Projection: An efficient solution algorithm for a class of quasi-birth-death processes*

Gianfranco Ciardo Evgenia Smirni

Department of Computer Science
College of William and Mary
Williamsburg, VA 23187-8795
{ciardo,esmirni}@cs.wm.edu

Over the last two decades, considerable effort has been put into the development of *matrix geometric techniques* [3] for the exact analysis of a general and frequently encountered class of queueing models that exhibit a regular structure. In these models, the embedded Markov chains are two-dimensional generalizations of elementary M/G/1 and G/M/1 queues [1]. The intersection of these two cases corresponds to the so-called quasi-birth-death (QBD) processes. In this work, we develop an efficient methodology that can be used to compute performance measures for a class of QBD processes.

Background: Matrix Geometric Approach

If a Markov process exhibits a matrix geometric structure, then its states can be partitioned into two categories: the boundary states $\mathcal{S}^{(0)} = \{s_1^{(0)}, \ldots, s_m^{(0)}\}$ and a sequence of sets of states $\mathcal{S}^{(j)} = \{s_1^{(j)}, \ldots, s_n^{(j)}\}$, for $j \geq 1$, corresponding to the repetitive portion of the chain. This repetitive structure allows for a recursive formulation of the stationary probabilities of the states in the chain and facilitates their computation. For a QBD process, the infinitesimal generator $\mathbf{Q}$ is expressed as [2]:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{L}^{(00)} & \mathbf{F}^{(01)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{B}^{(10)} & \mathbf{L}^{(11)} & \mathbf{F} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \mathbf{L} & \mathbf{F} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \qquad (1)$$

where $\mathbf{L}^{(00)} \in \mathbb{R}^{m\times m}$ represents the transition rates between states in $\mathcal{S}^{(0)}$, $\mathbf{F}^{(01)} \in \mathbb{R}^{m\times n}$ represents the transition rates from states in $\mathcal{S}^{(0)}$ to states in $\mathcal{S}^{(1)}$, $\mathbf{B}^{(10)} \in \mathbb{R}^{n\times m}$ represents the transition rates from states in $\mathcal{S}^{(1)}$ to states in $\mathcal{S}^{(0)}$, $\mathbf{F} \in \mathbb{R}^{n\times n}$ represents transition rates from states in $\mathcal{S}^{(j)}$ to states in $\mathcal{S}^{(j+1)}$, $\mathbf{L} \in \mathbb{R}^{n\times n}$ represents transition rates between states in $\mathcal{S}^{(j)}$, $\mathbf{B} \in \mathbb{R}^{n\times n}$ represents transition rates from states in $\mathcal{S}^{(j)}$ to states in $\mathcal{S}^{(j-1)}$ (for $j \geq 1$), and $\mathbf{0}$ is a matrix of all zeros of the appropriate dimension. Note that we use the mnemonics $\mathbf{B}$ (for backward), $\mathbf{L}$ (for local), and $\mathbf{F}$ (for forward), and that $\mathbf{L}^{(11)}$ is usually assumed to equal $\mathbf{L}$ (except possibly in the diagonal, since $\mathbf{Q}$ is an infinitesimal generator and $\mathbf{B}^{(10)} \cdot \mathbf{1}^T$ could differ from $\mathbf{B} \cdot \mathbf{1}^T$), but this is not required by our approach ($\mathbf{1}^T$ is a column vector of ones of the appropriate dimension).

*This work was partially supported by a Williamd & Mary Summer Research Grant.

Neuts [3] proposed an *iterative* algorithm that takes advantage of the geometric form of the repetitive portion of the chain to compute the stationary probability vector $\boldsymbol{\pi}$ of *all* states, that is, the solution of

$$\boldsymbol{\pi} \cdot \mathbf{Q} = \mathbf{0}.$$

We can partition $\boldsymbol{\pi}$ as

$$\boldsymbol{\pi} = \left[\boldsymbol{\pi}^{(0)}, \boldsymbol{\pi}^{(1)}, \boldsymbol{\pi}^{(2)}, \cdots\right]$$

Let $\boldsymbol{\pi}^{(j)}$ be the stationary probability vector for states in $\mathcal{S}^{(j)}$ (for $j \geq 0$). The equation for the repetitive portion of the process in block matrix form is

$$\boldsymbol{\pi}^{(j-1)} \cdot \mathbf{F} + \boldsymbol{\pi}^{(j)} \cdot \mathbf{L} + \boldsymbol{\pi}^{(j+1)} \cdot \mathbf{B} = 0, \quad j \geq 2$$

and since the values of $\boldsymbol{\pi}^{(j)}, j \geq 2$ have a matrix-geometric form,

$$\boldsymbol{\pi}^{(j)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{G}^{j-1}, \quad j \geq 2,$$

this implies that the matrix $\mathbf{G}$ is the minimal solution of

$$\mathbf{F} + \mathbf{G} \cdot \mathbf{L} + \mathbf{G}^2 \cdot \mathbf{B} = \mathbf{0}.$$

This quadratic equation in the matrix $\mathbf{G}$ is typically solved numerically using the following iterative procedure:

$$\begin{aligned} \mathbf{G}(0) &= 0 \\ \mathbf{G}(n+1) &= -\mathbf{F} \cdot \mathbf{L}^{-1} - \mathbf{G}^2(n) \cdot \mathbf{B} \cdot \mathbf{L}^{-1}, \quad n \geq 0. \end{aligned}$$

To determine the stationary probabilities, one can focus on the equations for the initial portion of the matrix. Observing that $\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{G}$, this can be written in matrix form as:

$$\left[\boldsymbol{\pi}^{(0)}, \boldsymbol{\pi}^{(1)}\right] \cdot \left[\begin{array}{cc} \mathbf{L}^{(00)} & \mathbf{F}^{(01)} \\ \mathbf{B}^{(10)} & \mathbf{L}^{(11)} + \mathbf{G} \cdot \mathbf{B} \end{array}\right] = \mathbf{0}$$

and, together with the normalization condition

$$1 = \boldsymbol{\pi}^{(0)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(1)} \cdot \sum_{j=1}^{\infty} \mathbf{G}^{j-1} \cdot \mathbf{1}^T = \boldsymbol{\pi}^{(0)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(1)} \cdot (\mathbf{I} - \mathbf{G})^{-1} \cdot \mathbf{1}^T \ ,$$

yields a unique solution. Subsequently, all types of interesting performance metrics (e.g., expected system utilization, throughput, queue length) can be derived.

The Projection Approach

We now discuss a class of Markov chains for which a *Projection* solution is a viable alternative to the classic iterative matrix geometric approach. The advantages of the approach we propose are its simplicity and efficiency. The solution of the system is derived from first principles using the classic Chapman-Kolmogorov [1] equations that relate the flow into and out of each state in equilibrium. Other explicit solution techniques that appear in the literature [4] for the type of chains we consider calculate the stationary probabilities using recursion (i.e., if one knows the stationary probability of state i then the stationary probability of state $i+1$ can be determined). The proposed method differs in that we do not evaluate the probability of *all* states in the chain, rather we compute the aggregate probability distribution for the repetitive portions of the chain. If the repetitive part of the chain is composed of

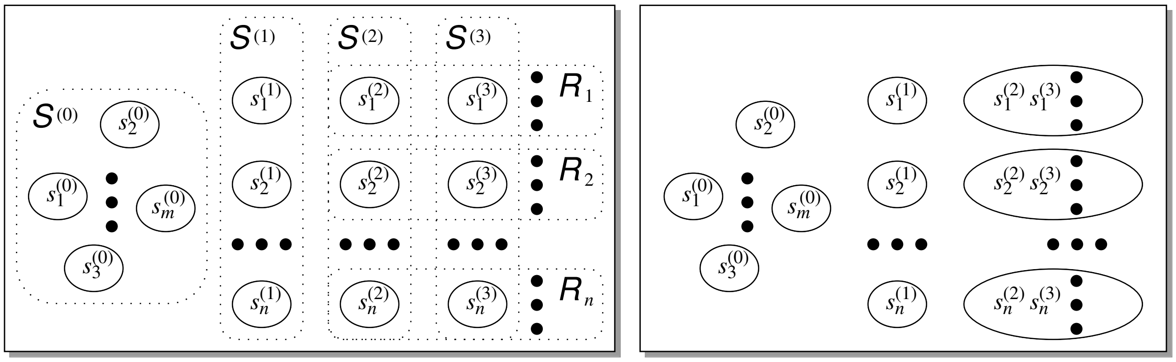

Figure 1: Partitioning of the state space (left) and its projection (right).

multiple subchains, we compute the aggregate probabilities of each infinite subchain in the system. In other words our approach explicitly computes the probability of being in each of the n sets $\mathcal{R}_i = \{s_i^{(j)} : j \geq 2\}$, for $1 \leq i \leq n$, in steady state (see Fig. 1).

More precisely, consider a CTMC exhibiting the QBD structure of Eq. 1. Then, we can write $\boldsymbol{\pi} \cdot \mathbf{Q} = \mathbf{0}$ as:

$$\begin{cases} \boldsymbol{\pi}^{(0)} \cdot \mathbf{L}^{(00)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{B}^{(10)} = \mathbf{0} \\ \boldsymbol{\pi}^{(0)} \cdot \mathbf{F}^{(01)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{L}^{(11)} + \boldsymbol{\pi}^{(2)} \cdot \mathbf{B} = \mathbf{0} \\ \boldsymbol{\pi}^{(1)} \cdot \mathbf{F} + \boldsymbol{\pi}^{(2)} \cdot \mathbf{L} + \boldsymbol{\pi}^{(3)} \cdot \mathbf{B} = \mathbf{0} \\ \boldsymbol{\pi}^{(2)} \cdot \mathbf{F} + \boldsymbol{\pi}^{(3)} \cdot \mathbf{L} + \boldsymbol{\pi}^{(4)} \cdot \mathbf{B} = \mathbf{0} \\ \cdots \end{cases} \quad (2)$$

Now, assume that $\mathbf{B}$ is a matrix of zeros except for the last column, that is all backward transitions from states in $\mathcal{S}^{(j+1)}$ to $\mathcal{S}^{(j)}$ are restricted to go to $s_n^{(j)}$, the last state in $\mathcal{S}^{(j)}$. Let $\mathbf{A}_{[r_1:r_2,c_1:c_2]}$ be the submatrix of $\mathbf{A}$ corresponding to the rows from r_1 to r_2 and the columns from c_1 to c_2.

We are going to show how to derive $m + 2n$ equations in $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and a new vector of n unknowns, $\boldsymbol{\pi}^{(s)} = \sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)}$.

- The normalization constraint offers one equation:

$$\boldsymbol{\pi}^{(0)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(1)} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(s)} \cdot \mathbf{1}^T = 1. \quad (3)$$

- The first row in (2) provides m equations:

$$\boldsymbol{\pi}^{(0)} \cdot \mathbf{L}^{(00)} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{B}^{(10)} = \mathbf{0}. \quad (4)$$

- Due to the structure of $\mathbf{B}$, we can observe that the first $n-1$ entries of $\boldsymbol{\pi}^{(2)} \cdot \mathbf{B}$ are null, hence the second row in (2) provides $n-1$ equations:

$$\boldsymbol{\pi}^{(0)} \cdot \mathbf{F}^{(01)}_{[1:m,1:n-1]} + \boldsymbol{\pi}^{(1)} \cdot \mathbf{L}^{(11)}_{[1:n,1:n-1]} = \mathbf{0}. \quad (5)$$

- If we sum all the remaining equations in (2) we obtain

$$\sum_{i=1}^{\infty} \boldsymbol{\pi}^{(i)} \cdot \mathbf{F} + \sum_{i=2}^{\infty} \boldsymbol{\pi}^{(i)} \cdot \mathbf{L} + \sum_{i=3}^{\infty} \boldsymbol{\pi}^{(i)} \cdot \mathbf{B} = \mathbf{0},$$

  which, again by using the structure of $\mathbf{B}$, results in another $n-1$ equations:

$$\boldsymbol{\pi}^{(1)} \cdot \mathbf{F}_{[1:n,1:n-1]} + \boldsymbol{\pi}^{(s)} \cdot (\mathbf{F} + \mathbf{L})_{[1:n,1:n-1]} = \mathbf{0}. \quad (6)$$

- Finally, we can consider the equations describing the balance of flow between successive sets of states, $\mathcal{S}^{(j)}$ and $\mathcal{S}^{(j+1)}$, for $j = 1, 2, \cdots$:

$$\left\{ \begin{array}{lcccl} \boldsymbol{\pi}^{(1)} \cdot \mathbf{F} \cdot \mathbf{1}^T & - & \boldsymbol{\pi}^{(2)} \cdot \mathbf{B} \cdot \mathbf{1}^T & = & \mathbf{0} \\ \boldsymbol{\pi}^{(2)} \cdot \mathbf{F} \cdot \mathbf{1}^T & - & \boldsymbol{\pi}^{(3)} \cdot \mathbf{B} \cdot \mathbf{1}^T & = & \mathbf{0} \\ \boldsymbol{\pi}^{(3)} \cdot \mathbf{F} \cdot \mathbf{1}^T & - & \boldsymbol{\pi}^{(4)} \cdot \mathbf{B} \cdot \mathbf{1}^T & = & \mathbf{0} \\ \cdots & & & & \end{array} \right. .$$

  By summing all these equations we obtain:

$$\boldsymbol{\pi}^{(1)} \cdot \mathbf{F} \cdot \mathbf{1}^T + \boldsymbol{\pi}^{(s)} \cdot (\mathbf{F} - \mathbf{B}) \cdot \mathbf{1}^T = 0. \tag{7}$$

The above equations are sufficient to compute $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{\pi}^{(1)}$, and $\boldsymbol{\pi}^{(s)}$. Thus, our approach only requires the solution of a linear system in $m + 2n$ variables, and although it does not compute the probability distribution of all individual states, it provides the necessary information to compute the expected stationary value of many Markov reward processes of interest with closed form formulas, such as the expected queue length.

Future Work

The condition for the applicability of the Projection approach can be easily described by the structure of matrix $\mathbf{B}$: it must consists of a single nonzero column. However, there are many cases where this condition does not (apparently) hold, but it can still be realized by an appropriate redefinition of the sets $\mathcal{S}^{(j)}$. To decide when this is possible, one only needs to consider the matrices $\mathbf{B}$, $\mathbf{L}$, and $\mathbf{F}$. One example is when $\mathbf{B}$ is strictly upper triangular except for a diagonal entry in the last column, and $\mathbf{L}$ is upper triangular, but many other combinations are possible. We are in the process of exploring efficient algorithms for automating this repartitioning of the states. An implementation of the Projection approach and a comparison with the Matrix Geometric approach will then be required for timing comparisons.

References

[1] L. Kleinrock. *Queueing Systems Volume I: Theory.* Wiley, 1975.

[2] R. Nelson. Matrix geometric solutions in Markov models: a mathematical tutorial. Research Report RC 16777 (#742931), IBM T.J. Watson Res. Center, Yorktown Heights, NY, Apr. 1991.

[3] M. Neuts. *Matrix-geometric solutions in stochastic models.* Johns Hopkins University Press, Baltimore, MD, 1981.

[4] P. Snyder and W. Stewart. Explicit and iterative numerical approaches to solving queueing models. *Operations Research*, 33(1):183–202, Jan-Feb. 1985.