University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Dendievel, S;Hautphenne, S;Latouche, G;Taylor, PG

Title:
The time-dependent expected reward and deviation matrix of a finite QBD process

Date:
2019-06-01

Citation:
Dendievel, S., Hautphenne, S., Latouche, G. & Taylor, P. G. (2019). The time-dependent expected reward and deviation matrix of a finite QBD process. Linear Algebra and Its Applications, 570, pp.61-92. https://doi.org/10.1016/j.laa.2019.02.002.

Persistent Link:
https://hdl.handle.net/11343/249540

The time-dependent expected reward and deviation matrix of a finite QBD process

Sarah Dendievel[*], Sophie Hautphenne[†], Guy Latouche[‡], and Peter Taylor[§]

January 14, 2019

Abstract

Deriving the time-dependent expected reward function associated with a continuous-time Markov chain involves the computation of its transient deviation matrix. In this paper we focus on the special case of a finite quasi-birth-and-death (QBD) process, motivated by the desire to compute the expected revenue lost in a MAP/PH/1/C queue.

We use two different approaches in this context. The first is based on the solution of a finite system of matrix difference equations; it provides an expression for the blocks of the expected reward vector, the deviation matrix, and the mean first passage time matrix. The second approach, based on some results in the perturbation theory of Markov chains, leads to a recursive method to compute the full deviation matrix of a finite QBD process. We compare the two approaches using some numerical examples.

Keywords: Finite quasi-birth-and-death process; expected reward; deviation matrix; matrix difference equations; perturbation theory.

AMS codes: 60J27, 60J99, 15A24, 47A55, 15A09.

[*]Ghent University, Department of Telecommunications and Information Processing, SMACS Research Group, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium, Sarah.Dendievel@UGent.be

[†]Ecole Polythechnique Fédérale de Lausanne, Institute of Mathematics, and The University of Melbourne, School of Mathematics and Statistics, sophiemh@unimelb.edu.au

[‡]Université libre de Bruxelles, Faculté des sciences, CP212, Boulevard du Triomphe 2, 1050 Bruxelles, Belgium, latouche@ulb.ac.be

[§]The University of Melbourne, School of Mathematics and Statistics, pgt@ms.unimelb.edu.au

1 Introduction

Our analysis of the expected reward in a finite QBD process is motivated by the problem of computing the expected amount of lost revenue in a *MAP/PH/1/C* queue over a finite time horizon $[0, t]$, given its initial occupancy. Since MAPs and PH distributions are the most general matrix extensions of Poisson processes and exponential distributions, respectively, we can think of this problem as a matrix generalisation of the similar analysis for the $M/M/C/C$ model considered in Chiera and Taylor [7] and the $M/M/1/C$ model in Braunsteins, Hautphenne and Taylor [5]. In these queueing models, customers are lost when they arrive to find C customers already present. Assuming that each arriving customer brings a certain amount of revenue, we are interested in calculating the expected amount of revenue that the queue will lose over a finite time horizon $[0, t]$, as well as exploring the limit of the rate of losing revenue in the asymptotic regime as $t \to \infty$.

Solving the expected lost revenue problem is important, for example, if we wish to find a way of managing a system where a number of, possibly different, queues share a number of servers. If it is feasible to reallocate servers from one queue to another every t time units, then a rational method for performing the allocation is for each queue to observe its occupancy at time 0, calculate the expected revenue lost in time $[0, t]$ for a range of capacities, given its initial occupancy, and then to allocate the servers to minimise the total expected amount of lost revenue over $[0, t]$. At time t, the calculation can be performed again, based upon the occupancies at that time and a reallocation performed if it is optimal to do so.

The *MAP/PH/1/C* queue can be modelled as a finite QBD process with generator matrix Q, and levels $0, 1, \ldots, C$ corresponding to the possible queue lengths. If $\boldsymbol{R}(t)$ is a vector containing the expected revenue lost in $[0, t]$ conditional on the initial state (level and phase) of the system, then computing $\boldsymbol{R}(t)$ reduces to solving a special case of the time-dependent version of Poisson's equation of the form

$$\begin{aligned} \boldsymbol{R}(0) &= \boldsymbol{0} \\ \boldsymbol{R}'(t) &= Q\boldsymbol{R}(t) + \boldsymbol{g}, \end{aligned} \tag{1}$$

where $\boldsymbol{0}$ is a column vector of 0's, and $\boldsymbol{g}$ is a column vector containing the reward (loss) per unit of time in each state of the system. Since the *MAP/PH/1/C* system loses revenue only when it is at full capacity, the only non-zero entries of $\boldsymbol{g}$ in our motivating example are those corresponding to level C.

The solution of (1), given by

$$\boldsymbol{R}(t) = (\boldsymbol{\pi g})\mathbf{1}\, t + D(t)\, \boldsymbol{g}, \tag{2}$$

where $\boldsymbol{\pi}$ is the stationary vector of the QBD process and $\mathbf{1}$ denotes the column vector of 1's, involves the transient deviation matrix,

$$D(t) = \int_0^t \left(e^{Qu} - \mathbf{1}\boldsymbol{\pi}\right) du, \tag{3}$$

see [5]. As $t \to \infty$, $D(t)$ converges to the deviation matrix D discussed in Coolen-Schrijner and van Doorn [8], which is the group inverse of $-Q$, and the expected lost revenue function has a linear asymptote, $\boldsymbol{R}(t) \sim (\boldsymbol{\pi g})\mathbf{1}\, t + D\boldsymbol{g}$.

After providing more detail on the reward function $\boldsymbol{R}(t)$ associated with a QBD process in Section 2, we tackle the computation of $\boldsymbol{R}(t)$ in transient and asymptotic regimes, and the corresponding matrices $D(t)$ and D, using two different approaches, each having some advantages in comparison to the other. In the first approach, developed in Section 3, we place ourselves in the general context where the reward vector $\boldsymbol{g}$ is not restricted to any particular structure. We use systems of matrix difference equations to gain insight into the block-structure of the vector $\boldsymbol{R}(t)$ and of the matrices $D(t)$ and D. In addition, this method also provides us with the blocks of the matrix of the mean first passage times in the QBD process. We obtain simple expressions for the relevant quantities that highlight the role of the maximal capacity C. We also derive limiting results when the maximal capacity increases to infinity. This approach is effective if one wants to focus on particular blocks of $\boldsymbol{R}(t)$, $D(t)$ and D rather than on the full matrices, as is the case in our motivating example. In practice, it also avoids dealing with large matrices: the size of the matrices involved in the expressions is at most twice the size of the phase space of the QBD process, regardless of the value of C.

The second approach, described in Section 4, relies on some elegant results from the perturbation theory of Markov chains, and leads to a recursive formula expressing the full deviation matrices $D(t)$ and D for a given value of C in terms of the corresponding matrices for a system with capacity $C-1$. Our result in Propositions 4.3 and 4.4 is an extension of Theorem 3.1 in Langville and Meyer [15], where the authors consider the modification of a single row, while we deal with a block modification.

We should mention that many updating formulas exist in the literature and are cited in [15]. In a recent paper [12], Hunter adapts the idea of [15] and devotes Section 5 to computing various key performance measures of a perturbed discrete-time Markov chain with transition matrix $\overline{P}$ expressed as the sum of an original transition matrix P and some perturbation matrix E. Amongst the measures of interest are the stationary distribution of $\overline{P}$, mean first passage times and various generalized inverses of $I - \overline{P}$, one of which is the deviation matrix. Section 6 of that paper provides algorithms for updating generalized inverses via successive row perturbations. One difference here is that we increase the state space of the QBD, in addition to perturbing the transition matrix, and we proceed in the spirit of the Sherman-Morrison-Woodbury (SMW) formula. We recall this formula in Section 4.1 and we give the details of our approach in Section 4.2.

Concerning perturbation analysis of block structured Markov chains, we may also refer to Avrachenkov *et al.* [2], and more recently to Jiang *et al.* [13] who build on the results of [2]. Both papers aim at extending general results which have long existed for finite Markov chains to the denumerable case. In both papers, there is no assumption about the structure of the perturbation matrix, but its elements are assumed to be small with respect to the non-zero elements of the unperturbed transition matrix. The main result in these papers is to express the stationary vector of the perturbed chain as a power series involving the stationary distribution and the deviation matrix of the unperturbed chain, and QBDs are used as illustrations. In contrast to [2] and [13], we look at *finite* QBDs with a very specific perturbation of the rank one type, of the same order of magnitude as the remainder of the transition matrices, and derive an expression for the deviation matrix of the perturbed process in terms of that of the unperturbed process.

The approaches described in Sections 3 and 4 complement each other and their use may depend on the context: the first method is algebraic because we solve matrix difference equations, but the solutions have a probabilistic interpretation, while the second method is probabilistic but leads to a solution with an algebraic flavour.

In Section 5, we provide numerical illustrations, starting with our motivating *MAP/PH/1/C* example which loses revenue only at full capacity, and moving on to look at a case where the reward function $\boldsymbol{g}$ is non-zero for all levels of the system. We compare the computational complexity of the two approaches applied to general examples and observe that there is a threshold value of C at which one method surpasses the other in terms of CPU time.

2 Background

Let $\{X(t) : t \geq 0\}$ be an ergodic continuous-time Markov chain on a finite state-space, with generator Q and stationary distribution $\boldsymbol{\pi}$. The *deviation matrix* of $\{X(t)\}$ is the matrix

$$D := \int_0^\infty \left(e^{Qu} - \mathbf{1}\boldsymbol{\pi}\right) du, \tag{4}$$

whose components may be written as $D_{ij} = \lim_{t\to\infty}[N_{ij}(t) - N_{\boldsymbol{\pi}j}(t)]$, where $N_{ij}(t)$ is the expected time spent in state j during the interval of time $[0, t]$ given that the initial state is i, and $N_{\boldsymbol{\pi}j}(t)$ is the same quantity but conditional on the initial state having the distribution $\boldsymbol{\pi}$ (see Da Silva Soares and Latouche [9]).

The *group inverse* $A^\#$ of a matrix A, if it exists, is defined as the unique solution to $AA^\#A = A$, $A^\#AA^\# = A^\#$, and $A^\#A = AA^\#$. From Campbell and Meyer [6, Theorem 8.5.5], the group inverse $Q^\#$ of the infinitesimal generator Q of any finite Markov chain is the unique solution of the system

$$QQ^\# = I - W, \tag{5}$$

$$WQ^\# = 0, \tag{6}$$

where $W := \lim_{t\to\infty}[\exp(Qt)]$. If Q has a unique stationary distribution $\boldsymbol{\pi}$, then $W = \mathbf{1}\boldsymbol{\pi}$, and so

$$QQ^\# = I - \mathbf{1}\boldsymbol{\pi}. \tag{7}$$

When it exists, the deviation matrix is related to the group inverse $Q^\#$ of Q by the relation

$$D = -Q^\#. \tag{8}$$

In addition, D also satisfies

$$D\mathbf{1} = \mathbf{0}; \tag{9}$$

see [8] for more detail on deviation matrices. We see that $D = \lim_{t\to\infty} D(t)$, where $D(t)$ is the transient deviation matrix defined by (3). The deviation matrices have the explicit forms

$$D(t) = (I - e^{Qt})(\mathbf{1}\boldsymbol{\pi} - Q)^{-1}, \qquad D = (\mathbf{1}\boldsymbol{\pi} - Q)^{-1} - \mathbf{1}\boldsymbol{\pi}. \tag{10}$$

More properties of $D(t)$ are discussed in [5].

Define the matrix $N(t) := (N_{ij}(t))$. Clearly,

$$N(t) = \int_0^t e^{Qu} du = \mathbf{1}\boldsymbol{\pi}\, t + D(t),$$

and $N(t)$ has the linear asymptote $\bar{N}(t) := \mathbf{1}\boldsymbol{\pi}\, t + D$. If we associate a reward (or loss) g_j per time unit when the Markov chain $\{X(t)\}$ occupies state j, and define the vector $\boldsymbol{g} := (g_j)$, then the expected cumulative reward up to time t, given that the chain starts in state i, is given by $R_i(t) := (N(t)\boldsymbol{g})_i$, so that the vector $\boldsymbol{R}(t) := (R_i(t))$ satisfies

$$\boldsymbol{R}(t) = (\boldsymbol{\pi g})\mathbf{1}\, t + D(t)\boldsymbol{g}, \tag{11}$$

and $\boldsymbol{R}(t)$ has the linear asymptote

$$\bar{\boldsymbol{R}}(t) := (\boldsymbol{\pi g})\mathbf{1}\, t + D\boldsymbol{g}. \tag{12}$$

Observe that (11) is the solution of a finite horizon version of Poisson's equation,

$$\boldsymbol{R}'(t) = Q\boldsymbol{R}(t) + \boldsymbol{g} \tag{13}$$

with $\boldsymbol{R}(0) = \mathbf{0}$. In the Laplace transform domain where, for Real$(s) > 0$,

$$\tilde{\boldsymbol{R}}(s) := \int_0^\infty e^{-st}\boldsymbol{R}(t)\, dt,$$

(11) and (13) become respectively

$$\tilde{\boldsymbol{R}}(s) = (1/s^2)\,(\boldsymbol{\pi g})\mathbf{1} + \tilde{D}(s)\,\boldsymbol{g} \tag{14}$$

$$= (sI - Q)^{-1}(1/s)\boldsymbol{g}, \tag{15}$$

where $\tilde{D}(s)$ is the Laplace transform of the transient deviation matrix, given by

$$\tilde{D}(s) = (1/s)(sI - Q)^{-1} - (1/s^2)\mathbf{1}\boldsymbol{\pi}. \tag{16}$$

Another equivalent expression, obtained after some algebraic manipulation, is given by

$$\tilde{D}(s) = (1/s)(sI - Q)^{-1}(I - \mathbf{1}\boldsymbol{\pi}).$$

An explicit solution for $\tilde{\boldsymbol{R}}(s)$ therefore exists, however, direct computation of the inverse $(sI - Q)^{-1}$ may be numerically costly for large systems. For structured Markov chains, efficient computational methods exploiting the specific structure can be developed, as we highlight in the next sections.

In this paper, we are interested in computing the expected reward function $\boldsymbol{R}(t)$, and the associated deviation matrices $D(t)$ and D, for any $t \geq 0$, when the Markov chain corresponds to a finite level-independent QBD process, that is, a two-dimensional continuous-time Markov chain $\{\boldsymbol{X}(t) = (J(t), \varphi(t)), t \geq 0\}$, where the variable $J(t)$, taking values in $\{0, 1, \ldots, C\}$, is called the *level* of the process at time t, and the variable $\varphi(t)$, taking values in $\{1, 2, \ldots, n\}$, is called the *phase* of the process at time t. The generator of the QBD process has a block-tridiagonal form given by

$$Q = \begin{array}{c} \scriptstyle 0 \\ \scriptstyle 1 \\ \scriptstyle 2 \\ \vdots \\ \\ \scriptstyle C-1 \\ \scriptstyle C \end{array} \overset{\begin{array}{ccccccc} \scriptstyle 0 & \scriptstyle 1 & \scriptstyle 2 & \scriptstyle \ldots & \scriptstyle C-2 & \scriptstyle C-1 & \scriptstyle C \end{array}}{\begin{bmatrix} B_0 & A_1 & & & & & \\ A_{-1} & A_0 & A_1 & & & & \\ & A_{-1} & A_0 & \ddots & & & \\ & & \ddots & \ddots & & & \\ & & & & & & \\ & & & & A_{-1} & A_0 & A_1 \\ & & & & & A_{-1} & C_0 \end{bmatrix}}, \tag{17}$$

where the block matrices are of size $n \times n$.

The vectors $\boldsymbol{R}(t)$ and $\boldsymbol{g}$ corresponding to the QBD process are of length $(C+1)n$ and can be structured into block sub-vectors $\boldsymbol{R}_k(t)$ and $\boldsymbol{g}_k$ of length n, for $0 \leq k \leq C$, corresponding to each level of the QBD process. The same block decomposition holds for the deviation matrices $D(t)$ and D, and we write $D_{k,\ell}(t) = \big(D_{(k,i)(\ell,j)}(t)\big)_{1\leq i,j\leq n}$ ($D_{k,\ell} = \big(D_{(k,i)(\ell,j)}\big)$ respectively) for the (k, ℓ)th block of these matrices.

The next two sections address the following questions:

(i) How do we use the structure of the QBD process to compute the blocks of $\boldsymbol{R}(t)$, $D(t)$ and D?

(ii) How do the deviation matrices $D(t)$ and D differ for two successive values of the capacity C?

3 A matrix difference equation approach

In this section, we answer Question *(i)*. We deal with the transient regime first, followed by the asymptotic regime.

3.1 Transient regime

We analyse the transient behaviour of the expected reward function in the Laplace transform domain. An explicit expression for $\tilde{\boldsymbol{R}}(s)$ is given in (15) but this expression does not give any insight in the structure of the blocks $\tilde{\boldsymbol{R}}_k(s)$, which are conditional on the initial queue length, nor in the role of the maximum capacity C in the solution. Instead, we rewrite (15) as a system of matrix second-order difference equations

$$\begin{aligned}(B_0 - sI)\tilde{\boldsymbol{R}}_0(s) + A_1\,\tilde{\boldsymbol{R}}_1(s) &= -\boldsymbol{g}_0(s) && (18)\\ A_{-1}\,\tilde{\boldsymbol{R}}_{k-1}(s) + (A_0 - sI)\tilde{\boldsymbol{R}}_k(s) + A_1\,\tilde{\boldsymbol{R}}_{k+1}(s) &= -\boldsymbol{g}_k(s),\ 1 \le k \le C-1 && (19)\\ A_{-1}\,\tilde{\boldsymbol{R}}_{C-1}(s) + (C_0 - sI)\tilde{\boldsymbol{R}}_C(s) &= -\boldsymbol{g}_C(s), && (20)\end{aligned}$$

where $\boldsymbol{g}_k(s) = (1/s)\,\boldsymbol{g}_k$, and we focus on the computation of the blocks $\tilde{\boldsymbol{R}}_k(s)$, for $0 \le k \le C$. We now introduce the ingredients of the solution.

For any $s \ge 0$, let $G(s)$ be the minimal nonnegative solution to the matrix quadratic equation

$$A_{-1} + (A_0 - sI)X + A_1 X^2 = 0, \tag{21}$$

and let $\hat{G}(s)$ be the minimal nonnegative solution to

$$A_1 + (A_0 - sI)X + A_{-1} X^2 = 0. \tag{22}$$

The matrices $G(s)$ and $\hat{G}(s)$ can be computed numerically using any of the linear or quadratic algorithms discussed in Latouche and Ramaswami [16]. In the doubly infinite QBD process $\{(Y(t), \phi(t)), t \ge 0\}$ with no boundary levels and generator

$$\begin{bmatrix} \ddots & \ddots & & & \\ \ddots & A_0 & A_1 & & \\ & A_{-1} & A_0 & A_1 & \\ & & A_{-1} & A_0 & \ddots \\ & & & \ddots & \ddots \end{bmatrix}, \tag{23}$$

the (i,j)th entry of the matrix $G(s)$ for $s > 0$ contains the Laplace transform

$$\mathbb{E}\left[e^{s\theta_{k-1}}\,\mathbb{1}\{\theta_{k-1} < \infty,\ \phi(\theta_{k-1}) = j\}\,|\,Y(0) = k, \phi(0) = i\right], \tag{24}$$

where $\theta_{k-1} := \inf_{t>0}\{Y(t) = k-1\}$ and $\mathbb{1}\{\cdot\}$ denotes the indicator function, and the (i,j)th entry of the matrix $\hat{G}(s)$ contains the Laplace transform

$$\mathbb{E}\left[e^{s\theta_{k+1}}\,\mathbb{1}\{\theta_{k+1}<\infty,\,\phi(\theta_{k+1})=j\}\,|\,Y(0)=k,\phi(0)=i\right]. \tag{25}$$

Both $G(s)$ and $\hat{G}(s)$ are sub-stochastic but not stochastic, and therefore have a spectral radius strictly less than 1, as long as $s > 0$. In the doubly-infinite QBD, the first passage probability matrices G and $\hat{G}$ correspond to $G(0)$ and $\hat{G}(0)$. When the doubly-infinite QBD process is transient and has positive drift, G is sub-stochastic but not stochastic and $\hat{G}$ is stochastic, and conversely, when it has negative drift, G is stochastic and $\hat{G}$ is sub-stochastic but not stochastic. Finally, we let

$$H_0(s) = -(A_0 - sI + A_1 G(s) + A_{-1}\hat{G}(s))^{-1},$$

which is well defined for any $s > 0$, and

$$H_0 = -(A_0 + A_1 G + A_{-1}\hat{G})^{-1}, \tag{26}$$

which is well defined except when the QBD with generator (23) is null-recurrent.

In the sequel, we use the convention that an empty sum, such as $\sum_{j=1}^{0}$ or $\sum_{j=0}^{-1}$, is zero.

Lemma 3.1 *For any $C \geq 2$, the solutions of the system of matrix second-order difference equations*

$$A_{-1}\,\boldsymbol{u}_{k-1} + A_0\,\boldsymbol{u}_k + A_1\,\boldsymbol{u}_{k+1} = -\boldsymbol{g}_k, \qquad 1 \leq k \leq C-1, \tag{27}$$

are given by vectors of the form

$$\boldsymbol{u}_k = G^k\,\boldsymbol{v} + \hat{G}^{C-k}\boldsymbol{w} + \boldsymbol{\nu}_k(C), \qquad 0 \leq k \leq C, \tag{28}$$

where $\boldsymbol{v}$ and $\boldsymbol{w}$ are arbitrary vectors in $\mathbb{C}^n$, and

$$\boldsymbol{\nu}_k(C) = \sum_{j=0}^{k-1} G^j H_0\,\boldsymbol{g}_{k-j} + \sum_{j=1}^{C-k} \hat{G}^j H_0\,\boldsymbol{g}_{k+j}.$$

Proof. First, using (21), (22) and (26) we can show that (28) is a solution to (27) for arbitrary vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{C}^n$. It remains to show that all solutions of (27) can be expressed in the form of (28).

Observe that there exists a nonsingular matrix M such that

$$\hat{G}M = MJ \tag{29}$$

with

$$J = \begin{bmatrix} V & 0 \\ 0 & W \end{bmatrix}, \tag{30}$$

where V is a non-singular square matrix of order p, and W is a square matrix of order q with $p + q = n$ and $\mathrm{sp}(W) = 0$. For instance, we may choose M such that J is the Jordan normal form of $\hat{G}$; in this case, V contains all the blocks for the non-zero eigenvalues of $\hat{G}$ (which lie within the unit circle), and W contains all the blocks for the zero eigenvalues. The matrices M and M^{-1} may be partitioned as

$$M = \left[\, L \mid K \,\right], \qquad M^{-1} = \left[\begin{array}{c} E \\ \hline F \end{array}\right], \tag{31}$$

where L has dimension $n \times p$, E has dimension $p \times n$, and $EL = I$, $FL = 0$. Lemma 10 in Bini *et al.* [3] states that the general solution of (27) is given, for $k \geq 1$, by

$$\boldsymbol{u}_k = G^k \boldsymbol{v} + LV^{-k} \boldsymbol{z} + \boldsymbol{\sigma}_k, \tag{32}$$

where $\boldsymbol{v} \in \mathbb{C}^n$ and $\boldsymbol{z} \in \mathbb{C}^p$ are arbitrary vectors, and

$$\boldsymbol{\sigma}_k = \sum_{j=0}^{k-1} (G^j - LV^{-j}E) H_0 \, \boldsymbol{g}_{k-j} + \boldsymbol{\tau}_k, \tag{33}$$

with

$$\boldsymbol{\tau}_k = \sum_{j=1}^{\nu-1} KW^j F H_0 \, \boldsymbol{g}_{k+j},$$

and ν the smallest integer such that $W^\nu = 0$ ($\boldsymbol{\tau}_k = \boldsymbol{0}$ if $\hat{G}$ is invertible).

Since the eigenvalues of V lie within the unit circle, the negative powers of V appearing in (32) and (33) have entries that can take unbounded values, rendering the expressions unsuitable for computational purposes. Therefore, we rewrite (32) in a more convenient form in order to get rid of any negative power of V. First observe that (32) is equivalent to

$$\boldsymbol{u}_k = G^k \boldsymbol{v} + \sum_{j=0}^{k-1} G^j H_0 \, \boldsymbol{g}_{k-j} + LV^{-k} (\boldsymbol{z} - \sum_{j=0}^{k-1} V^{k-j} E H_0 \, \boldsymbol{g}_{k-j}) + \boldsymbol{\tau}_k, \tag{34}$$

where the negative powers of V appear in the term

$$\boldsymbol{s}_k \;:=\; LV^{-k}\,(\boldsymbol{z} - \sum_{j=0}^{k-1} V^{k-j} EH_0\,\boldsymbol{g}_{k-j}) = LV^{-k}\,(\boldsymbol{z} - \sum_{j=1}^{k} V^{j} EH_0\,\boldsymbol{g}_{j}).$$

Letting $\boldsymbol{z} = \boldsymbol{y} + \sum_{j=1}^{C} V^j EH_0\,\boldsymbol{g}_j$, where $\boldsymbol{y} \in \mathbb{C}^p$ is an arbitrary vector, we obtain

$$\boldsymbol{s}_k \;= LV^{-k}\,\boldsymbol{y} + \textstyle\sum_{j=1}^{C-k} LV^j EH_0\,\boldsymbol{g}_{k+j}.$$

Let us fix a vector $\boldsymbol{y} \in \mathbb{C}^p$. Then, for any $k \le C$, $LV^{-k}\,\boldsymbol{y}$ can be equivalently written as $\hat{G}^{C-k}\boldsymbol{w}$ with $\boldsymbol{w} = LV^{-C}\boldsymbol{y}$. Indeed,

$$\begin{aligned} \hat{G}^{C-k}\boldsymbol{w} &= \hat{G}^{C-k}LV^{-C}\boldsymbol{y} \\ &= [LV^{C-k}E + KW^{C-k}F]LV^{-C}\boldsymbol{y} \\ &= LV^{C-k}V^{-C}\boldsymbol{y} \\ &= LV^{-k}\boldsymbol{y}. \end{aligned}$$

Therefore, we have

$$\boldsymbol{s}_k = \hat{G}^{C-k}\boldsymbol{w} + \sum_{j=1}^{C-k} LV^j EH_0\,\boldsymbol{g}_{k+j},$$

and the general solution (34) takes the form

$$\boldsymbol{u}_k = G^k\,\boldsymbol{v} + \hat{G}^{C-k}\boldsymbol{w} + \boldsymbol{\nu}_k(C),$$

where

$$\boldsymbol{\nu}_k(C) : \;=\; \sum_{j=0}^{k-1} G^j H_0\,\boldsymbol{g}_{k-j} + \sum_{j=1}^{C-k} LV^j EH_0\,\boldsymbol{g}_{k+j} + \sum_{j=1}^{\nu-1} KW^j FH_0\,\boldsymbol{g}_{k+j}.$$

Finally, note that $\boldsymbol{g}_\ell$ is defined for $0 \le \ell \le C$ only, so we can set $\boldsymbol{g}_{k+j} = \boldsymbol{0}$ for any $j > C - k$. With this, using (29), (30), and (31), we have

$$\begin{aligned} \sum_{j=1}^{C-k} LV^j EH_0\,\boldsymbol{g}_{k+j} + \sum_{j=1}^{\nu-1} KW^j FH_0\,\boldsymbol{g}_{k+j} &= \sum_{j=1}^{C-k}(LV^j E + KW^j F)H_0\,\boldsymbol{g}_{k+j} \\ &= \sum_{j=1}^{C-k} \hat{G}^j H_0\,\boldsymbol{g}_{k+j}, \end{aligned}$$

which shows that any solution to (27) can be written in the form (28). □

The advantage of the solution (28) over the solution (32) from [3] is that it does not require any spectral decomposition of the matrix $\hat{G}$, nor any matrix inversion (as long as $C \geq k$). Since the spectral radii of G and $\hat{G}$ are bounded by 1, all matrix powers involved in (28) are bounded, and the computation of the solution is therefore numerically stable.

Lemma 3.1 naturally extends to the Laplace transform domain, as stated in the next corollary.

Corollary 3.2 *For any $s > 0$, the solutions of the system of matrix second-order difference equations*

$$A_{-1}\,\boldsymbol{u}_{k-1}(s)+(A_0-sI)\,\boldsymbol{u}_k(s)+A_1\,\boldsymbol{u}_{k+1}(s) = -\boldsymbol{g}_k(s) \qquad 1 \leq k \leq C-1, \quad (35)$$

are given by vector of the form

$$\boldsymbol{u}_k(s) = G(s)^k\,\boldsymbol{v}(s) + \hat{G}(s)^{C-k}\,\boldsymbol{w}(s) + \boldsymbol{\nu}_k(s, C), \qquad 0 \leq k \leq C, \quad (36)$$

where $\boldsymbol{v}(s)$ and $\boldsymbol{w}(s)$ are arbitrary vectors, and

$$\boldsymbol{\nu}_k(s, C) = \sum_{j=0}^{k-1} G(s)^j H_0(s)\,\boldsymbol{g}_{k-j}(s) + \sum_{j=1}^{C-k} \hat{G}(s)^j H_0(s)\,\boldsymbol{g}_{k+j}(s). \quad (37)$$

□

We use Corollary 3.2 to obtain a closed-form expression for $\tilde{\boldsymbol{R}}_k(s)$ in terms of the matrices $G(s)$ and $\hat{G}(s)$ of the QBD process. The following proposition makes use of the generator of the transient Markov chain obtained from the QBD process in which absorption can happen from any state at rate $s > 0$, and censored to levels 0 and C. This generator is given by

$$\mathring{Q}(s, C) = \mathcal{A}(s) + \mathcal{B}(-\mathcal{C}(s))^{-1}\mathcal{D} \quad (38)$$

with

$$\mathcal{A}(s) = \begin{bmatrix} (B_0 - sI) & 0 \\ 0 & (C_0 - sI) \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & \dots & 0 & A_{-1} \end{bmatrix},$$

$$\mathcal{C}(s) = \begin{bmatrix} (A_0 - sI) & A_1 & & \\ A_{-1} & \ddots & & \\ & & \ddots & A_1 \\ & & A_{-1} & (A_0 - sI) \end{bmatrix}, \quad \mathcal{D} = \begin{bmatrix} A_{-1} & 0 \\ 0 & \vdots \\ \vdots & 0 \\ 0 & A_1 \end{bmatrix}.$$

Proposition 3.3 *For any $s > 0$, $C \geq 1$, and $0 \leq k \leq C$, the Laplace transform of the expected reward function, conditional on the initial level k, is given by*

$$\tilde{\boldsymbol{R}}_k(s) = G(s)^k \boldsymbol{v}(s,C) + \hat{G}(s)^{C-k} \boldsymbol{w}(s,C) + \boldsymbol{\nu}_k(s,C), \tag{39}$$

where $\boldsymbol{\nu}_k(s,C)$ is given by (37), *and*

$$\begin{bmatrix} \boldsymbol{v}(s,C) \\ \boldsymbol{w}(s,C) \end{bmatrix} = (-Z(s,C))^{-1} \begin{bmatrix} \boldsymbol{g}_0(s) + (B_0 - sI)\,\boldsymbol{\nu}_0(s) + A_1\,\boldsymbol{\nu}_1(s,C) \\ \boldsymbol{g}_C(s) + A_{-1}\,\boldsymbol{\nu}_{C-1}(s) + (C_0 - sI)\,\boldsymbol{\nu}_C(s,C) \end{bmatrix}$$

with

$$Z(s,C) = \begin{bmatrix} (B_0 - sI) + A_1 G(s) & ((B_0 - sI)\hat{G}(s) + A_1)\hat{G}(s)^{C-1} \\ (A_{-1} + (C_0 - sI)G(s))G(s)^{C-1} & A_{-1}\hat{G}(s) + (C_0 - sI) \end{bmatrix}. \tag{40}$$

In addition, $Z(s,C)$ can be written as

$$Z(s,C) = \mathring{Q}(s,C) \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix}.$$

Proof. We apply Corollary 3.2 to (19) and obtain that the general solution of the second-order difference equation is given by (39) for $1 \leq k \leq C-1$. We then specify the arbitrary vectors $\boldsymbol{v}(s,C)$ and $\boldsymbol{w}(s,C)$ using the boundary conditions (18) and (20). Injecting the general solution into (18) and (20) leads to the system of equations

$$\begin{aligned} -\boldsymbol{g}_0(s) &= (B_0 - sI)(\boldsymbol{v}(s,C) + \hat{G}(s)^C \boldsymbol{w}(s,C) + \boldsymbol{\nu}_0(s,C)) \\ &\quad + A_1(G(s)\boldsymbol{v}(s,C) + \hat{G}(s)^{C-1}\boldsymbol{w}(s,C) + \boldsymbol{\nu}_1(s,C)) \\ -\boldsymbol{g}_C(s) &= A_{-1}(G(s)^{C-1}\boldsymbol{v}(s,C) + \hat{G}(s)\boldsymbol{w}(s,C) + \boldsymbol{\nu}_{C-1}(s,C)) \\ &\quad + (C_0 - sI)(G(s)^C \boldsymbol{v}(s,C) + \boldsymbol{w}(s,C) + \boldsymbol{\nu}_C(s,C)), \end{aligned}$$

which can be rewritten as

$$Z(s,C) \begin{bmatrix} \boldsymbol{v}(s,C) \\ \boldsymbol{w}(s,C) \end{bmatrix} = - \begin{bmatrix} \boldsymbol{g}_0(s) + (B_0 - sI)\,\boldsymbol{\nu}_0(s) + A_1\,\boldsymbol{\nu}_1(s,C) \\ \boldsymbol{g}_C(s) + A_{-1}\,\boldsymbol{\nu}_{C-1}(s) + (C_0 - sI)\,\boldsymbol{\nu}_C(s,C) \end{bmatrix},$$

where $Z(s,C)$ is given by (40). To prove that the matrix $Z(s,C)$ is invertible, we now show that it can be written as the matrix product

$$Z(s,C) = \mathring{Q}(s,C) \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix}, \tag{41}$$

where $\mathring{Q}(s,C)$ is invertible by definition, and the matrix inverse

$$\begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix}^{-1}$$

exists because $\mathrm{sp}(G(s)) < 1$ and $\mathrm{sp}(\hat{G}(s)) < 1$ for any $s > 0$. Using (38), we have

$$\begin{aligned} &\mathring{Q}(s,C) \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix} \\ &= \mathcal{A}(s) \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix} + \mathcal{B}(-\mathcal{C}(s))^{-1}\mathcal{D} \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix}, \end{aligned}$$

and to show that this is equal to $Z(s,C)$ given by (40), it suffices to show that

$$\mathcal{B}(-\mathcal{C}(s))^{-1}\mathcal{D} \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix} = \begin{bmatrix} A_1 G(s) & A_1 \hat{G}(s)^{C-1} \\ A_{-1} G(s)^{C-1} & A_{-1}\hat{G}(s) \end{bmatrix}. \quad (42)$$

Observe that

$$\mathcal{D} \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix} = \begin{bmatrix} A_{-1} & A_{-1}\hat{G}(s)^C \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \\ A_1 G(s)^C & A_1 \end{bmatrix} = -\mathcal{C}(s) \begin{bmatrix} G(s) & \hat{G}(s)^{C-1} \\ G(s)^2 & \hat{G}(s)^{C-2} \\ \vdots & \vdots \\ G(s)^{C-2} & \hat{G}(s)^2 \\ G(s)^{C-1} & \hat{G}(s) \end{bmatrix} \quad (43)$$

where we have used (21) and (22). This directly leads to

$$(-\mathcal{C}(s))^{-1}\mathcal{D} \begin{bmatrix} I & \hat{G}(s)^C \\ G(s)^C & I \end{bmatrix} = \begin{bmatrix} G(s) & \hat{G}(s)^{C-1} \\ G(s)^2 & \hat{G}(s)^{C-2} \\ \vdots & \vdots \\ G(s)^{C-1} & \hat{G}(s) \end{bmatrix},$$

which, pre-multiplied by the matrix $\mathcal{B}$, provides (42). □

Observe that the expression (39) for $\tilde{\boldsymbol{R}}_k(s)$ involves matrices of size at most $2n$ (such as $Z(s,C)$). This method ensures a better computational stability than the direct inversion of a matrix of size $n(C+1)$ in (15) for large values of C. The function $\boldsymbol{R}_k(t)$ is obtained by taking the inverse Laplace transform of $\tilde{\boldsymbol{R}}_k(s)$, which can be done numerically using the method of Abate and Whitt [1] and the function ilaplace.m in Matlab.

The next corollary provides us with the limit as $C \to \infty$ of the result stated in Proposition 3.3. It is a direct consequence of the fact that the matrices $G(s)$ and $\hat{G}(s)$ are sub-stochastic but not stochastic for any $s > 0$.

Corollary 3.4 *Assume that the series* $\sum_{j=1}^{\infty} \hat{G}(s)^j H_0(s)\, \boldsymbol{g}_{k+j}(s)$ *converges for any* $k \geq 0$*. Then, for any* $s > 0$*, and* $k \geq 0$*, the limit as* $C \to \infty$ *of the Laplace transform of the expected reward function, conditional on the initial level* k*, is given by*

$$\tilde{\boldsymbol{R}}_k(s, \infty) = G(s)^k \boldsymbol{v}(s, \infty) + \boldsymbol{\nu}_k(s, \infty), \tag{44}$$

where

$$\boldsymbol{v}(s,\infty) = -((B_0 - sI) + A_1 G(s))^{-1}(\boldsymbol{g}_0(s) + (B_0 - sI)\,\boldsymbol{\nu}_0(s,\infty) + A_1\,\boldsymbol{\nu}_1(s,\infty)),$$

and

$$\boldsymbol{\nu}_k(s,\infty) \quad = \quad \sum_{j=0}^{k-1} G(s)^j H_0(s)\, \boldsymbol{g}_{k-j}(s) + \sum_{j=1}^{\infty} \hat{G}(s)^j H_0(s)\, \boldsymbol{g}_{k+j}(s), \quad k \geq 1.$$

Note that we can write $\boldsymbol{\nu}_0(s,\infty) = \hat{G}(s)\boldsymbol{\nu}_1(s,\infty)$.

In addition to the blocks $\tilde{\boldsymbol{R}}_k(s)$ of the expected reward function, Proposition 3.3 provides us with an expression for the blocks $\tilde{D}_{k,\ell}(s)$ of the Laplace transform of the transient deviation matrix in terms of the matrices $G(s)$ and $\hat{G}(s)$, as we show now. We partition the vector $\boldsymbol{\pi}$ by levels into subvectors $\boldsymbol{\pi}_\ell$, $0 \leq k \leq C$, where $\boldsymbol{\pi}_\ell$ has n components.

Proposition 3.5 *For* $0 \leq k \leq C$*,*

- *if* $\ell = 0$*,*

$$\tilde{D}_{k,0}(s) = G(s)^k V(s,C) + \hat{G}(s)^{C-k} W(s,C) - \mathbf{1}\boldsymbol{\pi}_0\,(1/s^2),$$

with

$$\begin{bmatrix} V(s,C) \\ W(s,C) \end{bmatrix} = (-s\,Z(s,C))^{-1} \begin{bmatrix} I \\ 0 \end{bmatrix},$$

- *if* $1 \le \ell \le C-1$,

$$\begin{aligned}\tilde{D}_{k,\ell}(s) &= \Big(G(s)^k V(s,C,\ell) + \hat{G}(s)^{C-k} W(s,C,\ell) \\ &\quad + (1/s) G(s)^{k-\ell} \mathbb{1}_{\{\ell \le k\}} + (1/s)\hat{G}(s)^{\ell-k}\mathbb{1}_{\{\ell > k\}}\Big) H_0(s) - \mathbf{1}\boldsymbol{\pi}_\ell \, (1/s^2),\end{aligned}$$

with

$$\left[\begin{array}{c} V(s,C,\ell) \\ W(s,C,\ell) \end{array}\right] = (-s\, Z(s,C))^{-1} \left[\begin{array}{c} (B_0 - sI)\hat{G}(s)^\ell + A_1 \hat{G}(s)^{\ell-1} \\ A_{-1} G(s)^{C-1-\ell} + (C_0 - sI) G(s)^{C-\ell} \end{array}\right],$$

- *if* $\ell = C$,

$$\begin{aligned}\tilde{D}_{k,C}(s) = &\Big(G(s)^k V(s,C) + \hat{G}(s)^{C-k}(W(s,C) + (1/s)I)\Big) H_0(s) \\ &- \mathbf{1}\boldsymbol{\pi}_C \, (1/s^2),\end{aligned}$$

with

$$\left[\begin{array}{c} V(s,C) \\ W(s,C) \end{array}\right] = (-s\, Z(s,C))^{-1} \left[\begin{array}{c} (B_0 - sI)\hat{G}(s)^C + A_1 \hat{G}(s)^{C-1} \\ (C_0 - A_0) - A_1 G(s) \end{array}\right],$$

where $Z(s,C)$ *is given by* (40).

Proof. We use (14) and Proposition 3.3 with, for $0 \le \ell \le C$ and $1 \le i \le n$, $\boldsymbol{g}_\ell = \boldsymbol{e}_i$ and $\boldsymbol{g}_j = \mathbf{0}$ for $j \neq \ell$. □

The limit as $C \to \infty$ of Proposition 3.5 again follows easily, and provides us with an expression for the blocks of the infinite matrix $\tilde{D}(s)$ corresponding to a QBD with no upper bound on the levels.

Corollary 3.6 *For* $k \ge 0$,

- *if* $\ell = 0$,

$$\tilde{D}_{k,0}(s,\infty) = G(s)^k (-s)^{-1}((B_0 - sI) + A_1 G(s))^{-1} - \mathbf{1}\boldsymbol{\pi}_0 \, (1/s^2),$$

- *if* $\ell \ge 1$,

$$\begin{aligned}\tilde{D}_{k,\ell}(s,\infty) &= \Big(G(s)^k V(s,\infty,\ell) + (1/s) G(s)^{k-\ell}\mathbb{1}_{\{\ell \le k\}} \\ &\quad + (1/s)\hat{G}(s)^{\ell-k}\mathbb{1}_{\{\ell > k\}}\Big) H_0(s) - \mathbf{1}\boldsymbol{\pi}_\ell \, (1/s^2),\end{aligned}$$

with

$$V(s,\infty,\ell) = (-s)^{-1}((B_0 - sI) + A_1 G(s))^{-1}((B_0 - sI)\hat{G}(s) + A_1)\hat{G}(s)^{\ell-1}.$$

3.2 Asymptotic regime

Next, we concentrate on the asymptotic properties of the expected reward function $\boldsymbol{R}_k(t)$ for large values of t. By decomposing (12) into blocks, the linear asymptote is given by

$$\bar{\boldsymbol{R}}_k(t) = \sum_{0\le\ell\le C} ((\boldsymbol{\pi}_\ell\, \boldsymbol{g}_\ell)\mathbf{1}\, t + D_{k,\ell}\, \boldsymbol{g}_\ell).$$

We determine now an explicit expression for the blocks $D_{k,\ell}$ of the deviation matrix for $0 \le k, \ell \le C$, together with the mean first passage times to any level ℓ in the QBD process. Our method is based on the relationship between the entries of the deviation matrix and mean first passage times, and involves the solution of finite systems of matrix difference equations similar to the ones we solved in Section 3.1. Indeed, we can write

$$D_{(k,i)(\ell,j)} = \pi_{(\ell,j)}\, [M_{\boldsymbol{\pi}(\ell,j)} - M_{(k,i)(\ell,j)}],$$

where $M_{(k,i)(\ell,j)}$ is the mean first entrance time to (ℓ, j) from (k, i), and $M_{\boldsymbol{\pi}(\ell,j)}$ is the mean first entrance time to (ℓ, j) if the state at time 0 has the stationary distribution $\boldsymbol{\pi}$. Define the block-matrices $M_{k,\ell} = \left(M_{(k,i)(\ell,j)}\right)_{1\le i,j\le n}$, for $0 \le k, \ell \le C$. In matrix form, we have

$$D_{k,\ell} = \left[\left(\mathbf{1}_n \otimes \sum_{0\le x\le C} \boldsymbol{\pi}_x\, M_{x,\ell}\right) - M_{k,\ell}\right] \mathrm{diag}(\boldsymbol{\pi}_\ell). \tag{45}$$

Let us fix level ℓ and define $\boldsymbol{m}^{(j)}_{k,\ell} = M_{k,\ell}\, \boldsymbol{e}_j$, the jth column of $M_{k,\ell}$, for $1 \le j \le n$ and $0 \le k \le C$. So $(\boldsymbol{m}^{(j)}_{k,\ell})_i = M_{(k,i)(\ell,j)}$. Since the diagonal of $M_{\ell,\ell}$ is null we must have $(\boldsymbol{m}^{(j)}_{\ell,\ell})_j = 0$. We introduce the notation $\bar{A}^{(j)}_{-1}$ and $\bar{A}^{(j)}_{1}$ for the matrices obtained by replacing the jth row in A_{-1} and A_1, respectively, by $\mathbf{0}^\top$, and $\bar{B}^{(j)}_0$, $\bar{A}^{(j)}_0$, $\bar{C}^{(j)}_0$ for the matrices obtained by replacing the jth row in B_0, A_0 and C_0, respectively, by $-\boldsymbol{e}_j^\top$.

Proposition 3.7 *For any fixed level* $0 \le \ell \le C$ *and any phase* $1 \le j \le n$, *the vectors* $\boldsymbol{m}^{(j)}_{k,\ell}$ *satisfy the system of matrix second-order difference equations*

$$A_{-1}\, \boldsymbol{m}^{(j)}_{k-1,\ell} + A_0\, \boldsymbol{m}^{(j)}_{k,\ell} + A_1\, \boldsymbol{m}^{(j)}_{k+1,\ell} \;\; = \;\; -\mathbf{1} \tag{46}$$

for $1 \le k \le \ell-1$ *and* $\ell+1 \le k \le C-1$, *with boundary conditions depending on the value of* ℓ*:*

- *for* $\ell = 0$,

$$\begin{aligned} \bar{B}_0^{(j)}\, \boldsymbol{m}_{0,\ell}^{(j)} + \bar{A}_1^{(j)}\, \boldsymbol{m}_{1,\ell}^{(j)} &= -\mathbf{1} + \boldsymbol{e}_j, & (47)\\ A_{-1}\, \boldsymbol{m}_{C-1,\ell}^{(j)} + C_0\, \boldsymbol{m}_{C,\ell}^{(j)} &= -\mathbf{1}, & (48) \end{aligned}$$

- *for* $1 \leq \ell \leq C-1$,

$$\begin{aligned} B_0\, \boldsymbol{m}_{0,\ell}^{(j)} + A_1\, \boldsymbol{m}_{1,\ell}^{(j)} &= -\mathbf{1}, & (49)\\ \bar{A}_{-1}^{(j)} \boldsymbol{m}_{\ell-1,\ell}^{(j)} + \bar{A}_0{}^{(j)} \boldsymbol{m}_{\ell,\ell}^{(j)} + \bar{A}_1^{(j)} \boldsymbol{m}_{\ell+1,\ell}^{(j)} &= -\mathbf{1} + \boldsymbol{e}_j & (50)\\ A_{-1}\, \boldsymbol{m}_{C-1,\ell}^{(j)} + C_0\, \boldsymbol{m}_{C,\ell}^{(j)} &= -\mathbf{1}, & (51) \end{aligned}$$

- *for* $\ell = C$,

$$\begin{aligned} B_0\, \boldsymbol{m}_{0,\ell}^{(j)} + A_1\, \boldsymbol{m}_{1,\ell}^{(j)} &= -\mathbf{1}, & (52)\\ \bar{A}_{-1}^{(j)}\, \boldsymbol{m}_{C-1,\ell}^{(j)} + \bar{C}_0^{(j)}\, \boldsymbol{m}_{C,\ell}^{(j)} &= -\mathbf{1} + \boldsymbol{e}_j. & (53) \end{aligned}$$

Proof. For a fixed value of ℓ, we let $m_{(k,i)}^{(j)} := (\boldsymbol{m}_{k,\ell}^{(j)})_i$, for $1 \leq i \leq n$. By conditioning on the epoch where the process first leaves state (k, i), we obtain for $1 \leq k \leq C-1$, $\ell + 1 \leq k \leq C-1$, and $1 \leq j \leq n$,

$$\begin{aligned} m_{(k,i)}^{(j)} &= \frac{1}{(-A_0)_{ii}} + \sum_{x \neq i} \frac{(A_0)_{ix}}{(-A_0)_{ii}} m_{(k,\ell)}^{(j)} \\ &\quad + \sum_x \frac{(A_1)_{ix}}{(-A_0)_{ii}} m_{(k+1,x)}^{(j)} + \sum_x \frac{(A_{-1})_{ix}}{(-A_0)_{ii}} m_{(k-1,x)}^{(j)}, \qquad (54) \end{aligned}$$

which in matrix form gives (46).

A similar argument leads to the boundary equations (48), (49), (51), and (52). Finally, the boundary equations (47), (50), and (53) are obtained by adding the constraint that when $k = \ell$, $m_{(\ell,j)}^{(j)} = 0$. □

For $0 \leq k \leq C$, we define the vectors

$$\boldsymbol{\mu}_k(C) := \sum_{j=0}^{k-1} G^j H_0\, \mathbf{1} + \sum_{j=1}^{C-k} \hat{G}^j H_0\, \mathbf{1}. \qquad (55)$$

The next proposition provides an explicit expression for the columns of the mean first passage time matrices $M_{k,\ell}$. In preparation for that proposition, we introduce the following notation: for any subset of levels $E \subseteq \{0, 1, \ldots, C\}$, and any state $(k,i) \in \{0,1,\dots,C\} \times \{1,2,\dots,n\}$, we denote by $K_E(k,i)$ the generator of the transient Markov chain obtained from the QBD process in which absorption can happen from state (k,i) at rate 1, and censored to the levels in E. So, if $Q^{(k,i)}$ denotes the generator Q of the QBD process modified such that absorption can happen from state (k,i) at rate 1, and whose levels are partitioned into the set E and its complement E^c as

$$Q^{(k,i)} = \begin{array}{c} \\ E \\ E^c \end{array} \begin{array}{c} \begin{array}{cc} E & E^c \end{array} \\ \left[\begin{array}{cc} Q^{(k,i)}_{EE} & Q^{(k,i)}_{EE^c} \\ Q^{(k,i)}_{E^cE} & Q^{(k,i)}_{E^cE^c} \end{array} \right] \end{array}, \tag{56}$$

then $K_E(k,i) = Q^{(k,i)}_{EE} + Q^{(k,i)}_{EE^c}(-Q^{(k,i)}_{E^cE^c})^{-1}Q^{(k,i)}_{E^cE}$.

Proposition 3.8 *In the non null-recurrent case, for any $0 \le k \le C$, the vector $\boldsymbol{m}^{(j)}_{k,\ell}$ takes the following expression, depending on the value of ℓ:*

- *if $\ell = 0$,*

$$\boldsymbol{m}^{(j)}_{k,0} = G^k \boldsymbol{v}^{(j)} + \hat{G}^{C-k}\boldsymbol{w}^{(j)} + \boldsymbol{\mu}_k(C),$$

  where $\boldsymbol{\mu}_k(C)$ is given by (55) and

$$\begin{bmatrix} \boldsymbol{v}^{(j)} \\ \boldsymbol{w}^{(j)} \end{bmatrix} = (-Z^{(j)}(C))^{-1} \begin{bmatrix} \mathbf{1} - \boldsymbol{e}_j + \bar{B}^{(j)}_0 \boldsymbol{\mu}_0(C) + \bar{A}^{(j)}_1 \boldsymbol{\mu}_1(C) \\ \mathbf{1} + A_{-1}\boldsymbol{\mu}_{C-1}(C) + C_0 \boldsymbol{\mu}_C(C) \end{bmatrix}$$

  with

$$\begin{aligned} Z^{(j)}(C) &= \begin{bmatrix} \bar{B}^{(j)}_0 + \bar{A}^{(j)}_1 G & (\bar{B}^{(j)}_0 \hat{G} + \bar{A}^{(j)}_1)\hat{G}^{C-1} \\ (A_{-1} + C_0 G)G^{C-1} & A_{-1}\hat{G} + C_0 \end{bmatrix} \\ &= K_E(0,j) \begin{bmatrix} I & \hat{G}^C \\ G^C & I \end{bmatrix}, \end{aligned}$$

  where $E = \{0, C\}$.

- *if $1 \le \ell \le C-2$,*

$$\begin{aligned} \boldsymbol{m}^{(j)}_{k,\ell} &= \left(G^k \boldsymbol{v}^{-(j)} + \hat{G}^{\ell-k}\boldsymbol{w}^{-(j)}\right) \mathbb{1}_{\{k \le \ell\}} \\ &\quad + \left(G^{k-\ell-1}\boldsymbol{v}^{+(j)} + \hat{G}^{C-k}\boldsymbol{w}^{+(j)}\right) \mathbb{1}_{\{k \ge \ell+1\}} \\ &\quad + \boldsymbol{\mu}_k(C), \end{aligned}$$

where $\boldsymbol{\mu}_k(C)$ *is given by* (55) *and*

$$\begin{bmatrix} \boldsymbol{v}^{-(j)} \\ \boldsymbol{w}^{-(j)} \\ \boldsymbol{v}^{+(j)} \\ \boldsymbol{w}^{+(j)} \end{bmatrix} = (-Z^{(j)}(\ell, C))^{-1} \begin{bmatrix} \mathbf{1} + B_0\boldsymbol{\mu}_0(C) + A_1\boldsymbol{\mu}_1(C) \\ \mathbf{1} - \boldsymbol{e}_j + \bar{A}_{-1}^{(j)}\boldsymbol{\mu}_{\ell-1}(C) + \bar{A}_0^{(j)}\boldsymbol{\mu}_\ell(C) + \bar{A}_1^{(j)}\boldsymbol{\mu}_{\ell+1}(C) \\ \mathbf{1} + A_{-1}\boldsymbol{\mu}_\ell(C) + A_0\boldsymbol{\mu}_{\ell+1}(C) + A_1\boldsymbol{\mu}_{\ell+2}(C) \\ \mathbf{1} + A_{-1}\boldsymbol{\mu}_{C-1}(C) + C_0\boldsymbol{\mu}_C(C) \end{bmatrix}$$

with

$$\begin{aligned} &Z^{(j)}(\ell, C) \\ &= \begin{bmatrix} B_0 + A_1G & (B_0\hat{G} + A_1)\hat{G}^{\ell-1} & 0 & 0 \\ (\bar{A}_{-1}^{(j)} + \bar{A}_0^{(j)}G)G^{\ell-1} & \bar{A}_{-1}^{(j)}\hat{G} + \bar{A}_0^{(j)} & \bar{A}_1^{(j)} & \bar{A}_1^{(j)}\hat{G}^{C-\ell-1} \\ A_{-1}G^\ell & A_{-1} & A_0 + A_1G & (A_0\hat{G} + A_1)\hat{G}^{C-\ell-2} \\ 0 & 0 & (C_0G + A_{-1})G^{C-\ell-2} & C_0 + A_{-1}\hat{G} \end{bmatrix} \\ &= K_E(\ell, j) \begin{bmatrix} I & \hat{G}^\ell & 0 & 0 \\ G^\ell & I & 0 & 0 \\ 0 & 0 & I & \hat{G}^{C-\ell-1} \\ 0 & 0 & G^{C-\ell-1} & I \end{bmatrix}, \end{aligned}$$

where $E = \{0, \ell, \ell+1, C\}$.

- *if* $\ell = C - 1$,

$$\boldsymbol{m}_{k,C-1}^{(j)} = \left(G^k\boldsymbol{v}^{(j)} + \hat{G}^{C-1-k}\boldsymbol{w}^{(j)}\right)\mathbb{1}_{\{k\le C-1\}} + \boldsymbol{x}^{(j)}\mathbb{1}_{\{k=C\}} + \boldsymbol{\mu}_k(C),$$

where $\boldsymbol{\mu}_k(C)$ *is given by* (55) *and*

$$\begin{bmatrix} \boldsymbol{v}^{(j)} \\ \boldsymbol{w}^{(j)} \\ \boldsymbol{x}^{(j)} \end{bmatrix} = (-Z^{(j)}(C))^{-1} \begin{bmatrix} \mathbf{1} + B_0\boldsymbol{\mu}_0(C) + A_1\boldsymbol{\mu}_1(C) \\ \mathbf{1} - \boldsymbol{e}_j + \bar{A}_{-1}^{(j)}\boldsymbol{\mu}_{C-2}(C) + \bar{A}_0^{(j)}\boldsymbol{\mu}_{C-1}(C) + \bar{A}_1^{(j)}\boldsymbol{\mu}_C(C) \\ \mathbf{1} + A_{-1}\boldsymbol{\mu}_{C-1}(C) + C_0\boldsymbol{\mu}_C(C) \end{bmatrix}$$

with

$$\begin{aligned} Z^{(j)}(C) &= \begin{bmatrix} B_0 + A_1G & (B_0\hat{G} + A_1)\hat{G}^{C-2} & 0 \\ (\bar{A}_{-1}^{(j)} + \bar{A}_0^{(j)}G)G^{C-2} & \bar{A}_{-1}^{(j)}\hat{G} + \bar{A}_0^{(j)} & \bar{A}_1^{(j)} \\ A_{-1}G^{C-1} & A_{-1} & C_0 \end{bmatrix} \\ &= K_E(C-1, j) \begin{bmatrix} I & \hat{G}^{C-1} & 0 \\ G^{C-1} & I & 0 \\ 0 & 0 & I \end{bmatrix}, \end{aligned}$$

where $E = \{0, C-1, C\}$.

- *if* $\ell = C$,

$$\boldsymbol{m}_{k,C}^{(j)} = G^k \boldsymbol{v}^{(j)} + \hat{G}^{C-k} \boldsymbol{w}^{(j)} + \boldsymbol{\mu}_k(C),$$

where $\boldsymbol{\mu}_k(C)$ *is given by* (55) *and*

$$\begin{bmatrix} \boldsymbol{v}^{(j)} \\ \boldsymbol{w}^{(j)} \end{bmatrix} = (-Z^{(j)}(C))^{-1} \begin{bmatrix} \mathbf{1} + B_0 \boldsymbol{\mu}_0(C) + A_1 \boldsymbol{\mu}_1(C) \\ \mathbf{1} - \boldsymbol{e}_j + \bar{A}_{-1}^{(j)} \boldsymbol{\mu}_{C-1}(C) + \bar{C}_0^{(j)} \boldsymbol{\mu}_C(C) \end{bmatrix}$$

with

$$\begin{aligned} Z^{(j)}(C) &= \begin{bmatrix} B_0 + A_1 G & (B_0 \hat{G} + A_1)\hat{G}^{C-1} \\ (\bar{A}_{-1}^{(j)} + \bar{C}_0^{(j)} G) G^{C-1} & \bar{A}_{-1}^{(j)} \hat{G} + \bar{C}_0^{(j)} \end{bmatrix} \\ &= K_E(C,j) \begin{bmatrix} I & \hat{G}^C \\ G^C & I \end{bmatrix}, \end{aligned}$$

where $E = \{0, C\}$.

Proof. The result follows from Lemma 3.1 applied to the system of difference equations (46). The arbitrary vectors are then determined using the boundary conditions given in Proposition 3.7.

When $\ell = 0$ and $\ell = C$, there are two boundary equations, determining two arbitrary vectors.

When $1 \le \ell \le C - 2$, the solution depends on whether $0 \le k \le \ell$ or $\ell + 1 \le k \le C$. There are four boundary conditions, namely the three equations described in Proposition 3.7, in addition to one boundary equation obtained by taking $k = \ell + 1$ in (46). This determines the four arbitrary vectors.

When $\ell = C - 1$ the solution depends on whether $0 \le k \le C - 1$ or $k = C$. There are three boundary equations, as described in Proposition 3.7, which determine the three arbitrary vectors.

The decomposition of the matrices $Z^{(j)}(C)$ and $Z^{(j)}(\ell, C)$ into the product of a non-conservative generator and the matrices involving G and $\hat{G}$ follows from the same arguments as those used in the proof of Proposition 3.3. Absorption from state (ℓ, j) at rate 1 comes from the definition of $\bar{B}_0^{(j)}$, $\bar{A}_0^{(j)}$, and $\bar{C}_0^{(j)}$. The matrices involving G and $\hat{G}$ are invertible because $\text{sp}(G) < 1$ or $\text{sp}(\hat{G}) < 1$ in the non null-recurrent case. $\square$

Observe that if $\text{sp}(\hat{G}) = 1$, the series $\sum_{j=1}^{\infty} \hat{G}^j H_0 \mathbf{1}$ diverges, and the limit as $C \to \infty$ of $\boldsymbol{\mu}_k(C)$ given in (55) is infinite, while if $\text{sp}(\hat{G}) < 1$,

$$\boldsymbol{\mu}_k(\infty) := \lim_{C \to \infty} \boldsymbol{\mu}_k(C) = \sum_{j=0}^{k-1} G^j H_0 \mathbf{1} + ((I - \hat{G})^{-1} - I) H_0 \mathbf{1}. \tag{57}$$

This leads to the following corollary to Proposition 3.8 for the mean first passage times in a positive recurrent QBD process with no upper bound on the levels:

Corollary 3.9 *In the positive recurrent case, for any $k \geq 0$, depending on the value of ℓ, the vector $\boldsymbol{m}_{k,\ell}^{(j)}$ has the explicit expression:*

- *if $\ell = 0$,*
$$\boldsymbol{m}_{k,0}^{(j)} = G^k \boldsymbol{v}^{(j)}(\infty) + \boldsymbol{\mu}_k(\infty),$$
where $\boldsymbol{\mu}_k(\infty)$ is given by (57), and
$$\boldsymbol{v}^{(j)}(\infty) = -(\bar{B}_0^{(j)} + \bar{A}_1^{(j)} G)^{-1}(\mathbf{1} - \boldsymbol{e}_j + \bar{B}_0^{(j)} \boldsymbol{\mu}_0(\infty) + \bar{A}_1^{(j)} \boldsymbol{\mu}_1(\infty))$$

- *if $\ell \geq 1$,*
$$\boldsymbol{m}_{k,\ell}^{(j)} = \left(G^k \boldsymbol{v}^{-(j)} + \hat{G}^{\ell-k} \boldsymbol{w}^{-(j)} \right) \mathbb{1}_{\{k \leq \ell\}} + G^{k-\ell-1} \boldsymbol{v}^{+(j)} \mathbb{1}_{\{k \geq \ell+1\}} + \boldsymbol{\mu}_k(\infty),$$
where $\boldsymbol{\mu}_k(\infty)$ is given by (57) and
$$\begin{bmatrix} \boldsymbol{v}^{-(j)} \\ \boldsymbol{w}^{-(j)} \\ \boldsymbol{v}^{+(j)} \end{bmatrix} = (-W^{(j)}(\ell))^{-1} \begin{bmatrix} \mathbf{1} + B_0 \boldsymbol{\mu}_0(\infty) + A_1 \boldsymbol{\mu}_1(\infty) \\ \mathbf{1} - \boldsymbol{e}_j + \bar{A}_{-1}^{(j)} \boldsymbol{\mu}_{\ell-1}(\infty) + \bar{A}_0^{(j)} \boldsymbol{\mu}_\ell(\infty) + \bar{A}_1^{(j)} \boldsymbol{\mu}_{\ell+1}(\infty) \\ \mathbf{1} + A_{-1} \boldsymbol{\mu}_\ell(\infty) + A_0 \boldsymbol{\mu}_{\ell+1}(\infty) + A_1 \boldsymbol{\mu}_{\ell+2}(\infty) \end{bmatrix}$$
with
$$W^{(j)}(\ell) = \begin{bmatrix} B_0 + A_1 G & (B_0 \hat{G} + A_1) \hat{G}^{\ell-1} & 0 \\ (\bar{A}_{-1}^{(j)} + \bar{A}_0^{(j)} G) G^{\ell-1} & \bar{A}_{-1}^{(j)} \hat{G} + \bar{A}_0^{(j)} & \bar{A}_1^{(j)} \\ A_{-1} G^\ell & A_{-1} & A_0 + A_1 G \end{bmatrix}.$$

To complete the characterisation of the block matrices $D_{k,\ell}$ using (45), it remains for us to compute the blocks $\boldsymbol{\pi}_x$ of the stationary distribution of the QBD process for $0 \leq x \leq C$. This can be done following Theorem 10.3.2 in [16] or Hajek [11], adapted to the continuous-time setting. This involves the matrices R and $\hat{R}$ of rates of sojourn in level $k+1$, respectively $k-1$, per unit of the local time in level k in the corresponding doubly-infinite QBD process. These matrices can be expressed in terms of G and $\hat{G}$ as $R = A_1(-(A_0 + A_1 G))^{-1}$ and $\hat{R} = A_{-1}(-(A_0 + A_{-1}\hat{G}))^{-1}$, see [16].

Proposition 3.10 *In the non null-recurrent case, the stationary distribution of the QBD process with transition matrix* (56) *is given by*

$$\boldsymbol{\pi}_k = \boldsymbol{v}_0\, R^k + \boldsymbol{v}_C\, \hat{R}^{C-k}, \quad 0 \le k \le C,$$

where $(\boldsymbol{v}_0, \boldsymbol{v}_C)$ *is the solution of the system*

$$(\boldsymbol{v}_0, \boldsymbol{v}_C) \begin{bmatrix} B_0 + R\, A_{-1} & R^{C-1}\,(RC_0 + A_1) \\ \hat{R}^{C-1}\,(\hat{R}B_0 + A_{-1}) & C_0 + \hat{R}\, A_1 \end{bmatrix} = \mathbf{0},$$

and

$$\boldsymbol{v}_0 \sum_{0\le i\le C} R^i\, \mathbf{1} + \boldsymbol{v}_C \sum_{0\le i\le C} \hat{R}^i\, \mathbf{1} = 1.$$

4 A perturbation approach

In this section, we answer Question *(ii)* at the end of Section 2 and we provide an expression for the deviation matrix of a QBD with maximal level C in terms of the deviation matrix of the same QBD with maximal level $C-1$, in transient and asymptotic regimes (Propositions 4.3 and 4.6). To highlight the number of levels in the QBDs, we shall denote by $Q^{(C)}$ the generator (56) of a QBD with maximum level C, and by $D^{(C)}$, $D^{(C)}(t)$ and $\tilde{D}^{(C)}(s)$, the corresponding deviation matrices, for $C \ge 1$.

We increase the number of states in two steps. In the first step, we start with the matrix $Q^{(C-1)}$, add a new level that consists of transient states, to obtain a matrix $T^{(C)}$ for which the deviation matrix is easily computed. In the second step, we adapt the Sherman-Morrison-Woodbury (SMW) formula ([10], page 65) to obtain $Q^{(C)}$ through a perturbation of $T^{(C)}$. The first step is described in Lemmas 4.1 and 4.2 while the second step is implemented in Propositions 4.3, 4.4 and 4.6.

The process for deriving the matrix $Q^{(C)}$ can be seen as a block version of the element-updating procedure described in Langville and Meyer [15]. Precisely, we write

$$Q^{(C)} = T^{(C)} + E^{(C)}_{C-1}\Delta^{(C)}, \tag{58}$$

where

$$
\begin{aligned}
T^{(C)} &= \left[\begin{array}{cccc|c} & & & & 0 \\ & & Q^{(C-1)} & & \vdots \\ & & & & 0 \\ \hline 0 & \cdots & 0 & A_{-1} & C_0 \end{array}\right], \qquad (59)\\
E^{(C)}_{C-1} &= \left[\begin{array}{ccccc} 0 & \cdots & 0 & I & 0 \end{array}\right]^{\top}, \\
\Delta^{(C)} &= [\begin{array}{ccccc} 0 & \cdots & 0 & A_0 - C_0 & A_1 \end{array}], \qquad (60)
\end{aligned}
$$

where $E^{(C)}_{C-1}$ has dimensions $n(C+1)\times n$ and $\Delta^{(C)}$ has dimensions $n\times n(C+1)$. Observe that $T^{(C)}$ is the generator of a reducible transient structured Markov chain, and its stationary distribution is given by

$$
\boldsymbol{\phi} = [\, \boldsymbol{\pi}^{(C-1)},\ \mathbf{0}\,],
$$

where $\boldsymbol{\pi}^{(C-1)}$ is the stationary distribution of $Q^{(C-1)}$. The group inverse of $T^{(C)}$ exists by [6, Theorem 8.5.5], and may be expressed as a function of $\boldsymbol{\pi}^{(C-1)}$ and the deviation matrix $D^{(C-1)} = -(Q^{(C-1)})^{\#}$, as shown in the next lemma.

Lemma 4.1 *The group inverse of $T^{(C)}$ is given by*

$$
(T^{(C)})^{\#} = \left[\begin{array}{c|c} -D^{(C-1)} & 0 \\ \hline C_0^{-1} M^{(C)} D^{(C-1)} + M^{(C)} \mathbf{1}\boldsymbol{\pi}^{(C-1)} & C_0^{-1} \end{array}\right], \qquad (61)
$$

where $M^{(C)}$ is an $n \times nC$ matrix defined as

$$
M^{(C)} = \left[\begin{array}{cccc} 0 & \cdots & 0 & A_{-1} \end{array}\right]. \qquad (62)
$$

Proof. We check that $(T^{(C)})^{\#}$ is the group inverse of $T^{(C)}$ by direct verification of (5) and (6). □

First, we use the decomposition (58) to derive an expression for $\boldsymbol{\pi}^{(C)}$ in terms of $\boldsymbol{\pi}^{(C-1)}$.

Lemma 4.2 *The stationary distribution of $Q^{(C)}$ may be expressed in terms of the stationary distribution of $Q^{(C-1)}$ as*

$$
\boldsymbol{\pi}^{(C)} = [\, \boldsymbol{\pi}^{(C-1)},\ \mathbf{0}\,](I + E^{(C)}_{C-1}\Delta^{(C)}(T^{(C)})^{\#})^{-1}.
$$

Proof. Our argument is a slight modification of the proof of Rising [18, Lemma 5.1], which deals with irreducible Markov matrices. Let A and $\widetilde{A}$ be two generators on the same state space, assume that each has a single recurrent class of states, not necessarily the same, so that each has a unique stationary probability vector. Denote the stationary distribution vectors as $\boldsymbol{\alpha}$ and $\widetilde{\boldsymbol{\alpha}}$, respectively.

By (7), $\widetilde{\boldsymbol{\alpha}} A A^{\#} = \widetilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$ and so

$$\widetilde{\boldsymbol{\alpha}}(I + (\widetilde{A} - A)A^{\#}) = \boldsymbol{\alpha}$$

since $\widetilde{\boldsymbol{\alpha}}\widetilde{A} = \mathbf{0}$. If A and $\widetilde{A}$ are irreducible, it follows from [18, Lemma 5.1] that $I + (\widetilde{A} - A)A^{\#}$ is non-singular. A key argument in [18] is that the kernel of both A and $\widetilde{A}$ is limited to Span($\{\mathbf{1}\}$). We can repeat it verbatim to conclude that $I + (\widetilde{A} - A)A^{\#}$ is non-singular in our case as well. Thus,

$$\widetilde{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(I + (\widetilde{A} - A)A^{\#})^{-1}, \tag{63}$$

and the lemma is proved once we set $A = T^{(C)}$ and $\widetilde{A} = Q^{(C)}$. □

We thus see that $\boldsymbol{\pi}^{(C)}$ can be expressed in terms of $\boldsymbol{\pi}^{(C-1)}$ and $D^{(C-1)}$ through $(T^{(C)})^{\#}$. In the next two sections, we use the decomposition (58) to derive recursive formulae for $\tilde{D}^{(C)}(s)$ and $D^{(C)}$.

4.1 Transient deviation matrix

Recall that the Laplace transform of the transient deviation matrix of $Q^{(C)}$ is given by

$$\tilde{D}^{(C)}(s) = (1/s)\left(sI - Q^{(C)}\right)^{-1} - (1/s^2)\mathbf{1}\boldsymbol{\pi}^{(C)}. \tag{64}$$

We can express $\tilde{D}^{(C)}(s)$ in terms of $\tilde{D}^{(C-1)}(s)$ thanks to the fact that $(sI - Q^{(C)})^{-1}$ can be expressed in terms of $(sI - Q^{(C-1)})^{-1}$ via the Sherman-Morrison-Woodbury (SMW) formula ([10], page 65), which we restate here.

Let M be a finite non-singular matrix of order m and let U and V be matrices of dimension $m \times k$. If $M + UV^{\top}$ is non-singular, then $I + V^{\top}M^{-1}U$ is non-singular as well and

$$(M + UV^{\top})^{-1} = M^{-1}\left(I - U(I + V^{\top}M^{-1}U)^{-1}V^{\top}M^{-1}\right). \tag{65}$$

As a consequence of the SMW formula, we have the following

Proposition 4.3 *The transient deviation matrix $\tilde{D}^{(1)}(s)$ of the QBD process with generator*

$$Q^{(1)} = \begin{bmatrix} B & A_1 \\ A_{-1} & C \end{bmatrix} \tag{66}$$

is given by

$$\tilde{D}^{(1)}(s) = (1/s)\left(sI - Q^{(1)}\right)^{-1} - (1/s^2)\mathbf{1}\boldsymbol{\pi}^{(1)}. \tag{67}$$

For $C \geq 2$, the transient deviation matrix $\tilde{D}^{(C)}(s)$ of the QBD process with generator $Q^{(C)}$ is recursively given by

$$\begin{aligned}
&\tilde{D}^{(C)}(s)\\
&= (1/s)(sI - T^{(C)})^{-1}\left(I + E^{(C)}_{C-1}(I - \Delta^{(C)}(sI - T^{(C)})^{-1}E^{(C)}_{C-1})^{-1}\Delta^{(C)}(sI - T^{(C)})^{-1}\right)\\
&\quad -(1/s^2)\mathbf{1}\boldsymbol{\pi}^{(C)},
\end{aligned}$$

where $(sI - T^{(C)})^{-1}$ is given in terms of $\tilde{D}^{(C-1)}(s)$ as

$$\begin{aligned}
&(sI - T^{(C)})^{-1}\\
&= \left[\begin{array}{c|c} I & 0 \\ \hline (sI - C_0)^{-1}M^{(C)} & (sI - C_0)^{-1} \end{array}\right]\left[\begin{array}{c|c} s\tilde{D}^{(C-1)}(s) + (1/s)\mathbf{1}\boldsymbol{\pi}^{(C-1)} & 0 \\ \hline 0 & I \end{array}\right],
\end{aligned}$$

with $M^{(C)}$ given in (62).

Proof. We start from (64) and we use the decomposition (58) to apply (65) with $M = sI - T^{(C)}$, $U = E^{(C)}_{C-1}$, and $V = -\Delta^{(C)\top}$. Next, we can verify that

$$(sI - T^{(C)})^{-1} = \left[\begin{array}{c|c} I & 0 \\ \hline (sI - C_0)^{-1}M^{(C)} & (sI - C_0)^{-1} \end{array}\right]\left[\begin{array}{c|c} (sI - Q^{(C-1)})^{-1} & 0 \\ \hline 0 & I \end{array}\right],$$

by checking $(sI - T^{(C)})^{-1}(sI - T^{(C)}) = I$ using (59). Finally, we use (64) again to replace $(sI - Q^{(C-1)})^{-1}$ by $s\tilde{D}^{(C-1)}(s) + (1/s)\mathbf{1}\boldsymbol{\pi}^{(C-1)}$ in the latter expression. □

Since $\boldsymbol{\pi}^{(C)}$ can be expressed in terms of $\boldsymbol{\pi}^{(C-1)}$ and $D^{(C-1)}$ by Lemmas 4.1 and 4.2, we conclude that $\tilde{D}^{(C)}(s)$ may be expressed in terms of $\tilde{D}^{(C-1)}(s)$, $\boldsymbol{\pi}^{(C-1)}$, and $D^{(C-1)}$. The recursive computation of the transient deviation matrices $\tilde{D}^{(C)}(s)$ may therefore be done together with the recursive computation of the stationary distribution vectors $\boldsymbol{\pi}^{(C)}$ and the deviation matrices $D^{(C)}$, for $C \geq 1$.

4.2 Deviation matrix

Now, we focus on obtaining an expression for $D^{(C)}$ in terms of $D^{(C-1)}$ using the decomposition (58). We start from Theorem 3.1 in [15], which gives an expression for the group inverse of a one-element updated Markov chain generator, and we extend it to a one-block update. Starting from a generator Q, we consider a new generator $\tilde{Q} = Q + A$ where A is a matrix with one non-zero block-row only. We denote by $E_K^{(C)} = [\, 0 \;\; \cdots \;\; 0 \;\; I \;\; \cdots \;\; 0 \,]^\top$ the block-column with identity as K-th block and C zero blocks elsewhere, and by P a block-row vector containing $C+1$ blocks, then $A = E_K^{(C)} P$ for some K.

Proposition 4.4 *Let Q be the generator of a Markov process with deviation matrix D and with stationary distribution $\boldsymbol{q}$. Suppose that $\tilde{Q} = Q + E_K^{(C)} P$ is the generator of an irreducible Markov process. The deviation matrix $\tilde{D}$ of $\tilde{Q}$ is given by*

$$\begin{aligned} \tilde{D} &= (I - \mathbf{1}\tilde{\boldsymbol{\pi}})\, D \left(I - E_K^{(C)} P D\right)^{-1} && (68) \\ &= (I - \mathbf{1}\tilde{\boldsymbol{\pi}})\, D \left(I + E_K^{(C)} (I - P D E_K^{(C)})^{-1} P D\right), && (69) \end{aligned}$$

where $\tilde{\boldsymbol{\pi}}$ is the stationary distribution of $\tilde{Q}$.

Proof. Observe first that the inverse in (68) is well defined by [18, Lemma 5.1]. We first show (68) by direct verification of (5) and (6) using (8). Replacing $\tilde{D}$ with the right-hand side of (68), we obtain

$$\begin{aligned} \tilde{Q}\tilde{D} &= \tilde{Q} D \left(I - E_K^{(C)} P D\right)^{-1} \\ &= (\mathbf{1}\boldsymbol{q} - I) \left(I - E_K^{(C)} P D\right)^{-1} + E_K^{(C)} P D \left(I - E_K^{(C)} P D\right)^{-1} \\ &= \mathbf{1}\tilde{\boldsymbol{\pi}} - I. \end{aligned}$$

The first equality holds since $\tilde{Q}\mathbf{1} = \mathbf{0}$, the second equality since $\tilde{Q} = Q + E_K^{(C)} P$ and $QD = \mathbf{1}\boldsymbol{q} - I$ by (7), (8), and the third equality since $\boldsymbol{q}(I - E_K^{(C)} P D)^{-1} = \tilde{\boldsymbol{\pi}}$, by (63), where A is replaced by Q and $\widetilde{A}$ by $Q + E_K^{(C)} P$. Next,

$$\left(I - E_K^{(C)} P D\right)^{-1} = I + E_K^{(C)} (I - P D E_K^{(C)})^{-1} P D \qquad (70)$$

by the SMW formula (65), and substituting (70) into (68) provides (69). □

Remark 4.5 As observed in [15], the cost of straightforward computation of (68) may be high if one updates more than one row, as it requires the inversion of the matrix $I - E_K^{(C)} PD$ whose size is the same as $\tilde{Q}$. Equation (69) reduces the computational cost because the size of $I - PDE_K^{(C)}$ is just one block.

Since the decomposition of $Q^{(C)}$ given in (58) is a block-element-updating of the matrix $T^{(C)}$ by $E_{C-1}^{(C)} \Delta^{(C)}$, the recursive formula for the deviation matrix $D^{(C)}$ of $Q^{(C)}$ follows from Proposition 4.4.

Proposition 4.6 *The deviation matrix $D^{(1)}$ of the QBD process with generator $Q^{(1)}$ given by (66) is*

$$D^{(1)} = (\mathbf{1}\boldsymbol{\pi}^{(1)} - Q^{(1)})^{-1} - \mathbf{1}\boldsymbol{\pi}^{(1)}. \tag{71}$$

For $C \geq 2$, the deviation matrix $D^{(C)}$ of the QBD process with generator $Q^{(C)}$ is recursively given by

$$D^{(C)} = \left(\mathbf{1}\boldsymbol{\pi}^{(C)} - I\right) (T^{(C)})^{\#} \left(I - E_{C-1}^{(C)} (I + \Delta^{(C)} (T^{(C)})^{\#} E_{C-1}^{(C)})^{-1} \Delta^{(C)} (T^{(C)})^{\#}\right), \tag{72}$$

where $(T^{(C)})^{\#}$ is given in terms of $D^{(C-1)}$ in (61).

Proof. The group inverse of $Q^{(1)}$ comes directly from the relationship between the group inverse of a generator and Kemeny and Snell's fundamental matrix [14], given in Theorem 3.1 in [17].

For any $C \geq 2$, Equation (72) corresponds to Equation (69) in Proposition 4.4 with $\tilde{D} = D^{(C)}$, $\tilde{\boldsymbol{\pi}} = \boldsymbol{\pi}^{(C)}$, $D = -(T^{(C)})^{\#}$, $K = C-1$, and $P = \Delta^{(C)}$. □

5 Numerical aspects

In this section, we compute the expected loss revenue function for various *MAP/PH/1/C* queues using the results of Section 3. We then compare the numerical complexity of the two different approaches developed in Sections 3 and 4 as a function of the number of levels and phases in the QBD.

5.1 The *MAP/PH/1/C* queue

Recall that our motivation behind the computation of the expected reward function and the deviation matrices of a finite QBD process stems from our desire to compute the expected amount of revenue lost in a *MAP/PH/1/C* queue. In such a process,

(i) the arrival *MAP* is characterised by the matrices D_0 and D_1 with n_1 phases;

(ii) the service time is distributed according to a $PH(\boldsymbol{\tau}, T)$ distribution of order n_2, with $\boldsymbol{t} = -T\mathbf{1}$.

The *MAP/PH/1/C* system is a finite QBD process $\{\boldsymbol{X}(t) = (J(t), \boldsymbol{\varphi}(t)), t \geq 0\}$ where

- $0 \leq J(t) \leq C$ represents the number of customers in the system at time t,
- $\boldsymbol{\varphi}(t) = (\varphi_1(t), \varphi_2(t))$ where $0 \leq \varphi_1(t) \leq n_1$ is the phase of the *MAP* at time t, and $0 \leq \varphi_2(t) \leq n_2$ is the phase of the *PH* distribution at time t.

The generator of that QBD process has the block-tridiagonal form in (56) where the block matrices are of size $n = n_1 n_2$ and are given by

$$A_{-1} = I \otimes \boldsymbol{t} \cdot \boldsymbol{\tau}, \qquad A_0 = D_0 \oplus T, \qquad A_1 = D_1 \otimes I,$$

and

$$B_0 = D_0 \otimes I, \qquad C_0 = (D_0 + D_1) \oplus T,$$

where $\otimes$ and $\oplus$ denote the Kronecker product and sum, respectively.

We consider a simple example where the *MAP* of arrivals corresponds to a *PH* renewal process with

$$D_0 = \begin{bmatrix} -10 & 2 \\ 1 & -6 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 8 \\ 5 \end{bmatrix} \cdot [0.8, 0.2],$$

and the service time is $PH(\boldsymbol{\tau}, T)$ with

$$\boldsymbol{\tau} = [0.4, 0.6], \quad T = \begin{bmatrix} -3 & 2 \\ 1 & -4 \end{bmatrix}.$$

There are four phases in the representation of this MAP/PH/1/C queue as a QBD process. Let $A = A_{-1} + A_0 + A_1$ be the phase transition matrix associated with the QBD process, and let $\boldsymbol{\alpha}$ be the stationary vector of A. Since in this example $\boldsymbol{\alpha} A_{-1}\mathbf{1} < \boldsymbol{\alpha} A_1\mathbf{1}$, we are in a *high-blocking system*. This would correspond to a transient QBD process if there were no bound on the number of levels [16].

We first assume that each customer accepted in the system generates θ units of revenue. The system loses revenue when it is in level C and a customer arrives, who is rejected, and does not lose any revenue otherwise. The expected reward function $\boldsymbol{R}(t)$ records the total expected amount of lost revenue over the finite time horizon $[0, t]$. In this particular case, the reward vector $\boldsymbol{g}$ records the *loss* per unit time, per state, and we have $\boldsymbol{g}_k = \mathbf{0}$ for $0 \leq k < C$ and $\boldsymbol{g}_C = \theta A_1 \mathbf{1}$. Results when $C = 5$ and $\theta = 1$ are shown in the upper panel of Figure 1 where we have assumed that the initial phase follows the distribution $\boldsymbol{\alpha}$.

Swapping the *PH* distribution of the arrival process and service time leads to a *low-blocking system*, whose results are shown in the lower panel of Figure 1. We clearly see that the expected lost revenue is much lower in the low-blocking system than in the high-blocking system, and in both cases it increases with the initial queue size, as expected.

We now consider a case where each customer brings a fixed amount of revenue γ per time unit when in the system, in addition to the fixed θ units of revenue when entering the system. The reward vector then becomes $\boldsymbol{g}_k = \theta A_1 \mathbf{1} + \gamma k \mathbf{1}$ for $0 \leq k \leq C - 1$ and $\boldsymbol{g}_C = \gamma C \mathbf{1}$, and we are now interested in the expected amount of revenue *gained* in $[0, t]$. The results are shown in Figure 2 for a high-blocking system (upper panel) and a low-blocking system (lower panel).

5.2 Computational efficiency

With the matrix difference approach of Section 3, the computation of the full vector $\tilde{\boldsymbol{R}}(s)$ for any given s (or equivalently the last column of the deviation matrix) using Proposition 3.3 has a complexity of order $Cn^3 + C^2 n^2$ floating point operations, and the computation of the whole transient deviation matrix using Proposition 3.5 has a complexity equal to $3C^2 n^3$ plus terms of lower order. A similar complexity is obtained when calculating the deviation matrix as detailed in Section 3.2. In evaluating the numerical complexity,

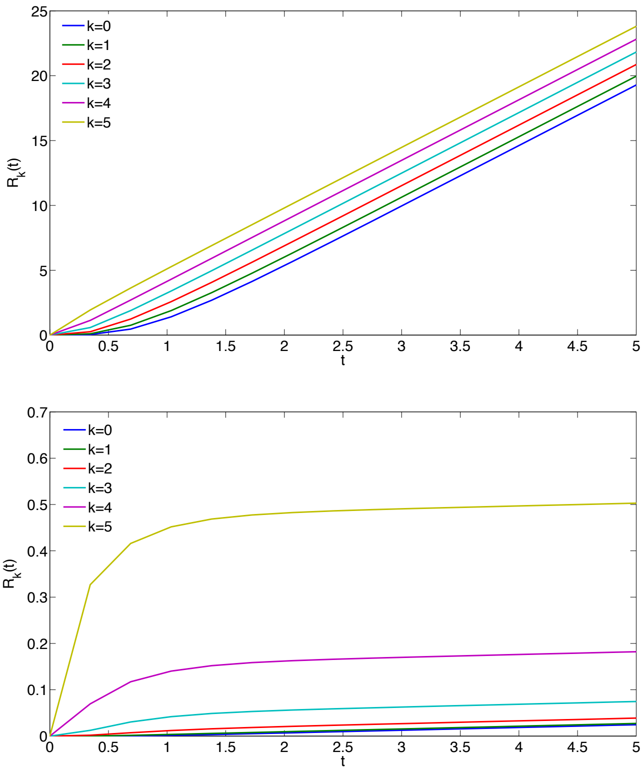

Figure 1: Expected lost revenue functions corresponding to a high-blocking system (top) and a low-blocking system (bottom) with $C = 5$ and $\theta = 1$, for different initial queue lengths k.

we assume that the implementation avoids obviously inefficient operations, such as repeatedly computing the inverse of a given matrix or performing a multiplication when it is known that one of the factors is equal to zero, and so on.

If $\tilde{\boldsymbol{R}}(s)$ is computed by solving the system (15), the complexity is of the order of C^3n^3. If one takes into account the tridiagonal structure of the system then the complexity is of the order of Cn^3. If we compute the whole deviation matrix, using the tridiagonal structure, the complexity would be of the order of C^2n^3. With the perturbation approach of Section 4, the numerical complexity using Proposition 4.6 is $2C^2n^3$ plus terms of lower order.

The conclusion is that the complexity of the different approaches (except for the brute-force computation of $(sI - Q)^{-1}$) are the same : quadratically

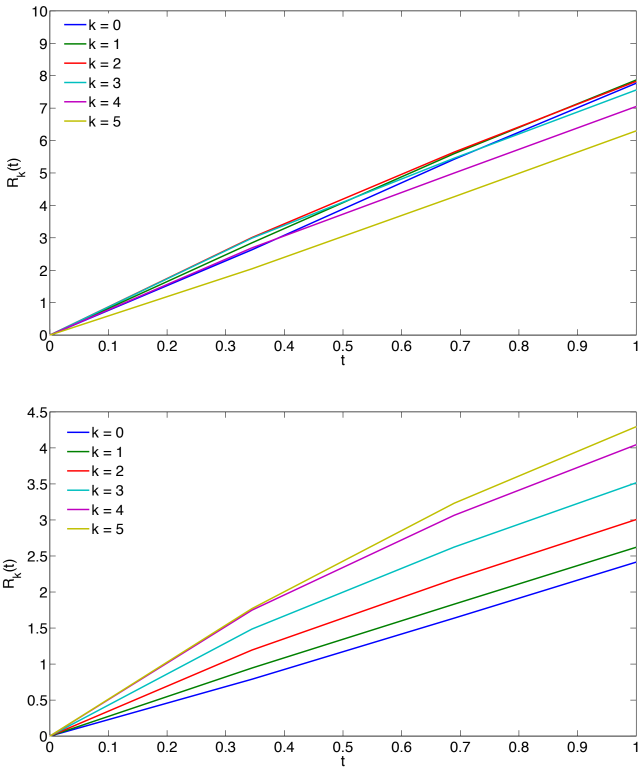

Figure 2: Expected gained revenue functions corresponding to a high-blocking system (top) and a low-blocking system (bottom) with $C = 5$, $\theta = 1$, and $\gamma = 1$, for different initial queue lengths k.

increasing like C^2, assuming that n is constant. This indicates that computations are feasible for large values of C.

We measured the actual processing time corresponding to the computation of the last block-column of the deviation matrix for arbitrary QBD processes by generating QBD processes with random entries, with $n = 2, \ldots, 5$ and $C = 1, \ldots, 500$, and for each process, running three programs 100 times and taking the average CPU time. In Figure 3, we compare the matrix difference approach with a straightforward implementation of the perturbation approach, for $C = 1, \ldots, 100$. In Figure 4, we compare the matrix difference approach with the direct computation via (10), for $C = 1, \ldots, 500$. We observe that there is always a threshold value of C such that the matrix difference approach is faster when C is greater than the threshold.

In conclusion, if one wishes to compute specific blocks of the (transient) deviation matrix, the matrix difference approach is the method of choice for large values of C, as we expected since it is especially developed for that purpose and involves matrices that are at most twice the size of the phase space.

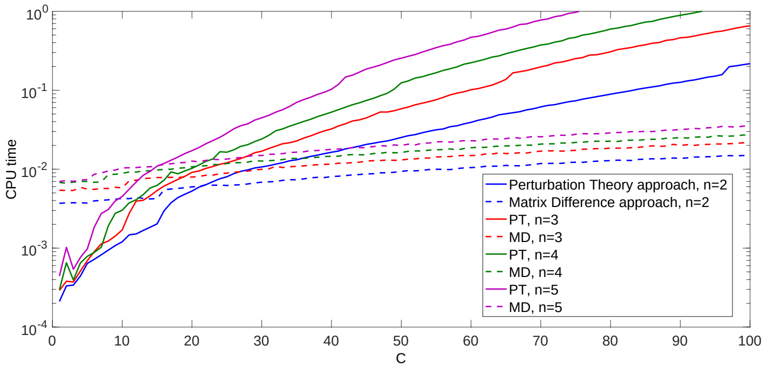

Figure 3: Comparison of the CPU time corresponding to the computation of the last block-column of the deviation matrix of a QBD process with n phases and maximum level C, using the perturbation approach and the matrix difference approach.

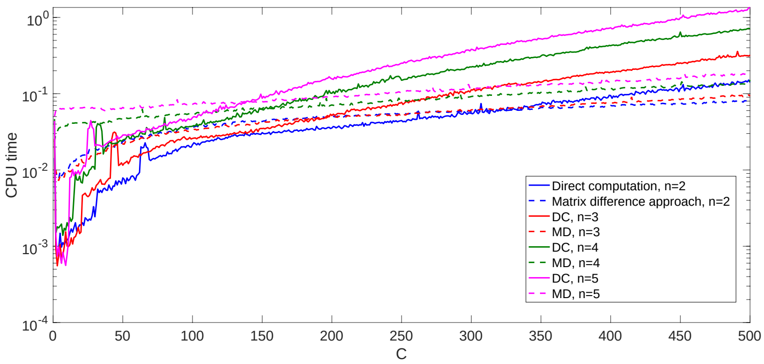

Figure 4: Comparison of the CPU time corresponding to the computation of the deviation matrix of a QBD process with n phases and maximum level C, using direct computation and the matrix difference approach.

Acknowledgments

Sophie Hautphenne and Peter Taylor would like to thank the Australian Research Council (ARC) for supporting this research through Discovery Early

Career Researcher Award DE150101044 and Laureate Fellowship FL130100039, respectively. In addition, they both acknowledge the support of the ARC Center of Excellence for Mathematical and Statistical Frontiers (ACEMS). Sarah Dendievel and Guy Latouche thank the Ministère de la Communauté française de Belgique for funding this research through the ARC grant AUWB-08/13-ULB 5. Sarah Dendievel would also like to thank the Methusalem program of the Flemish Community of Belgium for supporting this work in part.

References

[1] J. Abate and W. Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, 7(1):36–43, 1995.

[2] E. Altman, K. E. Avrachenkov, and R. Núnez-Queija. Perturbation analysis for denumerable Markov chains with application to queueing models. *Advances in Applied Probability*, 36(3):839-853, 2004.

[3] D. Bini, S. Dendievel, G. Latouche, and B. Meini. General solution of the Poisson equation for Quasi-Birth-and-Death processes. *SIAM J. Appl. Math.*, 76(6):2397–2417, 2016.

[4] D. Bini, B. Meini, and G. Latouche. *Numerical Methods for Structured Markov Chains*. Oxford Science Publications, 2005.

[5] P. T. Braunsteins, S. Hautphenne and P. G. Taylor. The roles of the deviation matrix and coupling in determining the value of capacity in $M/M/1/C$ queues. *Queueing Systems*, 83(1):157–179, 2016.

[6] S. L. Campbell and C. D. Meyer. Generalized inverses of linear transformations. Vol. 56 of Classics in Applied Mathematics. *SIAM*, 2009.

[7] B. A. Chiera and P. G. Taylor. What is a unit of capacity worth? *Probability in the Engineering and Informational Sciences*, 16(4):513–522, 2002.

[8] P. Coolen-Schrijner and E. A. van Doorn. The deviation matrix of a continuous-time Markov chain. *Probability in the Engineering and Informational Sciences*, 16(3):351–366, 2002.

[9] A. Da Silva Soares and G. Latouche. The group inverse of finite homogeneous QBD processes. *Stochastic Models*, 18(1):159–171, 2002.

[10] G. H. Golub, and C. F. Van Loan. Matrix computations. *The Johns Hopkins University Press*, 2012.

[11] B. Hajek. Birth-and-death processes on the integers with phases and general boundaries. *Journal of Applied Probability*, 19:488–499, 1982.

[12] J.J. Hunter. The computation of key properties of Markov chains via perturbations. *Linear Algebra and its Applications*, 511:176–202, 2016.

[13] S. Jiang, Y. Liu, and Y. Tang. A unified perturbation analysis framework for countable Markov chains. *Linear Algebra and its Applications*, Vol. 529, 413-440, 2017.

[14] J. G. Kemeny, and J. L. Snell. Finite Markov chains (Vol. 356). Princeton, NJ: van Nostrand. ISO 690, 1960.

[15] A. N. Langville and C. D. Meyer. Updating Markov chains with an eye on Google's PageRank. *SIAM Journal on Matrix Analysis and Applications*, 27(4), 968–987, 2006.

[16] G. Latouche and V. Ramaswami. Introduction to Matrix Analytic Methods in Stochastic Modeling. SIAM, 1999.

[17] C. D. Meyer. The role of the group generalized inverse in the theory of finite Markov chains. *SIAM Review*, 17(3), 443-464, 1975

[18] W. Rising. Applications of generalized inverses to Markov chains. *Advances in Applied Probability*, 293-302, 1991.