Matrix Geometric Solutions in Markov Models
A Mathematical Tutorial

Randolph Nelson
IBM Research Division
T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598

April 17, 1991

Abstract

We present a simple derivation of the matrix geometric solution form that is found in certain vector state Markov processes that have a repetitive structure. This solution form was pioneered by Marcel Neuts and has a wide application in computer performance modeling. Our approach is based on first principles and makes use of an analogy to corresponding solutions found in scalar state processes. The paper is self contained and an example is included that illustrates how one can apply these results in performance models.

1 Introduction

The essential problem in determining the stationary state probabilities of a Markov process is solving a set of linear, flow balance, equations. In this set, there is an equation associated with each state of the system. For systems with a large, or possibly infinite number of states exact solutions can only be obtained if one can exploit structural properties of the equations. For example, the solution for the infinite set of equations derived from a M/M/1 queue is easily determined because these equations have a repetitive structure. This repetition allows one to determine a recursive solution for the stationary state probabilities since it implies that if one knows the stationary probability for any state i then the stationary probability for state $i+1$ can be determined. The stationary state probabilities for the repeating portion of this process thus has a *geometric* form. This form is found in all scalar state processes that have a similar repetitive structure.

Neuts [7] developed a body of results that allows one to exploit repetitive structure more generally than described above. If the states of the Markov process can be grouped into vectors which possess a certain repetitive structure then a recursive procedure can be used to determine the stationary state probabilities of the $i+1$'st vector in terms of the probabilities for the i'th vector. As in the scalar case, the form of the solution for the stationary state probabilities is a generalization of that obtained in the scalar case, leading to Neuts' *matrix geometric* form.

There is a wide range of applications of matrix geometric results in computer performance modeling. Many computer models have a regular organization which leads to Markov models with a repetitive structure that fits within the matrix geometric framework. To describe one version of the matrix geometric form consider a Markov process with states $(m, \mathbf{n})$ where $m \geq 0$, and $\mathbf{n}$ is a vector. Assume that the number of possible values for m (resp. $\mathbf{n}$) is unbounded (resp. bounded), and that transitions can cause m to increase (resp. decrease) by at most 1 (resp. k) unit(s). If there exists a value $m^* \geq k$ so that for m', $m^*-k \leq m' \leq m^*+1$, transition rates between states $(m^*, \mathbf{n})$ and $(m', \mathbf{n}')$ are identical to transition rates between $(m^*+j, \mathbf{n})$ and $(m'+j, \mathbf{n}')$ for all $j \geq 0$, then the process is matrix geometric. A natural family of processes that satisfy these conditions are certain open queueing systems consisting of one queue with an infinite capacity. Many models can made to be approximately matrix geometric by truncating certain portions of the state space and assuming that subsequent state transitions repeat so as to satisfy the form. Additionally classical queueing systems can be solved using this method when service and arrival processes are given by phase distributions.

Considering the range of applications and the fact that the theoretical results for matrix geometric solutions are well established, it is surprising that these results are not more widely utilized by practicing performance modelers. We believe the reason for this might lie in the fact that initially the mathematics behind the method seems formidable and our goal in this tutorial is to make Neuts' results more accessible. We do this by presenting a simplified derivation of the matrix geometric solution that draws on the parallels between the scalar and vector cases as outlined above. This derivation is based on the fact that the linear equations that determine the stationary state probabilities have a unique solution and thus guessed values can be checked by establishing that they satisfy the equations. In Section 2 of the paper we establish preliminary properties of Markov processes and set up the correspondence between scalar and vector state processes. The matrix geometric form is derived in this section as well as some of its properties. A more detailed and formal exposition of these results is found in [7]. Section 3 presents an example of a performance model that uses these results and in Section 4 we present our conclusions.

2 Markov Processes

We begin by stating some preliminary results on Markov process in Section 2.1. We then consider a simple example in Section 2.2 of a scalar Markov process that has a geometric solution. This will be used as an analogy to derive a more general formulation in Section 2.3 for vector state processes which leads to a matrix geometric form. We then generalize the vector process to a more complete formulation in Section 2.4 and establish properties of matrix geometric solutions in Section 2.5.

2.1 Preliminary Definitions

We let $\mathcal{S}$ be a set of states and let $X(t), -\infty < t < \infty$, be a time homogeneous, irreducible and stationary Markov process defined on $\mathcal{S}$ [10]. We will sometimes suppress the time dependency in our notation of $X(t)$. Without loss of generality we assume that $\mathcal{S} = \{0, 1, \ldots\}$ and initially assume that time is continuous. We later show how the results here apply to discrete time processes. The *state transition rate* from state i to state j, $i, j \in \mathcal{S}$, is defined as

$$r(i,j) \equiv \begin{cases} \lim_{\tau \to 0} \frac{P[X(t+\tau)=j \mid X(t)=i]}{\tau}, & i \neq j, \\ 0, & i = j, \end{cases} \tag{1}$$

and we define the *total transition rate* from state i as

$$r(i) \equiv \sum_{j \in S} r(i,j). \tag{2}$$

The generator matrix of the process, denoted by Q, is the matrix formed from the transition rates, $Q = \{q(i,j)\}$ that satisfies

$$q(i,j) = \begin{cases} r(i,j), & i \neq j, \\ -r(i), & i = j. \end{cases} \tag{3}$$

The stationary distribution of X is denoted by $\pi_i, i \geq 0$, and is equal to the fraction of time that the process spends in state i. We let $\pi \equiv (\pi_0, \pi_1, \ldots)$ be the vector of stationary probabilities. The stationary distribution is the unique set of $\pi_i \geq 0, i \geq 0$, that solves

$$\begin{aligned} \pi Q &= 0, & (4) \\ \pi \underline{e} &= 1, & (5) \end{aligned}$$

where $\underline{e}$ throughout the paper denotes an appropriately dimensioned column vector of 1's. Observe that the j'th equation of (4) is given by

$$\sum_{i=0}^{\infty} \pi_i q(i,j) = 0, \quad j \geq 0, \tag{6}$$

and corresponds to a *global balance* equation that must hold for state j (i.e. the total probability flux into and out of state j must be equal). It is important to note that there are many solutions to (4) and it alone can only be used to determine the relative values of $\pi_i, i \geq 0$. Provided that the Markov process is ergodic, the normalization equations (5) are used to determine the *unique* stationary probabilities. Because the solution is unique, if one can *guess* a possible solution for the values of $\pi_i, i \geq 0$, then these values can can be shown to be correct by demonstrating that they satisfy (4) and (5).

There is a well known equivalence between the stationary probabilities of a continuous time process and the stationary probabilities for a corresponding discrete time version of the process that we will use. To establish this equivalence, we select a transition step size Δ as

$$\Delta \equiv \sup_{i \geq 0} \ r(i), \tag{7}$$

and then consider the following transition probability matrix $P = \{p(i,j)\}$ given by

$$p(i,j) \equiv \begin{cases} r(i,j)/\Delta, & i \neq j, \\ 1 - r(i)/\Delta, & i = j. \end{cases} \tag{8}$$

Observe that the rows P sum to 1 and also, written in matrix form, we have that $P = Q/\Delta + I$ where I is the identity matrix. The value $p(i,j)$ is the probability of a state transition from state i to state j at embedded transition epochs. Direct substitution into (4) and (5) shows that the the stationary probabilities of this discrete version are the same as for the continuous process and solve

$$\begin{aligned} \pi &= \pi P, & (9) \\ \pi \underline{e} &= 1. & (10) \end{aligned}$$

There is thus a direct mapping between continuous and discrete time versions of a Markov process. For certain results, derivations are simplified in the discrete time process because one is relieved of the need to account for the time the process stays in a given state (an example of this is the derivation of the meaning of the R matrix in Section 2.5.2).

2.2 Scalar State Process

Consider the Markov process with state transition diagram shown in figure 1 which corresponds to a modification of a M/M/1 queue. Customer arrivals to the system when the system is in state $i, i \geq 1$, are assumed to have exponentially distributed interarrival times at a rate of λ and customer service times are exponential with rate μ. Interarrival times to an empty system are exponential with rate λ'. The system is ergodic if $\lambda < \mu$. The generator matrix, Q, of the process is given by

$$Q = \begin{bmatrix} -\lambda' & \lambda' & 0 & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 & \cdots \\ 0 & 0 & \mu & -(\lambda+\mu) & \lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}. \tag{11}$$

Solution techniques for the stationary probabilities for such simple problems are well known but it is instructive to derive a solution from first principles. The techniques we use here require *guessing* the form of the solution and the derivation here will be used to guide the derivation of the solution for a vector state process.

The repetitive structure of matrix Q can be see in terms of its columns. Number columns starting from zero and observe that, the j'th column for $j \geq 2$ is the same as the second column except that it is shifted down by $j - 2$ steps. We call the linear equations that arise from these columns of the matrix, the *repeating portion* of the process since they have a repetitive structure. The remaining equations, i.e. those that result from columns 0 and 1, are called the *boundary portion.* We use the terms repeating and boundary to equivalently discuss states of the Markov process, i.e. states $j, j \geq 2$, are the repeating states and states 0 and 1 are the boundary states. We use the repeating structure of this matrix below in determining a solution for the stationary probabilities.

We first write down the the balance equation for the repeating portion of the process,

$$\pi_{j-1}\lambda - \pi_j(\lambda+\mu) + \pi_{j+1}\mu, \quad j \geq 2, \tag{12}$$

and make an important observation about states in this portion of the process. Since arrivals only cause transitions to neighboring states, it is reasonable to believe that, given the value of

$\pi_{j-1}, j \geq 2$, the value of π_j is a function of only the transition rates between state $j-1$ and state j. For the process we are considering, these transition rates do not depend upon the value of j, and consequently this suggests that there is some unknown *constant* ρ such that

$$\pi_j = \rho\pi_{j-1}, \quad j \geq 2. \tag{13}$$

This implies that the values of $\pi_j, j \geq 2$, satisfy the following *geometric* form,

$$\pi_j = \rho^{j-1}\pi_1, \quad j \geq 2. \tag{14}$$

To determine a value for ρ and to check that this is indeed the solution form, we substitute this *guess* into (12). This shows that

$$\pi_1\rho^{j-2}\lambda - \pi_1\rho^{j-1}(\lambda+\mu) + \pi_1\rho^j\mu = 0, \quad j \geq 2, \tag{15}$$

which, after simplification, shows that

$$\lambda - \rho(\lambda+\mu) + \rho^2\mu = 0. \tag{16}$$

This quadratic has two possible solutions $\rho = 1$ or $\rho = \lambda/\mu$. Although $\rho = 1$ is a solution to (4) it cannot also satisfy (5) since this requires that $\rho < 1$ for convergence. Thus we have that $\rho = \lambda/\mu$.

We can write equations for the initial portion of the matrix as

$$\begin{aligned} -\pi_0\lambda' + \pi_1\mu &= 0 \qquad (17)\\ \pi_0\lambda' - \pi_1(\lambda+\mu) + \pi_2\mu &= 0, \qquad (18) \end{aligned}$$

or in matrix form as

$$(\pi_0, \pi_1)\begin{bmatrix} -\lambda' & \lambda' \\ \mu & -\mu \end{bmatrix} = 0, \tag{19}$$

where we have used the fact that $\pi_2 = \rho\pi_1$ in equation (18). It is clear that (19) does not have a unique solution since since the rank of the matrix is one less than the number of unknowns. To determine the unique solution for these quantities we use the normalization condition (5) which shows that

$$1 = \pi_0 + \pi_1\sum_{j=1}^{\infty}\rho^{j-1} = \pi_0 + \pi_1(1-\rho)^{-1}. \tag{20}$$

This in combination with (19) yields a unique solution given by

$$\begin{aligned} \pi_0 &= \frac{1}{1+\rho'/(1-\rho)} \qquad (21)\\ \pi_1 &= \frac{\rho'}{1+\rho'/(1-\rho)} \qquad (22) \end{aligned}$$

with $\rho' = \lambda'/\mu$. Equations (21), (22) and (14) thus constitute a complete solution which can be checked by demonstrating that it solves (4) and (5).

We used the following three steps to derive the solution for this simple process which will form the basis for subsequent derivations.

1. Based on the repetitive structure of the process, we guessed a geometric solution form for the repeating portion of the Markov process. This form required us to calculate the value of an unknown constant.
2. To calculate the value of the constant, we substituted the geometric form for one of the balance equations in the repeating portion of the process and solved for the root of that equation.
3. The boundary portion of the process was then solved using the results for the repeating portion of the process and the normalization condition.

Most performance measures can be obtained as a linear combination of some state dependent function using the stationary distribution. Suppose, for example, that we wish to calculate the expected number of customers waiting in the queue for this model, denoted by $\overline{N_q}$. We can do this by assigning a value of $j-1$ to state j and calculating

$$\begin{aligned} \overline{N_q} &= \sum_{j=1}^{\infty}(j-1)\pi_j = \pi_1 \sum_{j=1}^{\infty}(j-1)\rho^{j-1} && (23) \\ &= \pi_1 \frac{\rho}{(1-\rho)^2} = \frac{\rho\rho'}{(1-\rho)(1-\rho+\rho')} && (24) \end{aligned}$$

Note that if $\rho' = \rho$ then $\overline{N_q} = \rho^2/(1-\rho)$ which is the expected number of customers queued in a M/M/1 queueing system. More complex performance measures can be similarly calculated as a weighted sum of the stationary probabilities. We now duplicate these steps for a vector state queueing system.

2.3 Vector State Process

Suppose now we assume that the service requirements of jobs arriving to the system is the sum of two exponential stages with rate μ_1 and μ_2, respectively. To analyze this system, we augment the state descriptor and let the state be given by $(i, s), i \geq 0, s = 0, 1, 2$, where i is the number of customers in the queue (not including any receiving service) and s is the current stage of service of the customer in service. By definition we set s to be equal to 0 if there are no customers in the system. The state transition diagram for this system is shown in figure 2 where we have grouped states according to the total number of customers in the queue. We order states lexigraphically, i.e. $(0,0),(0,1),(0,2),(1,1),(1,2),\ldots$, and let $\pi_{(i,s)}$ be the stationary probability of state (i,s). We shall say that states at *level* i are those states defined by $(i,0)$ and $(i,1)$. The generator matrix is given by

$$Q=\begin{bmatrix}
-\lambda' & \lambda' & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots\\
0 & -a_1 & \mu_1 & \lambda & 0 & 0 & 0 & 0 & \cdots & \\
\mu_2 & 0 & -a_2 & 0 & \lambda & 0 & 0 & 0 & 0 & \cdots\\
0 & 0 & 0 & -a_1 & \mu_1 & \lambda & 0 & 0 & 0 & \cdots\\
0 & \mu_2 & 0 & 0 & -a_2 & 0 & \lambda & 0 & 0 & \cdots\\
0 & 0 & 0 & 0 & 0 & -a_1 & \mu_1 & \lambda & 0 & \cdots\\
0 & 0 & 0 & \mu_2 & 0 & 0 & -a_2 & 0 & \lambda & \cdots\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{bmatrix} \tag{25}$$

where we have defined $a_i = \lambda + \mu_i, i = 1,2$.

The structure of this generator matrix is a generalization of that found in the previous example (11). This is most easily seen by grouping the entries of (25) according to the number of customers in the system. Let $\underline{\pi}_i \equiv (\pi_{(i,1)}, \pi_{(i,2)})$ for $i \geq 1$, $\underline{\pi}_0 \equiv (\pi_{(0,0)}, \pi_{(0,1)}, \pi_{(0,2)})$ and $\underline{\pi} \equiv (\underline{\pi}_0, \underline{\pi}_1, \underline{\pi}_2, \ldots)$. Define the following matrices

$$A_0 \equiv \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \quad A_1 \equiv \begin{bmatrix} -a_1 & \mu_1 \\ 0 & -a_2 \end{bmatrix}, \quad A_2 \equiv \begin{bmatrix} 0 & 0 \\ \mu_2 & 0 \end{bmatrix}. \tag{26}$$

Also define

$$B_{1,0} \equiv \begin{bmatrix} 0 & 0 & 0 \\ 0 & \mu_2 & 0 \end{bmatrix}, \quad B_{0,1} \equiv \begin{bmatrix} 0 & 0 \\ \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \quad B_{0,0} \equiv \begin{bmatrix} -\lambda' & \lambda' & 0 \\ 0 & -a_1 & \mu_1 \\ \mu_2 & 0 & -a_2 \end{bmatrix}. \tag{27}$$

With these definitions we can group the generator matrix into blocks as follows

$$Q = \begin{array}{c|ccc||cc||cc||cc||c|}
 & (0,0) & (0,1) & (0,2) & (1,1) & (1,2) & (2,1) & (2,2) & & & \\
(0,0) & -\lambda' & \lambda' & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots\\
(0,1) & 0 & -a_1 & \mu_1 & \lambda & 0 & 0 & 0 & 0 & 0 & \cdots\\
(0,2) & \mu_2 & 0 & -a_2 & 0 & \lambda & 0 & 0 & 0 & 0 & \cdots\\
\hline\hline
 & 0 & 0 & 0 & -a_1 & \mu_1 & \lambda & 0 & 0 & 0 & \cdots\\
 & 0 & \mu_2 & 0 & 0 & -a_2 & 0 & \lambda & 0 & 0 & \cdots\\
\hline\hline
 & 0 & 0 & 0 & 0 & 0 & -a_1 & \mu_1 & \lambda & 0 & \cdots\\
 & 0 & 0 & 0 & \mu_2 & 0 & 0 & -a_2 & 0 & \lambda & \cdots\\
\hline\hline
 & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{array} \tag{28}$$

$$= \begin{bmatrix} B_{0,0} & B_{0,1} & 0 & 0 & 0 & \dots \\ B_{1,0} & A_1 & A_0 & 0 & 0 & \dots \\ 0 & A_2 & A_1 & A_0 & 0 & \dots \\ 0 & 0 & A_2 & A_1 & A_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \tag{29}$$

where a 0 entry in (29) (and also in other matrices in this paper) is a matrix of all zeros of the appropriate dimension.

Comparing (11) and (29) shows that they have a similar structure with the exception that scalar entries in (11) have been replaced with matrix entries in (29). We again call the repeating portion of the process to be the set of linear equations starting at the second blocked column and the boundary portion to be the equations associated with the 0'th and 1'st blocked column (we similarly will call the associated states, the repeating and boundary states). We solve for the stationary probabilities in exactly the same manner as in the scalar example except that here we deal with vectors instead of scalars. First we write down the equation for the repeating portion of the process which given in block matrix form is

$$\underline{\pi}_{j-1}A_0 + \underline{\pi}_j A_1 + \underline{\pi}_{j+1}A_2 = 0, \quad j \geq 2. \tag{30}$$

Similarly to the scalar case, since state transitions are between nearest blocks, if we are given the value of $\underline{\pi}_{j-1}, j \geq 2$, then it would not be surprising to find that the value of $\underline{\pi}_j$ is a function only of the transition rates between states with $j-1$ queued customers and states with j queued customers. Since these transition rates do not depend upon the value of j, this suggests that there is some *constant matrix* R such that

$$\underline{\pi}_j = \underline{\pi}_{j-1}R, \quad j \geq 2, \tag{31}$$

and that the values of $\underline{\pi}_j, j \geq 2$, have a *matrix geometric* form, i.e.

$$\underline{\pi}_j = \underline{\pi}_1 R^{j-1}, \quad j \geq 2. \tag{32}$$

Substituting this *guess* into (30) shows that

$$\underline{\pi}_1 R^{j-2}A_0 + \underline{\pi}_1 R^{j-1}A_1 + \underline{\pi}_1 R^j A_2 = 0, \quad j \geq 2, \tag{33}$$

which on simplifying yields

$$A_0 + RA_1 + R^2A_2 = 0. \tag{34}$$

This is a quadratic in the matrix R which is typically solved numerically (we discuss computational aspects of the matrix geometric method in Section 2.5.3). Recall that in the quadratic equation for the scalar case there were two possible solutions only one of which could satisfy the normalization condition. Similar to the scalar case, we pick the minimal matrix R that satisfies (34). If the normalization constant is satisfied for the vector state process, it must be the case that $\underline{\pi}_1 \sum_{j=1}^{\infty} R^{j-1}\underline{e} < \infty$. Here the analogous criteria to $\rho < 1$ in the scalar case, is that the spectral radius of R must be less than 1. This follows from the fact that all eigenvalues of R must be less than 1 for the sum to converge [3].

Assume that we have such a solution to (34). To determine the stationary probabilities we continue as in scalar state case. The equations for the initial portion of the matrix are given by

$$\begin{aligned} \underline{\pi}_0 B_{0,0} + \underline{\pi}_1 B_{1,0} &= 0 && (35) \\ \underline{\pi}_0 B_{0,1} + \underline{\pi}_1 A_1 + \underline{\pi}_2 A_2 &= 0 && (36) \end{aligned}$$

which can be written in matrix form as

$$(\underline{\pi}_0, \underline{\pi}_1) \begin{bmatrix} B_{0,0} & B_{0,1} \\ B_{1,0} & A_1 + RA_2 \end{bmatrix} = 0 \qquad (37)$$

where we have used the fact that $\underline{\pi}_2 = R\underline{\pi}_1$ in (36). As in the scalar case, equations (37) are not sufficient to determine the values of $\underline{\pi}_0$ and $\underline{\pi}_1$ and we must use the normalization condition (5) to determine these values. This yields

$$1 = \underline{\pi}_0\underline{e} + \underline{\pi}_1 \sum_{j=1}^{\infty} R^{j-1}\underline{e} = \underline{\pi}_0\underline{e} + \underline{\pi}_1(I - R)^{-1}\underline{e} \qquad (38)$$

which together with (37) yields a unique solution.

Suppose, as in the scalar case, that we wish to calculate the expected the number of queued customers, denoted by $\overline{N_q}$. We do this by assigning the value of $j - 1$ to states $(j, 1)$ and $(j, 2)$ and then calculate an expectation of this value using the stationary distribution. This yields

$$\begin{aligned} \overline{N_q} &= \sum_{j=1}^{\infty}(j-1)\underline{\pi}_j\underline{e} = \underline{\pi}_1 \sum_{j=1}^{\infty}(j-1)R^{j-1}\underline{e} && (39) \\ &= \underline{\pi}_1 R(I - R)^{-2}\underline{e} && (40) \end{aligned}$$

Table 1 summarizes the analogous derivations for the scalar and vector state processes.

Table 1. Analogous Derivations

Step	Scalar process	Vector Process
j'th Balance Equation	$\pi_{j-1}\lambda - \pi_j(\lambda+\mu) + \pi_{j+1}\mu = 0$	$\underline{\pi}_{j-1}A_0 + \underline{\pi}_j A_1 + \underline{\pi}_{j+1}A_2 = 0$
Guessed Geometric Solution	$\pi_j = \rho^{j-1}\pi_1$	$\underline{\pi}_j = R^{j-1}\underline{\pi}_1$
Solution Repeating Portion	$\lambda - \rho(\lambda+\mu) + \rho^2\mu = 0$	$A_0 + RA_1 + R^2A_2 = 0$
Solution Initial Portion	$(\pi_0, \pi_1)\begin{bmatrix} -\lambda' & \lambda' \\ \mu & -\mu \end{bmatrix} = 0,$	$(\underline{\pi}_0, \underline{\pi}_1)\begin{bmatrix} B_{0,0} & B_{0,1} \\ B_{1,0} & A_1 + RA_2 \end{bmatrix} = 0$
Normalization Condition	$1 = \pi_0 + \pi_1(1-\rho)^{-1}$	$1 = \underline{\pi}_0\underline{e} + \underline{\pi}_1(I-R)^{-1}\underline{e}$
Performance Measure	$\overline{N_q} = \pi_1\rho(1-\rho)^{-2}$	$\overline{N_q} = \underline{\pi}_1 R(I-R)^{-2}\underline{e}$

2.4 Matrix Geometric Solutions

In this section we present the matrix geometric solution technique in a more general context than that previously considered. Since scalars are special cases of vectors, our results here also hold for more general scalar state processes. The solution techniques employed here are identical to those used in the simple cases considered in Section 2. Consider a Markov process that has a block generator matrix given by

$$Q = \begin{array}{c|cccccc|} & m_1-m & m & m & m & & \\ \hline m_1-m & B_{0,0} & B_{0,1} & 0 & 0 & 0 & \cdots \\ m & B_{1,0} & B_{1,1} & A_0 & 0 & 0 & \cdots \\ m & B_{2,0} & B_{2,1} & A_1 & A_0 & 0 & \cdots \\ & B_{3,0} & B_{3,1} & A_2 & A_1 & A_0 & \cdots \\ & B_{4,0} & B_{4,1} & A_3 & A_2 & A_1 & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \vdots & \end{array} \qquad (41)$$

where $B_{0,0}$ is a square matrix of dimension $(m_1 - m)$, $B_{0,1}$ is of dimension $(m_1 - m) \times m$, matrices $B_{k,0}$, $k \geq 1$, are of size $m \times (m_1 - m)$, and all other matrices are square of dimension m, where $m_1 \geq m$. The B matrices here correspond transitions from boundary states.

When referring to the states of the process it is convenient, as in the previous sections, to define *levels* of states. The 0'th level of the process associated with Q of equation (41) are the $m_1 - m$ boundary states. The first level are the next m states, and generally the i'th level for $i \geq 1$, are those states indexed by $m_1 - m + (i-1)m + l$, $l = 0, 1, \ldots, m-1$. We will index the repeating portion of the process by a tuple (i, j), $i \geq 1, 0 \leq j \leq m-1$, where i is the level of the state and j is called an *interlevel state*. The state of the original process indexed by (i, j) is $m_1 - m + (i-1)m + j$.

To write a general solution for the generator in (41) we proceed similarly to Sections 2.2 and 2.3. We first write a balance equation for the repeating portion of the process

$$\sum_{k=0}^{\infty} \underline{\pi}_{j-1+k} A_k = 0, \quad j \geq 2. \tag{42}$$

We then guess a geometric solution

$$\underline{\pi}_j = \underline{\pi}_{j-1} R, \quad j \geq 2, \tag{43}$$

or that

$$\underline{\pi}_j = \underline{\pi}_1 R^{j-1}, \quad j \geq 2. \tag{44}$$

Substituting this guess into (42) and simplifying shows that R solves

$$\sum_{k=0}^{\infty} R^k A_k = 0. \tag{45}$$

To determine a solution for the boundary states we must solve the following linear equations

$$(\underline{\pi}_0, \underline{\pi}_1) \left[\begin{array}{cc} B_{0,0} & B_{0,1} \\ \sum_{k=1}^{\infty} R^{k-1} B_{k,0} & \sum_{k=1}^{\infty} R^{k-1} B_{k,1} \end{array} \right] = 0. \tag{46}$$

These equations, however, are not of full rank and we must use the normalization condition to determine the solution. This is given by equation (38) which, together with (46), provides a unique solution. A way to compute this unique solution using a linear equation solver, is to replace one of the columns in the matrix in (46) with the column that represents the normalization condition and to appropriately change the right hand side of (46). Suppose, for example, we replace the first column of the matrix with the normalization condition. This yields

$$(\underline{\pi}_0, \underline{\pi}_1) \left[\begin{array}{ccc} \underline{e} & B_{0,0}^{\star} & B_{0,1} \\ (I-R)^{-1}\underline{e} & \left[\sum_{k=1}^{\infty} R^{k-1} B_{k,0}\right]^{\star} & \sum_{k=1}^{\infty} R^{k-1} B_{k,1} \end{array} \right] = [1, 0] \tag{47}$$

where $B^{\star}_{0,0}$ and $\left[\sum_{k=1}^{\infty} R^{k-1} B_{k,0}\right]^{\star}$ result from the removal of the first column from the matrix in (46) and $[1, 0]$ is a row vector of consisting of a 1 followed by $m_1 - 1$ zeros. Equation (47) is in the form suitable for most linear equation solvers and uniquely determines $\underline{\pi}_0$ and $\underline{\pi}_1$.

We end this section with some observations about how this general matrix form relates to problems often encountered in performance models. We first observe again that performance measures can then be calculated from the stationary probabilities. In practice one typically finds that matrices $B_{k,0}$ and $B_{k+1,1}$ are zero for values $k \geq K + 1$, $K \geq 1$. Similarly one often finds that that A_k are zero starting at $k > K + 1$. The value of K is the maximum number of levels that the process can jump down in the repeating portion of the process and for many performance models this value is small. For example the Markov process considered in Section 2.3 with generator matrix given in (29) has $m_1 = 5$ and $m = 2$, $B_{1,1} = A_1$, $B_{2,1} = A_2$ and and $K = 1$. Henceforth in the paper we will assume that K is finite.

A special case of matrix Q above that frequently occurs in practice is that where $m_1 = m$ which leads to a generator of the form

$$Q = \begin{bmatrix} B_0 & A_0 & 0 & 0 & 0 & \dots \\ B_1 & A_1 & A_0 & 0 & 0 & \dots \\ B_2 & A_2 & A_1 & A_0 & 0 & \dots \\ B_3 & A_3 & A_2 & A_1 & A_0 & \dots \\ B_4 & A_4 & A_3 & A_2 & A_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \tag{48}$$

where all matrices are square of dimension m.

2.5 Properties of Matrix Geometric Solutions

In this section we establish a condition for the stability of the process in terms of its block matrices, provide an interpretation for the R matrix, called the *rate matrix*, and discuss computational procedures associated with the method. We first start with an equation for the stability of the process.

2.5.1 Stability

We will interchangeably use the words ergodicity and stability. Loosely speaking, the ergodicity of a process depends upon the expected total *drift* of the process for states in the repeating portion of the process. For example, in the M/M/1 queueing model of figure 1, if we select a given state in the repeating portion then its expected drift towards higher states is given by λ (a rate of λ times a distance of plus 1 unit) and its expected drift towards lower states by

$-\mu$ (a rate of μ times a distance of minus 1 unit). The total drift of the process is given by $\lambda - \mu$ and the process is stable if this total drift is negative leading to the well known stability condition of $\lambda < \mu$. Intuitively this means that the expected direction of the process is towards lower valued states. If the total drift is positive, then the process tends to move towards higher valued states and is unstable.

In the calculation of the total drift of a process we include a notion of distance. For example, suppose in the scalar process transitions can go up by 1 step and down by at most K steps. Also suppose that the transition rate for l steps is given by $r(l), l = -K, -K-1, \ldots, -1, 1$. Then the drift upwards is given by $r(1)$, the drift downwards is given by $-\sum_{l=1}^{K} lr(-l)$ and the total drift by the sum of these two components. Stability of the process implies that

$$r(1) < \sum_{l=1}^{K} lr(-l). \tag{49}$$

We now investigate how we apply these concepts to processes with a matrix geometric structure.

Analogous to the scalar case, we will think of the drift of the process in terms of levels. Assume, as above, that transitions can go up one level or down by at most K levels. We wish to calculate the total drift from a level in the repeating portion of the process analogous to (49). To do this we consider the process for levels that are far from the boundary, i.e. level i such that $i >> 0$. It is clear that the stability of the system only depends upon the expected drift from these levels. Now there is one complication that arises when the process is of matrix geometric form. To see this consider a transition from level $i, i >> 0$, to level $i - k, 1 \leq k \leq K$, i.e. a distance of k downward. To calculate the drift for such a transition we must know which rate from level i to apply. If for example we knew that the transition was from interlevel state j, $0 \leq j \leq m-1$, then the component of the drift for this transition is given by

$$-k \sum_{l=0}^{m-1} A_{k+1}(j,l) \tag{50}$$

where $A_{k+1}(j,l)$ is the entry in position (j,l) of matrix A_{k+1} and is the transition rate from state j in level i to state l in level $i-k$. To define the average drift let f_j, $0 \leq j \leq m-1$, be the probability that the process is in interlevel state j of the repeating portion of the process for level very far from the boundary, i.e. for $i >> 0$ (intuitively speaking since these levels are far from the boundary, the values of f_j, $0 \leq j \leq m-1$, are independent of the level i). The average drift from level i to level $i-k$ then is given by

$$-k \sum_{j=0}^{m-1} f_j \sum_{l=0}^{m-1} A_{k+1}(j,l) \tag{51}$$

and to obtain the total rate we sum (51) over all $k, 0 \leq k \leq K+1$.

The difficulty now in performing this calculation is in determining the values of f_j, the probabilities that when the process is in interlevel state j. To determine these values we can consider a derived Markov process on state space $0, 1, \ldots, m-1$, which identifies transitions only in terms of their interlevel states. Specifically, if the original process has a transition from state (i, j) to state $(i-k, l)$, then in the derived process we consider that a transition occurs from interlevel state j to interlevel state l. It can be easily seen that the generator for this derived process is given by the matrix A where

$$A = \sum_{l=0}^{K+1} A_l, \tag{52}$$

and that the probabilities f_j are given by its stationary measure, i.e. they solve

$$\begin{aligned} \underline{f}A &= 0, & (53)\\ \underline{f}\underline{e} &= 1, & (54) \end{aligned}$$

where $\underline{f} \equiv (f_0, f_1, \ldots, f_{m-1})$. Using this in the above drift analysis and simplifying yields a stability condition given by

$$\underline{f}A_0\underline{e} < \sum_{k=2}^{K+1} (k-1)\underline{f}A_k\underline{e}. \tag{55}$$

2.5.2 Interpretation of the Rate Matrix

We now discuss an interpretation of the rate matrix. Our discussion for this is simplified if we first consider the discrete time case. We thus form the matrix P by using the construction $P = Q/\Delta + I$ from Section 2.1. The matrices for the repeating portion of the discrete process, denoted by A_k' are then given by

$$A_k' = \begin{cases} A_k/\Delta, & k \neq 1, \\ A_1/\Delta + I, & k = 1. \end{cases} \tag{56}$$

In a manner similar to that which lead to equations (45) we can show that the R matrix for the discrete version of the process, R_0, satisfies

$$R_0 = \sum_{l=0}^{K+1} R_0^l A_l'. \tag{57}$$

We now show how one can interpret the entries in R_0 (see [6] for the first derivation of these results). To do this we construct a transient Markov process. Consider the repeating portion of the process beginning with some given level and reindex levels from this point starting with the index 0. To start the process we select a *starting interlevel state*, say j, $0 \leq j \leq m-1$, in level 0 and let the process run according to state transitions given by $A'_k, 0 \leq k \leq K+1$. We stop the process (i.e. it *absorbs*) the first time it makes a transition to a state below level 1. Thus if the first transition of the process is not to level 1 (i.e. the first transition is not equal to one of the transitions $A_0(j,j')$ where $0 \leq j' \leq m-1$) then the process absorbs at the first step. Clearly the induced transient process depends upon the starting interlevel state j and we will denote the transient process that starts from level 0 in interlevel state j by Υ_j, $0 \leq j \leq m-1$.

Define

$$P_i^{(n)} = \{P_i^{(n)}(j,j'),\ i \geq 1,\ n \geq i,\ 0 \leq j,j' \leq m-1\} \tag{58}$$

to be a square matrix of dimension m where the value in position (j,j') is the probability that after n steps process Υ_j is in interlevel state j' of level i. We let $P_0^{(n)} = I$, (I is the identity matrix) for all n and note that since the process can go up by at most one level at each transition, it is clear that $P_i^{(n)} = 0$ for $i > n$. We also define

$$N_i = \{N_i(j,j'),\ i \geq 1,\ 0 \leq j,j' \leq m-1\} \tag{59}$$

to be a square matrix of dimension m where the value in position (j,j') is the expected number of visits made in Υ_j to interlevel state j' of level i before absorption. We set $N = N_1$. It is clear that

$$N_i = \sum_{n=i}^{\infty} P_i^{(n)}, \quad i \geq 1. \tag{60}$$

We now make the observation that if process Υ_j, $0 \leq j \leq m-1$, is in any interlevel state of level i, $i > 1$ at step $n, n \geq i$, it must have passed through level $i-1$. Suppose that at step $l, l < n$, Υ_j was in interlevel state $j', 0 \leq j' \leq m-1$, and that this was its *last visit* to level $i-1$ prior to step n. Then if we reindex levels so that level $i-1$ becomes level 0 it is clear that the future evolution of the process is statistically identical to $\Upsilon_{j'}$. This follows from the fact that step l is assumed to be the last visit to level $i-1$ and thus the process can be thought of stopping for any transition into levels less than i. These arguments permit us to write

$$P_i^{(n)} = \sum_{l=i-1}^{n-1} P_{i-1}^{(l)} P_1^{(n-l)}, \quad i \geq 2, \tag{61}$$

where $P_{i-1}^{(l)}$ in (61) is the probability that Υ_j visits level $i-1$ in l steps and the factor $P_1^{(n-l)}$ accounts for the fact that this is the last visit to level $i-1$ since it is the probability that Υ_j starting from level $i-1$ visits level i in $n-l$ steps without ever going below level i.

Summing this over all value n and using (60) yields

$$\begin{aligned} N_i &= \sum_{n=i}^{\infty} \sum_{l=i-1}^{n} P_{i-1}^{(l)} P_1^{(n-l)} \\ &= \sum_{l=i-1}^{\infty} P_{i-1}^{(l)} \sum_{n=l+1}^{\infty} P_1^{(n-l)} \\ &= N_{i-1} N, \qquad i \geq 2. \end{aligned}$$

which implies that

$$N_i = N^i, i \geq 2. \tag{62}$$

Assume now that we observe process Υ_j, $0 \leq j \leq m-1$, at step $n-1$ and it makes a transition into level 1. Conditioning on all possible such transitions yields

$$P_1^{(n)} = \sum_{l=0}^{K+1} P_l^{(n-1)} A_l'. \tag{63}$$

Summing this over n and using (62) yields

$$\begin{aligned} N &= \sum_{l=0}^{K+1} \sum_{n=1}^{\infty} P_l^{(n-1)} A_l' & (64) \\ &= \sum_{l=0}^{K+1} N^l A_l' & (65) \end{aligned}$$

where we select the minimal solution. Comparing (65) with (57) demonstrates that $R_0 = N$. Thus an interpretation for R_0 is that the entry in the (j, j') position is the expected number of visits to interlevel state j' of level 1 before absorption given that the process is started with an interlevel index of j of level 0.. In the case of continuous time processes, the interpretation of the R matrix is complicated by the fact that we must include the time the process spends in each state in the calculation. We let the *sojourn time* in interlevel state j be the value $-1/A_1(j,j)$. For the continuous case the (j, j') entry of R is the expected time (measured in interlevel state j's sojourn time) spent in interlevel state j of level 1 before absorption when started with an interlevel index of j'.

For some performance models the interpretation of the rate matrix can be used to determine explicit solutions for some of its elements. For example, if there is no path from interlevel state j at level i to interlevel state j' at level $i+1$ then $R(j, j') = 0$. We can also derive explicit expressions if interlevel j is *isolated* from other interlevels. To show this consider a simple case where $K = 1$ and there are transitions between interlevel states j, i.e. that $A_0(j,j) \neq 0$ and

$A_2(j,j) \neq 0$. Interlevel j is isolated if there are no transitions from it into other interlevels , i.e. if $A(j,j') = 0$ for $j \neq j'$ where A is given by (52). Here the entry in $R(j,j)$ is what would be obtained from a scalar state birth death process, i.e. $R(j,j) = A_0(j,j)/A_2(j,j)$. Other generalizations of isolation that lead to explicit entries in the rate matrix are clearly possible.

2.5.3 Computational Properties

One iterative procedure which is often used to solve for matrix R is given by

$$R(0) = 0, \tag{66}$$

$$R(n+1) = - \sum_{l=0,\ l\neq 1}^{\infty} R^l(n) A_l A_1^{-1}, \quad n \geq 0, \tag{67}$$

where the iteration halts when entries in $R(n+1)$ and $R(n)$ differ in absolute value by less than a given small constant. The equation in (67) is obtained by multiplying equation (45) from the right by A_1^{-1}. It can be shown that the entries in the sequence $\{R(n)\}$ are entry-wise nondecreasing and and converge monotonically to a non-negative matrix R which satisfies (45). Experimental results on other computational procedures for calculating the rate matrix can be found in [2, 7, 9]. See also [8] for a class of models where there is an explicit solution for the rate matrix and [11] for a comparison of matrix geometric methods with other methods that take advantage of structural features of the global balance equations.

We note that the number of iterations needed for convergence increases as the spectral radius of R increases. Similar to the scalar case where ρ played the part of R, the spectral radius of R in many performance models can be thought of as a measure of the utilization of the system. This implies that for these cases as the utilization of the system increases it becomes computationally more difficult to compute the entries in R. Most of the computational effort associated with the matrix geometric method is, in fact, expended in computing R. In our experience most problems were sufficiently small (typically $K \leq 5$ and $m \leq 50$) so that performance measures could be graphed interactively. The largest problem solved arose in a performance model of a parallel processing system where the state space grew exponentially since it consisted of a partition of the integers [5]. Here $K = 1$ and $m = 285$ and the solution was seriously compute and memory intensive. Performance measures were calculated in overnight executions with much less computational effort than would be required for simulations which has the disadvantage of only producing approximate results.

For many performance models the A_0 matrix is a diagonal where $A_0(j,j) = \lambda, 0 \leq j \leq m-1$, and λ is the arrival rate of customers to the system. This corresponds to a state independent Poisson arrival stream of customers. Here as λ increases the spectral radius of the R matrix also increases and performance measures become increasingly more compute intensive because (67) takes more iterations to converge. Typically one wishes to evaluate some performance measure, say expected response time, for many different arrival rates $0 \leq \lambda_1 < \lambda_2 < \ldots < \lambda_L$. Using the probabilistic interpretation for the R matrix, it is easy to see that for some models $R_l \leq R_{l+1}$, $1 \leq l \leq L-1$, where R_l denotes the rate matrix corresponding to an arrival rate of λ_l. Computational effort can be saved in these cases if one starts the iteration for $l > 1$ with the previously calculated rate matrix, i.e. using $R_l(0) = R_{l-1}$ rather than $R_l(0) = 0$ in (66).

3 An Example

In this section we provide an example of a problem that is solved using matrix geometric analysis. The model is a simplified version of a replicated data base that is analyzed in [4]. We consider the performance of a data base system where there are two replications of the data. Each replication is independently accessible and is modeled by a server. Requests to access the data base are assumed to queue in a central location and correspond to read or write operations. To preserve the integrity of both copies of the data base, we assume that write requests must wait until both copies of the data base are available before beginning execution. Both copies are then assumed to be updated in parallel and then released simultaneously. Read requests can be processed by any copy of the data base. Both types of requests are assumed to wait in the common queue in the order in which they arrive. We assume that requests arrive to the system from a Poisson point source with intensity of λ and that the probability a given request is a read (resp. write) is given by r (resp. $1-r$). Service times for both read and write requests are assumed to be exponential with a average value of μ^{-1}. Since we assume that writes are served in parallel the total service time for write requests is equal to the maximum of two exponential random variables with parameter μ.

We let I_t be the number of requests at time t that are waiting in the common queue and let $J_t, 0 \leq 2$, be the number of replications that are involved in a read or write operation at time t. Our assumptions above imply that (I_t, J_t) is a Markov process. The state transition diagram for the process is given in Figure 3. We explain some of the transitions from the repeating portion of the process. In state $(2,2)$ both servers are busy serving customers and the customer at the head of the queue is equivalent to an unexamined arrival. Thus it is a read (resp. write) request with probability r (resp. $(1-r)$). Upon service completion at rate 2μ, the next state will be $(1,2)$ with probability r and $(2,1)$ with probability $1-r$. The rest of the transitions can be explained similarly.

If we order the state lexigraphically, the generator matrix of the process is given by

$$Q = \begin{bmatrix} -\lambda & \lambda r & \lambda(1-r) & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda r & \lambda(1-r) & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 2\mu & -(\lambda+2\mu) & 0 & \lambda & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & \mu & -(\lambda+\mu) & 0 & \lambda & 0 & 0 & 0 & \dots \\ 0 & 0 & 2\mu r & 2\mu(1-r) & -(\lambda+2\mu) & 0 & \lambda & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \mu & -(\lambda+\mu) & 0 & \lambda & 0 & \dots \\ 0 & 0 & 0 & 0 & 2\mu r & 2\mu(1-r) & -(\lambda+2\mu) & 0 & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \tag{68}$$

This generator is of matrix geometric form. To easily see this identify the matrices of (41) as follows

$$B_{0,0} = \begin{bmatrix} -\lambda & \lambda r & \lambda(1-r) \\ \mu & -(\lambda+\mu) & \lambda r \\ 0 & 2\mu & -(\lambda+2\mu) \end{bmatrix}, \quad B_{0,1} = \begin{bmatrix} 0 & 0 \\ \lambda(1-r) & 0 \\ 0 & \lambda \end{bmatrix}, \tag{69}$$

$$B_{1,0} = \begin{bmatrix} 0 & 0 & \mu \\ 0 & 0 & 2\mu r \end{bmatrix}, \quad B_{1,1} = A_1 = \begin{bmatrix} -(\lambda+\mu) & 0 \\ 2\mu(1-r) & -(\lambda+2\mu) \end{bmatrix}, \tag{70}$$

$$A_0 \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}, \quad \text{and} \quad A_2 \begin{bmatrix} 0 & \mu \\ 0 & 2\mu r \end{bmatrix}. \tag{71}$$

The system thus has a matrix geometric form and we can use the procedure outlined in Section 2.4 to solve for its stationary distribution and performance measures.

To determine the stability of the system we first form the matrix A of equation (52) which is given by

$$A = \begin{bmatrix} -\mu & \mu \\ 2\mu(1-r) & -2\mu(1-r) \end{bmatrix}. \tag{72}$$

Solving equations (53) and (54) yields

$$f_1 = \frac{2(1-r)}{3-2r} \quad \text{and} \quad f_2 = \frac{1}{3-2r}, \tag{73}$$

which when used in (55) yields a stability condition given by

$$\lambda < \frac{2\mu}{3-2r}. \tag{74}$$

Observe that if $r = 1$ then the stability is identical to that obtained in an M/M/2 queue, i.e. $\lambda < 2\mu$ and for $r = 0$ it is identical to that obtained for a M/G/1 queue where the expected service time is the maximum of 2 exponentials at rate μ, i.e. $\lambda < 2\mu/3$.

4 Conclusions

In this tutorial we derived basic results of matrix geometric solutions. This solution method is applicable to Markov processes that have an infinite repetitive structure in terms of finite vectors of states (see [1] for a matrix geometric solution form for a process that has a finite repetitive structure). The tutorial only scratches the surface for results in this area and the reader is referred to Neuts' elegant book [7] for further reading. This tutorial, however, is complete in that it presents sufficient material for a modeler to create and solve performance models having a matrix geometric form.

Acknowledgements The author gratefully acknowledges comments received after careful readings of the paper by Cheng-Shang Chang and Mark Squillante. Incorporation of their suggestions led to an improved version of the paper.

References

[1] Gun, L. and Makowski, A., "Matrix-Geometric Solution for Finite Capacity Queues with Phase-Type Distributions", *Proceedings of Twelfth IFIP WG 7.3 International Symposium on Computer Performance Modeling*, Brussels, Dec. 7-9, 1987, ed. P.J. Courtois and G. Latouche, pp.269-282, 1987.

[2] Gun, L., "Experimental Results on Matrix-Analytical Solution Techniques", *Communications in Statistics. Stochastic Models*, Vol. 5, pp.669-682, 1989.

[3] Horn, R.A. and Johnson, C.R., "Matrix Analysis", *Cambridge*, 1987.

[4] Nelson, R. and Iyer, B.R., "Analysis of a Replicated Data Base", *Performance Evaluation*, Vol. 5, pp.133-148, 1985.

[5] R. Nelson, "A Performance Evaluation of a General Parallel Processing Model", *Performance Evaluation Review*, Vol 18, no 1, pp.13-26, 1990.

[6] Neuts, M.F., "The Probabilistic Significance of the Rate Matrix in Matrix-Geometric Invariant Vectors", *J. Appl. Prob.*, Vol. 17, pp.291-96, 1980.

[7] Neuts, M.F., "Matrix Geometric Solutions in Stochastic Models",*John Hopkins University Press*, 1981.

[8] Ramaswami, V. and Latouche, G., "A General Class of Markov Processes with Explicit Matrix-Geometric Solutions", *Operations Research Spektrum*, Vol 8, pp.209-218, 1986.

[9] Ramaswami, V. and Latouche, G., "An Experimental Evaluation of the Matrix-Geometric Method for the $GI/PH/1$ Queue", *Communications in Statistics. Stochastic Models*, Vol 5, pp.629-667, 1989.

[10] S. Ross, "Stochastic Processes", *Wiley*, 1983.

[11] Snyder, P.M. and Stewart, W.J., "Explicit and Iterative Numerical Approaches to Solving Queueing Models", *Operations Research*, Vol 33, pp.183-202, 1985.

Figure 1. Simple Scalar Process

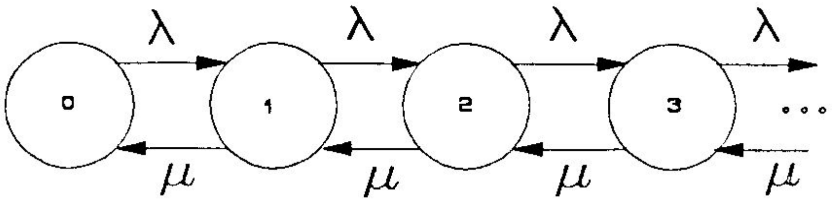


Figure 2. Simple Vector Process

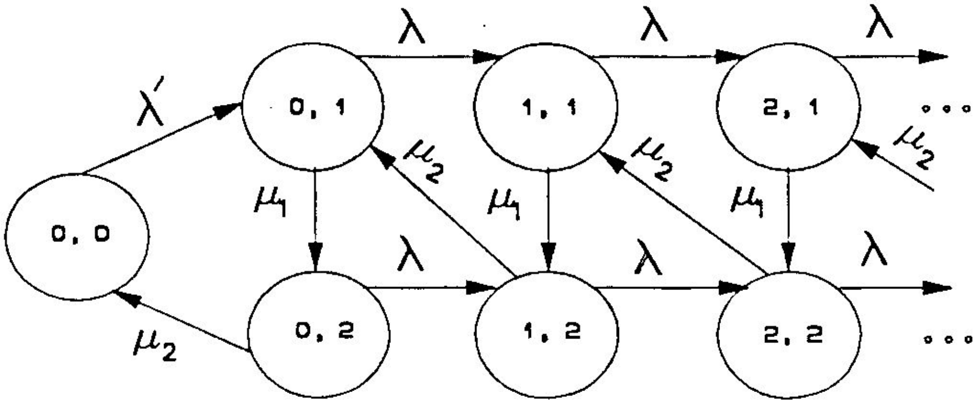
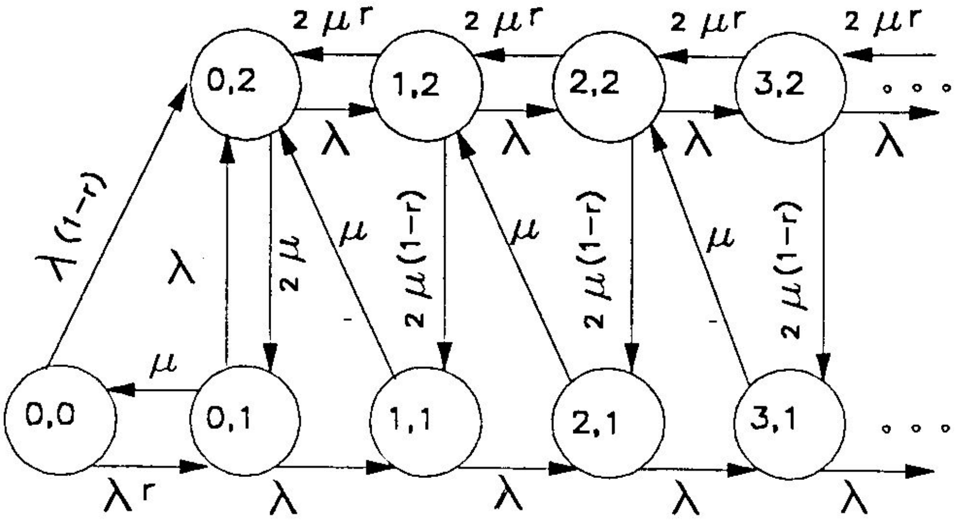


Figure 3. Replicated Data Base Model


2 μr
2 μr
2 μr
2 μr
0,2
1,2
2,2
3,2
λ
λ
λ
λ
λ(1-r)
λ
2 μ
μ
2 μ(1-r)
μ
2 μ(1-r)
μ
2 μ(1-r)
μ
0,0
0,1
1,1
2,1
3,1
λr
λ
λ
λ
λ