since \mathbf{w}_η is positive, implies that

$$\mathbf{R}^{(b)} \geq \eta^{-1} (\mathbf{\Delta}^{(b)})^{-1} \mathbf{G}^\top (\mathbf{\Delta}^{(b)}). \tag{16}$$

Now observe that

$$\begin{aligned} \sum_{\ell=0}^\infty (\eta^{-1} (\mathbf{\Delta}^{(b)})^{-1} \mathbf{G}^\top (\mathbf{\Delta}^{(b)}))^\ell \eta^{\ell-1} (\mathbf{\Delta}^{(b)})^{-1} \mathbf{A}_\ell^\top (\mathbf{\Delta}^{(b)}) &= \eta^{-1} (\mathbf{\Delta}^{(b)})^{-1} \left(\sum_{\ell=0}^\infty \mathbf{A}_\ell \mathbf{G}^\ell \right)^\top (\mathbf{\Delta}^{(b)}) \\ &= \mathbf{0}, \end{aligned}$$

where the last equality follows from (4). Thus, $\eta^{-1}(\mathbf{A}^{(b)})^{-1}\mathbf{G}^\top(\mathbf{A}^{(b)})$ is a nonnegative solution to (15) and, hence, since $\mathbf{R}^{(b)}$ is the minimal nonnegative solution to this equation, it follows that

$$\eta^{-1}(\mathbf{A}^{(b)})^{-1}\mathbf{G}^\top(\mathbf{A}^{(b)}) \geq \mathbf{R}^{(b)}. \tag{17}$$

Inequalities (16) and (17) together imply that (13) holds. An analogous argument establishes (14).

The GI/M/1-type process $\mathcal{G}\mathcal{I}(\mathbf{A}_\ell^{(b)})$ can be thought of as an alternative dual process to Ramaswami’s dual $\mathcal{G}\mathcal{I}(\mathbf{A}_\ell^{(r)})$. We will call this process the *Bright dual* of $\mathcal{M}(\mathbf{A}_\ell)$. It is clear from (13) that, denoting the eigenvalues of \mathbf{G} by $\tau_1, \dots, \tau_m = \eta < 1$, the eigenvalues of $\mathbf{R}^{(b)}$ are τ_i/η . Thus, the spectral radius $\eta^{(b)}$ of $\mathbf{R}^{(b)}$ is always equal to 1 and it follows that the Bright dual is either transient or null recurrent. Moreover, as $\det(\mathbf{A}^{(b)}(z))$ has exactly m zeros in the closed unit disk and any null recurrent process has at least $m + 1$ such zeros [11], the Bright dual must be transient. Similarly, we can define an M/G/1-type Bright dual $\mathcal{M}(\mathbf{A}_\ell^{(b)})$ of a positive recurrent $\mathcal{G}\mathcal{I}(\mathbf{A}_\ell)$ and by considering the eigenvalues of $\mathbf{G}^{(b)}$ we see that $\mathcal{M}(\mathbf{A}_\ell^{(b)})$ is always positive recurrent.

So far, we have defined an alternative dual process when the drift $\mu > 0$. Even though this will turn out to be the most relevant case from a numerical point of view (see Section 4), we can, in most cases, also construct an alternative dual when $\mu < 0$. In the construction of Proposition 1, we defined η by (11) and used the fact that η and \mathbf{w}_η satisfy (12) to prove the proposition. The algebra would go through identically for any other positive scalar and vector that satisfy (12).

In the context of GI/M/1-type Markov chains, Gail *et al.* [12] considered the properties of solutions to equations such as (12). In the discussion following Proposition 7 of [12] they showed that there can be at most two nonnegative real values of η such that the equation has a positive solution \mathbf{w}_η . In terms of the drift μ , their results show that, when $\mu > 0$, there are exactly two nonnegative real values of η (one equal to and one strictly less than 1) for which a positive solution exists to (12), and, for $\mu = 0$, exactly one such value of η (equal to 1). Hence, for $\mu > 0$, the Ramaswami and Bright duals are the only duals that can be constructed in this way, while the Ramaswami dual is unique when $\mu = 0$.

When $\mu < 0$, there can be either one or two real values of η for which a solution to (12) exists. There is only one such η when

- the matrix series $\mathbf{A}(z)$ has radius of convergence 1, or
- $\mathbf{A}(z)$ has a finite radius of convergence $\eta^* > 1$, if $\mathbf{A}(\eta^*)$ converges and if the dominant eigenvalue of $\mathbf{A}(\eta^*)$ is less than 0.

There are two values of η in all other cases, including the important cases where $\mathbf{A}(z)$ is a finite series, an entire function or a rational function, in which case the values of η for which a solution exists are 1 and strictly greater than 1. Examples of chains exhibiting both types of behavior are given after Theorem 5 of [12] (see also Figure 1 of the same paper).

3.2. Case 2: $\mu < 0$

A second solution to (12) exists if $a(z)$ has zeros outside the unit circle and then this solution has η given by

$$\xi = \min\{|z| : |z| > 1, a(z) = 0\}. \tag{18}$$

In this case, we can define the vector w_ξ by (12), and a second dual process, in exactly the same manner as we did in Proposition 1. Using reasoning similar to that used there, the process $\mathcal{G}\mathcal{I}(A_\ell)$ and its alternative dual $\mathcal{M}(A_\ell^{(b)})$ will be both transient, while the process $\mathcal{M}(A_\ell)$ and its alternative dual $\mathcal{G}\mathcal{I}(A_\ell^{(b)})$ will be both positive recurrent. It is worth noting that the asymptotic decay rate of the stationary vector π of $\mathcal{M}(A_\ell)$ equals ξ^{-1} [9]. We summarize this discussion, together with the results of Proposition 1, in the following theorem.

Theorem 2. *Let $\{A_\ell\}$ be such that $A = \sum_{\ell=0}^\infty A_\ell$ is an irreducible generator matrix such that $\mu \neq 0$. Let $\tau = \eta$ if $\mu > 0$, and $\tau = \xi$ if $\mu < 0$ and $a(z)$ has zeros outside the closed unit disk. Define*

$$A_\ell^{(b)} = \tau^{\ell-1}(\Delta^{(b)})^{-1} A_\ell^\top (\Delta^{(b)}),$$

with $\Delta^{(b)} = \text{diag}(w_\tau)$.

Then, the matrices $R^{(b)}$ and $G^{(b)}$ of $\mathcal{G}\mathcal{I}(A_\ell^{(b)})$ and $\mathcal{M}(A_\ell^{(b)})$ are given by

$$R^{(b)} = \tau^{-1}(\Delta^{(b)})^{-1} G^\top (\Delta^{(b)})$$

and

$$G^{(b)} = \tau^{-1}(\Delta^{(b)})^{-1} R^\top (\Delta^{(b)}),$$

respectively.

4. Computational properties of both duals for M/G/1- and GI/M/1-type Markov chains

In order to identify the possible benefits of the Bright dual, we start by taking a closer look at the inner workings of the cyclic reduction algorithm.

4.1. Computational aspects of cyclic reduction

The cyclic reduction (CR) algorithm [5] computes the matrix G of a continuous-time M/G/1-type Markov chain $\mathcal{M}(A_\ell)$, by first transforming it to a discrete-time Markov chain $\mathcal{M}(D_\ell)$ via a simple uniformization. Hence, the matrices A_ℓ for $\ell \geq 0$ are first replaced by $D_\ell = -A_\ell/\lambda$ for $\ell \neq 1$ and by $D_1 = I - A_1/\lambda$, where $\lambda < 0$ is the smallest value on the diagonal of A_1 and I is the identity matrix. Afterwards, the CR algorithm constructs a sequence of matrix power series $\hat{D}^{(n)}(z) = \sum_{\ell=0}^\infty \hat{D}_{\ell+1}^{(n)} z^\ell$ and $D^{(n)}(z) = \sum_{\ell=0}^\infty D_\ell^{(n)} z^\ell$, with $\hat{D}_\ell^{(0)} = D_\ell^{(0)} = D_\ell$ for $\ell \geq 1$, and $D_0^{(0)} = D_0$ and

$$\begin{aligned} \hat{D}^{(n+1)}(z) &= \hat{D}_{\text{even}}^{(n)}(z) + \hat{D}_{\text{odd}}^{(n)}(z)(I - D_{\text{odd}}^{(n)}(z))^{-1} D_{\text{even}}^{(n)}(z), \\ D^{(n+1)}(z) &= z D_{\text{odd}}^{(n)}(z) + D_{\text{even}}^{(n)}(z)(I - D_{\text{odd}}^{(n)}(z))^{-1} D_{\text{even}}^{(n)}(z), \end{aligned}$$

where $F_{\text{even}}(z) = \sum_{\ell=0}^\infty F_{2\ell} z^\ell$ and $F_{\text{odd}}(z) = \sum_{\ell=0}^\infty F_{2\ell+1} z^\ell$, with $F(z) = \sum_{\ell=0}^\infty F_\ell z^\ell$. The pointwise version of the CR algorithm computes $\hat{D}^{(n+1)}(z)$ and $D^{(n+1)}(z)$ by

- taking the fast Fourier transform (FFT) of each of the power series on the right-hand side, that is, we evaluate the series for $z = \exp(2\pi ik/r)$, $k = 0, \dots, r - 1$, with r a power of 2,
- computing the matrix inverses and products on the right-hand side for each $z = \exp(2\pi ik/r)$ to obtain $\hat{D}^{(n+1)}(\exp(2\pi ik/r))$ and $D^{(n+1)}(\exp(2\pi ik/r))$, and
- applying an inverse FFT to retrieve $\hat{D}^{(n+1)}(z)$ and $D^{(n+1)}(z)$.

An important property of this procedure is that, as the degrees of $\hat{\mathbf{D}}^{(n+1)}(z)$ and $\mathbf{D}^{(n+1)}(z)$ increase, we need to increase the number of roots required by the FFT and inverse FFT in order to get an accurate approximation. The CR algorithm starts with a (small) number of roots determined by the degrees of $\hat{\mathbf{D}}_{\text{even}}^{(n)}(z)$ and $\mathbf{D}_{\text{odd}}^{(n)}(z)$, and continues doubling the number of roots until the required accuracy is reached. Thus, the computation time per iteration very much depends on the degrees of the power series $\hat{\mathbf{D}}^{(n+1)}(z)$ and $\mathbf{D}^{(n+1)}(z)$, with smaller degrees resulting in faster computation times.

The matrix \mathbf{G} can be expressed via the series $\hat{\mathbf{D}}^{(n)}(z)$ as

$$\mathbf{G} = \left(\mathbf{I} - \sum_{\ell=0}^{\infty} \hat{\mathbf{D}}_{\ell+1}^{(n)} \mathbf{G}^{\ell 2^n} \right)^{-1} \mathbf{D}_0.$$

If the M/G/1-type Markov chain is positive recurrent (that is, $\mu < 0$), the spectral radius of \mathbf{G} is equal to 1, and so the powers $\mathbf{G}^{\ell 2^n}$ are also stochastic matrices. On the other hand, the degrees of $\hat{\mathbf{D}}^{(n)}(z)$ and $\mathbf{D}^{(n)}(z)$ decrease to 0 and 1 [5, Theorem 7.13], respectively, since, for some positive constant γ and $\ell \geq 2$,

$$\|\hat{\mathbf{D}}_{\ell}^{(n+1)}\| \leq \gamma(\xi - \varepsilon)^{-\ell 2^n}, \quad \|\mathbf{D}_{\ell}^{(n+1)}\| \leq \gamma(\xi - \varepsilon)^{-\ell 2^n},$$

where $\varepsilon > 0$ is small and ξ is defined in (18). Furthermore, the matrices $\hat{\mathbf{D}}_1^{(n)}$ converge to some $\hat{\mathbf{D}}_1^{(\infty)}$ (with a spectral radius less than 1), such that

$$\|\hat{\mathbf{D}}_1^{(n)} - \hat{\mathbf{D}}_1^{(\infty)}\| \leq \gamma \xi^{2^n} (\xi - \varepsilon)^{-2^{n+1}}.$$

Hence, the sum $\sum_{\ell=0}^{\infty} \hat{\mathbf{D}}_{\ell+1}^{(n)} \mathbf{G}^{\ell 2^n}$ converges to $\hat{\mathbf{D}}_1^{(\infty)}$ and

$$\|\mathbf{G} - \mathbf{G}^{(n)}\| \leq \gamma \xi^{2^n} (\xi - \varepsilon)^{-2^{n+1}},$$

with $\mathbf{G}^{(n)} = (\mathbf{I} - \hat{\mathbf{D}}_1^{(n)})^{-1} \mathbf{D}_0$. For the transient case [5, Theorem 7.14], the degrees of $\hat{\mathbf{D}}^{(n)}(z)$ and $\mathbf{D}^{(n)}(z)$ do not decrease to 0 and 1 in general. However, the spectral radius η of \mathbf{G} is less than 1 and, therefore, the powers $\mathbf{G}^{\ell 2^n}$ decrease to 0, that is, $\|\mathbf{G}^{\ell 2^n}\| \leq \gamma'(\eta + \varepsilon)^{2^n}$ for some constant $\gamma' > 0$, and so

$$\|\mathbf{G} - \mathbf{G}^{(n)}\| \leq \gamma'(\eta + \varepsilon)^{2^n}.$$

In short, for a positive-recurrent continuous-time M/G/1-type Markov chain, the speed of convergence of the pointwise CR algorithm is determined by ξ and the amount of work required at each iteration tends to decrease as the degrees of $\hat{\mathbf{D}}^{(n)}(z)$ and $\mathbf{D}^{(n)}(z)$ converge to 0 and 1, respectively. For a transient Markov chain, we find that the speed of convergence is determined by η , while the computation time in general does not decrease with each iteration. As such, provided that $1/\xi$ and η , which determine the speed of convergence, are (nearly) identical, there is a computational advantage in having a positive-recurrent chain.

4.2. Which dual is best?

We are now in a position to identify the possible computational advantage of the Bright dual when computing the \mathbf{R} matrix of a GI/M/1-type Markov chain characterized by $(A_{\ell})_{\ell \geq 0}$.

Positive recurrent case ($\mu > 0$). We start with $\mathcal{G}\mathcal{I}(A_{\ell})$ positive recurrent, implying that η is the dominant eigenvalue of \mathbf{R} . As discussed in Section 2, its Ramaswami dual $\mathcal{M}(A_{\ell}^{(r)})$ is a transient M/G/1-type Markov chain. As such, the speed of convergence is determined by η , while there

is typically no reduction in the computation time of successive iterations. For the Bright dual $\mathcal{M}(A_\ell^{(b)})$ on the other hand, all the zeros of $\det(A^{(b)}(z))$ are identical to those of $a(z)$ divided by $\eta < 1$. In other words, $\xi^{(b)} = \min\{|z| : |z| > 1, \det(A^{(b)}(z)) = 0\} = 1/\eta$. Therefore, running the CR algorithm on $\mathcal{M}(A_\ell^{(b)})$ results in the same asymptotic speed of convergence as the Ramaswami dual; however, as $\mathcal{M}(A_\ell^{(b)})$ is a positive-recurrent Markov chain, the degrees of the series $(\hat{D}^{(b)})^{(n)}(z)$ and $(D^{(b)})^{(n)}(z)$ now do converge to 0 and 1, resulting in a significant reduction in the runtime of the CR algorithm.

Transient case ($\mu < 0$). If the original chain $\mathcal{G}\mathcal{I}(A_\ell)$ is transient, the situation is reversed. We would still find the speed of convergence to be determined by ξ , but now $(\hat{D}^{(r)})^{(n)}(z)$ and $(D^{(r)})^{(n)}(z)$ would converge to degree 0 and 1 polynomials, respectively. Therefore, the Ramaswami dual is superior whenever the chain is transient.

4.3. Combining the Bright and Ramaswami duals

Taking the dual of the dual typically results in the original Markov chain; however, looking at the properties on the recurrence or transience of both duals, the Bright dual of the Ramaswami dual cannot correspond to the original Markov chain. In this section we take a closer look at the resulting Markov chain and investigate its usefulness in computing the G matrix of an M/G/1-type Markov chain.

Transient case ($\mu > 0$). Assume that we wish to compute the G matrix of a transient M/G/1-type Markov chain $\mathcal{M}(A_\ell)$. This can be done directly using the CR algorithm, the convergence speed of which is determined by η . As the chain is transient, the degrees of the series $\hat{D}^{(n)}(z)$ and $D^{(n)}(z)$ do not converge to 0 and 1, respectively.

If we first take the Ramaswami dual of this chain to obtain a positive-recurrent GI/M/1-type Markov chain $\mathcal{G}\mathcal{I}(A_\ell^{(r)})$ and subsequently take the Bright dual of the Ramaswami dual, we end up with a positive-recurrent M/G/1-type Markov chain $\mathcal{M}(A_\ell^{(rb)})$ with

$$A_\ell^{(rb)} = (\Delta^{(rb)})^{-1}(\Delta^{(r)})A_\ell(\Delta^{(r)})^{-1}(\Delta^{(rb)})\eta^{\ell-1},$$

whose G matrix, denoted as $G^{(rb)}$, is related to G by

$$G = (\Delta^{(r)})^{-1}(\Delta^{(rb)})G^{(rb)}(\Delta^{(rb)})^{-1}(\Delta^{(r)})\eta,$$

where $\Delta^{(rb)} = \text{diag}(w^{(r)})$ with $w^{(r)}A^{(r)}(\eta) = \mathbf{0}$ and η is the spectral radius of G .

It is readily verified that $w^{(r)}$ is proportional to $(\Delta^{(r)}v)^\top$, with v such that $A(\eta)v = \mathbf{0}$, as the transpose of $(\Delta^{(r)}v)^\top A^{(r)}(\eta)$ can be written as

$$\sum_{\ell=0}^{\infty} (\Delta^{(r)})A_\ell\eta^\ell(\Delta^{(r)})^{-1}(\Delta^{(r)}v) = (\Delta^{(r)})\left(\sum_{\ell=0}^{\infty} A_\ell\eta^\ell\right)v = \mathbf{0}.$$

This simplifies the above expressions to

$$A_\ell^{(rb)} = \text{diag}(v)^{-1}A_\ell\text{diag}(v)\eta^{\ell-1} \tag{19}$$

and

$$G = \text{diag}(v)G^{(rb)}\text{diag}(v)^{-1}\eta. \tag{20}$$

The zeros of $\det(A^{(rb)}(z))$ are identical to those of $\det(A(z))$ divided by $\eta < 1$, implying that $\xi^{(rb)} = \min\{|z| : |z| > 1, \det(A^{(rb)}(z)) = 0\} = 1/\eta$. Hence, the speeds of convergence of the

transient chain $\mathcal{M}(A_\ell)$ and the positive-recurrent chain $\mathcal{M}(A_\ell^{(rb)})$ are the same. However, the series $(\hat{D}^{(rb)})^{(n)}(z)$ and $(D^{(rb)})^{(n)}(z)$ in the positive-recurrent case converge to degree 0 and 1 polynomials, resulting in a substantial gain in the computation time of each iteration.

It is also possible to use the reverse order, that is, by starting with the Bright dual and subsequently taking the Ramaswami dual. This affects the equations somewhat as

$$A_\ell^{(br)} = (\Delta^{(br)})^{-1}(\Delta^{(b)})A_\ell(\Delta^{(b)})^{-1}(\Delta^{(br)})\eta^{\ell-1}$$

and

$$G = (\Delta^{(b)})^{-1}(\Delta^{(br)})G^{(br)}(\Delta^{(br)})^{-1}(\Delta^{(b)})\eta,$$

where $\Delta^{(b)} = \text{diag}(w_\eta)$ with $w_\eta A(\eta) = \mathbf{0}$, $\Delta^{(br)} = \text{diag}(a^{(b)})$ with $a^{(b)} A^{(b)} = \mathbf{0}$, and η is the spectral radius of G . By checking that $a^{(b)}$ is proportional to $(\Delta^{(b)}v)^\top$, these equations reduce to (19) and (20), and so both orders are equivalent.

Positive recurrent case ($\mu < 0$). Suppose that $\mathcal{M}(A_\ell)$ is a positive-recurrent M/G/1-type Markov chain and that we wish to determine the matrix G . This can be done directly using the CR algorithm, the convergence speed of which is determined by ξ . As the chain is positive recurrent, the degrees of the series $\hat{D}^{(n)}(z)$ and $D^{(n)}(z)$ now converge to 0 and 1, respectively.

As in the transient case, the combination of both duals can be simplified to

$$A_\ell^{(rb)} = \text{diag}(v)^{-1}A_\ell \text{diag}(v)\xi^{\ell-1} \tag{21}$$

and

$$G = \text{diag}(v)G^{(rb)} \text{diag}(v)^{-1}\xi,$$

with v such that $A(\xi)v = \mathbf{0}$. The zeros of $\det(A^{(rb)}(z))$ are identical to those of $a(z)$ divided by $\xi > 1$, implying that

$$\eta^{(rb)} = \max\{|z|: |z| < 1, \det(A^{(rb)}(z)) = 0\} = 1/\xi.$$

Hence, the speeds of convergence of the positive-recurrent chain $\mathcal{M}(A_\ell)$ and the transient chain $\mathcal{M}(A_\ell^{(rb)})$ are the same. However, the series $(\hat{D}^{(rb)})^{(n)}(z)$ and $(D^{(rb)})^{(n)}(z)$ do not converge to degree 0 and 1 polynomials. In this case, it is better to compute G directly as opposed to via $G^{(rb)}$.

Remark. The transformation given in (21) was also used in [6] to transform a positive recurrent M/G/1-type Markov chain into a transient chain (though it was not identified as a combined Bright/Ramaswami dual). As shown in [6], this transformation greatly simplifies the proof of convergence of the cyclic reduction algorithm as well as the convergence of some of the matrix series involved to a matrix polynomial of degree at most 1, when the chain is positive recurrent. Thus, when proving theorems, it is easier to work with a transient chain, while, from a computational point of view, the reverse is true.

4.4. Impact of the shift technique

The shift technique [5, Section 8.2] allows us to accelerate the speed of convergence of the CR algorithm discussed in Section 4.1 as applied to chains of M/G/1 type. Without going into detail, a new sequence of matrices $(\hat{D}_\ell)_{\ell \geq 0}$ is constructed from $(D_\ell)_{\ell \geq 0}$ such that all the zeros of $\det zI - D(z)$ and $\det zI - \hat{D}(z)$ coincide, except for one, say $z = v$, which is shifted to 0 or ∞ (depending on whether the chain is transient or recurrent). In principle, any zero can be

shifted; however, to construct $(\tilde{D}_\ell)_{\ell \geq 0}$, we need to determine ν accurately (and in an efficient manner). The most obvious candidate for ν is $\nu = 1$ as $D(1)e = e$. Depending on the value of the drift μ , we can also easily determine

- $\mu < 0$: the asymptotic decay rate ξ^{-1} of the stationary distribution, using a bisection algorithm on $(1, \infty)$ to find ξ ,
- $\mu > 0$: the spectral radius η of the G matrix, using a bisection algorithm on $(0, 1)$.

In other words, apart from $\nu = 1$, we can also shift the zero $z = \tau$ as defined in Theorem 2. Moreover, shifting both zeros, that is, a double (dbl) shift, causes the strongest convergence acceleration. Table 1 lists the impact of the various shift operations on the speed of convergence of $G^{(n)}$ to G , as well as the rates at which $\hat{D}_\ell^{(n)}(z)$ and $\tilde{D}_\ell^{(n)}(z)$ converge to degree 0 and 1 polynomials (if appropriate), with $\xi_2 = \min\{|z| : |z| > \xi, a(z) = 0\}$ and $\eta_2 = \max\{|z| : |z| < \eta, a(z) = 0\}$. For the ‘one’ shift, the result holds for $\mu \neq 0$. It is important to note that the results in Table 1 apply only if $a(z)$ has no other zeros with modulus 1 and ξ (for $\mu < 0$) and modulus 1 and η (for $\mu > 0$). Otherwise the shift operation will be less effective. By Remark 4.25 of [5], for the first of these cases, it suffices that the doubly infinite M/G/1-type Markov chain characterized by (8) has only one final class and all other states are on a path to the final class. This condition is also sufficient for the $\mu > 0$ case, as all the zeros within the unit disk are eigenvalues of the nonnegative matrices R and G . By Theorem 4.21 of [5], this condition is somewhat stronger than demanding that A has only one final class (which is clearly the case for A irreducible). Next, we argue that the conclusions drawn in Sections 4.2 and 4.3 still hold irrespective of the type of shifting performed.

Which dual is best, revisited. In Section 4.2 we noted that the Bright dual should be preferred when computing the R matrix of a positive-recurrent GI/M/1-type Markov chain, while in the transient case the Ramaswami dual is superior. Recall that the zeros of $\det(A^{(r)}(z))$ are identical to those of $a(z)$, while the zeros of $\det(A^{(b)}(z))$ are found by dividing the zeros of $a(z)$ by τ . In the positive-recurrent case the three shift operations (one, tau, and dbl) therefore

TABLE 1: Impact of the various shift operations.

Operation	$\ G^{(n)} - G\ $	$\ \hat{D}_\ell^{(n)}\ , \ \tilde{D}_\ell^{(n)}\ , \ell \geq 2$
No shift, $\mu < 0$	$\leq \gamma_1 \xi^{2n} (\xi - \varepsilon)^{-2n+1}$	$\leq \gamma_1 \xi^{2n} (\xi - \varepsilon)^{-(\ell+1)2^n}$
No shift, $\mu > 0$	$\leq \gamma_2 (\eta + \varepsilon)^{2n}$	–
Shift $z = 1$	$\leq \gamma_3 \left(\frac{\eta + \varepsilon}{\xi - \varepsilon}\right)^{2n}$	$\leq c_\ell (\xi - \varepsilon)^{-\ell 2^n}$
Shift $z = \tau, \mu < 0$	$\leq \gamma_4 \xi_2^{2n} (\xi_2 - \varepsilon)^{-2n+1}$	$\leq d_\ell \xi_2^{2n} (\xi_2 - \varepsilon)^{-(\ell+1)2^n}$
Shift $z = \tau, \mu > 0$	$\leq \gamma_5 (\eta_2 + \varepsilon)^{2n}$	–
Double shift, $\mu < 0$	$\leq \gamma_6 \left(\frac{\eta + \varepsilon}{\xi_2 - \varepsilon}\right)^{2n}$	$\leq e_\ell (\xi_2 - \varepsilon)^{-\ell 2^n}$
Double shift, $\mu > 0$	$\leq \gamma_7 \left(\frac{\eta_2 + \varepsilon}{\xi - \varepsilon}\right)^{2n}$	$\leq f_\ell (\xi - \varepsilon)^{-\ell 2^n}$

achieve the following rates of convergence (where we neglect the influence of small ε).

- Ramaswami dual ($\mu^{(r)} > 0$): η (no shift), η/ξ (one), η_2 (tau), and η_2/ξ (dbl).
- Bright dual ($\mu^{(b)} < 0$): $1/\xi^{(b)} = 1/(1/\eta) = \eta$ (no shift), $\eta^{(b)}/\xi^{(b)} = (\eta_2/\eta)/(1/\eta) = \eta_2$ (one), $1/(\xi_2)^{(b)} = 1/(\xi/\eta) = \eta/\xi$ (tau), and $\eta^{(b)}/(\xi_2)^{(b)} = (\eta_2/\eta)/(\xi/\eta) = \eta_2/\xi$ (dbl).

Hence, the asymptotic rate of convergence is the same for both duals, except that the effect of shifting the zero in $z = 1$ and $z = \tau$ is reversed by both duals. The alternative dual however has a computational advantage when we look at how quickly $\hat{\mathbf{D}}^{(n)}(z)$ and $\tilde{\mathbf{D}}^{(n)}(z)$ converge to degree 0 and 1 polynomials.

- Ramaswami dual ($\mu^{(r)} > 0$): no convergence (no shift), $1/\xi$ (one), no convergence (tau), and $1/\xi$ (dbl).
- Bright dual ($\mu^{(b)} < 0$): $1/\xi^{(b)} = 1/(1/\eta) = \eta$ (no shift), $1/\xi^{(b)} = 1/(1/\eta) = \eta$ (one), $1/(\xi_2)^{(b)} = 1/(\xi/\eta) = \eta/\xi$ (tau), and $1/(\xi_2)^{(b)} = 1/(\xi/\eta) = \eta/\xi$ (dbl).

In other words, the Bright dual achieves a convergence acceleration, the magnitude of which is determined by $\eta < 1$.

For a transient GI/M/1-type Markov chain, we also observe the same rates of convergence for both duals (where the $z = 1$ and $z = \tau$ shifts are reversed by both duals), that is, $1/\xi$ (no shift), η/ξ (one^(r)/tau^(b)), $1/\xi_2$ (tau^(r)/one^(b)), and η/ξ_2 (dbl). However, this time ($\hat{\mathbf{D}}^{(b)}(z)$) and ($\tilde{\mathbf{D}}^{(b)}(z)$) only converge to degree 0 and 1 polynomials when performing the $z = 1$ and double shift, and overall the Ramaswami dual now achieves an acceleration determined by $1/\xi < 1$.

Combining the Bright and Ramaswami duals, revisited. In Section 4.3 we demonstrated that combining both duals allows us to transform a transient M/G/1-type Markov chain into a positive recurrent one (and vice versa). This is useful as the speed of convergence (mainly determining the number of iterations required) remained identical, while positive recurrence implies that the degrees of $\mathbf{D}^{(n)}(z)$ and $\hat{\mathbf{D}}^{(n)}(z)$ converge, resulting in a lower computational complexity per iteration. When either the $z = 1$ or double shift operation is performed, the degrees of $\tilde{\mathbf{D}}^{(n)}(z)$ and $\hat{\hat{\mathbf{D}}}(z)$ also converge, even if $A(z)$ corresponds to a transient M/G/1-type Markov chain. However, as in the previous paragraph, it is readily verified that $\mathbf{G}^{(n)}$ and $(\mathbf{G}^{(rb)})^{(n)}$ achieve the same rate of convergence for the various shift operations (again, with the $z = 1$ and $z = \tau$ reversed), while, for the convergence of $\mathbf{D}^{(n)}(z)$ and $\hat{\mathbf{D}}^{(n)}(z)$, an acceleration determined by η is realized.

5. Numerical examples

In order to perform these numerical experiments, we extended the functionality of the SMCSolver tool [7] such that it supports the Bright and Ramaswami duals, as well as the double (dbl) shift operation. Here we report the outcome of applying the various algorithms to a randomly generated GI/M/1-type Markov chain characterized by 41 size 20 matrices A_0, A_1, \dots, A_{40} that exhibits the typical performance that we saw when comparing both duals for a positive-recurrent GI/M/1-type Markov chain. We observed cases where the difference between both duals was more dramatic, but we also saw cases where the difference was less pronounced.

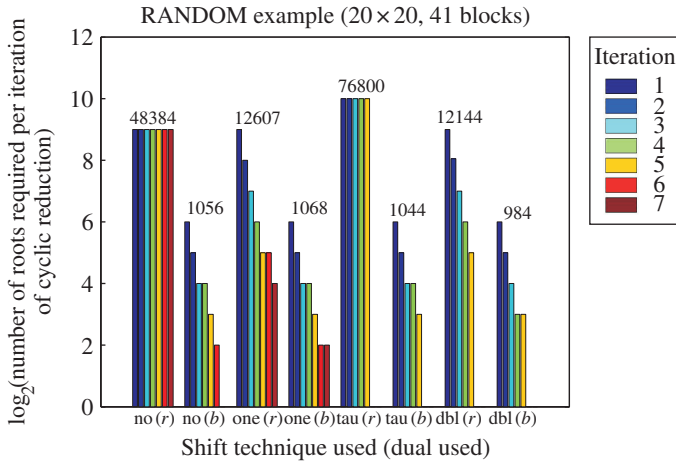


FIGURE 1: Graphical representation of the Bright versus Ramaswami performance for a positive-recurrent GI/M/1-type Markov chain. Each line in a block represents an iteration of the algorithm. The height of the line represents the number of roots required in the corresponding iteration.

Figure 1 depicts the behavior of the CR algorithm for four types of shifting (no shift, one shift, tau shift, and dbl shift), and this for both the Ramaswami (r) and Bright (b) duals. Each iteration is represented by a single bar, the height of which corresponds to (the second logarithm of) the number of roots required to perform the FFTs. The number of roots required typically decreases after each iteration, except for the Ramaswami dual when the no shift or tau shift operation is selected, which is in correspondence with our earlier discussions. The number of roots required per iteration for the tau shift operation may be larger than when no shift is used, while the number of iterations drops from seven to five.

Instead of providing a machine-dependent measurement of the required computation time, we added an estimate for the number of flops needed to perform a single FFT of the required size for each iteration (above the bars), where we estimated the number of flops of a size N FFT as $3/2N \log N$ (the exact number depends on the FFT variant selected and requires at least $N \log N$ multiplications). Thus, the ratio between two such numbers indicates how much faster/slower the computation will be. For instance, without shifting, the Ramaswami dual is about 45 times slower than the Bright dual, on this particular example, which corresponds roughly to the observed computation time, where the Ramaswami dual ran for about 20 seconds, while the Bright dual finished in approximately 0.5 seconds. As we incorporate the shift operations (except for the tau shift), this ratio decreases to about 12, meaning that the difference is less pronounced, but still very substantial.

For the case where the initial GI/M/1-type Markov chain is transient, we made similar observations, but now the Ramaswami dual clearly outperformed the Bright dual. Also, for transient M/G/1-type Markov chains, the combined Ramaswami and Bright duals can also easily achieve performance improvements of a factor 10 or more.

Acknowledgements

The first author gratefully acknowledges the support of the Australian Research Council Centre of Excellence for the Mathematics and Statistics of Complex Systems. Both authors would like to thank an anonymous referee for some valuable suggestions.

References

- [1] AKAR, N. AND SOHRABY, K. (1997). An invariant subspace approach in M/G/1 and G/M/1 type Markov chains. *Commun. Statist. Stoch. Models* **13**, 381–416.
- [2] ASMUSSEN, S. (1989). Aspects of matrix Wiener–Hopf factorisation in applied probability. *Math. Scientist* **14**, 101–116.
- [3] ASMUSSEN, S. AND RAMASWAMI, V. (1990). Probabilistic interpretations of some duality results for the matrix paradigms in queueing theory. *Commun. Statist. Stoch. Models* **6**, 715–733.
- [4] BINI, D. A. AND MEINI, B. (1995). On cyclic reduction applied to a class of Toeplitz-like matrices arising in queueing problems. In *Computations with Markov Chains*, ed. W. J. Stewart, Kluwer, Boston. pp. 21–28.
- [5] BINI, D. A., LATOUCHE, G. AND MEINI, B. (2005). *Numerical Methods for Structured Markov Chains*. Oxford University Press.
- [6] BINI D.A., MEINI B. AND RAMASWAMI V. (2008). A probabilistic interpretation of cyclic reduction and its relationships with logarithmic reduction. *Calcolo* **45**, 207–216.
- [7] BINI, D. A., MEINI, B., STEFFÉ, S. AND VAN HOUTD, B. (2006). Structured Markov chains solver: software tools. In *Proc. SMCTools Workshop* (Pisa, 2006; ACM Internat. Conf. Proc. Ser. **201**), ACM Press, New York.
- [8] BRIGHT, L. (1996). Matrix-analytic methods in applied probability. Doctoral Thesis, University of Adelaide.
- [9] FALKENBERG, E. (1994). On the asymptotic behaviour of the stationary distribution of Markov chains of M/G/1-type. *Commun. Statist. Stoch. Models* **10**, 75–97.
- [10] GAIL, H. R., HANTLER, S. L. AND TAYLOR, B. A. (1994). Solutions of the basic matrix equation for M/G/1 and G/M/1 type Markov chains. *Commun. Statist. Stoch. Models* **10**, 1–44.
- [11] GAIL, H. R., HANTLER, S. L. AND TAYLOR, B. A. (1996). Spectral analysis of M/G/1 and G/M/1 type Markov chains. *Adv. Appl. Prob.* **28**, 114–165.
- [12] GAIL, H. R., HANTLER, S. L. AND TAYLOR, B. A. (1998). Matrix-geometric invariant measures for G/M/1 type Markov chains. *Commun. Statist. Stoch. Models* **14**, 537–569.
- [13] KELLY, F. P. (1983). Invariant measures and the q-matrix. In *Probability, Statistics and Analysis* (London Math. Soc. Lecture Notes **79**), eds J. F. C. Kingman and G. E. H. Reuter, Cambridge University Press, pp. 143–160.
- [14] LATOUCHE, G. (1993). Algorithms for infinite Markov chains with repeating columns. In *Linear Algebra, Markov Chains, and Queueing Models* (IMA Vol. Math. Appl. **48**), Springer, New York, pp. 231–265.
- [15] LATOUCHE, G. (1994). Newton’s iteration for non-linear equations in Markov chains. *IMA J. Numer. Anal.* **14**, 583–598.
- [16] LATOUCHE, G. AND RAMASWAMI, V. (1993). A logarithmic reduction algorithm for quasi-birth-and-death processes. *J. Appl. Prob.* **30**, 650–674.
- [17] LATOUCHE, G. AND RAMASWAMI, V. (1999). *Introduction to Matrix Analytic Methods in Stochastic Modeling*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- [18] NEUTS, M. F. (1978). Markov chains with applications in queueing theory, which have a matrix-geometric invariant probability vector. *Adv. Appl. Prob.* **10**, 185–212.
- [19] NEUTS, M. F. (1981). *Matrix-Geometric Solutions in Stochastic Models*. John Hopkins University Press, Baltimore, MD.
- [20] NEUTS, M. F. (1989). *Structured Stochastic Matrices of M/G/1 and Their Applications*. Marcel Dekker, New York.
- [21] RAMASWAMI, V. (1990). A duality theorem for the matrix paradigm in queueing theory. *Commun. Statist. Stoch. Models* **6**, 151–161.