

# **SERIE RESEARCH MEMORANDA**

TRUNCATION OF MARKOV CHAINS WITH  
APPLICATIONS TO QUEUEING

Nico M. van Dijk

Research Memorandum 1988-53      Dec. '88



VRIJE UNIVERSITEIT  
FACULTEIT DER ECONOMISCHE WETENSCHAPPEN  
EN ECONOMETRIE  
AMSTERDAM



TRUNCATION OF MARKOV CHAINS  
WITH  
APPLICATIONS TO QUEUEING

Nico M. van Dijk  
Free University, Amsterdam, The Netherlands

**Abstract**

State space truncation is frequently demanded for practical computations of large or infinite Markov chains. Conditions are given which guarantee an error bound or rate of convergence. Concrete truncations and explicit error bounds are obtained for two non-product form queueing applications: an overflow model and a tandem queue with blocking.

**Keywords** Markov chain \* truncation \* error bound \* overflow model \*  
tandem queue



## 1 Introduction

Markov chain theory has proven to be a powerful tool for both modeling and computational purposes in practice. A variety of practical models have so been analyzed such as in communication, manufacturing and reliability (e.g. [18]). Particularly, the steady state behavior is frequently of main interest. Analytical expressions to this end are available only in a limited number of applications, so that computational procedures such as successive approximation are to be used. In practice, however, one often encounters large or infinite state spaces such as in an infinite server queue or a maintenance system with an a priori unbounded lifetime. Truncation of the state space then becomes necessary.

Though the technique of state space truncation is a common feature in practice, theoretical support in terms of orders of accuracy or rates of convergence seems hardly available. Convergence proofs as the truncation size tends to infinity have already been investigated in the early fifties by Savynsakov and were cristallized most notably by Seneta (1967), (1968) with reference to private communications with Kendall. A detailed study of these convergence results as well as an extensive list of related literature can be found in Seneta (1980). In this latter reference, also simple error bounds are provided (cf. theorem 6.4 and its corollary, p.215), but these are just robust bounds and do not secure an order of accuracy.

A different though related line of research is that of approximating or perturbed Markov chains (cf. Schweitzer (1968), Hinderer (1978), Whitt (1978), Meyer (1980), Van Dijk and Puterman (1988), Van Dijk (1988)). These references pay attention also to orders of accuracy. However, none of these results can be directly adopted for truncation purposes, as they basically require one and the same or at most a minorly perturbed state space.

The present paper therefore is concerned with a simple but general criterion which guarantees an order of accuracy of truncated Markov chains. To this end, two concrete conditions are provided. The underlying insight of these conditions is essentially based upon bounded bias terms for total reward structures or relatedly of bounded mean first passage times. The results typically seem to be applicable to random walk or spatial birth-

death type structures such as of queueing systems for which recursive proof-techniques of monotonicity results such as developed in [18], [19], [21] or [23] can be exploited. Much of the attention of this paper therefore is devoted to applying such techniques for two concrete multi-dimensional non-product form queueing examples: an overflow model and a tandem queue with blocking.

The conditions are verifiable in two essentially different ways. Either by showing that sufficiently large states are reached only with sufficiently small probability or by using that transition probabilities or rates for upward changes tend to 0 as the state label tends to infinity. Either way is illustrated by one of the queueing examples. Explicit error bounds for these two non-product form queueing examples are obtained. These bounds guarantee a rate of convergence as the truncation limit tends to infinity and enable one to determine an à priori truncation limit for any desired accuracy.

## 2 Model and sufficient conditions

Consider a Markov chain  $\{X_t, t=0,1,2,\dots\}$  with state space  $\mathbb{N}=(1,2,\dots)$  and one-step transition probability matrix  $P=(p(i,j))$ . Without loss of generality assume that this Markov chain is irreducible.

Let  $L$  be a fixed number and define the truncated Markov chain  $\{\bar{X}_t, t=0,1,2,\dots\}$  by the one-step transition probability matrix  $\bar{P}=(\bar{p}(i,j))$  with

$$\begin{cases} \bar{p}(i,j) = 0, & j > L, i \leq L, \\ \bar{p}(i,j) = p(i,j), & j \leq t[i], j \leq L, i \leq L \\ \bar{p}(i,t[i]) = p(i,t[i]) + \sum_{j > L} p(i,j) & i \leq L \end{cases} \quad (2.1)$$

where  $t[i] \leq L$  is some given "state of truncation" for any  $i \leq L$ . Assume that this truncated chain is also irreducible. Throughout we use the upper bar "-"-symbol to indicate the truncated model while the symbol " $(-)$ " is used when expressions are to be read both with and without the upper bar. In order to compare the original and truncated chain, define expectation operators  $T_t^{(-)}$ ,  $t=0,1,2,\dots$  upon real-valued functions by:

$$\begin{cases} \overline{T}_0 = I \\ \overline{T}_{t+1} = \overline{T} \overline{T}_t, \quad t=0,1,2,\dots \\ \overline{T} f(i) = \sum_j \overline{p}_{ij} f(j) \quad (vi) \end{cases} \quad (2.2)$$

and for given real-valued function  $r$  let the functions  $\overline{V}_N$ ,  $N=0,1,2,\dots$  be given by:

$$\overline{V}_N = \sum_{t=0}^{N-1} \overline{T}_t r \quad (2.3)$$

In words that is,  $\overline{V}_N(i)$  represents the total expected reward over  $N$  periods when starting in state  $i$  at time  $t=0$  and receiving a one step reward  $r(X_t)$  at time  $t=0,1,\dots,N-1$ . Then

$$\overline{g} = \lim_{N \rightarrow \infty} \frac{1}{N} \overline{V}_N \quad (2.4)$$

is the expected average reward provided this limit exists componentwise. Based upon the irreducibility assumptions, moreover, we may assume these limits to be the same for all initial states. In the sequel, therefore, the symbol  $\overline{g}$  can be read both as a constant function and a single real value. The following theorem is the key-result of this paper. It provides a pair of conditions that guarantees an error bound for the accuracy of the truncated average reward. These conditions will be argued and illustrated later on.

**Theorem 2.1** Suppose that for some function  $\mu$ , initial state  $l$ , constants  $\epsilon_1, \epsilon_2 > 0$ , all  $t > 0$  and  $i \leq L$ :

$$\overline{T}_t \mu(l) \leq \epsilon_1 \quad (2.5)$$

$$\left| \sum_{j>l} p(i,j) [V_t(j) - V_t(i)] \right| \leq \epsilon_2 \mu(i) \quad (2.6)$$

Then

$$|\overline{V}_N(l) - V_N(l)| \leq \epsilon_1 \epsilon_2 N \quad (2.7)$$

and

$$|\overline{g} - g| \leq \epsilon_1 \epsilon_2 \quad (2.8)$$

Proof As by (2.3)

$$\langle \bar{V}_{t+1} \rangle = r + \langle \bar{T} \rangle \langle \bar{V}_t \rangle \quad (t \geq 0) \quad (2.9)$$

while  $\bar{T}$  remains restricted to  $\{1, 2, \dots, L\}$ , we can write

$$\begin{aligned} (\bar{V}_N - V_N)(\ell) &= (\bar{T} \bar{V}_{N-1} - T V_{N-1})(\ell) \\ &= (\bar{T} - T) V_{N-1}(\ell) + \bar{T} (\bar{V}_{N-1} - V_{N-1})(\ell) \\ &= \dots = \sum_{t=0}^{N-1} \bar{T}_t [(\bar{T} - T) V_{N-t-1}](\ell), \end{aligned} \quad (2.10)$$

where the last term follows by iteration and the fact that  $\langle \bar{V} \rangle(\cdot) = 0$ . From (2.1) and (2.2), however, we readily obtain for any  $i \leq L$ :

$$(\bar{T} - T) V_t(i) = \sum_j [\bar{p}(i, j) - p(i, j)] V_t(j) - \sum_{j > L} p(i, j) \{V_t(j) - V_t(t\{i\})\}. \quad (2.11)$$

By substituting expression (2.11) in (2.10), taking absolute values and noting that expectation operators are monotone operators, the proof of (2.7) now directly follows from (2.5) and (2.6). Inequality (2.8) is an immediate consequence of (2.4) and (2.7).  $\square$

Let us briefly comment on the conditions (2.5) and (2.6). First of all, the initial state  $\ell$  will usually represent the origin, but some other state might be appropriate. As for the constants  $\epsilon_1$  and  $\epsilon_2$ , one must typically think of either  $\epsilon_1$  or  $\epsilon_2$  tending to 0 as the truncation limit  $L$  tends to  $\infty$ . This is related to a twofold manner in which the conditions and the  $\mu$ -functions can be utilized as will be described below.

Remark 2.2 (Twofold verification)

(i) ( $\epsilon_1$ -small) Mos: naturally, the  $\mu$ -function can be an indicator function for sufficient large numbers. By requiring as by (2.5) that the probability of being in such states at some arbitrary fixed epoch is small, it then suffices as in (2.6) only to require that in these states the difference in the one-step transition structures of both models remains



(ii) ( $\epsilon_2$ -small) Another way of using the  $\mu$ -function is to let it represent some bounding function such as the constant unit function or some polynomial in  $i$ . Condition (2.5) then simply requires boundedness under this function over time, while condition (2.6) requires transition probabilities for upward changes to become small. This can be useful when it is hard to prove (2.5) for small  $\epsilon_1$ . This will be illustrated in section 3.2.

Most essentially in either case, however, is the fact that differences of the form  $|V_t(j) - V_t(i)|$  can usually be estimated from above by a constant independently of  $t$ , while the functions  $V_t(\cdot)$  themselves grow linearly in  $t$ . Lemma 2.3 below, adopted from [4] and related to standard results in the theory of dynamic programming (e.g. [13]), provides explicit estimates of these differences for the bounded reward case. For the unbounded reward case, a similar though slightly more complicated result can be given (cf. [5]). First, we define  $R_{i,j}$  as the expected number of transitions needed to reach state  $j$  from state  $i$ , also known as the mean-first passage time. Then the following lemma is proven in [4].

Lemma 2.3 If  $|r(i)| \leq B$  for all  $i$  and some  $B$ , then for all  $i, j$  and  $t$ :

$$|V_t(j) - V_t(i)| \leq 2B \min [R_{i,j}, R_{j,i}] \quad (2.12)$$

As in [4], the above lemma can be fruitfully exploited for one-dimensional queueing applications, for which the mean first passage time  $R_{i,j}$  can usually be estimated from above based upon random walk type results such as developed in [10]. In the present paper, however, we wish to apply theorem 2.1 to more complicated two-dimensional queueing models so that random walk estimates are not readily applicable. Most of section 3, therefore, is concerned with estimating the differences  $V_t(j) - V_t(i)$  in a direct manner.

Remark 2.4 (Other truncations). The truncation (2.1) is a natural one as it corresponds to the original model as long as the truncation limit  $L$  is not exceeded. Clearly, similar conditions can be provided for other types of truncations. For example, rather than letting a transition  $i \rightarrow j$  for all  $j > L$  transform into one and the same state  $t[i]$ , we can also let it transform into different states in a randomized manner. Then, theorem 2.1 remains valid provided condition (2.6) is modified correspondingly.

Remark 2.5 ((Un)bounded rewards). Note that no conditions are imposed upon the one-step reward function  $r(\cdot)$  other than that we implicitly assume the average rewards  $g$  and  $\bar{g}$  to be well-defined. The example of section 3.1, for instance, covers a linear one-step reward function (see remark 3.1.4) so as to compute a mean queue length of an infinite system. As a particular application of the bounded reward case,  $g$  represents the steady state probability of a set  $G$  if we choose:

$$r(i) = \begin{cases} 1 & \text{for } i \in G \\ 0 & \text{otherwise} \end{cases}.$$

Remark 2.6 (State labeling). For expository convenience the states were labeled in a countable manner. Clearly, for more-dimensional applications such as in section 3 we thus have to label the states in an appropriate manner such that under the given truncation no states with label exceeding  $L$  can be reached (see the proofs of theorems 3.1.3 and 3.2.2). Particularly, in such applications the need for different "states of truncation"  $t[i]$  rather than one fixed state naturally comes up.

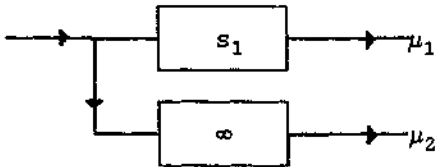
### 3 Applications

This section contains two applications which will illustrate how the conditions of theorem 2.1 can be verified. Both applications concern a two-dimensional queueing model which does not exhibit an explicit product form expression for the steady state distribution. Numerical computation is thus required for evaluating a performance measure such as a steady-state probability, mean queue length or throughput. As the number of states is infinite numerical truncation will then be necessary.

As in the queueing examples the jump rates are bounded, the standard uniformization technique can be applied (cf. [20], p.110) in order to obtain a discrete-time formulation. The crucial step is the estimation of differences of the form  $|V_n(j) - V_n(i)|$ , where it is to be realized that the states are of a two-dimensional form (see lemma 3.1.1 and lemma 3.2.1). To this end, a recursive proof-technique will be applied as based upon Markov reward theory. Throughout, we let  $1_{\{A\}}$  denote an indicator of an event  $A$ , i.e.  $1_{\{A\}} = 1$  if  $A$  is satisfied and 0 otherwise, and  $1_{\{A\}}(\cdot)$  the corresponding indicator function of an event  $A$ .

### 3.1 Overflow model

Consider a service system which consists of an Erlang loss station (station 1) with  $s_1$  servers and an additional Erlang-delay overflow station (station 2) with  $s_2$  servers. More precisely, a customer which arrives is assigned a free server at station 1 when available and otherwise routed to the overflow station 2, which behaves as a standard delay station with  $s_2$  servers and a first-come first-served infinite waiting room. Further, customer switching from station 2 to station 1 is prohibited. The services are assumed to be exponential with parameter  $\mu_i$  at station  $i$ . The customer arrival process is Poisson with parameter  $\lambda$ . It is assumed that  $\lambda < \mu_2$ .



The system under consideration has no product-form solution for the steady-state joint queue size distribution (cf. [8]). The overflow stream is known to be hyperexponential (cf. [3]), so that the overflow station can be analyzed as a  $GI|M|s_2$ -system. This, however, would still require complex computational procedures for large  $s_2$ -values (cf. [2], p. 270-275). Moreover, here we are interested in a performance measure that can depend on both queue sizes, such as the total number of jobs present, where it is noted that  $\mu_1 \neq \mu_2$  is allowed. For expository convenience from now on we assume  $s_1 = s_2 = 1$ .

Let  $[i, j]$  denote the number of customers  $i$  at station 1 and  $j$  at station 2, and consider the embedded chain by inspecting the process at exponential times with parameter  $M = (\lambda + \mu_1 + \mu_2)$ . Then by virtue of the uniformization technique (cf. [9], [18], p.110), the one-step transition probabilities of this chain  $p([i, j] \rightarrow [m, n])$  for transitions from a state  $[i, j]$  into a state  $[m, n]$  become:

$$\left\{ \begin{array}{l} p([0,j] \rightarrow [1,j]) = \lambda/M \\ p([1,j] \rightarrow [1,j+1]) = \lambda/M \\ p([i,j] \rightarrow [i-1,j]) = \mu_1/M \\ p([i,j] \rightarrow [i,j-1]) = \mu_2/M \end{array} \right. \quad (3.1.1)$$

where  $i-1$  and  $j-1$  are to be read as 0 for  $i=0$  respectively  $j=0$ . Consider an arbitrary reward rate  $\tilde{r}([i,j])$  for the original queueing process whenever the system is in state  $[i,j]$ . Further assume that for some constant  $H$  and all  $j$  and  $(i,j)$ :

$$|\tilde{r}(1,j) - \tilde{r}(0,j)| \leq H \quad (3.1.2)$$

$$|\tilde{r}(i,j+1) - \tilde{r}(i,j)| \leq H.$$

Let  $r = \tilde{r}/M$  and consider the expected average rewards  $g$  and  $\tilde{g}$  of the embedded process with one step reward  $r$  and the original queueing process with reward rate  $\tilde{r}$ . Then by the uniformization technique we also conclude

$$\tilde{g} = g M. \quad (3.1.3)$$

We may thus restrict our attention to  $g$  for evaluating the performance measure  $\tilde{g}$ . To this end, for some constant  $Q$  the following truncation is proposed:

$$\left\{ \begin{array}{l} \bar{p}([1,Q] \rightarrow [1,Q]) = 1 - [\mu_1 + \mu_2]/M \\ \bar{p}([i,j] \rightarrow [m,n]) = p([i,j] \rightarrow [m,n]) \text{ otherwise.} \end{array} \right. \quad (3.1.4)$$

In words that is, we truncate the queue size of station 2 at level  $Q$  by rejecting arrivals whenever  $j=Q$ . The state space is thus restricted to  $S_Q = \{(i,j) | i=0,1; 0 \leq j \leq Q\}$ . In order to apply theorem 2.1 the following lemma is essential.

Lemma 3.1 With  $C = \max[H/\mu_1, H/(\mu_2 - \lambda)]$ , we have for all  $n$  and  $j$ :

$$|V_n(0, j) - V_n(1, j)| \leq (j+1)C \quad (3.1.5)$$

$$|V_n(1, j) - V_n(1, j+1)| \leq (j+1)C. \quad (3.1.6)$$

**Proof** We apply induction to  $n$ . For  $n=0$  it trivially holds as  $V_0(\cdot)=0$ . Suppose that (3.1.5) and (3.1.6) hold for  $n \leq m$ . Then by virtue of (2.11),

$$\begin{aligned} V_{m+1}(0, j) - V_{m+1}(1, j) &= r(0, j) - r(1, j) \\ &+ [\lambda/M][V_m(1, j) - V_m(1, j+1)] \\ &+ [\mu_1/M][V_m(0, j) - V_m(0, j)] \\ &+ [\mu_2/M][V_m(0, j-1) - V_m(1, j-1)] 1_{\{j>0\}} \\ &+ [\mu_2/M][V_m(0, 0) - V_m(1, 0)] 1_{\{j=0\}}. \end{aligned}$$

Taking absolute values per term and substituting (3.1.5) and (3.1.6), we find:

$$|V_{m+1}(0, j) - V_{m+1}(1, j)| \leq H/M + [(\lambda + \mu_1 + \mu_2)/M](j+1)C - [\mu_1/M]C.$$

Recalling  $C \geq H/\mu_2$  and  $M = \lambda + \mu_1 + \mu_2$ , we have thus proven (3.1.5) for  $n=m+1$ . Similarly,

$$\begin{aligned} V_{m+1}(i, j) - V_{m+1}(i, j+1) &= r(i, j) - r(i, j+1) \\ &+ [\lambda/M][V_m(1, j) - V_m(1, j+1)] 1_{\{i=0\}} \\ &+ [\lambda/M][V_m(1, j+1) - V_m(1, j+2)] 1_{\{i=1\}} \\ &+ [\mu_1/M][V_m(0, j) - V_m(0, j+1)] \\ &+ [\mu_2/M][V_m(i, j-1) - V_m(i, j)] 1_{\{j>0\}} \\ &+ [\mu_2/M][V_m(i, 0) - V_m(i, 0)] 1_{\{j=0\}}. \end{aligned}$$

Taking absolute values per term and substituting (3.1.6), we now obtain

$$|V_{m+1}(i,j) - V_{m+1}(i,j+1)| \leq H/M + [\lambda/M]C + [(\lambda + \mu_1 + \mu_2)/M](j+1)C - [\mu_2/M]C.$$

With  $C \geq H/(\mu_2 - \lambda)$  and  $M = \lambda + \mu_1 + \mu_2$ , (3.1.6) is hereby proven for  $n = m+1$ .

□

Lemma 3.1.2 For all  $t \geq 0$ , we have

$$\bar{T}_t 1_{\{(1,Q)\}}(0,0) \leq (\lambda/\mu_2)^Q / [1 - \lambda/\mu_2]. \quad (3.1.7)$$

**Proof** Clearly, by directly routing all customers to the second station, the steady state-probability of  $Q$  customers at the second station is estimated from above by the steady-state probability that there are at least  $Q$  customers in an  $M|M|1|_\infty$ -system with arrival rate  $\lambda$  and service rate  $\mu_2$ . This probability is equal to the right hand side of (3.1.7). It thus suffices to prove

$$[\bar{T}_{t+1} - \bar{T}_t]f(0,0) \geq 0 \quad (3.1.8)$$

for all  $t$  and with  $f(\dots) = 1_{\{(1,Q)\}}(\dots)$ . We will inductively prove (3.1.8) for any function  $f(\dots)$  which is non-decreasing in each argument, i.e., satisfying

$$\begin{cases} f(1,j) - f(0,j) \geq 0 & (j \leq Q) \\ f(i,j+1) - f(i,j) \geq 0 & (i = 0,1) (j < Q). \end{cases} \quad (3.1.9)$$

For  $t=0$ , (3.1.8) is readily verified by

$$\bar{T}_1 f(0,0) - f(0,0) = [\lambda/M][f(1,0) - f(0,0)].$$

Assume that (3.1.8) holds for  $t \leq n$ . Then for  $t = n+1$

$$[\bar{T}_{n+2} - \bar{T}_{n+1}]f(0,0) = [\bar{T}_{n+1} - \bar{T}_n](\bar{T}f)(0,0),$$

so that by induction hypothesis (3.1.8) for  $t=n$ , (3.1.8) is proven provided (3.1.9) is satisfied if we substitute  $f = \bar{T}f$ . This in turn is guaranteed by:

$$\begin{aligned}
 (\bar{T}f)(1,j) - (\bar{T}f)(0,j) &= [\lambda/M][f(1,j+1) - f(1,j)]1_{\{j < Q\}} \\
 &+ [\mu_1/M][f(0,j) - f(0,j)] \\
 &+ [\mu_2/M][f(1,j-1) - f(0,j-1)]1_{\{j > 0\}} \geq 0,
 \end{aligned}$$

and

$$\begin{aligned}
 (\bar{T}f)(i,j+1) - (\bar{T}f)(i,j) &= [\lambda/M][f(1,j+1) - f(1,j)]1_{\{i=0\}} \\
 &+ [\lambda/M][f(1,j+2) - f(1,j+1)]1_{\{i=1\}}1_{\{j+1 < Q\}} \\
 &+ [\mu_1/M][f(0,j+1) - f(0,j)] \\
 &+ [\mu_2/M][f(i,j) - f(i,j-1)]1_{\{j > 0\}} \geq 0.
 \end{aligned}$$

Finally, by noting that the indicator function  $1_{\{(1,Q)\}}(\dots)$  satisfies (3.1.9), the proof is concluded.  $\square$

**Theorem 3.1.3** With the truncation (3.1.4) for truncation limit  $Q$ ,  $\bar{g}$  the corresponding average reward,  $H$  given by (3.1.2) and  $M = \lambda + \mu_1 + \mu_2$ , we have:

$$|\bar{g} - g| \leq [Q+1](\lambda/\mu_2)^Q [\lambda\mu_2 H/M] \max[\mu_1^{-1}, (\mu_2 - \lambda)^{-1}] (\mu_2 - \lambda)^{-1} \quad (3.1.10)$$

**Proof** Label state  $[m,n]$  by  $k=n$  voor  $m=0$  and  $k=(Q+2)+n$  for  $m=1$  for all states with  $m=0,1$  and  $n \leq Q$  and let  $p(i,j)$  and  $V_t(i)$  be defined correspondingly. Now consider (2.6) with  $i$  and  $j$  representing labels of different states of the form  $[m,n]$  and  $L=2(Q+1)$ . Then by virtue of (3.1.1), (3.1.4) and (3.1.6), for any state  $[m,n]$  condition (2.6) transfers in:

$$\begin{aligned}
 \sum_{\{(1,s) | s > Q\}} p([m,n] \rightarrow [1,s]) [V_t(1,s) - V_t(1,Q)] = \\
 1_{\{(1,Q)\}}(m,n) [\lambda/M] [V_t(1,Q+1) - V_t(1,Q)] \leq \\
 1_{\{(1,Q)\}}(m,n) [\lambda/M] [Q+1] \max[H/\mu_1, H/(\mu_2 - \lambda)].
 \end{aligned}$$

Taking  $\mu(m,n) = 1_{\{(1,Q)\}}(m,n)$  and noting that condition (2.5) with reference state  $l = (0,0)$  is then given by (3.1.7), application of theorem 2.1 completes the proof.  $\square$

Remark 3.1.4 Note that (3.1.10) holds for any reward rate  $\tilde{r}$  satisfying (3.1.2). For example, a stationary tail probability is computed by:

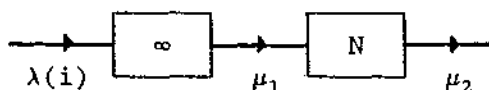
$$\tilde{r}(i,j) = 1_{\{j>s\}}(i,j) \quad (3.1.11)$$

and the mean system load by:

$$\tilde{r}(i,j) = i + j \quad (3.1.12)$$

### 3.2 Tandem queue

Consider a tandem queue of an  $M|M|1|\infty$  and  $M|M|1|N$  queue. That is, at the first station an infinite number of customers is allowed but at the second no more than  $N$ . When the second station is saturated the servicing at the first station is stopped (communication protocol). Both stations have a single server and queueing is assumed to be first-come first-served. The service requirements are exponential with parameters  $\mu_1$  and  $\mu_2$  at stations 1 and 2 respectively. The interarrival times are also exponential but with a state dependent parameter  $\lambda(i)$  when  $i$  customers are present at station 1.



The steady-state joint queue size distribution of this tandem queue system does not exhibit a closed product form expression (cf. [8]). Numerical studies and approximation procedures have therefore been widely investigated (e.g., [1], [6], [11]). Error bounds, however, have therein not been included. Now assume that for some constant  $\lambda$ :

$$\begin{aligned} \lambda(i) &\leq \lambda \\ \lambda(i) &\rightarrow 0 \quad \text{as } i \rightarrow \infty \\ \lambda(i) &: \text{non-increasing in } i, \end{aligned}$$

Let  $\{i,j\}$  denote the number of customers  $i$  at station 1 and  $j$  at station 2 and consider the embedded chain by inspecting the process at exponential times with parameter  $M = (\lambda + \mu_1 + \mu_2)$ . The uniformization technique (cf. [20], p.110) then leads to the one-step transition probabilities:



$$\left\{ \begin{array}{l} p([i,j] \rightarrow [i+1,j]) = [\lambda(i)/M] \\ p([i,j] \rightarrow [i,j-1]) = [\mu_2/M] 1_{\{j>0\}} \\ p([i,j] \rightarrow [i-1,j+1]) = [\mu_1/M] 1_{\{i>0\}} 1_{\{j<N\}} \\ p([i,j] \rightarrow [i,j]) = 1 - [\lambda(i) + \mu_1 1_{\{i>0\}} 1_{\{j<N\}} + \mu_2 1_{\{j>0\}}] / M. \end{array} \right. \quad (3.2.1)$$

We aim to evaluate the throughput  $\tilde{g}$ . To this end, let

$$r(i,j) = 1_{\{j>0\}}(i,j) \quad (3.2.2)$$

Then with  $g$  the corresponding average reward of this embedded Markov chain, the throughput of the original system is given by

$$\tilde{g} = g \mu_2 \quad (3.2.3)$$

To compute this relevant performance measure we can thus restrict our attention to evaluating  $g$ . To this end, for some constant  $Q$  the following truncation is proposed:

$$\left\{ \begin{array}{l} \bar{p}([Q,j] \rightarrow [Q,j]) = 1 - [\mu_1 1_{\{i>0\}} 1_{\{j<N\}} + \mu_2 1_{\{j>0\}}] / M \\ \bar{p}([i,j] \rightarrow [m,n]) = p([i,j] \rightarrow [m,n]) \text{ otherwise.} \end{array} \right. \quad (3.2.4)$$

In words that is, the queue size at station 1 is truncated at level  $Q$  by rejecting arrivals whenever  $i=Q$ . The state space is thus restricted to  $S_Q = \{(i,j) | 0 \leq i \leq Q, 0 \leq j \leq N\}$ . In order to apply theorem 2.1, the following lemma is crucial.

**Lemma 3.2.1** For all  $n$  and restricted to  $S_Q$ :

$$0 \leq \Delta_1 V_n(i,j) = V_n(i+1,j) - V_n(i,j) \leq [M/\mu_1] \quad (3.2.5)$$

$$0 \leq \Delta_2 V_n(i,j) = V_n(i,j+1) - V_n(i,j) \leq [M/\mu_1] \quad (3.2.6)$$

$$0 \leq \Delta_3 V_n(i,j) = V_n(i-1,j+1) - V_n(i,j) \leq [M/\mu_1] \quad (3.2.7)$$

**Proof** We will apply induction to  $n$ . For  $n=0$ , these estimates trivially follow from  $V_0(\dots)=0$ . Suppose that (3.2.5), (3.2.6) and (3.2.7) hold for  $n=m$ . Then we will separately verify (3.2.5), (3.2.6) and (3.2.7) for  $n=m+1$  under (i), (ii) and (iii) below. Though the technique in each of these is the same it turns out that the technicalities involved are different and therefore are to be studied in full detail.

(i) (3.2.5) for  $n=m+1$  From (2.11), (3.2.1) and (3.2.2) we find:

$$\begin{aligned}
 \Delta_1 V_{m+1}(i,j) = & \\
 & ([\lambda(i+1)/M]V_m(i+2,j) + \\
 & + [\mu_1/M] 1_{\{j < N\}} V_m(i,j+1) + [\mu_2/M] 1_{\{j > 0\}} V_m(i+1,j-1) \\
 & + [1 - [\lambda(i+1) + \mu_1 1_{\{j < N\}} + \mu_2 1_{\{j > 0\}}]/M]V_m(i,j)) - \\
 & ([\lambda(i)/M]V_m(i+1,j) + \\
 & + [\mu_1/M] 1_{\{i > 0\}} 1_{\{j < N\}} V_m(i-1,j+1) + [\mu_2/M] 1_{\{j > 0\}} V_m(i,j-1) \\
 & + [1 - [\lambda(i) + \mu_1 1_{\{i > 0\}} 1_{\{j < N\}} + \mu_2 1_{\{j > 0\}}]/M]V_m(i,j)) = \\
 & [\lambda(i+1)/M]\Delta_1 V_m(i+1,j) + \{[\lambda(i) - \lambda(i+1)]/M\}[V_m(i+1,j) - V_m(i+1,j)] \\
 & + [\mu_1/M] 1_{\{j < N\}} \{1_{\{i > 0\}} \Delta_1 V_m(i-1,j+1) + 1_{\{i=0\}} [V_m(0,j+1) - V_m(0,j)]\} \\
 & + [\mu_2/M] 1_{\{j > 0\}} \Delta_1 V_m(i,j-1) + [1 - (\lambda(i) + \mu_1 1_{\{j < N\}} + \mu_2 1_{\{j > 0\}})]/M \Delta_1 V_m(i,j)
 \end{aligned}
 \tag{3.2.8}$$

By noting that the second term is equal to 0 and substituting the induction hypotheses (3.2.5) and (3.2.6) for  $n=m$ , we have thus verified (3.2.5) also for  $n=m+1$ .

(ii) (3.2.6) for  $n=m+1$  Similarly to (3.2.8) and noting that  $r(i, j+1) - r(i, j) = 1_{\{j=0\}}$ , we now obtain:

$$\begin{aligned}
 \Delta_2 V_{m+1}(i, j) &= 1_{\{j=0\}} \\
 &+ [\lambda(i)/M] \Delta_2 V_m(i+1, j) \\
 &+ [\mu_1/M] 1_{\{i>0\}} 1_{\{j+1 \leq N\}} \Delta_2 V_m(i-1, j+1) \\
 &+ [\mu_1/M] 1_{\{i>0\}} 1_{\{j+1=N\}} [V_m(i, j+1) - V_m(i-1, j+1)] \\
 &+ [\mu_2/M] 1_{\{j>0\}} \Delta_2 V_m(i, j-1) \\
 &+ [\mu_2/M] 1_{\{j=0\}} [V_m(i, 0) - V_m(i, 0)] + \\
 &+ [1 - [\lambda(i) + \mu_1 1_{\{i>0\}} + \mu_2]/M] \Delta_2 V_m(i, j) \tag{3.2.9}
 \end{aligned}$$

By substituting the induction hypothesis  $\Delta_1 V_m \geq 0$  in the fourth term and  $\Delta_2 V_m \geq 0$  in the other terms, we directly verify  $\Delta_2 V_{m+1} \geq 0$ . By rewriting the first term as  $[\mu_2/M] 1_{\{j=0\}} [M/\mu_2]$ , noting that the one but last term is equal to 0 and substituting the hypotheses  $\Delta_1 V_m \leq [M/\mu_2]$  and  $\Delta_2 V_m \leq [M/\mu_2]$  in the other terms, we also conclude  $\Delta_2 V_{m+1} \leq [M/\mu_2]$ , which proves (3.2.6) for  $n=m+1$ .

(iii) (3.2.7) for  $n=m+1$  Again, similarly to (3.2.8) and noting that  $r(i-1, j+1) - r(i, j) = 1_{\{j=0\}}$ , we now find:

$$\begin{aligned}
 \Delta_3 V_{m+1}(i, j) &= 1_{\{j=0\}} \\
 &+ [\lambda(i)/M] \Delta_3 V_m(i+1, j) \\
 &+ [(\lambda(i-1) - \lambda(i))/M] [V_m(i, j+1) - V_m(i, j)] \\
 &+ [\mu_1/M] 1_{\{i-1>0\}} 1_{\{j+1<N\}} \Delta_3 V_m(i-1, j+1) \\
 &+ [\mu_1/M] 1_{\{i-1=0\}} 1_{\{j+1<N\}} [V_m(0, j+1) - V_m(0, j+1)] \\
 &+ [\mu_1/M] 1_{\{j+1=N\}} [V_m(i-1, j+1) - V_m(i-1, j+1)] \\
 &+ [\mu_2/M] 1_{\{j>0\}} \Delta_3 V_m(i, j-1) \\
 &+ [\mu_2/M] 1_{\{j=0\}} [V_m(i-1, j) - V_m(i, j)] \\
 &+ [1 - \{\lambda(i-1) + \mu_1 + \mu_2\}/M] \Delta_3 V_m(i, j) \tag{3.2.10}
 \end{aligned}$$

Now note that the one but last term is non-positive by induction hypothesis. However, also by hypothesis we can estimate this term from below by  $-[\mu_2/M] 1_{\{j=0\}} [M/\mu_2] = -1_{\{j=0\}}$ , so that the first and this one but last term together are estimated from below by 0. Substituting the hypotheses  $\Delta_2 V_m \geq 0$  and  $\Delta_3 V_m \geq 0$  in the remaining terms, we have thus shown  $\Delta_3 V_{m+1} \geq 0$ . Conversely, by deleting this nonpositive one but last term, rewriting the first term as  $[\mu_2/M] 1_{\{j=0\}} [M/\mu_2]$  and substituting the hypotheses  $\Delta_2 V_m \leq [M/\mu_2]$  and  $\Delta_3 V_m \leq [M/\mu_2]$  we conclude  $\Delta_3 V_{m+1} \leq [M/\mu_2]$ . Also (3.2.7) for  $n=m+1$  is hereby proven, which completes the proof of the lemma.

□

**Theorem 3.2.2** With the truncation (3.2.4) for truncation limit  $Q$ ,  $\bar{g}$  the corresponding average reward and  $M = \lambda + \mu_1 + \mu_2$ , we have

$$|\bar{g} - g| \leq \lambda(Q)/\mu_2 \quad (3.2.11)$$

**Proof** Label state  $[m,n]$  by  $k = (1+m)(1+n)$  for all  $m=0,1,\dots,N$  and  $n=0,1,\dots,Q$  and set  $L = (N+1)(Q+1)$ . Further, let  $p(i,j)$  and  $V_t(i)$  be defined correspondingly. Now consider condition (2.6) with  $i$  and  $j$  representing labels of different states of the form  $(m,n)$  and  $L = (N+1)(Q+1)$ . Then by virtue of (3.2.1), (3.2.4) and (3.2.5), for any  $[m,n]$  condition (2.6) transfers in:

$$\begin{aligned} \sum_{\{[s,v] | s > Q\}} p([m,n] \rightarrow [s,v]) [V_t(s,v) - V_t(Q,n)] = \\ 1_{\{m=Q\}}(m,n) [\lambda(Q)/M] [V_t(Q+1,n) - V_t(Q,n)] \leq \\ 1_{\{m=Q\}}(m,n) [\lambda(Q)/M] [M/\mu_2]. \end{aligned}$$

Choosing  $\mu(m,n)=1$  for all  $(m,n)$ , so that condition (2.5) is trivially satisfied with  $B=1$  for all  $\ell$ , application of theorem 2.1 completes the proof.

**Corollary 3.2.3** The truncation (3.2.4) secures an exact order of accuracy  $\lambda(Q)$  for the computations of the system throughput  $\bar{g} = g\mu_2$ .

**Evaluation** Truncation of state spaces is investigated so as to enable a numerical computation of infinite Markov chains. In order to obtain an error bound of the accuracy a pair of conditions is provided that appears applicable in two characteristic situations: (i) when the probability of being in sufficiently large states can be shown to be small, or (ii) when the probability for state increases tends to 0. Essential to either case, moreover, is the estimation from above of bias terms of total reward functions. To this end, inductive proof-techniques for monotonicity results appear to be useful. The results typically apply to spatial birth-death type processes such as queueing networks.

## References

- [1] BOXMA, O.J. and A.G. KONHEIM (1981), Approximate Analysis of Exponential Queueing Systems with Blocking, *Acta Inform.* 15, 19-66.
- [2] COOPER, R.B. (1984), *Introduction to Queueing Theory*. North-Holland (1984).
- [3] VAN DOORN, E.A. (1984), On the overflow process from a finite Markovian queue. *Performance Eval.* 4, 233-240.
- [4] VAN DIJK, N.M. and PUTERMAN, M.L. (1988), Perturbation theory for Markov reward processes with applications to queueing systems. *Adv. Appl. Prob.* 20, 79-98.
- [5] VAN DIJK, N.M. (1988), Perturbation theory for unbounded Markov reward processes with applications to queueing. *Adv. Appl. Prob.* 20, 99-111.
- [6] HILLIER, F.S. and R.W. BOLING (1967), Finite Queues in Series with Exponential or Erlang Service Times - A Numerical Approach. *Oper.Res.* 15, 286-303.
- [7] HINDERER, K. (1978), On approximate solutions of finite-stage dynamic programs. *Dynamic Programming and its Applications*, ed. M.L. Puterman, Academic Press, New York.
- [8] HORDIJK, A. and N.M. van DIJK (1981), Networks of Queues with Blocking. In *Performance '81*, pp. 51-65. North-Holland, Amsterdam.
- [9] JENSEN, A. (1953), Markov Chains as an Aid in the Study of Markov Processes. *Skand. Aktuartidskr.* 36, 87-91.
- [10] KEMENY, J.G., SNELL, J.L. and KNAPP, A.W. (1966), *Denumerable Markov Chains*. Van Nostrand, Princeton, NJ.
- [11] LATOUCHE, G. and M.F. NEUTS (1980), Efficient Algorithmic Solutions to Exponential Tandem Queues with Blocking. *SIAM J. Alg. Disc. Meth.* 1, 93-105.
- [12] MEYER, C.D. Jr. (1980), The condition of a finite Markov chain and perturbation bounds for the limiting probabilities. *SIAM J. Alg. Disc. Math.* 1, 273-283.
- [13] ROSS, S.M. (1970), *Applied Probability Models with Optimization Applications*. Holden Day, San Francisco.
- [14] SCHWEITZER, P.J. (1968), Perturbation theory and finite Markov chains. *J. Appl. Prob.* 5, 401-413.

- [15] SENETA, E. (1967), Finite approximations to infinite non-negative matrices. *Proc. Cambridge, Phil. Soc.* 63, 983-992.
- [16] SENETA, E. (1968), The principles of truncations in applied probability. *Comm. Math. Univ. Carolina* 9, 533-539.
- [17] SENETA, E. (1980), *Non-negative matrices and Markov chains*, Springer-Verlag.
- [18] SHANTHIKUMAR, J.G. and YAO, .D. (1987), Stochastic monotonicity of the queue lengths in closed queueing networks, *Oper. Res.* 35, 583-588.
- [19] STOYAN, D. (1983), *Comparison methods for queues and other stochastic models*, Wiley, New York.
- [20] TIJMS, H.C. (1986), *Stochastic Modelling and Analysis. A computational Approach*. Wiley, New York.
- [21] VAN DIJK, N.M. and VAN DER WAL, J. (1987), Simple bounds and monotonicity results for multi-server exponential tandem queues, To appear: *Queueing Systems*.
- [22] WHITT, W. (1978), Approximations of dynamic programs I. *Math. Oper. Res.* 3, 231-243.
- [23] WHITT, W. (1981), Comparing counting processes and queues, *Adv. Appl. Prob.* 13, 207-220.

|         |   |  |
|---------|---|--|
| 1988-1  | H. Visser   | Austrian thinking on international economics   |
| 1988-2  | A.H.Q.M. Merkies<br>T. van der Meer               | Theoretical foundations for the 3-C model  |
| 1988-3  | H.J. Bierens<br>J. Hartog                         | Nonlinear regression with discrete explanatory variables, with an application to the earnings function |
| 1988-4  | N.M.van Dijk                                      | On Jackson's product form with 'jump-over' blocking  |
| 1988-5  | N.M.van Dijk<br>M.Rumsewicz                       | Networks of queues with service anticipating routing   |
| 1988-6  | H.Linneman<br>C.P.van Beers                       | Commodity Composition of Trade in Manufactures and South-South Trade Potential                         |
| 1988-7  | N.M.van Dijk                                      | A LCFS finite buffer model with batch input and non-exponential services                               |
| 1988-8  | J.C.W.van Ommeren                                 | Simple approximations for the batch-arrival $M^X/G/1$ queue  |
| 1988-9  | H.C.Tijms   | Algorithms and approximations for batch-arrival queues   |
| 1988-10 | J.P.de Groot<br>H.Clemens                         | Export Agriculture and Labour Market in Nicaragua  |
| 1988-11 | H.Verbruggen<br>J.Wuijts                          | Patterns of South-South trade in manufactures  |
| 1988-12 | H.C.Tijms<br>J.C.W.van Ommeren                    | Asymptotic analysis for buffer behaviour in communication systems                                      |
| 1988-13 | N.M.van Dijk<br>E.Smeitink                        | A non-exponential queueing system with batch servicing   |
| 1988-14 | J.Rouwendal                                       | Existence and uniqueness of stochastic price equilibria in heterogeneous markets                       |
| 1988-15 | H.Verbruggen                                      | GSTP, the structure of protection and South-South trade in manufactures                                |
| 1988-16 | Mevr.H.Weijland<br>Mevr.R Herweijer<br>J.de Groot | Female participation in agriculture in the Dominican Republic  |
| 1988-17 | N.M.van Dijk                                      | Product Forms for Random Access Schemes  |
| 1988-18 | A.H.Q.M.Merkies<br>I.J.Steyn                      | Adaptive Forecasting with Hyperfilters   |
| 1988-19 | J.Rouwendal                                       | Specification and Estimation of a Logit Model for Housing Choice in the Netherlands                    |
| 1988-20 | J.C.W.van Ommeren<br>R.D. Nobel                   | An elementary proof of a basic result for the GI/G/1 queue   |