# Asymptotics of the First Hitting Times of Markov Jump Processes with Applications to ATM

by

Kun Qian

A thesis submitted to
the School of Graduate Studies and Research
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Mathematics*

University of Ottawa
Ottawa, Ontario

*The Ph.D. Program is a joint program with
Carleton University, administered by
the Ottawa-Carleton Institute of Mathematics and Statistics

ISBN 0-315-83836-1

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

*To my parents*

# Abstract

This dissertation has three parts. The first part (Chapter 2) is about the asymptotics of the distribution of the first hitting time of a forbidden set by a Markov jump process. Explicit error bounds for the departure of the hitting time distribution from exponentiality are provided. The second part (Chapter 3 and Chapter 4, joint with Ian Iscoe and David McDonald) discusses the capacity of an ATM multiplexor in terms of the probability distribution of the time until the first occurrence of an excessive demand for bandwidth. In the third part (Chapter 5), the problem of the buffer overflow of an ATM multiplexor is studied. The methods developed give an excellent approximation for the steady-state probabilities of the contents of a buffer driven by heterogeneous sources.

# Acknowledgements

I am grateful to my supervisor Dr. David McDonald for his valuable guidance and advice which made this thesis possible. His effort and patience in helping me to prepare the thesis is highly appreciated. I also thank Dr. Ian Iscoe with whom I have being closely working throughout these years.

I wish to thank the University of Ottawa and the Mathematics Department for providing me with financial support.

My special thanks go to my wife and my parents for their understanding, encouragement and support.

# Contents

# Chapter 1

# Introduction

## 1.1 The First Hitting Time Problem

### 1.1.1 The Exponential Property

Many problems which arise in several areas of theoretical and applied probability may be described by the first hitting time of a rarely-visited set of states by a Markov jump process. In modern telecommunication networks, for example, hitting some bad states may correspond to buffers being filled beyond capacity, in which case messages are truncated or lost. These large deviations can be engineered to be highly improbable but not impossible. The distribution of, and the mean time between such large deviations are important design features. Exact solutions for these problems are very hard to obtain in general, but a number of approximation methods are quite useful. It has been known for quite a while that under very general circumstances the first hitting time of a continuous time Markov chain is approximately exponentially distributed (Keilson (1966)). There are at least two intuitive explanations of this fact. One of them led to the so called *regeneration method* (Keilson (1979), Aldous (1982)). Suppose $B^c$ (the complement of $B$) is a

set of some rarely-visited states while $i_0$ is frequently visited by a Markov chain $X_t$. This suggests that the probability $p$ of starting from $i_0$ and hitting $B^c$ before first returning to $i_0$ is small. So, for a chain starting at $i_0$, the path until $\tau_B$, the first hitting time on $B^c$, consists of a geometric (mean $p^{-1}-1$) number of excursions from $i_0$ which return to $i_0$ without hitting $B^c$, followed by part of an excursion which hits $B^c$. Apart from this final excursion, $\tau_B$ is a large geometric sum of *i.i.d.* variables, and this sum should be almost exponential. The other explanation (Aldous (1982)), led to the *mixing technique* which supposes a rapid mixing property of the chain in the sense that it converges to stationarity rapidly. Then the distribution of $X_t$ given $\tau_B > t$ will stay near the stationary distribution $\pi$, because the tendency to drift away (due to paths hitting $B^c$ being eliminated) is offset by the rapid mixing. Therefore $P(\tau_B > t+s|\tau_B > t)$ will be approximately $P_\pi(\tau_B > s)$, and this makes $\tau_B$ almost exponential.

There are at least four different ways to study exponential limit theorems for hitting times on rare sets (see Aldous (1989)). They are *the small parameter method, the eigenvalue method* and the two mentioned in the last paragraph. Our special interest is in the eigenvalue method and the mixing technique. We give a detailed discussion on these methods in the following and in the next section.

**The Eigenvalue Method:** Consider a discrete time Markov chain with state space $J$ and transition matrix $P$. Let $B$ be a finite subset of $J$ and let $\hat{P}$ be $P$ restricted to $B$ and define $\tau_B = min\{n : X_n \in B^c\}$. Then

$$P(X_n = j, \tau_B > n | X_0 = i) = \hat{P}^n(i,j) \sim \theta^n \beta_i \alpha_j \tag{1.1.1}$$

as $n \to \infty$ according to the following (see Aldous (1989)).

**Theorem 1.1.1** *Let $\hat{P}$ be a finite substochastic matrix which is irreducible and aperiodic. Then $\hat{P}^n(i,j) \sim \theta^n \beta_i \alpha_j$ as $n \to \infty$, where:*

- *$\theta$ is the Perron-Frobenius eigenvalue of $\hat{P}$, i.e. $|\theta|$ is the largest.*
- *$\theta$ is real: $\theta = 1$ if $\hat{P}$ is stochastic, $0 < \theta < 1$ otherwise.*
- *$\alpha$ and $\beta$ are the corresponding eigenvectors $\alpha\hat{P} = \theta\alpha$, $\hat{P}\beta = \theta\beta$, normalized so that $\sum_i \alpha_i = 1$ and $\sum_i \alpha_i\beta_i = 1$.*

Here *substochastic* means $\hat{P}(i,j) \geq 0, \sum_j \hat{P}(i,j) \leq 1$, and *irreducible, aperiodic* are defined as in Seneta(1981).

Notice that (1.1.1) presents the *exponential tail property* for the first hitting time $\tau_B$, where only large values of $n$ (or $t$ in continuous time) are concerned. For the lossy Markov chains in continuous-time, Keilson (1979) gives the exponential tail property for the transition probability matrix on finite set $B$ (see Keilson (1979) Theorem 6.6.2. A). This property is different from the *approximately exponential distribution property* which provides approximate exponentiality for all values of $n$ (or $t$). However, it is clear that the eigenvalue method is also useful for proving exponentiality—the question which concerns us most. In fact, it helps in identifying the exponential parameter, as we can see from the recent work by Aldous and Brown (1991) and Iscoe and McDonald (1992).

### 1.1.2 A Review of Related Recent Work

From a practical point of view, it is particularly interesting to provide calculable estimates, e.g. lower or upper bounds, for the distribution and the mean of the first hitting time. It is also essential to quantify departure from exponentiality via error bounds (Keilson (1979)). These are the central topics of Chapter 2. Some recent results are reviewed in the following.

Let $(X_t; t \geq 0)$ be an irreducible finite-state *reversible* Markov chain in continuous time. The state space is $S$ and the transition rate matrix (or the infinitesimal

generator) is $-L = (q(i,j): i,j \in S)$ where $q(i,i) = -\sum_{j \neq i} q(i,j)$. Let $\pi$ be the stationary distribution. Because of the reversibility of the chain, the matrix $L$ is symmetrizable and therefore has real eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$. Since $\lambda_1$ is the distance between the eigenvalue 0 and the rest of the spectrum of $L$, it is called the *gap* and is denoted by *Gap(L)* or simply *Gap*. Let $\tau_B$ be the first hitting time on $B^c$ which is a proper non-empty subset of $S$. So $0 < E_\pi \tau_B < \infty$. Let $-L^B$ be the transition rate matrix $-L$ restricted to $B$ and $\lambda_0(B)(> 0)$ be the smallest eigenvalue of $L^B$. Then we have

**Theorem 1.1.2** *(Aldous and Brown (1991) Theorem 3)*

$$P_\pi(\tau_B > t) \geq (1 - \frac{\lambda_0(B)}{Gap}) \exp(-\lambda_0(B)t), \qquad t > 0. \tag{1.1.2}$$

This theorem provides a lower bound of the probability of $\tau_B > t$ in terms of two parameters. It is therefore easy to use since we have various methods to get bounds or estimates for these parameters. We will discuss some of the methods at the end of this section.

The error bound for the departure of $\tau_B$ from exponentiality is given by

**Theorem 1.1.3** *(Aldous and Brown (1991) Theorem 1)*

$$|P_\pi(\tau_B / E_\pi \tau_B > t) - \exp(-t)| \leq \frac{\lambda_0(B)/Gap}{1 + \lambda_0(B)/Gap} \leq \lambda_0(B)/Gap \tag{1.1.3}$$

*for all* $t > 0$.

Again this result is easy to use. But, since the bounds do not depend on $t$, they are not good enough for large values of $t$.

**The Mixing Technique:** Aldous (1982) proved that if a Markov chain converges rapidly to stationarity, then the time until the first hit on a rarely-visited

set of states is approximately exponentially distributed. As a matter of fact, a parameter which measures the time taken for the chain to approach stationarity was defined as:

$$d(e) := \min\{t : \sum_j |P_i(X_t = j) - \pi(j)| \leq 1/e \text{ for all } i\},$$

where $X_t$ is an irreducible continuous-time Markov chain with stationary distribution $\pi$, and $1/e$ is an arbitrary constant. Then for any subset $B$ of the state space and the first hitting time $\tau_B$ on $B^c$, we have

**Theorem 1.1.4** *(Aldous (1982; 1983a))*

$$\sup_{t\geq 0} |P_\pi(\tau_B > t) - exp(-t/E_\pi\tau_B)| \leq \psi(\frac{d}{E_\pi\tau_B}), \tag{1.1.4}$$

*where $\psi(x) \to 0$ as $x \to 0$ is an absolute function, not depending on the chain.*

It is noted that the error bound given above, which is valid for all chains, might not be numerically useful in particular cases. Rather, the theorem indicates which parameters need to be small in order to justify an exponential approximation. The *Gap* in Theorem 1.1.2 is such an important parameter since a strictly positive *Gap* makes the chain converge to stationarity exponentially fast (see Liggett (1989)). We will concentrate on the cases where the *Gap*'s are strictly positive. With this standing assumption, Iscoe and McDonald (1992) give an explicit error bound for the exponential approximations to either reversible and non-reversible Markov jump processes. The main result is

**Theorem 1.1.5** *(Iscoe and McDonald (1992) Theorem 2.8) If $\pi(B^c)$ is sufficiently small then for all $t \geq 0$*

$$|P_\pi(\tau_B > t) - e^{-\lambda_0(B)t}| \leq \beta(B)e^{-\lambda_0(B)t} \tag{1.1.5}$$

*where*

$$\mathcal{J}(B) := \frac{4}{(\mathrm{Gap}(L)-\overline{\kappa})^2 - 4\kappa_1\kappa_2}\left[1 + \frac{\sqrt{(\mathrm{Gap}(L)-\overline{\kappa})^2 + 4\kappa_2^2}}{\mathrm{Gap}(L)-\overline{\kappa}}\right]\kappa_1\kappa_2.$$

While the exact definitions of the parameters $\overline{\kappa}, \kappa_1$ and $\kappa_2$ will be given in the next chapter, here we simply point out that they tend to 0 as $B \rightarrow S$ given that the infinitesimal generator $-L$ is a bounded operator.

As was mentioned earlier, there are two major parameters, i.e. $\lambda_0(B)$ and *Gap*, which are important in the error estimation of the approximation and in bounding the probability of $\tau_B > t$. There is a fairly large literature concerning bounding *Gap*. For instance, Lawler and Sokal (1988) provide bounds for *Gap* via a generalization of Cheeger's constant, while Diaconis and Stroock (1991) discuss the geometric bounds of *Gap* for the discrete-time Markov chains. It is noted that the uniformization technique (see Keilson (1979)) permits the continuous time processes to be related directly to the discrete time chains and to exploit the results and numerical matrix techniques for the maximal eigenvalue available in discrete time.

Iscoe and McDonald (1992) use the mean killing rate as the asymptotic estimate of $\lambda_0(B)$. The error, in general, is of the same order as $\lambda_0$ while in some special cases this estimate is asymptotically accurate. Some of the results are reviewed in Chapter 2, especially in the last section.

### 1.1.3 New Results on the First Hitting Times

In Chapter 2, we present an improvement of Theorem 1.1.5, which also enables us to obtain an extension and generalization of Theorem 1.1.2 and Theorem 1.1.3. Specifically, we show a higher order error bound in terms of $t$ for the exponential approximation:

$$\left|P_\pi(\tau_B > t) - c(B)e^{-\lambda_0(B)t}\right| \leq \frac{\lambda_0(B)}{Gap}(1 + ||\rho_0 - \phi_0||_\pi)e^{-\Gamma_m t}$$

$$\leq \left(\frac{\lambda_0(B)}{Gap} + o(\lambda_0(B))\right) e^{-(Gap-\lambda_0(B))t}$$

where $\phi_0$ and $\rho_0$ are the eigenvectors of $L^B$ and its adjoint $(L^B)^*$ (in $L^2(\pi)$) respectively, corresponding to to $\lambda_0(B)$, $c(B) := (\rho_0, \mathbf{1}_B)_\pi(\phi_0, \mathbf{1}_B)_\pi$ and $o(\lambda_0(B)) := \lambda_0(B)||\rho_0 - \phi_0||_\pi/Gap$. The extension of Theorem 1.1.2 to the *nonreversible* Markov jump process is the following:

$$P_\pi(\tau_B > t) \geq \left(1 - \frac{\lambda_0(B)}{Gap}\right) e^{-\lambda_0(B)t} - \left(\frac{\lambda_0(B)}{Gap} + o(\lambda_0(B))\right) e^{-(Gap-\lambda_0(B))t}.$$

The lower and upper bounds for the expectation of $\tau_B$ are also obtained:

$$\frac{1}{\lambda_0(B)} - \frac{1}{Gap} - \mathcal{O}(\lambda_0(B)) \leq E_\pi \tau_B \leq \frac{||\rho_0||_\pi}{\lambda_0(B)} + \mathcal{O}(\lambda_0(B))$$

where $\mathcal{O}(\lambda_0(B)) := \lambda_0(B)(1 + ||\rho_0 - \phi_0||_\pi)/Gap(Gap - \lambda_0(B))$. Notice that the lower bound provided for $P_\pi(\tau_B > t)$ and $E_\pi \tau_B$ depend essentially on two parameters, i.e. $\lambda_0(B)$ and $Gap$, except for higher order error terms.

In the *reversible* cases, things are even simpler. First, we obtain the same lower bound as given by Theorem 1.1.2 for a general Markov jump process. Secondly, we show an improvement of the error bound of Theorem 1.1.3: for all $t > 0$,

$$\left| P_\pi(\frac{\tau_B}{E_\pi \tau_B} > t) - e^{-t} \right| \leq \max\{\frac{\lambda_0(B)}{Gap}, \exp\{\frac{\lambda_0(B)t}{Gap}\} - 1\} e^{-t}.$$

The following *exponential tail property* of $\tau_B$ is the continuous-time version of Theorem 1.1.1:

$$P(X_t = j, \tau_B > t | X_0 = i) \sim \phi_0(i)\rho_0(j)\pi(j)e^{-\lambda_0(B)t}, \qquad as \qquad t \to \infty.$$

This result comes from a natural generalization of the exponential approximation of the distribution of $\tau_B$. Details are also given in Chapter 2.

We present a simple formula for $\lambda_0(B)$ based on the observation that, if the process has no long-range interaction, the value of $\lambda_0(B)$ will very much depend

on the values of the stationary distribution and the values of the corresponding eigenvector on the boundary of $B$. We show by two examples (the $M/M/1$ queue and the $M/M/\infty$ queue) that this formula is useful for obtaining the asymptotics of $\lambda_0(B)$. As a matter of fact, the formula also helped in finding the asymptotics of $\lambda_0(B)$ for a more complicated Markov jump process in Chapter 4. Another result we get in the last section of Chapter 2 is the exact value of *Gap* for the $M/M/\infty$ queue. This allows us to apply the theorems obtained in Chapter 2 to the ATM (asynchronous transfer mode) link capacity problem. In that calculation, we employed *the generating function method* for the eigenvectors of the infinitesimal generator.

## 1.2 Capacity of an ATM Multiplexor

### 1.2.1 Description and Previous Work

Chapter 3 and 5 are devoted to applications to telecommunication networks, especially to problems concerning the capacity of an ATM multiplexor and the size of its buffer. A simplified diagram of the model of an ATM multiplexor is given in Figure 1.1. There, a number of categories of bursty traffic sources are statistically multiplexed over a common link. All traffic sources produce fixed length information packets called ATM cells. Sources within a category produce cells at the same rate (cells per second) when they are *active* (or *on*). The link has a certain capacity to handle the information packets being multiplexed over it. Once the link capacity is exceeded, the excess packets are stored in the buffer shared by all traffic sources. If the buffer is overloaded, information will be lost.

Statistical multiplexing and buffer overflow are very important issues in network design and engineering. They have been investigated for years by many authors.

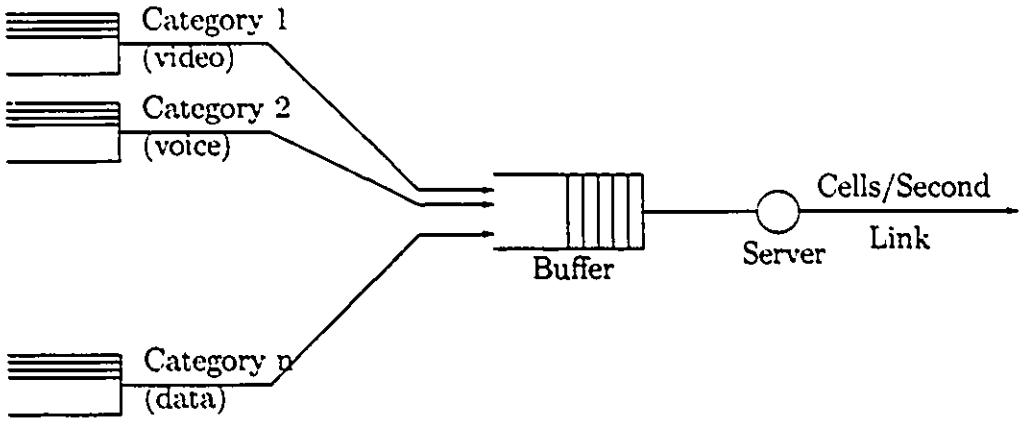


Figure 1.1: The model of an ATM multiplexor

The reader can find interesting recent work and a comprehensive list of references on these problems under the ATM environment in Norros, etc (1991). We will also discuss some related work on the buffer overflow problem in the next section and in Chapter 5. Most previous work was done by steady-state analysis or by simulation. We estimated the probability that the link capacity is exceeded during a fixed time interval along with the mean time until this happens. This provides the network designer with a degree of assurance about the probability and frequency of overloads. We compared a number of different methods to achieve this, including the theory developed in Chapter 2.

### 1.2.2 A New Approximation Technique

**This part is joint with Ian Iscoe and David McDonald.**

In Chapter 3, the ATM multiplexor is modeled as follows. Assume that there are $n$ different categories of bursty traffic sources. Within each category there are infinite many sources which are identical and mutually independent. Active sources in category $i$ produce cells at a rate of $d_i$ cells per second. Arrivals of active sources

from category $i$ subject to a Poisson process with a rate of $a_i$ arrivals per second and the duration of active period is independent and exponentially distributed with a mean length of $1/b_i$. Let $N_i(t)$ represent the number of active sources from category $i$ being multiplexed at the link at time $t$. The total load at time $t$ may be represented by

$$N(t) := \sum_{i=1}^{n} d_i N_i(t).$$

Assuming that the link rate is $L-1$ cells per second, we may define the first overload time by

$$\tau = \inf\{t \geq 0 : N(t) \geq L\}.$$

Notice that $N_i(t)$ is statistically equivalent to an $M/M/\infty$ queue with arrival rate $a_i$ and service rate $b_i$ . Thus, the multidimensional Markov process

$$\vec{N}(t) := (N_1(t), \cdots, N_n(t))$$

(defined on the state space $S := \{0, 1, 2, \cdots\}^n$) is reversible with respect to the stationary distribution $\pi$. Let $\mathcal{L}$ be the infinitesimal generator of $\vec{N}(t)$ and define $\alpha(\vec{x}) = E_{\vec{x}}\tau$, then $\alpha$ satisfies

$$\mathcal{L}\alpha(\vec{x}) = -1 \quad \text{for } \vec{x} \in B$$

$$\alpha(\vec{x}) = 0 \qquad \text{for } \vec{x} \in B^c$$

where $B^c := \{(s_1, s_2, \cdots, s_n) \in S : \sum_{i=1}^{n} d_i s_i \geq L\}$. Define $\kappa_\theta(\vec{x}) = E_{\vec{x}} \exp(-\theta\tau)$, then

$$\mathcal{L}\kappa_\theta(\vec{x}) = \theta\kappa_\theta(\vec{x}) \quad \text{for } \vec{x} \in B$$

$$\kappa_\theta(\vec{x}) = 1 \qquad \text{for } \vec{x} \in B^c.$$

The Laplace transform of $\tau$ determines an upper bound for the probability distribution function of $\tau$. However, the exact solutions to these linear systems are intractable for a fairly large model.

We use the theory of *induced Dirichlet forms* to generate a one-dimensional Markov jump process $N^*(t)$, with the induced stationary distribution $\pi^*$, from the original multidimensional Markov process. Define the first hitting time $\tau^* := \inf\{t \geq 0 : N^*(t) \geq L\}$, then for all $\theta \geq 0$,

$$E_\pi \exp(-\theta\tau) \leq E_{\pi^*} \exp(-\theta\tau^*)$$

and

$$E_{\pi^*}\tau^* \leq E_\pi\tau.$$

We may solve

$$\begin{aligned} \mathcal{L}^*\alpha^*(r) &= -1 \quad \text{for } r \in B^* \\ \alpha^*(r) &= 0 \qquad \text{for } r \in (B^*)^c \end{aligned}$$

for $\alpha^*(r) = E_r\tau^*$, and

$$\begin{aligned} \mathcal{L}^*\kappa_\theta^*(r) &= \theta\kappa_\theta^*(r) \quad \text{for } r \in B^* \\ \kappa_\theta^*(r) &= 1 \qquad\qquad \text{for } r \in (B^*)^c \end{aligned} \tag{1.2.6}$$

for $\kappa_\theta^*(r) = E_r \exp(-\theta\tau^*)$, where $(B^*)^c := \{L, L+1, \cdots\}$. The point is that the sizes of the linear systems associated with $N^*(t)$ are much smaller than those associated with $\vec{N}(t)$. An engineering approach might be to estimate any probability associated with the process $\vec{N}(t)$ by the corresponding probability for the induced process $N^*(t)$. Numerical comparisons of the mean and Laplace transform of $\tau$ and $\tau^*$ for different models show that the estimates are extremely close.

In Chapter 4, we try to give some explanation for the success of the induced Dirichlet forms method applied to the ATM multiplexor problem. By the theory presented in Chapter 2 we know that the principal eigenvalue $\lambda_0(B)$ ($\lambda_0^*(B^*)$) of $\mathcal{L}^B$ ($(\mathcal{L}^B)^*$) determines the asymptotic behaviour of the mean of $\tau$ ($\tau^*$). For the ATM

model we show $\lambda_0(B)/\lambda_0^*(B^*) \to 1$, as $B \to S$, where $S$ is the state space of $\vec{N}(t)$. This eventually says that the means of $\tau$ and $\tau^*$ are asymptotically the same as $B \to S$.

## 1.3 Buffer Overflow of an ATM Multiplexor

### 1.3.1 Markov-Modulated Fluid Model

The buffer size problem is studied in Chapter 5. There, the buffer content is described by a Markov-modulated fluid model. Again, assume that there are $n$ independent categories of bursty traffic sources being served by a link of capacity $L$. The traffic state $\vec{N}(t) := (N_1(t), N_2(t), \cdots, N_n(t))$ is modeled as a finite-state reversible Markov process. Let $r(\vec{N}(t))$ be the superposition of the rates at which ATM cells are produced. In the Markov-modulated fluid model, the buffer content $X(t)$ is a continuous random variable which satisfies the following equation: for $X > 0$,

$$\frac{dX}{dt} = r(\vec{N}(t)) - L.$$

Clearly, $(\vec{N}(t), X(t))$ is a Markov process. Let $F(\vec{n}, x)$ be the equilibrium probability that while in state $\vec{n}$ the buffer content is less than or equal to $x$. Then the following equation is satisfied:

$$\mathbf{D}\frac{d\mathbf{F}}{dx} = \mathbf{G}^T\mathbf{F}$$

where $\mathbf{D} = diag\{r(\vec{n})\}$, $\mathbf{F}$ is the vector with entries $F(\vec{n}, x)$ and $\mathbf{G}$ is the generator matrix of the process $\vec{N}(t)$. The solution of the above equation takes the form:

$$\mathbf{F}(x) = \sum_i a_i \exp(z_i x)\phi(i).$$

The procedure involves: (1) finding the eigenvalues and eigenvectors $\{z_i, \phi(i)\}$ of the matrix $\mathbf{D}^{-1}\mathbf{G}^T$. (2) calculating the coefficients $\{a_i\}$ which are determined by *implicit* boundary conditions.

### 1.3.2 Some Earlier Work

Anick etc. (1982) gave a complete solution for the case of a finite number of independent and identical ON/OFF sources. In this case, the total number of ON sources is a one-dimensional Markov process. They were able to construct closed-form formulas for calculating the eigensystem and the coefficients. Kosten (1984) considered systems composed of a superposition of several independent subsystems, each subsystem being of the same type as Anick etc. (1982). By decomposing the big system into small subsystems one may get the eigensystem. But in general, it is numerically intractable since finding the coefficients of the eigenfunction expansion of $\mathbf{F}$ involves solving a large linear system. Stern and Elwalid (1991) extended Kosten (1984) to a very general model known as the *separable Markov modulated rate model*. The method is again a decomposition technique. Certain lower and upper bounds for the coefficients were provided.

### 1.3.3 Equivalent Bandwidth Methods

In the final chapter, we present a new approximation method for the Markov-modulated fluid models. This is based on the idea that we can create a one-dimensional Markov process $r^*(t)$ which has a bandwidth approximately equivalent to the rate process $r(\vec{N}(t))$ (non-Markovian in general). Within this method, all states $\vec{s} = (s_1, s_2, \cdots, s_n)$ which satisfy $r(\vec{s}) := \sum_{i=1}^{m} d_i s_i = k$ are collapsed into a single state $k$ of $r^*(t)$. Assume that $X^*(t)$ is the content of a virtual buffer modulated by $r^*(t)$. Then the system of differential equations governing the steady state distribution of the generated Markov process $(r^*(t), X^*(t))$ can be dramatically smaller than that of $(\vec{N}(t), X(t))$. Numerical results show that the eigenvalues of the original system are very well interlaced with those of the reduced virtual system, and the

estimates for the coefficients are uniformly good for all systems considered. We also prove that the probability of overflow of the virtual buffer is asymptotically a lower bound for that of the original buffer. It is observed that the more homogeneous the traffic characteristics the better the approximations. We believe that this fact is related to the goodness of approximating a non-Markov rate process by a Markov one.

# Chapter 2

# Asymptotics of the First Hitting Times

## 2.1 General Assumptions and Notation

Let $(X_t; t \geq 0)$ be a positive-recurrent continuous-time Markov jump process with a general measurable state space $(S, \mathcal{S})$, transition semi-group $T_t$ and invariant probability measure $\pi$, where $T_t$ operates on $L^2(S, \pi)$. Denote the probability transition kernel of the $X_t$ by $P_t(x, dy)$, so that

$$T_t f(x) = \int_S P_t(x, dy) f(y); \qquad t \geq 0, \qquad f \in L^2(S, \pi). \tag{2.1.1}$$

Note that $T_t \mathbf{1} = T_t^* \mathbf{1} = \mathbf{1}$, for all $t \geq 0$, where $\mathbf{1}_B$ always denotes the indicator function of $B \subseteq S$ with $\mathbf{1} := \mathbf{1}_S$, and $^*$ denotes the adjoint; so

$$\int T_t f(x) \pi(dx) = \int f(x) \pi(dx), \qquad f \in L^2(S, \pi).$$

Unless mentioned otherwise, if $\mu$ is a measure on $B$, a measurable subset of $S$, $L^2(B, \mu)$ will be considered as a real vector space of real-valued functions. One important exception will be when we discuss spectral properties of operators related to $T_t$. In this case we of course must consider $L^2(B, \mu)$ as a complex vector space of complex-valued functions; we adopt the following convention:

$$(f, g)_\mu = \int \overline{f(x)} g(x) \mu(dx), \qquad f, g \in L^2(B, \mu).$$

The nature of $L^2(B,\mu)$ will always be clear from the context, so we shall not adopt separate notations to distinguish the two possibilities. Note that $T_t$ and all other related operators (such as $L$, below) have obvious, natural extensions from *real* $L^2(\mu)$ to *complex* $L^2(\mu)$ given formally by (2.1.1) and (2.1.3). It should be noted that the norms of these extended operators do not change (see Lemma 7.5 in Davies (1980)).

Denote by $P_\mu$ the law of $(X_t; t \geq 0)$ initially (at $t = 0$) distributed according to the probability $\mu$ on $(S,\mathcal{S})$; $E_\mu$ will denote the corresponding expectation. Thus the process is stationary with respect to $P_\pi$.

We denote the probability transition rate kernel by $J(x,dy)$ and consider only processes in which the transition rates are essentially bounded, i.e.

$$\pi - \operatorname{ess\,sup}_{x\in S} J(x,\{x\}^c) \leq M < \infty. \tag{2.1.2}$$

Then the infinitesimal generator $-L$ of this jump process, which is given by

$$Lf(x) = \int J(x,dy)|f(x) - f(y)|. \tag{2.1.3}$$

defines a bounded linear operator (of norm $\leq 2M$) on $L^2(S,\pi)$ (and in fact on all the spaces $L^p(S,\pi)$). The constant function $\mathbf{1}$ is an eigenvector of $L$ (and its adjoint $L^*$) with eigenvalue 0. So, for all $f \in L^2(S,\pi)$, $\int Lf\pi(dx) = 0$; or in terms of $J$:

$$\int\int J(x,dy)f(y)\pi(dx) = \int\int J(x,dy)f(x)\pi(dx). \tag{2.1.4}$$

A jump process is called *reversible* if

$$\pi(dx)J(x,dy) = \pi(dy)J(y,dx), \tag{2.1.5}$$

or equivalently if the generator $-L$ on $L^2(S,\pi)$ is selfadjoint.

Let $B \in \mathcal{S}$ denote those states outside the *bad region* $B^c$ and denote by $T_t^B$ the semi-group of the process killed outside $B$; that is the process killed when it enters

$B^c$. For $f \in L^2(B, \pi_B)$ this semigroup is defined by

$$T_t^B f(x) := E_x[f(X_t); \tau > t]; \qquad x \in B$$

$$\tau \equiv \tau^B := \inf\{t > 0 : X_t \notin B\}. \tag{2.1.6}$$

Let $-L^B$ denote the infinitesimal generator associated with $T_t^B$. Then the associated transition rate kernel and killing rates are given by, for $x \in B$,

$$J^B(x, dy) := 1_B(y)J(x, dy), \qquad K^B(x) := J(x, B^c).$$

The infinitesimal generator $-L^B$ expressed through

$$L^B f(x) = \int_B J^B(x, dy)[f(x) - f(y)] + K^B(x)f(x),$$

defines a bounded linear operator (of norm $\leq 2M$) on $L^2(B, \pi_B)$. Note that $K^B = L^B 1_B$.

We will only consider the cases where $B \in \mathcal{S}$ is a non-trivial subset of $S$ in the sense that $0 < \pi(B) < 1$. We use $\pi_B$ for $\pi$ restricted to $B$ and $\hat{\pi}$ for its normalized version, i.e. $\hat{\pi} := \pi_B/\pi(B)$. Notice that $-L^B$ is selfadjoint on $L^2(B, \pi_B)$ or $L^2(B, \hat{\pi})$ if $-L$ is selfadjoint on $L^2(S, \pi)$.

We wish to estimate the tail of the distribution of $\tau$ starting from the stationary distribution or $P_\pi(\tau > t)$. One of our main goals is to estimate $P_\pi(\tau > t)$ by $c(B)\exp(-\lambda_0(B)t)$ as $B \to S$, where $c(B)$ is some constant, $0 \leq \lambda_0(B) := \inf \Re e\{\sigma(L^B)\}$, the bottom of the real part of the spectrum of $L^B$; and "$B \to S$" is a shorthand for the convergence: $\pi(B) \to 1$ (Strictly speaking, the convergence statements should involve a *sequence* of sets $(B_n)_{n\in\mathbf{N}}$ for which $\lim_{n\to\infty} \pi(B_n^c) = 0$). This in turn will imply that $\tau$ is approximately exponentially distributed

with mean $1/\lambda_0(B)$. To establish this result, it is essential to know that $\lambda_0(B) > 0$. Under certain conditions it is ensured by the fact that $\lambda_0(B)$ is an isolated point

of the spectrum of $L^B$ and is positive-valued. This and other eigenvalue properties, are discussed in the next section.

## 2.2 Spectral Properties

This section is devoted to various spectral properties of the operators $L$ and $L^B$, respectively associated with the Markov jump process $X_t$ and the one killed outside $B$. There are two parameters which are central to this work. One is $\lambda_0(B) := \inf \Re e\{\sigma(L^B)\}$, the bottom of the real part of the spectrum $\sigma(L^B)$ of $L^B$. The other is defined as follows.

**Definition 2.2.1** *For $L$ as defined in (2.1.3)*

$$Gap(L) := \inf\{(f, Lf)_\pi : f \in L^2(S,\pi), ||f||_\pi = 1, (f,\mathbf{1})_\pi = 0\}.$$

The reason for this name is that if $L$ is selfadjoint then $Gap(L)$ is the gap between the simple eigenvalue 0 and the rest of the spectrum.

A strictly positive $Gap(L)$ ensures exponential convergence of the process to its stationary distribution in $L^2$ norm (details will be given in the last section of this chapter). This, in turn, ensures the weak convergence of the law of the first hitting time $\tau$ to that of an exponential random variable. As was mentioned at the end of the last section, a positive $\lambda_0(L)$ is also essential to the exponential approximation problem. Techniques for bounding $Gap(L)$ will be discussed in the last section, while theorems about the existence of a positive $\lambda_0(B)$ are presented next. Unless mentioned otherwise, the following will be standing assumptions throughout this chapter:

**Assumption 1:** $Gap(L) > 0$.

**Assumption 2:** $\lambda_0(B) > 0$. *Moreover, the eigenvectors of $L^B$ and $(L^B)^*$ corresponding to $\lambda_0(B)$ are real and nonnegative vectors.*

**Assumption 3:** *$\pi(B^c)$ is small enough so that $\lambda_0(B) < Gap(L)$.*

**Assumption 4:** The infinitesimal generator $-L$ of the jump process is essentially bounded.

### 2.2.1 About Assumption 2

An important special case for which **Assumption 2** is true is when $B$ is a finite set, so $L^B$ has transition rates bounded above by $M$. Hence $-L^B + MI$ is a positive matrix, where $I$ is the identity matrix. By the Perron-Frobenius theorem (see Seneta (1981)) this matrix has a maximum positive eigenvalue $M + \lambda_0(B)$ of multiplicity one associated with the nonnegative right eigenvector $\phi_0$ and the nonnegative left eigenvector $\rho_0$. This eigenvector pair is also an eigenvector pair for $L^B$ and is associated with the least eigenvalue $\lambda_0(B) \geq 0$ of multiplicity one. From the irreducibility assumption of the Markov process $X_t$ it follows that $\lambda_0(B)$ is strictly positive.

For a countable infinite subset $B$, an extension of the Perron-Frobenius theorem (see Vere-Jones (1967)) can be applied in the same way as above as long as $L^B$ is bounded. Nummelin (1984) gives an extension of the Perron-Frobenius theorem in the context of more general non-negative kernels on a general measurable state space, which is also useful for justifying **Assumption 2**.

Since our principal interest is in the asymptotic behavior as $B \to S$ of the distribution of $\tau$, a result by Iscoe and McDonald (1992) is quite useful. This gives the closeness of the Perron-Frobenius eigenvector to $\mathbf{1}_B$. Details are as follows.

There are three quantities related to the kernel $J$ which will appear in the hypotheses and in the results of this and the following sections. They are the mean killing rate, the standard deviations of the killing rate and resuscitation rate with

respect to the probability $\hat{\pi}$ on $B$, viz.

$$(2.2.7) \quad \overline{\kappa} \equiv \overline{\kappa}^B := \int_B K^B(x)\hat{\pi}(dx), \kappa_1 := \| K^B - \overline{\kappa} \|_{\hat{\pi}}, \kappa_2 := \| R^B - \overline{\kappa} \|_{\hat{\pi}},$$

where the resuscitation rate $R^B$ is defined to be the Radon-Nikodym derivative of the measure $\mu(dy) := \int_{B^c} \pi(dx)J(x, dy)$ with respect to $\pi|B$, $\pi$ restricted to $B$.

The following simple estimates for $\overline{\kappa}$, $\kappa_1$ and $\kappa_2$ are immediate and serve to make the hypotheses for the results more explicit and perhaps more practical, albeit less sharp.

**Lemma 2.2.2** *(Iscoe and McDonald (1992) Lemma 2.5) Assume that the transition rates are essentially bounded, i.e.*

$$(2.2.8) \qquad \pi - ess\sup_{x \in S} J(x, \{x\}^c) \leq M < \infty.$$

*If $\pi(B^c)$ is sufficiently small, then*

$$(i) \quad \overline{\kappa} \leq M\frac{\pi(B^c)}{\pi(B)},$$

$$(ii) \quad \kappa_1 \leq M\left[\frac{\pi(B^c)}{\pi(B)}\right]^{1/2},$$

$$(iii) \quad \kappa_2 \leq M\left[\frac{\pi(B^c)}{\pi(B)}\right]^{1/2}.$$

*So, $\overline{\kappa}, \kappa_1, \kappa_2 \to 0$ as $\pi(B^c) \to 0$.*

Using a theorem in Stewart (1971), Iscoe and McDonald (1992) showed the following:

**Theorem 2.2.3** *(Iscoe and McDonald (1992) Theorem 2.7) Assume that the infinitesimal generator $-L^B$ is bounded. If the quantities $\overline{\kappa}, \kappa_1$ and $\kappa_2$ satisfy*

$$\overline{\kappa} < Gap(L), \qquad 4\kappa_1\kappa_2 < |Gap(L) - \overline{\kappa}|^2,$$

*then* $L^B$ *and* $(L^B)^*$ *have a common non-negative, isolated eigenvalue* $\lambda_0(B)$ *and associated real eigenvectors* $\phi_0$ *and* $\rho_0$, *respectively, belonging to* $L^2(B,\hat{\pi})$ *such that* $\int_B \rho_0 d\hat{\pi} = 1$ *and* $\int_B \phi_0\rho_0 d\hat{\pi} = 1$.

*Moreover,* $\lambda_0(B) = \inf \Re e\{\sigma(L^B)\}$, $\inf|\Re e\{\sigma(L^B)\backslash\{\lambda_0(B)\}\}| > 0$, *and*

$$(i) \quad |\lambda_0(L) - \overline{\kappa}| \leq 2\kappa_1\kappa_2/|Gap(L) - \overline{\kappa}|,$$

$$(ii) \quad ||\rho_0 - \mathbf{1}_B||_{\hat{\pi}} \leq 2\kappa_2/|Gap(L) - \overline{\kappa}|.$$

$$(iii) \quad ||\phi_0 - \mathbf{1}_B||_{\hat{\pi}} \leq \frac{2\kappa_1\sqrt{(Gap(L) - \overline{\kappa})^2 + 4\kappa_2^2}}{(Gap(L) - \overline{\kappa})^2 - 4\kappa_1\kappa_2}.$$

$$(iv) \quad |\int_B \phi_0 d\hat{\pi} - 1| \leq \frac{4\kappa_1\kappa_2}{(Gap(L) - \overline{\kappa})^2 - 4\kappa_1\kappa_2}.$$

### 2.2.2 Inequalities

We begin by recalling the notion of the numerical range of a bounded operator $A$ on a complex Hilbert space $\mathcal{H}$ . The *numerical range,* $W(A)$, of $A$ is defined by:

$$W(A) := \{(f, Af) : f \in \mathcal{H}, \|f\| = 1\}.$$

We recall the following well-known facts.

- $W(A)$ is convex.
- The spectrum of $A$, $\sigma(A)$, is contained in $\overline{W(A)}$, the closure of $W(A)$.
- If $A$ is normal then $\overline{W(A)}$ is the convex hull of $\sigma(A)$. However, in general $\overline{W(A)}$ can be much larger than the convex hull of $\sigma(A)$; e.g. if $A$ is nilpotent.
- If $\lambda \notin \overline{W(A)}$ then $||(\lambda - A)^{-1}|| \leq [\text{dist}(\lambda, W(A))]^{-1}$.

Now consider a complex Hilbert space $H$, a bounded operator $A$ on $H$, and a fixed vector $\rho \in H$. Set $\mathcal{H} = \rho^{\perp} \equiv \{f \in H : (f, \rho) = 0\}$ and denote by $Y$ the injection of $\mathcal{H}$ into $H$; $Y^*$ is the adjoint projection. Finally set $\tilde{A} = Y^* A Y : \mathcal{H} \to \mathcal{H}$, the *compression* of $A$ into $\mathcal{H}$.

**Definition 2.2.4**

$$\Gamma_\rho \equiv \Gamma_\rho(A) := w(\tilde{A}) \equiv \inf\{(f, Af) : \|f\| = 1, (f, \rho) = 0\}.$$

In particular $\Re e\{\sigma(\tilde{A})\} \subset [\Gamma_\rho, \infty)$; and in the case that $\rho$ is an eigenvector of $A$ with eigenvalue $\lambda$ such that $\Re e\lambda < \Gamma_\rho$, it is easy to check that $\sigma(A)/\{\lambda\}$ does not meet the strip: $\{z \in \mathbf{C} : \Re e\lambda < \Re e z < \Gamma_\rho\}$. In this case it is appropriate to refer to $\Gamma_\rho - \Re e\lambda$ as the *numerical gap* of $A$ (with respect to $\rho$ and $\lambda$). It may be smaller than the actual gap in the real part of the spectrum of $A$.

In the following lemma we normalize the eigenvectors $\phi_0$ and $\rho_0$ in a different way from that in Theorem 2.2.3. We also use the simplified notation: $Gap := Gap(L), \lambda_0 := \lambda_0(B)$.

**Lemma 2.2.5** *Assume that $\phi_0$ and $\rho_0$ are the eigenvectors of $L^B$ and $(L^B)^*$, respectively, which correspond to $\lambda_0$ such that $||\phi_0||_\pi = 1$ and $(\rho_0, \phi_0)_\pi = 1$. Under* **Assumptions 1, 2** *and* **4**, *we have*

$$(i) \quad 1 - \frac{\lambda_0}{Gap} \le (\phi_0, \mathbf{1}_B)_\pi^2 \le \pi(B).$$

$$(ii) \quad 1 - \frac{\lambda_0}{Gap} \le \frac{(\rho_0, \mathbf{1}_B)_\pi^2}{\|\rho_0\|_\pi^2} \le \pi(B).$$

$$(iii) \quad 1 - \frac{\lambda_0}{Gap} \le (\phi_0, \mathbf{1}_B)_\pi (\rho_0, \mathbf{1}_B)_\pi \le \pi(B)\|\rho_0\|_\pi.$$

**Proof**: $(i)$ By definition,

$$Gap = \inf_u \left\{ \frac{(u, Lu)_\pi}{(u, u)_\pi}, u \in L^2(S, \pi), (u, \mathbf{1})_\pi = 0 \right\}$$

and

$$\lambda_0 = (\phi_0, L^B\phi_0)_\pi.$$

We make extensions of $\phi_0$ and $\rho_0$ so that they take 0 values on $B^c$. The extended vectors are also denoted by $\phi_0$ and $\rho_0$ respectively, and they belong to $L^2(S,\pi)$.

Let $c := (\phi_0, \mathbf{1})_\pi$ and $u_0 := \phi_0 - c\mathbf{1}$, then

$$|c|^2 = |(\phi_0, \mathbf{1}_B)_\pi|^2 \le ||\phi_0||_\pi^2 ||\mathbf{1}_B||_\pi^2 = \pi(B) < 1. \tag{2.2.9}$$

Moreover,

$$(u_0, \mathbf{1})_\pi = (\phi_0, \mathbf{1})_\pi - c(\mathbf{1}, \mathbf{1})_\pi = 0.$$

and

$$(u_0, u_0)_\pi = (\phi_0, \phi_0)_\pi - 2c(\phi_0, \mathbf{1})_\pi + c^2 = 1 - c^2.$$

On the other hand,

$$\begin{aligned}
(u_0, Lu_0)_\pi &= (\phi_0 - c\mathbf{1}, L\phi_0 - cL\mathbf{1})_\pi = (\phi_0 - c\mathbf{1}, L\phi_0)_\pi \qquad (\text{since } L\mathbf{1} = 0) \\
&= (L^*\phi_0 - cL^*\mathbf{1}, \phi_0)_\pi = (\phi_0, L\phi_0)_\pi \qquad (\text{since } L^*\mathbf{1} = 0) \\
&= (\phi_0, L^B\phi_0)_\pi = \lambda_0.
\end{aligned}$$

Therefore

$$Gap \le \frac{(u_0, Lu_0)_\pi}{(u_0, u_0)_\pi} = \frac{\lambda_0}{1 - c^2}.$$

This yields

$$1 - \frac{\lambda_0}{Gap} \le c^2 = (\phi_0, \mathbf{1}_B)_\pi^2 \le \pi(B). \tag{2.2.10}$$

(*ii*) Taking $c := (\rho_0, \mathbf{1})_\pi$ and

$$u_0 = \rho_0 - c\mathbf{1},$$

the same procedure as in (i) gives the proof.

(*iii*) This is proved by using (*i*) , (*ii*) and the following:

$$1 = (\phi_0, \rho_0)_\pi \leq \|\phi_0\|_\pi \|\rho_0\|_\pi = \|\rho_0\|_\pi.$$

Therefore

$$\pi(B)\|\rho_0\|_\pi \geq (\phi_0, \mathbf{1}_B)_\pi (\rho_0, \mathbf{1}_B)_\pi \geq (\phi_0, \mathbf{1}_B)_\pi \frac{(\rho_0, \mathbf{1}_B)_\pi}{\|\rho_0\|_\pi} \geq 1 - \frac{\lambda_0}{Gap}.$$

■

**Remark.** In the proof of Lemma 2.2.5, **Assumption 4** (i.e., $L$ is bounded) is used to ensure that the domain $\mathcal{D}(L)$ of $L$, which equals to $L^2(S, \pi)$, contains the zero-extensions of the eigenvectors $\phi_0$ and $\rho_0$. It is clear that this lemma remains valid for processes with unbounded generator $-L$ given that the zero-extensions of $\phi_0$ and $\rho_0$ belong to $\mathcal{D}(L)$. In some cases, as we will see later, the latter assumption can be verified directly. ■

Lemma 2.2.5 is mainly used for bounding $(\phi_0, \mathbf{1}_B)_\pi$ and $(\rho_0, \mathbf{1}_B)_\pi$ in the sequel. But it also provides an interesting relation between $Gap$ and $\lambda_0$ as presented by

**Corollary 2.2.6**

$$Gap(L) \leq \inf_B \frac{\lambda_0(B)}{\pi(B^c)}. \tag{2.2.11}$$

*where the infimum is taken over all those $B$'s such that the conditions given in Lemma 2.2.5 are satisfied.*

**Proof:** By the proof of Lemma 2.2.5, for any $B$ satisfying the given conditions, we have (see (2.2.10))

$$1 - \frac{\lambda_0}{Gap} \leq \pi(B).$$

This is the same as

$$Gap(L) \leq \frac{\lambda_0(B)}{\pi(B^c)}.$$

So, we can take the infimum on all those $B$'s in the above inequality. ■

We will use this result to estimate $Gap(L)$ or $\lambda_0(B)$ for some specific processes at the end of this chapter.

The relation between $Gap(L)$ and $\Gamma_{\phi_0} := \Gamma_{\phi_0}(L^B)$ (or $\Gamma_{\rho_0} := \Gamma_{\rho_0}(L^B)$) is also interesting and useful. In the following lemma, we show this relation and a condition for $\lambda_0$ to be an isolated eigenvalue.

**Lemma 2.2.7** *With the assumptions given in Lemma 2.2.5, we have*

*(i)* $\Gamma_{\phi_0} \geq Gap - \lambda_0$, $\Gamma_{\rho_0} \geq Gap - \lambda_0$.

*(ii)* $\lambda_0$ *is an isolated eigenvalue if* $Gap > 2\lambda_0$.

**Proof**: $(i)$ We only give the proof for the first inequality. The proof for the second is the same.

By Definition 2.2.4, we have

$$\Gamma_{\phi_0} = \inf\{(f, L^B f)_\pi : f \in L^2(B, \pi_B), \|f\|_\pi = 1, (f, \phi_0)_\pi = 0\}.$$

Therefore, for any $\epsilon > 0$, there is a $g_0 \in L^2(B, \pi_B)$ such that $||g_0||_\pi = 1$ , $(g_0, \phi_0)_\pi = 0$ and $\Gamma_{\phi_0} = (g_0, L^B g_0)_\pi - \epsilon$. We make the extension of $g_0$ (also denoted by $g_0$) such that it takes 0 values on $B^c$ and therefore belongs to $L^2(S, \pi)$. We also extend $\phi_0$ in the same way. Let $f = g_0 - (g_0, 1)_\pi 1$, then $f \neq 0$.

Otherwise, $g_0 = (g_0, 1)_\pi 1$, and $(g_0, \phi_0)_\pi = (g_0, 1)_\pi(\phi_0, 1)_\pi$. So $(g_0, \phi_0)_\pi = 0$ implies that $(g_0, 1)_\pi = 0$ or $(\phi_0, 1)_\pi = 0$. But $(\phi_0, 1)_\pi^2 = (\phi_0, 1_B)_\pi^2 \geq 1 - \lambda_0/Gap > 0$, by Lemma 2.2.5 and the assumption that $\lambda_0 < Gap$, and $(g_0, 1)_\pi = 0$ implies $g_0 \equiv 0$ which is impossible because $||g_0||_\pi = 1$.

Next, $(f,f)_\pi = ||g_0||_\pi^2 - (g_0,\mathbf{1})_\pi^2 = 1 - (g_0,\mathbf{1})_\pi^2$ and

$$\begin{aligned}(g_0,\mathbf{1})_\pi^2 &= (g_0,\mathbf{1}-(\phi_0,\mathbf{1})_\pi\phi_0)_\pi^2, \qquad (\text{since } (g_0,\phi_0)_\pi = 0)\\ &\le ||g_0||_\pi^2||\mathbf{1}-(\phi_0,\mathbf{1})\phi_0||_\pi^2 = \mathbf{1}^T - (\phi_0,\mathbf{1})_\pi^2\\ &\le \frac{\lambda_0}{Gap} \qquad (\text{by Lemma 2.2.5}).\end{aligned}$$

By the definition of $Gap$ and the fact that $L\mathbf{1} = L^*\mathbf{1} = 0$, we have

$$\begin{aligned}Gap &\le \frac{(f,Lf)_\pi}{(f,f)_\pi} = \frac{(g_0,L^Bg_0)_\pi}{1-(g_0,\mathbf{1})_\pi^2}\\ &\le \frac{\Gamma_{\phi_0}+\epsilon}{1-\lambda_0/Gap}.\end{aligned}$$

Therefore

$$Gap - \lambda_0 \le \Gamma_{\phi_0} + \epsilon.$$

The proof is completed by letting $\epsilon \to 0$ in the above inequality.

*(ii)* If $Gap > 2\lambda_0$, then, according to *(i)*, $\Gamma_{\phi_0} > \lambda_0$ which implies that $\lambda_0$ is isolated from the remained spectrum of $L^B$ since $\sigma(L^B)\backslash\{\lambda_0\} \subset [\Gamma_{\phi_0},\infty)$. ■

## 2.3 Asymptotics of the First Hitting Times

For the convenience of reading the present section, we summarize the notation and definitions used in the previous sections as follows:

- $-L$ : the generator of the process.
- $-L^B$ : the generator of the killed process.
- $-(L^B)^*$: the adjoint operator of $-L^B$.
- $T_t^B = e^{-L^Bt}$.

- $\lambda_0 := \lambda_0(B)$: the smallest eigenvalue of $L^B$.
- $L^B\phi_0 = \lambda_0\phi_0; (L^B)^*\rho_0 = \lambda_0\rho_0$ and $\|\phi_0\|_\pi = 1; (\phi_0, \rho_0)_\pi = 1$.
- $\Gamma_\phi = \inf\{(u, L^B u)_\pi : \|u\|_\pi = 1, (u, \phi)_\pi = 0\}$.
- $Gap(L) = \inf\{(u, Lu)_\pi : u \in L^2(\pi); \|u\|_\pi = 1; (u, 1)_\pi = 0\}$.
- $W(A) = \{(f, Af) : f \in \mathcal{H}, \|f\| = 1\}$: the numerical range of a bounded linear operator $A$ on Hilbert space $\mathcal{H}$.
- $1_B(x) = 1, x \in B; = 0,$ otherwise.

The following proposition is a variant of a part of Theorem 2.3 of Liggett (1989); the proof is essentially the same.

**Proposition 2.3.1** *(Iscoe and McDonald (1992) Proposition 4.4) Let $(T_t; t \geq 0)$ be a strongly continuous semigroup on a complex Hilbert space $H$ with bounded infinitesimal generator $-A$. Then for all $f \in H$*

$$\|T_t f\| \leq e^{-w(A)t}\|f\|, \qquad t \geq 0,$$

*where $w(A) := \inf \Re e W(A)$. In particular if $\rho$ is an eigenvector of $A^*$, then for all $f \in H$ such that $f \perp \rho$:*

$$\|T_t f\| \leq e^{-\Gamma_\rho t}\|f\|. \qquad t \geq 0.$$

**Proof:** Let $f \in H$ such that $\|f\| = 1$; w.l.o.g. assume $\|T_t f\| \neq 0$. Then

$$\frac{d}{dt}\|T_t f\|^2 = -2(T_t f, AT_t f) \leq -2w(A) \cdot \|T_t f\|^2.$$

Therefore $\|T_t f\|^2 \leq e^{-2w(A)\cdot t}\|f\|^2$. For the second part, simply apply the first part to $\mathcal{H} \equiv \rho^\perp$ and $\bar{A} = A|\mathcal{H} : \mathcal{H} \to \mathcal{H}$ (since if $(f, \rho) = 0$ then $(Af, \rho) = (f, A^*\rho) =$

$\lambda(f,\rho) = 0$, if $A^*\rho = \lambda\rho$; similarly $T_t(\mathcal{H}) \subset \mathcal{H}$ since $T_t^*\rho = e^{At}\rho$), recalling that $\Gamma_\rho = w(\tilde{A})$. ∎

It was noticed (Iscoe and McDonald (1992) Remark 4.5) that Proposition 2.3.1 remains valid on a real Hilbert space. In this case, $w(A)$ and $\Gamma_\rho(A)$ can be calculated as infimums using real vectors given that $\rho$ is real. This proposition also remains valid (with the same proof) in the case that $A$ is an unbounded operator and operates on $f$ which belongs to the domain $\mathcal{D}(A)$.

We are now able to present the main results about the exponential approximation of the distribution of the first hitting time $\tau$.

**Theorem 2.3.2** *Assume all notation and definitions given at the beginning of this section. Then*

$$\left|P_\pi(\tau > t) - c(B)e^{-\lambda_0 t}\right| \leq \|\mathbf{1}_B - (\rho_0, \mathbf{1}_B)_\pi \phi_0\|_\pi \|\mathbf{1}_B - \frac{(\rho_0, \mathbf{1}_B)_\pi}{\|\rho_0\|_\pi^2}\rho_0\|_\pi e^{-\Gamma_{\rho_0} t}$$

*where* $c(B) := (\rho_0, \mathbf{1}_B)_\pi(\phi_0, \mathbf{1}_B)_\pi$.

**Proof**: For any arbitrary constant $\alpha$,

$$\begin{aligned} P_\pi(\tau > t) &= \int_B T_t^B \mathbf{1}_B \pi(dx) = (\mathbf{1}_B, T_t^B \mathbf{1}_B)_\pi \\ &= (\mathbf{1}_B, T_t^B(\mathbf{1}_B - (\rho_0, \mathbf{1})_\pi\phi_0))_\pi + (\mathbf{1}_B, T_t^B(\rho_0, \mathbf{1})_\pi\phi_0)_\pi \\ &= (\mathbf{1}_B - \alpha\rho_0, T_t^B(\mathbf{1}_B - (\rho_0, \mathbf{1}_B)_\pi\phi_0))_\pi + (\rho_0, \mathbf{1}_B)_\pi(\phi_0, \mathbf{1}_B)_\pi e^{-\lambda_0 t}. \end{aligned}$$

Here we have used the fact that $\alpha\rho_0$ is an eigenvector of $(T_t^B)^*$ and is orthogonal to $\mathbf{1}_B - (\rho_0, \mathbf{1}_B)_\pi\phi_0$ since $(\rho_0, \phi_0)_\pi = 1$. Using this fact and Proposition 2.3.1, we get

$$\left|P_\pi(\tau > t) - c(B)e^{-\lambda_0 t}\right| = \left|P_\pi(\tau > t) - (\rho_0, \mathbf{1}_B)_\pi(\phi_0, \mathbf{1}_B)_\pi e^{-\lambda_0 t}\right|$$

$$= \left|(\mathbf{1}_B - \alpha\rho_0, T_t^B(\mathbf{1}_B - (\rho_0, \mathbf{1}_B)_\pi \phi_0))_\pi\right|$$

$$\leq ||\mathbf{1}_B - \alpha\rho_0||_\pi ||\mathbf{1}_B - (\rho_0, \mathbf{1}_B)_\pi \phi_0||_\pi e^{-\Gamma_{\rho_0} t}.$$

The theorem is then proved by minimizing $||\mathbf{1}_B - \alpha\rho_0||_\pi$ via a standard argument.

Let $f(\alpha) = ||\mathbf{1}_B - \alpha\rho_0||_\pi^2$, then

$$f'(\alpha) = 2\sum_{x\in B}(1 - \alpha\rho_0(x))\rho_0(x)\pi(x)$$

and

$$f'(\alpha) = 0 \Longrightarrow \alpha = \frac{(\rho_0, \mathbf{1}_B)_\pi}{||\rho_0||_\pi^2}.$$

■

The following corollary provides error bounds for the exponential approximation and lower and upper bounds for $P_\pi(\tau > t)$ and $E_\pi\tau$ in terms of simple parameters which essentially do not require information about eigenvectors.

**Corollary 2.3.3** *As $B$ — $S$.*

$$\left|P_\pi(\tau > t) - c(B)e^{-\lambda_0 t}\right| \leq \frac{\lambda_0}{Gap}(1 + ||\rho_0 - \phi_0||_\pi)e^{-\Gamma_{\rho_0} t}$$

$$\leq \left(\frac{\lambda_0}{Gap} + o(\lambda_0)\right) e^{-(Gap - \lambda_0)t} \tag{2.3.12}$$

*where $c(B) := (\rho_0, \mathbf{1}_B)_\pi(\phi_0, \mathbf{1}_B)_\pi$ and $o(\lambda_0) := \lambda_0||\rho_0 - \phi_0||_\pi/Gap$. Moreover, we have the following inequalities:*

$$(i) \qquad P_\pi(\tau > t) \geq \left(1 - \frac{\lambda_0}{Gap}\right) e^{-\lambda_0 t} - \left(\frac{\lambda_0}{Gap} + o(\lambda_0)\right) e^{-(Gap - \lambda_0)t}$$

$$(ii) \qquad P_\pi(\tau > t) \;\le\; \|\rho_0\|_\pi e^{-\lambda_0 t} + \left(\frac{\lambda_0}{Gap} + o(\lambda_0)\right) e^{-(Gap-\lambda_0)t}$$

$$(iii) \qquad \frac{1}{\lambda_0} \;-\; \frac{1}{Gap} - \mathcal{O}(\lambda_0) \le E_\pi \tau \le \frac{\|\rho_0\|_\pi}{\lambda_0} + \mathcal{O}(\lambda_0)$$

*where* $\mathcal{O}(\lambda_0) := \lambda_0(1 + \|\rho_0 - \phi_0\|_\pi)/Gap(Gap - \lambda_0)$.

**Proof**: First, by Lemma 2.2.5

$$\begin{aligned}
\|\mathbf{1}_B - \frac{(\rho_0, \mathbf{1}_B)_\pi}{\|\rho_0\|_\pi^2}\rho_0\|_\pi^2 &= (\mathbf{1}_B, \mathbf{1}_B)_\pi - \frac{(\rho_0, \mathbf{1}_B)_\pi^2}{\|\rho_0\|_\pi^2} \\
&\le 1 - \frac{(\rho_0, \mathbf{1}_B)_\pi^2}{\|\rho_0\|_\pi^2} \\
&\le \frac{\lambda_0}{Gap}.
\end{aligned}$$

Secondly, since $(\rho_0 - \phi_0, \phi_0)_\pi = 0$,

$$\begin{aligned}
\|\mathbf{1}_B - (\rho_0, \mathbf{1}_B)_\pi \phi_0\|_\pi &= \|\mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0 + (\phi_0, \mathbf{1}_B)_\pi \phi_0 - (\rho_0, \mathbf{1}_B)_\pi \phi_0\|_\pi \\
&\le \|\mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0\|_\pi + \|\phi_0\|_\pi \|(\rho_0 - \phi_0, \mathbf{1}_B)_\pi\| \\
&\le \left(\frac{\lambda_0}{Gap}\right)^{1/2} + \|(\rho_0 - \phi_0, \mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0)_\pi\| \\
&\le \left(\frac{\lambda_0}{Gap}\right)^{1/2} + \left(\frac{\lambda_0}{Gap}\right)^{1/2} \|\rho_0 - \phi_0\|_\pi.
\end{aligned}$$

By Lemma 2.2.7, $\Gamma_{\rho_0} \ge Gap - \lambda_0$, we get (2.3.12).

$(i)$ and $(ii)$ are proved by using (2.3.12) and Lemma 2.2.5 which gives the inequality:

$$1 - \frac{\lambda_1}{Gap} \leq c(B) \leq \pi(B)\|\rho_1\|_\pi.$$

Taking integrals over $(i)$ and $(ii)$ from 0 to $\infty$, we get $(iii)$. ■

**Remark.** By Theorem 2.2.3, $\|\phi_0 - \mathbf{1}_B\|_\pi \to 0$ and $\|\rho_0 - \mathbf{1}_B\|_\pi \to 0$ as $B \to S$ given that $\kappa_1, \kappa_2 \to 0$. So, under the latter condition, we have $\|\rho_0 - \phi_0\|_\pi \to 0$ as $B \to S$ by the triangle inequality. This explains the use of the notation of $o(\lambda_0)$. Since $o(\lambda_0)$ is smaller in order than $\lambda_0$, it is negligible from a practical point of view. A sufficient condition for $\kappa_1, \kappa_2 \to 0$ as $B \to S$ is that the full infinitesimal generator $-L$ is bounded (see Lemma 2.2.2). In some other cases, we can directly calculate $\kappa_1$ and $\kappa_2$ and show that they tend to 0 as $B \to S$. ■

The following theorem is a natural generalization of Theorem 2.3.2.

**Theorem 2.3.4** *For any $f, g \in L^2(\pi)$*

$$\left|(g, T_t^B f)_\pi - (\rho_0, f)_\pi(\phi_0, g)_\pi e^{-\lambda_0 t}\right| \leq \|f - (\rho_0, f)_\pi\phi_0\|_\pi \|g - \frac{(\rho_0, g)_\pi}{\|\rho_0\|_\pi^2}\rho_0\|_\pi e^{-\Gamma_{\rho_0} t}.$$

**Proof**: Noticing that $(f - (\rho_0, f)_\pi\phi_0, \rho_0)_\pi = 0$, we have

$$\begin{aligned}
\left|(g, T_t^B f)_\pi - (\rho_0, f)_\pi(\phi_0, g)_\pi e^{-\lambda_0 t}\right| &= \left|(g, T_t^B(f - (\rho_0, f)_\pi\phi_0))_\pi\right| \\
&= \left|(g - c\rho_0, T_t^B(f - (\rho_0, f)_\pi\phi_0))_\pi\right| \\
&\leq \|g - c\rho_0\|_\pi \|T_t^B(f - (\rho_0, f)_\pi\phi_0)\|_\pi \\
&\leq \|g - c\rho_0\|_\pi \|f - (\rho_0, f)_\pi\phi_0\|_\pi e^{-\Gamma_{\rho_0} t}.
\end{aligned}$$

The theorem is proved by minimizing $||g - c\rho_0||_\pi$ in $c$. ■

With this generalization we can give estimates for some other interesting parameters. For example, if we take $g = \mathbf{1}_{\{i\}}$ and $f = \mathbf{1}_{\{j\}}$ with $i, j \in B$, then Theorem 2.3.4 yields

$$\left|P_i(X_t = j, \tau > t) - \phi_0(i)\rho_0(j)\pi(j)e^{-\lambda_0 t}\right|$$
$$\leq \|\mathbf{1}_{\{j\}} - \rho_0(j)\pi(j)\phi_0\|_\pi \|\mathbf{1}_{\{i\}} - \frac{\rho_0(i)\pi(i)}{\|\rho_0\|_\pi^2}\rho_0\|_\pi \frac{e^{-\Gamma_{\rho_0} t}}{\pi(i)}.$$

Letting $t \to \infty$, we get the *exponential tail property* of $\tau$:

$$P(X_t = j, \tau_B > t | X_0 = i) \sim \phi_0(i)\rho_0(j)\pi(j)e^{-\lambda_0(B)t}, \qquad as \qquad t \to \infty.$$

This result can be compared with Theorem 1.1.2 which gives the discrete-time version of the same property.

When the Markov jump process $X_t$ is reversible, the lower and upper bounds given in Corollary 2.3.3 simplify. In the following discussion, we assume reversibility of $X_t$ and give extensions and improvements of some of the results given by Aldous and Brown (1991).

As $L^B$ is a bounded selfadjoint operator in $L^2(B, \pi_B)$, the principal eigenvectors $\phi_0$ and $\rho_0$ (corresponding to $\lambda_0$) are equal. So immediately we have, by Corollary 2.3.12,

$$\left|P_\pi(\tau > t) - (\phi_0, \mathbf{1}_B)_\pi^2 e^{-\lambda_0 t}\right| \leq \frac{\lambda_0}{Gap} e^{-\Gamma_{\phi_0} t}. \tag{2.3.13}$$

In addition, $L^B$ is also positive in the sense that $(f, L^B f)_\pi \geq 0$ for any $f \in L^2(\pi)$ since its spectrum is contained in the positive half of the real axis (See Kreyszig (1978) for more discussion on positive operators). So

$$P_\pi(\tau > t) = \int_B T_t^B \mathbf{1}_B \pi(dx) = (\mathbf{1}_B, T_t^B \mathbf{1}_B)_\pi$$

$$= (\mathbf{1}_B, T_t^B(\mathbf{1}_B - (\phi_0, \mathbf{1})_\pi \phi_0))_\pi + (\mathbf{1}_B, T_t^B(\phi_0, \mathbf{1})_\pi \phi_0)_\pi$$

$$= (\mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0, T_t^B(\mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0))_\pi + (\phi_0, \mathbf{1}_B)_\pi^2 e^{-\lambda_0 t}.$$

Here we have used the fact that $\phi_0$ is an eigenvector of the selfadjoint operator $T_t^B$ and is orthogonal to $\mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0$ since $||\phi_0||_\pi = 1$. As $T_t^B := e^{-L^B t}$ is positive, we have

$$(\mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0, T_t^B(\mathbf{1}_B - (\phi_0, \mathbf{1}_B)_\pi \phi_0))_\pi \geq 0.$$

This means

$$P_\pi(\tau > t) \geq (\phi_0, \mathbf{1}_B)_\pi^2 e^{-\lambda_0 t}.$$

On the other hand, it is clear that $P_\pi(\tau > t) = (\mathbf{1}_B, T_t^B \mathbf{1}_B)_\pi \leq ||\mathbf{1}_B||_\pi e^{-\lambda_0 t} \leq e^{-\lambda_0 t}$. Using the lower bound for $(\phi_0, \mathbf{1}_B)_\pi^2$ given in Lemma 2.2.5, we get

**Theorem 2.3.5** *If $X_t$ is reversible and $\pi(B^c)$ is sufficiently small, then for all $t > 0$*

$$(i) \quad \left| P_\pi(\tau > t) - (\phi_0, \mathbf{1}_B)_\pi^2 e^{-\lambda_0 t} \right| \leq \frac{\lambda_0}{Gap} e^{-\Gamma_{\phi_0} t}.$$

$$(ii) \quad \left(1 - \frac{\lambda_0}{Gap}\right) e^{-\lambda_0 t} \leq P_\pi(\tau > t) \leq e^{-\lambda_0 t}.$$

$$(iii) \quad \frac{1}{\lambda_0} - \frac{1}{Gap} \leq E_\pi \tau \leq \frac{1}{\lambda_0}.$$

Here $(iii)$ is obtained by taking integrals from 0 to $\infty$ over the inequalities in $(ii)$.

The following result is an improvement of Theorem 1.1.3 by Aldous and Brown (1991).

**Corollary 2.3.6** *If $X_t$ is reversible and $\pi(B^c)$ is sufficiently small, then for all $t > 0$*

$$\left| P_\pi(\frac{\tau}{E_\pi \tau} > t) - e^{-t} \right| \leq \max\{\frac{\lambda_0}{Gap}, \exp\{\frac{\lambda_0 t}{Gap}\} - 1\} e^{-t}.$$

**Proof**: Substitute $t$ by $tE_\pi\tau$ in the inequalities in $(ii)$ of

Theorem 2.3.5 we get

$$e^{-\lambda_0 t E_\pi \tau} \geq P_\pi(\frac{\tau}{E_\pi \tau} > t) \geq \left(1 - \frac{\lambda_0}{Gap}\right) e^{-\lambda_0 t E_\pi \tau}.$$

Using the inequalities for the mean exit time given by $(iii)$ in

Theorem 2.3.5,

$$1 - \frac{\lambda_0}{Gap} \leq \lambda_0 E_\pi \tau \leq 1,$$

we obtain

$$\exp\{-(1 - \frac{\lambda_0}{Gap})t\} \geq P_\pi(\frac{\tau}{E_\pi \tau} > t) \geq \left(1 - \frac{\lambda_0}{Gap}\right) e^{-t}.$$

Therefore

$$\left(\exp\{\frac{\lambda_0}{Gap} t\} - 1\right) e^{-t} \geq P_\pi(\frac{\tau}{E_\pi \tau} > t) - e^{-t} \geq -\frac{\lambda_0}{Gap} e^{-t}.$$

This proves the corollary. ∎

## 2.4 $Gap(L)$ and $\lambda_0(B)$ via Examples

Any time one has a Markov process with a finite invariant measure $\pi$, a natural problem is to determine rates of convergence to equilibrium. It is of particular interest to determine when this convergence occurs exponentially rapidly. The precise form of the solution to such a problem depends on the nature of the Markov process. One can have exponential convergence in the uniform norm, or in $L^2(\pi)$ norm, for example. Let $T_t$ and $-L$ be the semigroup and generator of the process. We say

that the process converges exponentially in the uniform norm if there are positive constants $C$ and $\epsilon$ so that for each function $f$ in a sufficiently rich class, there is a constant $B(f)$ so that

$$\sup_{x,y} |T_t f(x) - T_t f(y)| \leq CB(f)e^{-\epsilon t}$$

for all $t \geq 0$. The process converges exponentially in $L^2(\pi)$ norm if there is a positive $\epsilon$ so that for all $f \in L^2(\pi)$,

$$||T_t f - \int f d\pi|| \leq e^{-\epsilon t}||f - \int f d\pi||,$$

where $||\cdot||$ denotes the $L^2(\pi)$ norm. The largest $\epsilon$ with this (latter) property is called $Gap(L)$. Thus by definition, exponential $L^2$ convergence occurs if and only if $Gap(L) > 0$. So, it is of essential importance to know the value of $Gap(L)$ when we talk about $L^2$ convergence.

In this section, we collect some elementary results about $Gap(L)$ and $\lambda_0(B)$. As examples, we calculate or estimate the $Gap(L)$'s and $\lambda_0(B)$'s for the $M/M/1$ and $M/M/\infty$ queues. These techniques and results will also be used in the application to the ATM multiplexor model in the later chapters.

Liggett (1989) (Theorem 2.3) proved that the definition of $Gap(L)$ given above is coincident with the one given by Definition 2.2.1 in the

previous sections, i.e.

$$Gap(L) := \inf\{(f, Lf)_\pi : f \in L^2(S, \pi), ||f||_\pi = 1, (f, 1)_\pi = 0\}.$$

Since $Gap(L)$ is an infimum, upper bounds for it are relatively easy to obtain by using special choices of the function $f$. For example, in the case of a positive recurrent continuous-time Markov chain on $Z^+ = \{0, 1, 2, \cdots\}$ with no instantaneous states and with transition rates $q(x, y)$ for $x \neq y$, Liggett (1989) gives

**Theorem 2.4.1** *(Liggett (1989) Theorem 3.4)*

$$Gap(L) \leq \frac{1}{2} \inf_{n \geq 0} \frac{\sum_{x \leq n < y} |\pi(x)q(x,y) + \pi(y)q(y,x)|}{\sum_{x \leq n} \pi(x) \sum_{x > n} \pi(x)}$$

*where $\pi$ is the stationary distribution of the process.*

This result is easy to use but is coarse in general. A similar upper bound and a lower bound are also given by Lawler and Sokal (1988) in terms of Cheeger's constant, which represents the rate of probability flow from a set $A$ to its complement $A^c$ normalized by the stationary distribution $\pi$. The precise definition of Cheeger's constant $k$ for a positive recurrent continuous-time Markov jump process with generator $-L$ and transition rate kernel $J(x, dy)$ is

**Definition 2.4.2**

$$k := \inf_A k(A)$$

*with*

$$k(A) := \frac{(\mathbf{1_A}, L\mathbf{1_A})_\pi}{\pi(A)\pi(A^c)},$$

*where the infimum is taken over all measurable set $A$ such that $0 < \pi(A) < 1$.*

**Theorem 2.4.3** *(Lawler and Sokal (1988) Theorem 2.1) Let $L$ be a bounded self-adjoint operator on $L^2(\pi)$ such that*

$$\pi - ess \sup_x J(x, \{x\}^c) \leq M < \infty.$$

*Define*

$$\kappa := \inf_c \sup \frac{(E|(X+c)^2 - (Y+c)^2|)^2}{E[(X+c)^2]},$$

*where the infimum is taken over all distributions of i.i.d. real-valued random variables $(X, Y)$ with variance 1. Then*

$$\kappa k^2/8M \leq Gap(L) \leq k.$$

**Remark:** The above result is given for reversible processes, but by the definition of $Gap(L)$ we can easily see that:

$$Gap(L) = Gap((L + L^*)/2),$$

where $L^*$ is the adjoint operator of $L$. Since $(L + L^*)/2$ is always selfadjoint, this theorem can be used for non-reversible processes.

There is a useful result regarding $Gap(L)$ of a multi-dimensional Markov process which is also provided by Liggett (1989). Let $-L$ be the generator of a vector Markov process whose components are independent Markov processes with generators $-L_k$ and stationary distribution $\pi_k$. Assume that the stationary distribution $\pi$ of the vector process is the product of the $\pi_k$'s, then

**Theorem 2.4.4** *(Liggett (1989) Theorem 2.6)*

$$Gap(L) = \inf_k Gap(L_k).$$

The following theorem is provided by Lawler and Sokal (1988) which shows an interesting relation between $Gap(L)$ and the principal eigenvalue $\lambda_0(B)$ of $L^B$ corresponding to the process killed outside $B$.

**Theorem 2.4.5** *(Lawler and Sokal (1988) Proposition 3.3) Let $L$ be a bounded selfadjoint operator on $L^2(\pi)$. Then*

$$Gap(L) \geq \inf_B max[\lambda_0(B), \lambda_0(B^c)]. \tag{2.4.14}$$

We should compare the above theorem with Corollary 2.2.6 which shows, for both selfadjoint and nonselfadjoint cases:

$$Gap(L) \leq \inf_B \frac{\lambda_0(B)}{\pi(B^c)}, \tag{2.4.15}$$

where the infimum is taken over all those $B$'s such that the conditions given in Lemma 2.2.5 are satisfied. These results show that $\lambda_0(B)$ plays a role in estimating $Gap(L)$ and vice versa.

Determining the asymptotic behavior of $\lambda_0(B)$ is important since it is equivalent to determining that of the mean first hitting time $E_\pi\tau$. The following simple formula is observed:

**Proposition 2.4.6** *Let $-L^B$ be a bounded infinitesimal generator on $L^2(B,\pi)$ and $\rho_0$ be the nonnegative eigenvector of $(L^B)^*$ corresponding to the principal eigenvalue $\lambda_0(B)$ such that $(\rho_0, \mathbf{1}_B)_\pi = 1$. Then*

$$\lambda_0(B) = (\rho_0, L^B\mathbf{1}_B)_\pi = (\rho_0, K^B)_\pi, \tag{2.4.16}$$

*where $K^B(x) := \mathbf{1}_B(x)J(x, B^c)$ is the killing rate.*

**Proof:** By definition.

$$\begin{aligned}
\lambda_0(B) &= \frac{((L^B)^*\rho_0, \mathbf{1}_B)_\pi}{(\rho_0, \mathbf{1}_B)_\pi} \\
&= (\rho_0, L^B\mathbf{1}_B)_\pi \qquad (\text{since } (\rho_0, \mathbf{1}_B)_\pi = 1) \\
&= (\rho_0, K^B)_\pi. \qquad (\text{since } L^B\mathbf{1}_B = K^B) \ \blacksquare
\end{aligned}$$

This result is especially useful if the process has no long-range interaction, e.g. the birth and death process. In this case, the killing rates are zero except at the boundary of $B$ where they are just the birth rates. So this proposition says that $\lambda_0$ very much depends on the values of the stationary distribution and the eigenvector at the boundary. Therefore, estimating the boundary behavior of those values will help in getting information about $\lambda_0$. We show here by two concrete examples how this idea is to be used. More sophisticated applications can be found in Chapter 4.

**Example 1. The $M/M/1$ queue.** This is the continuous-time Markov chain $X_t$ on states $\{0, 1, 2, \cdots\}$ with transition rates

$$i \rightarrow i+1, \text{ (birth) rate } a,$$

$$i \rightarrow i-1, (i \geq 1) \text{ (death) rate } b(> a).$$

The stationary distribution $\pi$ has the form

$$\pi(i) = \left(1 - \frac{a}{b}\right)\left(\frac{a}{b}\right)^i.$$

$$\pi[i, \infty) = \left(\frac{a}{b}\right)^i.$$

For $\ell$ large, consider the rare set $B^c := [\ell, \infty)$ and

the first hitting time

$$\tau_\ell := \inf\{t : X_t \geq \ell\}.$$

Let $-L^\ell$ be the generator matrix of the process killed at $\ell$, and $\lambda_0(\ell)$ be the principal eigenvalue of $L^\ell$ corresponding to the eigenvector $\phi^\ell$ (which is nonnegative by the Perron-Frobenius theorem). Let $\phi^\ell$ be normalized so that $(\phi^\ell, \mathbf{1}_B)_\pi = 1$. Then

$$\begin{cases} a\phi^\ell(0) - a\phi^\ell(1) = \lambda_0(\ell)\phi^\ell(0) & \text{equation(0)} \\ -b\phi^\ell(k-1) + (a+b)\phi^\ell(k) - a\phi^\ell(k+1) = \lambda_0(\ell)\phi^\ell(k) & \text{equation(k)} \\ -b\phi^\ell(\ell-2) + (a+b)\phi^\ell(\ell-1) = \lambda_0(\ell)\phi^\ell(\ell-1) & \text{equation}(\ell\text{-}1) \end{cases}$$

where $1 \leq k \leq \ell - 2$. Define the generating function of $\phi^\ell$ as $\Phi(s) := \sum_{k=0}^{\ell-1} \phi^\ell(k)s^k$. For each $k \in \{0, 1, \cdots, \ell-1\}$, multiply $s^k$ on both sides of equation($k$), sum over all and we get

$$(a + b - \lambda_0(\ell))\Phi(s) - b\phi^\ell(0) = \frac{a}{s}[\Phi(s) - \Pi^{1/2}hi^\ell(0)] + bs[\Phi(s) - \phi^\ell(\ell-1)s^{\ell-1}].$$

After some rearrangement, we end up with

$$(bs^2 - (a + b - \lambda_0(\ell))s + a)\Phi(s) = b\phi^\ell(\ell - 1)s^{\ell+1} - b\phi^\ell(0)s + a\phi^\ell(0).$$

Let $s = 1$ in the above equation and we obtain

$$\sum_{k=0}^{\ell-1} \phi^\ell(k) = \Phi(1) = \frac{\phi^\ell(\ell - 1) - (1 - a/b)\phi(0)}{\lambda_0(\ell)/b}. \tag{2.4.17}$$

Now, since $\lambda_0(\ell) > 0$ (by the irreducibility of the process) and $\phi^\ell$ is nonnegative, we conclude from (2.4.17) that

$$\phi^\ell(\ell - 1) \geq \left(1 - \frac{a}{b}\right)\phi^\ell(0).$$

Then, by Proposition 2.4.6 and the fact that the killing occurs only at $\ell - 1$,

$$\lambda_0(\ell) = a\phi^\ell(\ell - 1)\pi(\ell - 1) \sim O\left(\left(\frac{a}{b}\right)^{\ell-1}\right).$$

Here we also used the fact that $\phi^\ell(0) > 1$ since $\phi^\ell(k)$ is a strictly decreasing function of $k$ (by Proposition 6.1 in Iscoe and McDonald (1991)) and $\sum_{k=0}^{\ell-1}\phi^\ell(k)\pi(k) = 1$. It is easy to see that the mean killing rate $\overline{\kappa} = a\pi(\ell - 1)/\pi[0, \ell) \to 0$ and

$$\kappa_1^2 = [\sum_{x=0}^{\ell-2} \overline{\kappa}^2\pi(x) + (a - \overline{\kappa})^2\pi(\ell - 1)]/\pi[0, \ell) \to 0,$$

as $\ell \to \infty$. Applying Theorem 2.2.3 we get $\|\phi^\ell - 1_B\|_\pi \to 0$ and therefore $\phi^\ell(0) \to 1$ as $\ell \to \infty$. Then, again by the decreasing property of $\phi^\ell$, we have $\sum_{k=0}^{\ell-1}\phi^\ell(k) \leq \ell\phi^\ell(0)$, so

$$\begin{aligned}\phi^\ell(\ell - 1) - \left(1 - \frac{a}{b}\right)\phi^\ell(0) &\leq \ell\phi^\ell(0)\frac{\lambda_0(\ell)}{b} \qquad \text{(by (2.4.17))} \\ &\sim \ell\phi^\ell(0)O\left(\left(\frac{a}{b}\right)^{\ell-1}\right) \to 0, \text{ as } \ell \to \infty.\end{aligned}$$

We then conclude that

$$\phi^\ell(\ell - 1) \to 1 - \frac{a}{b}, \text{ as } \ell \to \infty.$$

This yields the following asymptotic result about $\lambda_0(\ell)$:

**Theorem 2.4.7** *As* $\ell \to \infty$,

$$\lambda_0(\ell) \sim a\left(1-\frac{a}{b}\right)\pi(\ell-1) = b\left(1-\frac{a}{b}\right)^2\left(\frac{a}{b}\right)^{\ell} \equiv \hat{\lambda}(\ell). \tag{2.4.18}$$

Applying the theorem of asymptotics of the first hitting time to this $M/M/1$ queue, we can get

$$\sup_t |P_\pi(\tau_\ell > t) - \exp(-\hat{\lambda}(\ell)t)| \to 0, \text{as } \ell \to \infty.$$

This is to be compared with a similar result provided by A1 in Aldous (1989).

In general, it is not easy to obtain the infimum in the upper bound for $Gap(L)$ given by Corollary 2.2.6 (see (2.4.15)) because $\lambda_0(B)$ as a function of $B$ is not given in an analytic form, e.g. $\lambda_0(\ell)$ in the $M/M/1$ queue as a function of $\ell$. The asymptotics of $\lambda_0(\ell)$, on the other hand, do provide a closed form formula so that we can give the following upper bound of $Gap(L)$ for the $M/M/1$ queue:

$$Gap(L) \le \inf_\ell \frac{\lambda_0(\ell)}{\pi[\ell,\infty)} \le \left(1-\frac{a}{b}\right)^2 b.$$

Although this is not the infimum, it is better than the upper bound $b-a$ provided by Theorem 2.4.1 that is derived as follows:

$$\begin{aligned} Gap(L) &\le \frac{1}{2}\inf_{n\ge 0}\frac{\sum_{x\le n<y}[\pi(x)q(x,y)+\pi(y)q(y,x)]}{\sum_{x\le n}\pi(x)\sum_{x>n}\pi(x)} \\ &= \inf_{n\ge 0}\frac{\pi(n)a}{\pi[0,n]\cdot\pi[n+1,\infty)} \\ &= \inf_{n\ge 0}\frac{a(1-a/b)(a/b)^n}{\pi[0,n](a/b)^{n+1}} \\ &= \inf_{n\ge 0}\frac{b-a}{\pi[0,n]} \\ &= b-a. \end{aligned}$$

**Example 2. The $M/M/\infty$ queue.** For the purpose of applying the theory developed in Chapter 2 to the ATM multiplexor model, we calculate $Gap(L)$ and

analyze the asymptotics of $\lambda_0(B)$ for the $M/M/\infty$ queue. This is the continuous-time Markov chain $X_t$ on states $\{0, 1, 2, \cdots\}$ with transition rates

$$i \rightarrow i+1, \text{ (birth) rate } a,$$

$$i \rightarrow i-1, (i \geq 1) \text{ (death) rate } bi.$$

Let $\mathcal{D}_0$ denote the set of functions which are constant off a finite subset of $S := \{0, 1, 2, \cdots\}$. Then $\mathcal{D}_0$ forms a core for the infinitesimal generator $-L$ of this Markov chain $X_t$. (a subspace $\mathcal{D}_0$ of $\mathcal{D}(L)$ is a core of $L$ if the closure of the restriction of $L$ to $\mathcal{D}_0$ is equal to $L$. See Ethier and Kurtz (1986)). It is clear that $X_t$ is reversible with respect to the stationary distribution $\pi$ which has the form: $\pi(k) = \exp(-\lambda)\lambda^k/k!$ $(\lambda := a/b)$.

Let $\alpha$ be an eigenvalue of the generator $-L$ and $\phi \in L^2(S, \pi)$ be the corresponding right eigenvector. Then,

$$(\alpha + a + bk)\phi(k) = a\phi(k+1) + bk\phi(k-1). \tag{2.4.19}$$

Define the weighted generating function of $\phi$ as following

$$\Phi(z) := \sum_{k=0}^{\infty} \phi(k) z^k \pi(k).$$

Using the Cauchy-Schwarz inequality,

$$|\Phi(z)|^2 \leq \sum_{k=0}^{\infty} |\phi(k)|^2 \pi(k) \cdot \sum_{k=0}^{\infty} |z|^{2k} e^{-\lambda} \frac{\lambda^k}{k!}.$$

Since $\phi \in L^2(\pi)$ it is clear that $\Phi$ is entire. Multiply $z^k\pi(k)$ on both sides of equation (2.4.19) and sum over $k$ from 0 to $\infty$ we get

$$(\alpha + a)\Phi(z) + bz\Phi'(z) = \frac{a}{\lambda}\Phi'(z) + b\lambda z\Phi(z).$$

This is equivalent to

$$\frac{\Phi'(z)}{\Phi(z)} = \lambda + \frac{\alpha}{b}\frac{1}{1-z}$$

The only solutions to this equation, normalized so that $\Phi(0) = 1$, are

$$\Phi(z) = e^{\lambda z}|1 - z|^{\alpha/b},$$

where $\alpha/b$ must be a non-negative integer. Hence, reversing the above argument, the spectrum of $L$ is $\sigma(L) = \{0, b, 2b, 3b, ...\}$. In particular, we have

**Theorem 2.4.8** *For the $M/M/\infty$ queue,*

$$Gap(L) = b.$$

Let $B := \{0, 1, 2, \cdots, \ell - 1\}$ and $-L^B$ be the generator of the $M/M/\infty$ queue killed outside $B$. Then $L^B$ is bounded.

Since $\mathcal{D}_0$ is a core for the generator $-L$ and $B$ is a finite subset, the zero-extensions of the eigenvectors of $L^B$ and $(L^B)^*$ corresponding to $\lambda_0(B)$ are contained in the domain of $L$. Therefore, some of the results obtained in Chapter 2, especially Lemma 2.2.5 and Proposition 2.4.6, can be applied (see the remark after the proof of Lemma 2.2.5).

We now turn to the asymptotics of the principal eigenvalue $\lambda_0(\ell) := \lambda_0(B)$ with corresponding eigenvector $\phi^\ell$ which is normalized so that $\phi^\ell(0) = 1$. Then

$$\begin{aligned}
\lambda_0(\ell) &= (\phi^\ell, L^B\phi^\ell)_\pi / \|\phi^\ell\|_\pi^2 \\
&= \sum_{k=0}^{\ell-1} \frac{1}{2}\left[(\phi^\ell(k+1) - \phi^\ell(k))^2 a + (\phi^\ell(k-1) - \phi^\ell(k))^2 kb\right] \pi(k) / \|\phi^\ell\|_\pi^2 \\
&= \sum_{k=0}^{\ell-1} (\phi^\ell(k+1) - \phi^\ell(k))^2 a\pi(k) / \|\phi^\ell\|_\pi^2, \quad \text{(by reversibility)} \qquad (2.4.20)
\end{aligned}$$

where $\phi^\ell$ is extended so that $\phi^\ell(\ell) = \phi^\ell(-1) = 0$. Since $\phi^\ell(k)$ is a decreasing function of $k$ (by Proposition 6.1 in Iscoe and McDonald (1991)) and $\phi^\ell(0) = 1$, we know that $\phi^\ell$ is uniformly bounded, i.e. $\phi^\ell(k) \leq 1$ for all $k$ and $\ell$. In addition,

$$\|\phi^\ell\|_\pi^2 = \sum_{k=0}^{\ell-1} (\phi^\ell(k))^2 \pi(k) \leq \pi[0, \ell - 1] < 1.$$

Therefore

$$0 < (\phi^\ell, \mathbf{1}_B)_\pi^2 \le \|\phi^\ell\|_\pi^2 \|\mathbf{1}_B\|_\pi^2 \le 1.$$

On the other hand, by Lemma 2.2.5

$$1 \ge \frac{(\phi^\ell, \mathbf{1}_B)_\pi^2}{\|\phi^\ell\|_\pi^2} \ge 1 - \frac{\lambda_0(\ell)}{Gap(L)} \to 1, \text{ as } \ell \to \infty$$

since $Gap(L) = b$ as given in Theorem 2.4.8 and $\lambda_0(\ell) \le \overline{\kappa} \sim a\pi(\ell-1) \to 0$. Applying Proposition 2.4.6 we get

$$\lambda_0(\ell) = \frac{(\phi^\ell, K_B)_\pi}{(\phi^\ell, \mathbf{1}_B)_\pi} \sim \frac{a\phi^\ell(\ell-1)\pi(\ell-1)}{\|\phi^\ell\|_\pi}, \text{ as } \ell \to \infty.$$

So

$$\frac{\lambda_0(\ell)\|\phi^\ell\|_\pi^2}{a\pi(\ell-1)} \sim \phi^\ell(\ell-1)\|\phi^\ell\|_\pi, \text{ as } \ell \to \infty.$$

By (2.4.20) and $\phi^\ell(\ell-1)\|\phi^\ell\|_\pi \le 1$, we conclude that there is a constant $M$ such that, for $\ell$ large enough,

$$\sum_{k=0}^{\ell-1}(\phi^\ell(k+1) - \phi^\ell(k))^2 \frac{\pi(k)}{\pi(\ell-1)} = \frac{\lambda_0(\ell)\|\phi^\ell\|_\pi^2}{a\pi(\ell-1)} \le M < \infty,$$

and, as $\pi(k)$ is decreasing in $k$ after some large value,

$$\sum_{k=0}^{\ell-2}(\phi^\ell(k+1) - \phi^\ell(k))^2 \left(\frac{\pi(\ell-2)}{\pi(\ell-1)}\right) \le \sum_{k=0}^{\ell-2}(\phi^\ell(k+1) - \phi^\ell(k))^2 \frac{\pi(k)}{\pi(\ell-1)} \le M.$$

Since $\pi(\ell-2)/\pi(\ell-1) = (\ell-1)/\lambda \to \infty$ as $\ell \to \infty$, we have

$$\sum_{k=0}^{\ell-2}(\phi^\ell(k+1) - \phi^\ell(k))^2 \to 0. \qquad \text{as } \ell \to \infty.$$

So, uniformly for $k = 0, 1, \cdots, \ell-2$,

$$|\phi^\ell(k+1) - \phi^\ell(k)| \to 0. \qquad \text{as } \ell \to \infty.$$

Also

$$\sum_{k=0}^{\ell-3}(\phi^\ell(k+1) - \phi^\ell(k))^2 \left(\frac{\pi(\ell-3)}{\pi(\ell-1)}\right) \le \sum_{k=0}^{\ell-3}(\phi^\ell(k+1) - \phi^\ell(k))^2 \frac{\pi(k)}{\pi(\ell-1)} \le M,$$

and

$$\begin{aligned}
(\phi^\ell(\ell-2)-1)^2 &= (\phi^\ell(\ell-2)-\phi^\ell(0))^2 \\
&= \left[\sum_{k=0}^{\ell-3}(\phi^\ell(k+1)-\phi^\ell(k))\right]^2 \\
&\le (\ell-2)\sum_{k=0}^{\ell-3}(\phi^\ell(k+1)-\phi^\ell(k))^2 \text{ (Cauchy-Schwarz)} \\
&\le (\ell-2)M\frac{\pi(\ell-1)}{\pi(\ell-3)} \\
&= M\frac{\lambda^2}{\ell-1} \to 0, \qquad \text{as } \ell \to \infty.
\end{aligned}$$

Finally,

$$|\phi^\ell(\ell-1)-1| \le |\phi^\ell(\ell-1)-\phi^\ell(\ell-2)| + |\phi^\ell(\ell-2)-1| \to 0, \qquad \text{as } \ell \to \infty.$$

We conclude that

$$\lim_{\ell\to\infty}\max_{0\le k<\ell}|\phi^\ell(k)-1| = 0,$$

and therefore $\|\phi^\ell\|_\pi \to 1$ as $\ell \to \infty$. These then yield the asymptotics of $\lambda_0(\ell)$:

**Theorem 2.4.9** $\lambda_0(\ell) \sim a\pi(\ell-1), \qquad as\ \ell \to \infty.$

# Chapter 3

# Capacity of an ATM Multiplexor

## 3.1 The Model

The asynchronous transfer mode (ATM) is currently being considered as the preferred transport method for the broadband integrated services digital network (see Woodruff and Kositpaiboon (1990) for a general overview). ATM is suitable for multimedia traffic because it offers greater flexibility in bandwidth allocation by transmitting information in fixed length packets, called cells, through virtual network connections.

To achieve maximum bandwidth efficiency, bursty traffic is statistically multiplexed. When traffic sources are statistically multiplexed over a common link, the sum of the peak rates of the sources, in cells per second, exceeds the throughput of the link. The excess cells may be stored in a buffer but when this overflows, cells are lost. The results in Li (1989) suggest, moreover, that when transmission rates are high, no practical buffering will prevent the loss of cells when the link rate is exceeded. When a source is bursty, cells are generated at the peak rate only for very short periods of time. Immediately afterwards the source becomes idle and generates no cells. Since the sources are independent, the chance that many sources

transmit simultaneously at the peak rate is small.

We assume that the link rate is $\ell - 1$ cells per second. Further we assume that traffic sources belong to $n$ distinct, independent service categories ( voice, text, video, etc.) and that traffic sources in category $i$ may be described as an alternating series of idle and bursty periods. A burst from a source in category $i$ produces cells at a rate of $d_i$ cells per second. This means that during a burst from source $i$, approximately every $(\ell - 1)/d_i$'th cell leaving the link comes from source $i$. We must say approximately because cells which arrive simultaneously at the link from different sources must be slotted one after the other. This is accomplished by buffering and it results in jitter or a slight delay in the arrival of one cell relative to others in the burst. We assume that bursts of category $i$ arrive according to a Poisson process having a rate of $a_i$ bursts per second. We also assume that the burst periods are independent (and independent of the arrival process) and are exponentially distributed with a mean burst length of $1/b_i$.

The aggregate of the $n$ different source categories represents the total load at the link. In particular, if we let $N_i(t)$ represent the number of bursts from category $i$ sources being multiplexed at the link at time $t$, then the total load at time $t$ may be represented by

$$N(t) := \sum_{i=1}^{n} d_i N_i(t).$$

When the load exceeds the link rate we say the multiplexor is congested. Define

$$\tau = \inf\{t \geq 0 : N(t) \geq \ell\}$$

so $\tau$ is the time until congestion occurs.

Note that $N_i(t)$ is statistically equivalent, up to time $\tau$, to an $M/M/\infty$ queue (which we still denote by $N_i(t)$) with arrival rate $a_i$ and service rate $b_i$ so, assuming each category is in equilibrium, the mean load is $\sum_{i=1}^{n} d_i a_i/b_i$. To characterize $\tau$,

the time until congestion, we describe the traffic at the multiplexor by the Markov process $\vec{N}(t) := (N_1(t), \cdots, N_n(t))$ defined on the state space $S := \{0, 1, 2, \cdots\}^n$. Let $\mathcal{D}_0$ denote those real-valued functions which are constant outside a finite subset of $S$. $\vec{N}$ has infinitesimal generator $-L$ , having $\mathcal{D}_0$ as a core, given at $u \in \mathcal{D}_0$ by

$$\begin{aligned} -Lu(\vec{x}) &= \sum_{i=1}^{n} [(u(\vec{x} + \delta_i) - u(\vec{x}))a_i + (u(\vec{x} - \delta_i) - u(\vec{x}))x_i b_i], \\ \vec{x} &= (x_1, x_2, \cdots, x_n) \in S \end{aligned}$$

where $\delta_i$ is the $i^{th}$ basis vector in $S$ having all its components equal to 0 except the $i^{th}$ which is 1; $\tau$ is the first time the process $\vec{N}(t)$ reaches the bad region $B^c := \{\vec{x} \in S : \sum_i^n d_i x_i \geq \ell\}$.

We first remark that the $N_i$ are independent and each is reversible with respect to the stationary Poisson measure having mean $\lambda_i := a_i/b_i$. Hence $\vec{N}(t)$ is also reversible with respect to the stationary product measure $\pi$ given by

$$\pi(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i}. \tag{3.1.1}$$

The reversibility of $\vec{N}(t)$ with respect to $\pi$ means that for all $1 \leq i \leq n$,

$$b_i x_i \pi(\vec{x}) = a_i \pi(\vec{x} - \delta_i) \text{ if } x_i > 0.$$

Since the equilibrium distribution is known and is given by (3.1.1), the steady-state problem is of no more interest. The problem with which we are concerned here is the transient behavior of the process, especially the distribution and the mean of $\tau$. Let $-L^B$ be the generator of the process $\vec{N}(t)$ killed outside $B$, then, since $B$ here is a finite subset of the state space, $-L^B$ can be represented by a finite matrix which we also denoted by $-L^B$. We know that $-L^B$ is selfadjoint on $L^2(\pi_B)$ so that it is diagonizable. The spectral theory can be applied to calculate the exact distribution of $\tau$, or equivalently $P_\pi(\tau > t)$. But from practical point of view, an

exact solution is either impossible or unnecessary because the state space is usually huge. For the purpose of studying the quality of various estimation methods, we will need to calculate the exact solutions for some small problems. The precise procedure for obtaining exact solutions and many estimating techniques are presented in the following section.

## 3.2 Methods in General

This section is devoted to a general discussion on the methods we will use to get solutions to the first hitting time problems for the ATM multiplexor model introduced in the last section.

1. **The exact solution method.**

**Theorem 3.2.1** *Assume that the dimension of $-L^B$ is $m+1$, then the killed generator $-L^B$ which is selfadjoint in $L^2(\pi_B)$ can be decomposed as*

$$L^B = PDP^{-1}$$

*where $D = diag(\lambda_0, \lambda_1, ..., \lambda_m)$, $P = (p_0, p_1, ..., p_m)$ with $\lambda_i$ the ith smallest eigenvalue of $L^B$ and $p_i$ its corresponding eigenvector such that $\|p_i\|_\pi = 1$ for all $i$. Therefore, for $t > 0$,*

$$P_\pi(\tau > t) = \sum_{i=0}^{m} (\pi_B' p_i)^2 e^{-\lambda_i t}, \tag{3.2.2}$$

$$E_\pi \tau = \sum_{i=0}^{m} (\pi_B' p_i)^2 / \lambda_i, \tag{3.2.3}$$

*where $\pi_B$ is also used to denote the vector which has elements $\pi(x)$, for $x \in B$.*

**Proof**: This is a standard spectral argument. Define the symmetric version of $L^B$ as

$$A = \Pi_B^{1/2} L^B \Pi_B^{-1/2}$$

where $\Pi_B^{1/2}$ is defined as the diagonal matrix with $\sqrt{\pi(x)}$, $x \in B$, on the diagonal. Then there exists an orthonormal matrix $Q$ such that

$$A = QDQ'.$$

So

$$L^B = \Pi_B^{-1/2} A \Pi_B^{1/2} = \Pi_B^{-1/2} Q D Q' \Pi_B^{1/2}.$$

Define $P = \Pi_B^{-1/2} Q$ or $\Pi_B^{1/2} P = Q$, then

$$L^B = PDP^{-1}$$

and

$$P^{-1} = Q^{-1}\Pi_B^{1/2} = Q'\Pi_B^{1/2} = (P'\Pi_B^{1/2})\Pi_B^{1/2} = P'\Pi_B$$

or equivalently

$$P'\Pi_B P = I.$$

From the above equations, we know that $p_i$ is the normalized ($||p_i||_\pi = 1$) eigenvector corresponding to $\lambda_i$. Moreover,

$$\begin{aligned} P_\pi(\tau > t) &= (\mathbf{1}_B, e^{-L^B t}\mathbf{1}_B)_\pi = \pi_B' e^{-PDP^{-1}t}\mathbf{1}_B \\ &= \pi_B' P e^{-Dt} P^{-1}\mathbf{1}_B = \pi_B' P e^{-Dt} P'\pi_B \\ &= \pi_B' \sum_{i=1}^{m} p_i p_i' e^{-\lambda_i t}\pi_B = \sum_{i=0}^{m} (\pi_B' p_i)^2 e^{-\lambda_i t}. \end{aligned}$$

Taking integrals from 0 to $\infty$ over the above equations yields the expression for $E_\pi \tau$. This completes the proof. ■

In order to use the above theorem, we have to calculate all eigenvalues and eigenvectors of $L^B$, which is usually impossible for a fairly large system. But if we know two parameters $\lambda_0(B)$ and $Gap(L)$, we can still get some good lower and

upper bounds for the distribution and the mean of $\tau$. This theory was developed in Chapter 2.

**2. Aldous and Brown's method**

By definition, $B$ is a finite subset of $S$ and $\mathcal{D}_0$ forms a core for the generator $-L$, so the zero-extensions of the eigenvectors of $L^B$ are contained in the domain of $L$. Therefore, we can apply Theorem 2.3.5 to the process $\vec{N}(t) := (N_1(t), \cdots, N_n(t))$ (see a similar argument in **Example 2** at the end of Chapter 2). We may also reduce to a finite state space as in Aldous and Brown (1991) by grouping all the points in $B^c$ to one point and then apply Theorem 2.3.5 to this finite-state Markov chain (see Iscoe, McDonald and Qian (1992a) for details). Next, we show $Gap(L) > 0$.

Since $N_i(t)$'s are independent one from the others and the stationary distribution $\pi$ is of product form, by Theorem 2.4.4, $Gap(L) = \min_{j=1,\cdots,n} Gap(L_j)$ where $-L_j$ is the generator of $N_j(t)$. But $N_j(t)$ is equivalent to an $M/M/\infty$-queue with arrival rate $a_j$ and service rate $b_j$, by the example of the $M/M/\infty$ queue in Chapter 2. $Gap(L_j) = b_j$. So we conclude that

$$Gap(L) = \min_{j=1,\cdots,n} b_j > 0.$$

On the other hand, the mean killing rate $\overline{\kappa}$ is an upper bound on $\lambda_0$ by the Rayleigh-Ritz principle; that is

$$\begin{aligned} \lambda_0 &= \inf\{(u, L^B u)_\pi : \|u\|_\pi = 1\} \\ &\leq (1_B, L^B 1_B)_\pi / \|1_B\|_\pi^2 \\ &\leq \sum_{i=1}^{n} \sum_{\vec{x} \in \partial B} \frac{a_i \pi(\vec{x})}{\pi(B)} \equiv \overline{\kappa} \end{aligned}$$

where $\partial B := \{\vec{x} \in B : \vec{x} + \delta_i \in B^c, \text{for some } i = 1, \cdots, n\}$ is the boundary of $B$. The following is then an immediate result from Theorem 2.3.5 and will be applied to the ATM multiplexor model.

**Corollary 3.2.2** *(Iscoe, McDonald and Qian (1992a) Theorem 1.5)*

*For the first hitting time $\tau$, we have*

$$P_\pi(\tau \le T) \le 1 - (1 - \frac{\overline{\kappa}}{Gap(L)})e^{-\overline{\kappa}T},$$

$$E_\pi\tau \ge \frac{1}{\overline{\kappa}} - \frac{1}{Gap(L)}$$

*where* $Gap(L) = \min\{b_i; i = 1, \cdots, n\}$.

**3. The linear system methods.**

Define $\alpha(\vec{x}) = E_{\vec{x}}\tau$, which represents the mean time to reach the bad region $B^c$ starting at $\vec{x} \in S$. Then it is known (see Dynkin (1965)) that $\alpha$ satisfies

$$\begin{aligned} L\alpha(\vec{x}) &= 1 \quad \text{for } \vec{x} \in B \\ \alpha(\vec{x}) &= 0 \quad \text{for } \vec{x} \in B^c. \end{aligned} \tag{3.2.4}$$

After solving the linear system, we can get $E_\pi\tau$ by the formula:

$$E_\pi\tau = \sum_{\vec{x}\in B} \alpha(\vec{x})\pi(\vec{x}).$$

This linear system can be solved but the number of variables is of the order $\ell^n$, so large systems are intractable.

Similarly define $\kappa_\theta(\vec{x}) = E_{\vec{x}}\exp(-\theta\tau)$ to be the Laplace transform of $\tau$. Then we have (Dynkin (1965)).

$$\begin{aligned} -L\kappa_\theta(\vec{x}) &= \theta\kappa_\theta(\vec{x}) \quad \text{for } \vec{x} \in B \\ \kappa_\theta(\vec{x}) &= 1 \quad \text{for } \vec{x} \in B^c. \end{aligned} \tag{3.2.5}$$

The solutions of the linear system are used to give an upper bound for the distribution of $\tau$ in the following way:

$$P_\pi(\tau \le T) = P_\pi(e^{-\theta\tau} \ge e^{-\theta T})$$

$$\leq e^{\theta T} E_\pi e^{-\theta\tau} \qquad (\text{ by Chebyshev's inequality })$$

$$= e^{\theta T} \sum_{\vec{x}\in B} \kappa_\theta(\vec{x})\pi(\vec{x}).$$

But again this linear system is only tractable for small $n$. In the next section we introduce the so-called *induced Dirichlet forms method* which makes it possible for us to reduce the sizes of the linear systems described above and provides excellent approximations to the parameters we are concerned.

Before we close this section, one more solution to our problem is worthwhile to mention.

**4. Saddle point approximation method**

For larger problems, a saddle point approximation could be tried (see Daniels (1954)). Define $G_\pi(\theta) = E_\pi \exp(-\theta\tau)$ and $\mu_\pi(\theta) := \log(G_\pi(\theta))$. Then

$$P_\pi(\tau \leq T) \approx \frac{1}{\sqrt{2\pi\theta_0^2(\mu_\pi)''(\theta_0)}} \exp(\theta_0 T + \mu_\pi(\theta_0))$$

for $(\mu_\pi)'(\theta_0) = -T$. The existence of $\theta_0$ and its numerical calculation are discussed later.

## 3.3 The Induced Dirichlet Forms Method

For simplicity of notation, we first assume that the rates at which ATM cells are produced by sources in different categories are distinct, i.e. $d_i \neq d_j$ if $i \neq j$. The special case where all $d_i$s are identical will be discussed later in a separate section.

Recall that the $N_i(t)$'s are independent and each is reversible with respect to the stationary Poisson measure having mean $\lambda_i := a_i/b_i$, and $\vec{N}(t)$ is also reversible with respect to the stationary product measure $\pi$ given by

$$\pi(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i}.$$

Define the Dirichlet (zero) form

$$\begin{aligned}
\mathcal{E}(u,u) &:= (u, Lu)_\pi, \text{ for } u \in \mathcal{D}_0 \\
&= \sum_{\vec{x}\in S} u(\vec{x}) \cdot Lu(\vec{x})\pi(\vec{x}) \\
&= \sum_{\vec{x}\in S}\sum_{i=1}^{n} \frac{1}{2}\left[(u(\vec{x}+\delta_i)-u(\vec{x}))^2 a_i + (u(\vec{x}-\delta_i)-u(\vec{x}))^2 x_i b_i\right]\pi(\vec{x}) \\
&= \sum_{\vec{x}\in S}\sum_{i=1}^{n} [u(\vec{x}+\delta_i)-u(\vec{x})]^2 a_i \pi(\vec{x}) \qquad (3.3.6)
\end{aligned}$$

and define

$$\mathcal{E}_\theta(u,u) := \mathcal{E}(u,u) + \theta \sum_{\vec{x}\in S} u(\vec{x})^2\pi(\vec{x}),$$

$$\mathcal{A}(u,u) := \frac{1}{2}\mathcal{E}(u,u) - \sum_{\vec{x}\in S} u(\vec{x})g(\vec{x})\pi(\vec{x}).$$

Let $\mathcal{H}$ be the set of functions defined on $S$ which equal 1 on $B^c$. Since $\vec{N}(t)$ is reversible, the function $\kappa_\theta(\vec{x}) = E_{\vec{x}}\exp(-\theta\tau)$ satisfies a variational principle.

**Theorem 3.3.1** *(Iscoe, McDonald and Qian (1992a) Theorem 1.1)*

*Among $u \in \mathcal{H}$, $\kappa_\theta$ minimizes $\mathcal{E}_\theta(u,u)$. Moreover*

$$\begin{aligned}
Cap_\theta(B^c) &:= \inf\{\mathcal{E}_\theta(u,u); u \in \mathcal{H}\} \\
&= \theta E_\pi \exp(-\theta\tau) \\
&\equiv \theta \sum_{\vec{x}\in S} \kappa_\theta(\vec{x})\pi(\vec{x}).
\end{aligned}$$

**Proof:** For a proof which is valid for a general reversible process see Fukushima (1980): Lemma 3.1.1 and Lemma 4.3.1. For an elementary proof differentiate the form $\mathcal{E}_\theta(u,u)$ at any $u(\cdot)$ and follow Liggett (1985). ∎

Let $\mathcal{K}$ be the convex set of functions defined on $S$ which equal 0 on $B^c$. Again since $\vec{N}(t)$ is reversible, the function $\alpha$ satisfies a variational principle.

**Theorem 3.3.2** *(Iscoe, McDonald and Qian (1992a) Theorem 1.2)*

*Among $u \in \mathcal{K}$, $\alpha$ minimizes $\mathcal{A}(u,u)$. Moreover*

$$\inf\{\mathcal{A}(u,u); u \in \mathcal{K}\} = -\frac{1}{2}\sum_{\vec{x}\in S} \alpha(\vec{x})\pi(\vec{x}) = -\frac{1}{2}E_\pi \tau$$

**Proof:** Note that on $\mathcal{K}$ there exists an $m > 0$ such that $\mathcal{E}(u,u) \geq m$ if $\|u\|_\pi = 1$. For otherwise, if $\mathcal{E}(u,u) = 0$ then, by (3.3.6), necessarily $u$ is constant and hence 0 by the boundary condition. This gives the coercivity of $\mathcal{A}$. To find the minimum, differentiate $\mathcal{A}(u,u)$ at any $u(\cdot)$. ∎

We now map these complicated minimization problems onto simpler ones. Specifically define the map $f$ from $S$ into $\mathcal{R}_+$ by $f(x_1, x_2, \cdots, x_n) = \sum_{j=1}^n d_j x_j$. First $f$ induces a measure $\pi^*$ having (countable) support $S^* \subset \mathcal{R}_+$ defined by

$$\pi^*(r) \equiv \pi(\{\vec{x} : f(\vec{x}) = r\}) = \sum_{\vec{x}:\sum_{j=1}^n d_j x_j = r} \pi(\vec{x})$$

By Corollary 1.12 in Iscoe and McDonald (1990) this map also induces a regular Dirichlet form $\mathcal{E}^*$ on $S^*$. For any function $h \in \mathcal{D}_0^*$, where $\mathcal{D}_0^*$ is the set of real-valued functions defined on $S^*$ which are constant outside a finite subset of $S^*$:

$$\begin{aligned}
\mathcal{E}^*(h,h) &:= \mathcal{E}(h\circ f, h\circ f) \\
&= \frac{1}{2}\sum_{r\in S^*}\sum_{f(\vec{x})=r}\sum_{i=1}^n \left[(h(r+d_i)-h(r))^2 a_i + (h(r-d_i)-h(r))^2 x_i b_i\right]\pi(\vec{x}) \\
&= \frac{1}{2}\sum_{r\in S^*}\sum_{i=1}^n (h(r+d_i)-h(r))^2 a_i \pi^*(r) \\
&\quad + \frac{1}{2}\sum_{r\in S^*}\sum_{i=1}^n (h(r-d_i)-h(r))^2 \left[\sum_{\vec{x}\in S:\sum_{j=1}^n d_j x_j = r} x_i b_i \frac{\pi(\vec{x})}{\pi^*(r)}\right]\pi^*(r) \\
&= \frac{1}{2}\sum_{r\in S^*}\sum_{i=1}^n (h(r+d_i)-h(r))^2 a_i \pi^*(r) \\
&\quad + \frac{1}{2}\sum_{r\in S^*}\sum_{i=1}^n (h(r-d_i)-h(r))^2 \left[\frac{a_i \pi^*(r-d_i)}{\pi^*(r)}\right]\pi^*(r)
\end{aligned}$$

$$(3.3.7) \qquad = \sum_{r\in S^*}\sum_{i=1}^{n}(h(r+d_i)-h(r))^2 a_i\pi^*(r).$$

The form $\mathcal{E}^*$ is associated with a Markov jump process $N^*(t)$ on $S^*$, having stationary measure $\pi^*$ and generator $-L^*$, which jumps from $r \in S^*$ to the right to $r+d_i \in S^*$ with intensity $a_i$ and to the left to $r-d_i$ with intensity $a_i\pi^*(r-d_i)/\pi^*(r)$. It is not in general the same process as $N(t)$ since the latter is not typically Markovian.

Now let $h$ be any function in $\mathcal{H}^*$, those functions on $S^*$ taking the value 1 on the image of the bad region $(B^c)^* \equiv (B^*)^c = f(B^c) \subseteq \{\ell, \ell+1, \cdots\}$. Clearly,

$$\begin{aligned} Cap_\theta(B^c) &= \inf_{u\in\mathcal{H}} \mathcal{E}_\theta(u,u) \\ &\le \inf_{h\in\mathcal{H}^*} \mathcal{E}_\theta(h\circ f, h\circ f) \\ &= \inf_{h\in\mathcal{H}^*}\left[\mathcal{E}^*(h,h) + \theta\sum_{r\in S^*} h(r)^2\pi^*(r)\right]. \end{aligned}$$

Hence, defining

$$\mathcal{E}^*_\theta(h,h) := \mathcal{E}^*(h,h) + \theta\sum_{r\in S^*} h(r)^2\pi^*(r)$$

and

$$Cap^*_\theta((B^*)^c) := \inf_{h\in\mathcal{H}^*} \mathcal{E}^*_\theta(h,h)$$

we have

$$Cap_\theta(B^c) \le Cap^*_\theta((B^*)^c).$$

If we define

$$\tau^* := \inf\{t \ge 0 : N^*(t) \ge \ell\}$$

and we denote by $E_{\pi^*}$ the expectation associated with $N^*$ started with its stationary measure $\pi^*$, then the analogue of Theorem 3.3.1 is valid for the Markov process $N^*(t)$, and we have the following result.

**Proposition 3.3.3** *(Iscoe, McDonald and Qian (1992a) Proposition 1.3)*

*For all* $\theta \geq 0$

$$E_\pi \exp(-\theta\tau) \leq E_{\pi^*} \exp(-\theta\tau^*)$$

*and*

$$E_{\pi^*}\tau^* \leq E_\pi\tau.$$

**Proof:** The second inequality follows from the first by subtracting 1 from both sides, dividing by $\theta$ and letting $\theta$ tend to 0. ■

Using Chebyshev's inequality (as we did in the last section), we immediately have an upper bound on the probability of congestion occurring in a fixed time interval $[0, T]$.

**Corollary 3.3.4** *(Iscoe, McDonald and Qian (1992a) Corollary 1.4)*

*For any* $\theta > 0$:

$$P_\pi(\tau \leq T) \leq T\theta^{-1}e^{\theta}Cap^*_{\theta/T}((B^c)^*).$$

Here the central idea is to estimate any parameter related with the *original* multi-dimensional process by its counterpart associated with the *induced* one-dimensional jump process. All those methods presented in the last section for obtaining the distribution and the mean (exact or estimated) of $\tau$ can be used for obtaining those of $\tau^*$. Those methods not tractable for large multi-dimensional processes might be usable for the relatively small induced processes. For example, applying the *linear system method* to the induced process gives:

$$L^*\alpha^*(r) = 1 \quad \text{for } r \in B^*$$

$$\alpha^*(r) = 0 \quad \text{for } r \in (B^c)^*$$

for $\alpha^*(r) = E_r\tau^*$, and

$$
\begin{aligned}
-L^*\kappa_\theta^*(r) &= \theta\kappa_\theta^*(r) \quad \text{for } r \in B^* \\
\kappa_\theta^*(r) &= 1 \qquad\qquad\quad \text{for } r \in (B^c)^*
\end{aligned}
\tag{3.3.8}
$$

for $\kappa_\theta^*(r) = E_r \exp(-\theta\tau^*)$. From this exact solution, $Cap_\theta^*((B^c)^*)$ and $E_{\pi^*}\tau^*$ follow. This is feasible even if $n$ is arbitrarily large since the above systems are at most of dimension $\ell$.

In the next section, we compare numerically whatever we can get for $\tau$ from those methods given in the previous section with that for $\tau^*$.

## 3.4 Numerical Comparison

For simplicity we usually assume the $d_i$ are integer-valued for otherwise we can round up to the next integer and this has the effect of reducing $\tau$. This is acceptable since we are looking for underestimates of $E_\pi\tau$ and overestimates of $P_\pi(\tau \le T)$.

First we wish to point out that, in real computation, it is unnecessary to generate the huge generator matrix $L^B$ to obtain the induced generator $(L^*)^{B^*}$. What we do need to calculate is the induced distribution $\pi^*$. Unfortunately $\pi^*$ is quite complicated in general. The asymptotic behavior and some other properties are analyzed in the next chapter where the following recursion relation is also shown:

$$
\pi^*(r) = \frac{1}{r}\sum_{j=1}^{n} d_j\lambda_j\pi^*(r-d_j), \quad r \in S^*\backslash\{0\} \tag{3.4.9}
$$

with $\pi^*(0) = \exp(-A)$ where $A := \sum_{i=1}^{n}\lambda_j$. Note that $\pi^*(r-d_j)$ is 0 in the recursion if $r - d_j$ is not in $S^*$. This recursion provides a practical means of calculating $\pi^*$. With it we may evaluate the the jump rates for the induced Markov process having Dirichlet form (3.3.7)

The models we used for the numerical comparison are as follows:

| Model | Queue 1 | | Queue 2 | | Queue 3 | |
|---|---|---|---|---|---|---|
| | $a_1$ | $b_1$ | $a_2$ | $b_2$ | $a_3$ | $b_3$ |
| 1 | .1 | 12 | .2 | 20 | .3 | 30 |
| 2 | .3 | 6 | .2 | 9 | .1 | 10 |
| 3 | .3 | 12 | .2 | 18 | .1 | 20 |
| 4 | .3 | 16 | .2 | 22 | .1 | 30 |

Table 3.1: Rates for four different models.

**Notation and parameters:**

$n = 3$ : Number of queues in the system.

$\ell = 11$ : Maximum link capacity.

$d_i$ : Burst rate of queue $i$: $d_1 = 1$, $d_2 = 3$, $d_3 = 5$ are fixed.

$a_i$ : Arrival rate of queue $i$.

$b_i$ : Service rate of queue $i$.

$T = 10$ : Time.

The first thing we compared is the capacity. For three different values of $\theta$, we calculated $Cap_\theta$ and $Cap_\theta^*$. This is achieved by using the *linear system method* and the results are given in Table 3.2. We used the *exact solution method* to calculate $P_\pi(\tau \leq T)$ and $P_{\pi^*}(\tau^* \leq T)$ which are shown in Table 3.3. In the same table we also compared results from the *saddle point approximation* for both the original and the induced processes. Under Approximation 2, we provide numerical results by using the asymptotic theory about $P_\pi(\tau \leq T)$ studied in Chapter 2, which says, as

| Model | $\theta_1 = .05$ | | $\theta_2 = .1$ | | $\theta_3 = .15$ | |
|---|---|---|---|---|---|---|
| | original | induced | original | induced | original | induced |
| 1 | .00011609 | .00011609 | .00011630 | .00011630 | .00011643 | .00011643 |
| 2 | .00016153 | .00016154 | .00016210 | .00016211 | .00016250 | .00016250 |
| 3 | .00004161 | .00004161 | .00004167 | .00004167 | .00004171 | .00004171 |
| 4 | .00002235 | .00002235 | .00002237 | .00002237 | .00002238 | .00002238 |

Table 3.2: Capacities at different $\theta$'s.

| Model | Exact Probability | | Saddle Point Approx. | | Approximation 2 | |
|---|---|---|---|---|---|---|
| | $P_\pi(\tau \leq T)$ | $P_{\pi^*}(\tau^* \leq T)$ | original | induced | $T/E_\pi\tau$ | $T/E_{\pi^*}\tau^*$ |
| 1 | .00116369 | .00116370 | .00108643 | .00108643 | .00116279 | .00116280 |
| 2 | .00162234 | .00162236 | .00151901 | .00151903 | .00161749 | .00161752 |
| 3 | .00041676 | .00041676 | .00038878 | .00038878 | .00041607 | .00041607 |
| 4 | .00022368 | .00022368 | .00020850 | .00020850 | .00022341 | .00022341 |

Table 3.3: Exact probabilities and approximations.

| Model | Upper Bound 1 | Upper Bound 2 | | Lower Bound | |
|---|---|---|---|---|---|
| | | original | induced | $1-e^{-\lambda_0 T}$ | $1-e^{-\lambda_0^* T}$ |
| 1 | .00118269 | .00314619 | .00314619 | .00116212 | .00116212 |
| 2 | .00166685 | .00438615 | .00438622 | .00161618 | .00161620 |
| 3 | .00042244 | .00112713 | .00112714 | .00041598 | .00041598 |
| 4 | .00022586 | .00060497 | .00060497 | .00022338 | .00022338 |

Table 3.4: Upper bounds and lower bounds for the probabilities.

| Model | Mean | | Lower Bound | $\lambda_0$ | $\lambda_0^*$ | $\pi(B^c)$ |
|---|---|---|---|---|---|---|
| | $E_\pi \tau$ | $E_{\pi^*} \tau^*$ | | | | |
| 1 | 8599.98 | 8599.96 | 8520.63 | 1.16279E-4 | 1.16279E-4 | 1.56E-6 |
| 2 | 6182.41 | 6182.31 | 6094.06 | 1.61748E-4 | 1.61751E-4 | 6.04E-6 |
| 3 | 24034.6 | 24034.4 | 23864.4 | 4.16067E-5 | 4.16070E-5 | 7.68E-7 |
| 4 | 44761.5 | 44761.4 | 44546.1 | 2.23406E-5 | 2.23407E-5 | 2.95E-7 |

Table 3.5: Means and their lower bounds, principal eigenvalues.

$t \to \infty$.

$$P_\pi(\tau \le T) \sim 1 - \exp(-\lambda_0 T) \sim \lambda_0 T \sim \frac{T}{E_\pi \tau},$$

where $E_\pi \tau$ is obtained by the *linear system method.*

To apply Corollary 3.3.4 we may simply set $\theta = 1$ and, by defining $\theta' = 1/T$, the upper bound may be written as $(e/\theta')Cap^*_{\theta'}((B^c)^*)$. Alternatively, we may optimize in $\theta$. Let $G_{\pi^*}(\theta) = E_{\pi^*} \exp(-\theta\tau^*)$ and define $\mu^*_{\pi^*}(\theta) := \log(G_{\pi^*}(\theta))$. The upper bound is now equivalent to $P_\pi(\tau \le T) \le \exp(\theta T + \mu^*_{\pi^*}(\theta))$. For $T \le -(\mu^*_{\pi^*})'(0) = E_{\pi^*}\tau^*$, the value $\theta$ that produces the tightest upper bound is found by solving $(\mu^*_{\pi^*})'(\theta_0) = -T$. This equation has a unique solution because $\mu^*_{\pi^*}(\theta)$ is strictly convex, strictly decreasing and analytic on the interior of its domain of convergence, $(\overline{\theta}, \infty)$ for some $\overline{\theta} \le 0$. In Table 3.4, under the columns labeled Upper Bound 2, we find this minimum numerically. This involves fitting a quadratic to the function $\theta T + \mu^*_{\pi^*}(\theta)$ and finding the minimum for this quadratic. Each evaluation of this function at a given $\theta$ involves the solution of the induced linear system (3.3.8). This is the same procedure to obtain $\theta_0$ when we try the *saddle point approximation method.* We see this is moderately successful but while the Laplace transform $E_\pi \exp(-\theta\tau)$ is well approximated, the Chebyshev inequality is rather coarse. Better upper bounds and lower bounds on $P_\pi(\tau \le T)$ are found by applying the *Aldous and Brown's method.* The results are given in Table 3.4 under the column Upper Bound 1 and Lower Bound. These are seen to be excellent. The corresponding lower bound on $E_\pi \tau$ are given in Table 3.5 under the column Lower Bound. Also in Table 3.5, the exact values of $E_\pi \tau$ and $E_{\pi^*}\tau^*$ are compared, and we see that the principal eigenvalues $\lambda_0$ and $\lambda_0^*$ are extremely close. The latter fact is believed to be the most important factor for the success of the *induced Dirichlet forms method* applied to the ATM problem. We can show that $\lambda_0(B)$ and $\lambda_0^*(B^*)$ are asymptotically the same when $B \to S$. This is the main topic of the next chapter.

## 3.5 Special Cases

In the case where the burst rates $d_j$ from all sources are identical (and without loss of generality equal to 1), everything simplifies. From (3.3.7), $\mathcal{E}^*$ becomes ($S^* = \mathcal{N}$, the set of natural numbers):

$$\mathcal{E}^*(h,h) = \sum_{r\in S^*} \left[(h(r+1) - h(r))^2 B\pi^*(r)\right]$$

where $B = \sum_{i=1}^n a_i$. Since $\sum_{i=1}^n N_i(t)$ is a Poisson random variable with mean $A = \sum_{j=1}^n \lambda_j$, it follows that $\pi^*(r) = \exp(-A)A^r/r!$. By reversibility, it follows that the jump rate from $r$ to $r-1$ is

$$\frac{B\pi^*(r-1)}{\pi^*(r)} = \frac{Br}{A}.$$

We conclude that the induced form is that of a $M/M/\infty$ queue having constant birth rate $B$, linear death rate $Br/A$ and equilibrium measure $\pi^*(r) = \exp(-A)A^r/r!$.

Consider any positive recurrent, irreducible, birth and death process $(X(t); t \geq 0)$ with generator $-L$ with birth rates $B(x) > 0$, $x \in \mathcal{N}$, and death rates $D(x) > 0$, $x \in \mathcal{N}\backslash\{0\}$, $D(0) = 0$; and stationary measure $\pi^*$ given by

$$\pi^*(x) = \pi_0^*(x)/\sum_{y=0}^{\infty} \pi_0^*(y), \qquad \pi_0^*(x) := \begin{cases} 1 & \text{if } x = 0, \\ \frac{B(0)\cdots B(x-1)}{D(1)\cdots D(x)} & \text{if } x \geq 1. \end{cases}$$

The reversibility property is

$$B(x)\pi^*(x) = D(x+1)\pi^*(x+1). \tag{3.5.10}$$

As usual we set

$$\tau_\ell = \min\{t \geq 0 : X(t) \geq \ell\}. \tag{3.5.11}$$

Let $\alpha(x) = E_x\tau_\ell$. Then $\alpha$ satisfies

$$\text{(3.5.12)} \qquad \begin{cases} L\alpha(x) = 1 & \text{for } 0 \le x \le \ell - 1 \\ \alpha(x) = 0 & \text{for } x \ge \ell. \end{cases}$$

Also define

$$M^*(r) = \sum_{s=0}^{r} \pi^*(s) \text{ and } \nu_1(x) = \sum_{r=0}^{x-1} \frac{M^*(r)}{B(r)\pi(r)}.$$

It is straightforward to verify directly that $\nu_1(\ell) - \nu_1(x)$ solves the problem (3.5.12). Consequently we have that

$$\text{(3.5.13)} \qquad E_x\tau_\ell = \nu_1(\ell) - \nu_1(x)$$

and in particular $\nu_1(\ell) = E_0\tau_\ell$. Also,

$$\begin{aligned} E_{\pi^*}\tau_\ell &= \sum_{x=0}^{\ell-1} E_x\tau_\ell \pi^*(x) = \nu_1(\ell)M^*(\ell-1) - \sum_{x=0}^{\ell-1} \nu_1(x)\pi^*(x) \\ &= \sum_{x=0}^{\ell-1} \frac{M^*(x)^2}{B(x)\pi^*(x)}. \qquad \text{(3.5.14)} \end{aligned}$$

The last equality follows from a summation by parts. Alternative results are given in Karlin and Taylor (1975).

For the $M/M/\infty$ queue above this gives

$$\alpha(r) = \sum_{k=r}^{\ell-1} \frac{k!}{B} A^k \exp(A) M^*(k)$$

where $M^*(k) := \sum_{i=0}^{k} \pi^*(i) = \exp(-A)\sum_{i=0}^{k} \frac{1}{i!}A^i$. By Proposition 3.3.3, it follows that

$$\begin{aligned} E_\pi\tau \ge E_{\pi^*}\tau^* &= \frac{1}{B}\exp(A)\sum_{k=0}^{\ell-1} A^{-k}k!M^*(k)^2 \qquad \text{(3.5.15)} \\ &\ge \frac{M^*(\ell-1)^2}{D\ell\pi^*(\ell)}. \qquad \text{(3.5.16)} \end{aligned}$$

## 3.6 Comments

We see from the numerical comparisons that the induced process always provides excellent approximations to all exact solutions or estimates of parameters regarding the distribution and the mean of the first hitting time of the multi-dimensional process. As far as possible, we should solve the problem for the much smaller induced process. But we remark that when $\ell$ is large, even solving the induced problem may become troublesome. In that case, the asymptotic results depending only on *Gap* and $\lambda_0$ should be used.

The induced process is the aggregated process which arises in the aggregation-disaggregation method for finding the steady state of a large Markov system (see Schweitzer (1991)). The engineering approach, then, is to replace any quantity associated with the process $\vec{N}(t)$ by the corresponding quantity for the induced process $N^*(t)$. In Chapter 5, we use the small induced process rather than the huge original process to drive a buffer in the multiplexor and approximate the original buffer content by the new one. Hong and Perros (1992) similarly use the induced or aggregated processes associated with a number of interrupted Bernoulli processes to drive a multiplexor buffer. If one takes this approach, we should estimate $P_\pi(\tau \le T)$ by $P_{\pi^*}(\tau^* \le T)$. For small problems we may calculate the latter directly, as in Table 3.3, under Exact Probability. The relation between $P_\pi(\tau \le T)$ and $P_{\pi^*}(\tau^* \le T)$ can be described as follows. For any fixed subset $B$, there is a $T_0 > 0$ such that

$$P_\pi(\tau \le T) \le P_{\pi^*}(\tau^* \le T), \qquad \text{for } T > T_0. \tag{3.6.17}$$

This is because when $T$ is large, $P_\pi(\tau \le T)$ and $P_{\pi^*}(\tau^* \le T)$ are asymptotically equal to $1 - \exp\{-\lambda_0 T\}$ and $1 - \exp\{-\lambda_0^* T\}$ respectively, and by the Rayleigh-Ritz principle we know $\lambda_0 \le \lambda_0^*$.

# Chapter 4

# Asymptotics of $\lambda_0$ in the ATM Model

We first recall some definitions, notation and results given in the ATM multiplexor model investigated in the last chapter. It is assumed that the link rate of the multiplexor is $\ell - 1$ cells per second, and that $n$ distinct independent traffic categories are multiplexed together at the switch. Traffic sources in category $i$ may be described as an alternating series of idle and bursty periods. A burst from a source in category $i$ produces cells at a rate of $d_i$ cells per second. We assume that bursts of category $i$ arrive according to a Poisson process having a rate of $a_i$ bursts per second. We also assume that the burst periods are independent (and independent of the arrival process) and are exponentially distributed with a mean burst length of $1/b_i$. We describe the traffic at the multiplexor by the Markov process $\vec{N}(t) := (N_1(t), \cdots, N_n(t))$ defined on the state space $S := \{0, 1, \cdots\}^n$ where $N_i(t)$ represents the number of bursts from category $i$ sources being multiplexed at time $t$. As such, $\tau$ is the first hitting time the process $\vec{N}(t)$ reaches the bad region $B^c := \{\vec{x} \in S : f(\vec{x}) \geq \ell\}$ where $f(\vec{x}) := \sum_{i=1}^n d_i x_i$.

We know that for each $1 \leq i \leq n$, $N_i(t)$ is reversible with respect to the stationary Poisson measure having mean $\lambda_i := a_i/b_i$. Moreover the $(N_i(t); 1 \leq i \leq n)$ are independent. Hence $\vec{N}(t)$ is also reversible, with respect to the stationary product

measure $\pi$ given by

$$\pi(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \frac{\lambda_i^{x_i}}{x_i!} e^{-\lambda_i}.$$

The technique of induced Dirichlet forms was used to generate a reversible one-dimensional induced Markov process $(N_t^*; t \geq 0)$ with induced stationary measure $\pi^*$ having (countable) support $S^* \subset \mathcal{R}_+$, defined by

$$\pi^*(r) \equiv \pi(\{\vec{x} : f(\vec{x}) = r\}) = \sum_{\vec{x}: \sum_{j=1}^{n} d_j x_j = r} \pi(\vec{x}), \text{ where } f(\vec{x}) := \sum_{j=1}^{n} d_j x_j.$$

The generator $-L^*$ of the induced process $N_t^*$ is given by

$$\begin{aligned} -L^*u(r) &:= \int J^*(r, ds)[u(s) - u(r)] \\ &= \sum_{r \in S^*} \sum_{i=1}^{n} (u(r + d_i) - u(r)) a_i \pi^*(r) \\ &+ \sum_{r \in S^*} \sum_{i=1}^{n} (u(r - d_i) - u(r)) \left[ \frac{a_i \pi^*(r - d_i)}{\pi^*(r)} \right] \pi^*(r) \end{aligned}$$

for $u$ in the core $\mathcal{D}_0^*$ of real-valued functions which are constant outside a finite subset of the non-negative integers. As above let $-(L^B)^* \equiv -(L^*)^{B^*}$ denote the generator of the process killed on the induced bad region $(B^*)^c \subseteq \{\ell, \ell + 1, \cdots\}$. Denote by $[K^*]^{B^*}(x) := 1_{B^*}(x) J^*(x, (B^*)^c)$ the killing rate of the induced process. We note in passing that the mean killing rate

$$\overline{\kappa}^* = \int_{B^*} [K^*]^{B^*}(r) \pi^*(dr) / \pi^*(B^*) = \int_{B^*} [K^*]^{B^*}(r) \dot{\pi}^*(dr)$$

where $\dot{\pi}^*$ is the truncated probability measure of $\pi^*$ on $B^*$. It should be noticed that the mean killing rates are the same for both the original and the induced processes, i.e. $\overline{\kappa} = \overline{\kappa}^*$.

The first hitting time $\tau^*$ of $N_t^*$ into $(B^*)^c$ provides a stochastic bound on the first hitting time $\tau$ of $N_t$ into $B^c$: $\tau^*$ is stochastically smaller than $\tau$ in the weak

sense that, for all $\theta \geq 0$,

$$E_\pi \exp(-\theta\tau) \leq E_{\pi^*} \exp(-\theta\tau^*)$$

and

$$E_{\pi^*}\tau \leq E_\pi\tau.$$

The numerical results in the previous chapter yielded virtually identical values for $E_\pi\tau$ and $E_{\pi^*}\tau^*$, far better than just a lower bound. The goal of this chapter is to explain this extremely good approximation. The explanation lies in the fact that $E_\pi\tau$ is closely approximated by $1/\lambda_0$, while $E_{\pi^*}\tau^*$ is closely approximated by $1/\lambda_0^*$, where $\lambda_0$ and $\lambda_0^*$ are the *principal* or Perron-Frobenius eigenvalues corresponding to $L^B$ and $L^{B^*}$ (also denoted by $L^*(\ell)$) respectively. We use the Temple-Kato theorem (see Kato (1949)) to show that as $B \to S$ (or equivalently, $\ell \to \infty$), $\lambda_0/\lambda_0^* \to 1$. This is the main result of this chapter and is presented in Proposition 4.2.8.

## 4.1 Properties of the Induced Stationary Measure

In this section we derive the recursion relation (3.4.9) for the induced probability $\pi^*$ and study its asymptotic behavior. For each $t \geq 0$, the weighted sum, $\sum_{i=1}^n d_i N_i(t)$, of independent Poisson random variables, where $N_i(t)$ has mean $\lambda_i$, has a compound Poisson distribution with characteristic function

$$\phi(t) := \exp\left(\sum_{j=1}^n \lambda_j |e^{itd_j} - 1|\right).$$

For any $r \in S^*$,

$$\begin{aligned}\pi^*(r) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-irt}\phi(t)dt \\ &= \frac{1}{2\pi i}\oint z^{-(r+1)} \exp\{\sum_{j=1}^n \lambda_j |z^{d_j} - 1|\}dz\end{aligned}$$

after substituting $z = \exp(it)$ where $\oint$ denotes complex integration around the unit circle. Next, integrating by parts, we get

$$\begin{aligned}\pi^*(r) &= \frac{1}{2\pi i r}\oint z^{-(r+1)}\exp\{\sum_{j=1}^{n}\lambda_j[z^{d_j}-1]\}\sum_{j=1}^{n}\lambda_j d_j z^{d_j}dz,\ r>0\\ &= \frac{1}{r}\sum_{j=1}^{n}\lambda_j d_j \pi^*(r-d_j).\end{aligned}$$

This recursion, with $\pi^*(0) = \exp(-A)$ where $A := \sum_{i=1}^{n}\lambda_i$, concentrated on $S^* \subset \mathcal{R}_+$ provides a practical means of calculating $\pi^*$.

We may, moreover, derive the asymptotics of $\pi^*$.

**Theorem 4.1.1** *(Iscoe, McDonald and Qian (1992a) Theorem 6.1)*

*If the $d_j$'s are aperiodic then as $r \to \infty$*

$$\pi^*(r) \sim \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\sum_{j=1}^{n} d_j^2\lambda_j s^{d_j}}}s^{-r}\exp\left(\sum_{j=1}^{n}\lambda_j(s^{d_j}-1)\right)$$

*where $s = s(r)$ is the positive solution of $r = \sum_{j=1}^{n} d_j\lambda_j s^{d_j}$: so $s \sim (r/\lambda d)^{1/d}$.*

**Proof:** See Moser and Wyman (1956), expansion (3.49). ■

**Lemma 4.1.2** *(Iscoe, McDonald and Qian (1992a) Lemma 6.2)*

*If the $d_j$'s are aperiodic, $d := \max\{d_i; i = 1,\cdots,n\}$ and $\lambda := \sum_{j:d_j=d}\lambda_j$, then*

$$\lim_{r\to\infty}\frac{\pi^*(r)r^{1/d}}{\pi^*(r-1)} = (d\lambda)^{1/d}.$$

**Proof:** Using the previous theorem we have

$$\begin{aligned}&\frac{\pi^*(r)s(r)}{\pi^*(r-1)}\\ &\sim \left(\frac{s(r-1)}{s(r)}\right)^{r-1}\frac{\exp(\sum_{j=1}^{n}\lambda_j s(r)^{d_j})}{\exp(\sum_{j=1}^{n}\lambda_j s(r-1)^{d_j})}\sqrt{\frac{\sum_{j=1}^{n} d_j^2\lambda_j s(r)^{d_j}}{\sum_{j=1}^{n} d_j^2\lambda_j s(r-1)^{d_j}}}.\end{aligned} \tag{4.1.1}$$

Let $r = f(s) := \sum_{j=1}^{n} d_j \lambda_j s^{d_j}$. Clearly $f(s)/s^d \to d\lambda$ so $s(r) \sim (r/d\lambda)^{1/d}$. The result will follow if we show the right-hand side of (4.1.1) tends to 1.

Expanding $s$ around $r$ we get

$$s(r-1) = s(r) - s'(r) + \frac{1}{2}s''(\tilde{r})$$

where $r - 1 < \tilde{r} < r$. Now

$$s'(r) = \frac{1}{f'(s)} = \frac{s}{\sum_{j=1}^{n} d_j^2 \lambda_j s^{d_j}}$$

and

$$s''(r) = \frac{d}{ds}\left(\frac{s}{\sum_{j=1}^{n} d_j^2 \lambda_j s^{d_j}} s'(r)\right)$$

so $|s''(r)| = \mathcal{O}(s(r)/r^2)$. Hence,

$$\frac{s(r-1)}{s(r)} = 1 - \frac{1}{d^2 \lambda s^d (1 + \mathcal{O}(1/s))} + \mathcal{O}(1/r^2) = 1 - \frac{1}{d^2 \lambda s^d}(1 + \mathcal{O}(1/s)).$$

We conclude

$$\begin{aligned}
\left(\frac{s(r-1)}{s(r)}\right)^r &= \prod_{j=1}^{n} \left(\frac{s(r-1)}{s(r)}\right)^{d_j \lambda_j s^{d_j}} \\
&\sim \left(1 - \frac{1}{d^2 \lambda s^d}(1 + \mathcal{O}(1/s))\right)^{d\lambda s^d} \\
&\to \exp(-1/d)
\end{aligned}$$

as $s \to \infty$. Also,

$$\begin{aligned}
\frac{\exp(\sum_{j=1}^{n} \lambda_j s(r)^{d_j})}{\exp(\sum_{j=1}^{n} \lambda_j s(r-1)^{d_j})} &= \exp\left(\sum_{j=1}^{n} \lambda_j s(r)^{d_j} \left[1 - \left(\frac{s(r-1)}{s(r)}\right)^{d_j}\right]\right) \\
&= \exp\left(\sum_{j=1}^{n} \lambda_j s(r)^{d_j} \frac{d_j}{d^2 \lambda s^d}(1 + \mathcal{O}(1/s))\right) \\
&\to \exp(1/d).
\end{aligned}$$

Finally, the last term on the right-hand side of (4.1.1) tends to 1 so the proof is complete. ■

**Corollary 4.1.3** *(Iscoe, McDonald and Qian (1992a) Corollary 6.3)*

*If the $d_j$'s are aperiodic then there is an $R$ such that $r\pi^*(r)$ is decreasing for $r \geq R-d$ . $R$ may be identified as the smallest value $r$ such that $\pi^*(r)$ is decreasing on $[r-d, r]$.*

**Proof:** By the previous lemma we have that, for $r$ sufficiently large, $\pi^*(r)/\pi^*(r-1) < 1$; so there is an $R$ such that $\pi^*(r)$ is decreasing for $r \geq R-d$. By the recursion formula for $\pi^*$, if $r \geq R$, then

$$\begin{aligned}(r+1)\pi^*(r+1) &= \sum_{j=1}^{n} \lambda_j d_j \pi^*(r-d_j) \\ &\leq \sum_{j=1}^{n} \lambda_j d_j \pi^*(r-d_j-1) \\ &= r\pi^*(r).\end{aligned}$$

We can also see from the above that if $\pi^*(r)$ is decreasing on $[r-d, r]$ . then it also is decreasing on $[r+1, \infty)$. ■

## 4.2 Asymptotics of $\lambda_0(B)$

Since $B^*$ is finite and $(L^B)^*$ is selfadjoint, it has positive eigenvalue $\lambda_0^*(\ell)$ and associated right (and left) nonnegative eigenvectors $\phi^* \equiv \rho^*$ belonging to $L^2(\pi^*)$.

**Lemma 4.2.1** *(Iscoe, McDonald and Qian (1992b) Lemma 3.1)*

*For $d = \max_i d_i$,*

$$\lambda_0^*(\ell) = \mathcal{O}(\pi^*(\ell-d)), \quad as\ \ell \to \infty.$$

**Proof:** By the Rayleigh-Ritz principle and the definition of the mean killing rate,

$$\lambda_0^*(\ell) \leq \overline{\kappa}^*$$

$$
\begin{aligned}
&= \sum_{k=0}^{\ell-1} J^*(k,[\ell,\infty])\pi^*(k)/\pi^*[0,\ell-1] \\
&= \sum_{k=\ell-d}^{\ell-1} \sum_{i:d_i\geq \ell-k} a_i\,\pi^*(k)/\pi^*[0,\ell-1] \\
&= \sum_{j=1}^{d} \left(\sum_{i:d_i\geq j} a_i\right) \pi^*(\ell-j)/\pi^*[0,\ell-1] \\
&\sim \left(\sum_{i:d_i=d} a_i\right) \pi^*(\ell-d) \qquad \text{(by Lemma 4.1.2)}
\end{aligned}
$$

since $\pi^*[0,\ell-1] \to 1$ as $\ell \to \infty$. ■

In the last section of Chapter 2 we observed that the principal eigenvalue is very much dependent on the values of the stationary distribution and the corresponding eigenvector on the 'boundary' of the rare set (see Proposition 2.4.6). In order to study the asymptotic behavior of the principal eigenvalue, we need first to find out the asymptotics of the eigenvector. This is sometimes achievable via the generating function of the eigenvector (see the example on $M/M/1$ queue) or the analysis of the stationary distribution (see the example on $M/M/\infty$ queue). For the induced Markov jump process, we do the same as we did in the case of $M/M/\infty$ queue. We first show that the eigenvector $\rho^*$ converges uniformly to 1 as $\ell \to \infty$.

**Lemma 4.2.2** *(Iscoe, McDonald and Qian (1992b) Lemma 3.2)*

*Let $\rho^*$ be normalized such that $||\rho^*||_{\pi^*} = 1$. Then*

$$\lim_{\ell\to\infty} \max_{0\leq k\leq \ell-1} |\rho^*(k) - 1| = 0.$$

**Proof:** By Lemma 4.2.1 there is a constant $M$ such that $\lambda_0^*(\ell) \leq M\pi^*(\ell-d)$. Set $a = \sum_{i:d_i=d} a_i$. Then for all large $\ell$

$$
\begin{aligned}
M &\geq \lambda_0^*(\ell)/\pi^*(\ell-d) \\
&= ((L^B)^*\rho^*, \rho^*)_{\pi^*}/\pi^*(\ell-d)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \sum_{k=0}^{\ell-1} a_i |\rho^*(k+d_i) - \rho^*(k)|^2 \frac{\pi^*(k)}{\pi^*(\ell-d)} \\
&\geq \sum_{k=0}^{\ell-d-1} |\rho^*(k+d) - \rho^*(k)|^2 a \frac{\pi^*(k)}{\pi^*(\ell-d)} \\
&\geq \left\{ \sum_{k=0}^{\ell-d-1} |\rho^*(k+d) - \rho^*(k)|^2 \right\} a \frac{\pi^*(\ell-d-1)}{\pi^*(\ell-d)}
\end{aligned}
$$

Since, by Lemma 4.1.2 $\lim_{\ell\to\infty} \pi^*(\ell-d-1)/\pi^*(\ell-d) = +\infty$, we conclude that

$$
\lim_{\ell\to\infty} \sum_{k=0}^{\ell-d-1} |\rho^*(k+d) - \rho^*(k)|^2 = 0. \tag{4.2.2}
$$

Given $\ell - d \leq k \leq \ell - 1$, write $k = md + r$, with $0 \leq r < d$. Then

$$
\begin{aligned}
|\rho^*(k) - 1| \;\leq\; & |\rho^*(k) - \rho^*(k-d)| \\
& + \sum_{j=1}^{m-1} |\rho^*(k-jd) - \rho^*(k-[j+1]d)| \;+\; |\rho^*(r) - 1|.
\end{aligned} \tag{4.2.3}
$$

Now $\lim_{\ell\to\infty} |\rho^*(r) - 1| = 0$ for each $0 \leq r \leq d-1$; and by (4.2.2)

$$
\lim_{\ell\to\infty} \max_{d\leq k\leq \ell-1} |\rho^*(k) - \rho^*(k-d)| = 0.
$$

Finally, with $C$ denoting a generic constant (depending on $d, \lambda$, etc.) whose value varies from line to line, we have, by the Cauchy-Schwarz inequality:

$$
\begin{aligned}
\Big(\sum_{j=1}^{m-1} |\rho^*(k-jd) \;-\; \rho^*(k-[j+1]d)|\Big)^2 &\leq \frac{\ell}{d} \sum_{i=0}^{\ell-2d-1} |\rho^*(i+d) - \rho^*(i)|^2 \\
&\leq C\,\ell \sum_{i=0}^{\ell-2d-1} |\rho^*(i+d) - \rho^*(i)|^2 a \frac{\pi^*(i)}{\pi^*(\ell-2d-1)} \\
&\leq C\,\ell\,\lambda_0^*(\ell)/\pi^*(\ell-2d-1) \\
&\leq C\,\ell\,\pi^*(\ell-d)/\pi^*(\ell-2d-1) \\
&\leq C\,\ell\,\ell^{-(d+1)/d} = C\ell^{-1/d} \to 0
\end{aligned}
$$

as $\ell \to \infty$; the last inequality following from Lemma 4.1.2.

Note that for $k < \ell - d$, we can include the first term on the right-hand side of (4.2.3) in the the summation following it. ■

We now show that the estimate obtained in the proof of Lemma 4.2.1 is actually asymptotically sharp.

**Proposition 4.2.3** *(Iscoe, McDonald and Qian (1992b) Proposition 3.3)*

*Let* $a = \sum_{i:d_i=d} a_i$. *Then* $\lambda_0^*(\ell) \sim a\pi^*(\ell - d)$, *as* $\ell \to \infty$.

**Proof:** By the Rayleigh-Ritz principle, it is clear that $Gap(L^*) \geq Gap(L) > 0$. So, we see from Lemma 2.2.5 that

$$(\rho^*, \mathbf{1}_{B^*})_{\pi^*} \geq 1 - \frac{\lambda_0^*(\ell)}{Gap(L^*)} \to 1, \qquad \text{as } \ell \to \infty.$$

Then applying Proposition 2.4.6, we have

$$\begin{aligned}
\lambda_0^*(\ell) &\sim \sum_{k=0}^{\ell-1} \rho^*(k) K^*(k) \pi^*(k) \\
&\sim \sum_{k=0}^{\ell-1} K^*(k) \pi^*(k), \text{ (by Lemma 4.2.2)} \\
&= \sum_{j=1}^{d} \left( \sum_{i:d_i \geq j} a_i \right) \pi^*(\ell - j) \\
&\sim a\pi^*(\ell - d)
\end{aligned}$$

by Lemma 4.1.2. ■

Using Proposition 4.2.3, we can sharpen the analysis in the beginning of the proof of Lemma 4.2.2 to obtain the following estimate, which will be used in the proof of Theorem 4.2.8.

**Corollary 4.2.4** *(Iscoe, McDonald and Qian (1992b) Corollary 3.4)*

*Let $\rho^*$ be normalized such that $||\rho^*||_{\pi^*} = 1$. Then*

$$\sum_{i=1}^{n} \sum_{k=0}^{\ell-d_i-1} a_i[\rho^*(k+d_i) - \rho^*(k)]^2 \pi^*(k) = o(\lambda_0^*(\ell)), \quad as\ \ell \to \infty.$$

**Proof:**

$$\begin{aligned}
\lambda_0^*(\ell) &= \sum_{i=1}^{n} \sum_{k=0}^{\ell-1} a_i[\rho^*(k+d_i) - \rho^*(k)]^2 \pi^*(k) \\
&\geq \sum_{i=1}^{n} \sum_{k=0}^{\ell-d_i-1} a_i[\rho^*(k+d_i) - \rho^*(k)]^2 \pi^*(k) + a[\rho^*(\ell) - \rho^*(\ell-d)]^2 \pi^*(\ell-d)
\end{aligned}$$

Therefore

$$\begin{aligned}
\lambda_0^*(\ell)^{-1} \sum_{i=1}^{n} \sum_{k=0}^{\ell-d_i-1} a_i[\rho^*(k+d_i) - \rho^*(k)]^2 \pi^*(k) &\leq 1 - [\rho^*(\ell-d)]^2 a\pi^*(\ell-d)\lambda_0^*(\ell)^{-1} \\
&\to 0, \quad \text{as } \ell \to \infty.
\end{aligned}$$

by Lemma 4.2.1 and Proposition 4.2.3. ■

The next two lemmas will also be used in the proof of Proposition 4.2.8.

**Lemma 4.2.5** *(Iscoe, McDonald and Qian (1992b) Lemma 3.5)*

*Let $X = (X_1, \ldots, X_n)$ be a random vector with law $\pi$. Set $\lambda_i = a_i/b_i$ and, for $k \in S^*$, set*

$$E^{(k)}X_i = E[X_i \mid \sum_{j=1}^{n} d_j X_j = k] \quad and \quad Var^{(k)}X_i = Var[X_i \mid \sum_{j=1}^{n} d_j X_j = k].$$

*Then*

$$\begin{aligned}
E^{(k)}X_i &= \lambda_i \frac{\pi^*(k-d_i)}{\pi^*(k)} \\
Var^{(k)}X_i &= \lambda_i^2 \frac{\pi^*(k-2d_i)}{\pi^*(k)} + \lambda_i \frac{\pi^*(k-d_i)}{\pi^*(k)} - \lambda_i^2 \left[\frac{\pi^*(k-d_i)}{\pi^*(k)}\right]^2
\end{aligned}$$

**Proof:**

$$\begin{aligned} E^{(k)}X_i &= \sum_{x:f(x)=k} x_i \frac{\pi(x)}{\pi^*(k)} \\ &= \frac{1}{\pi^*(k)} \sum_{x:f(x)=k} b_i x_i \pi(x)/b_i \\ &= \frac{1}{\pi^*(k)} \sum_{x:f(x)=k} a_i \pi(x-\delta_i)/b_i \qquad (4.2.4) \\ &= \frac{1}{\pi^*(k)} a_i \pi^*(k-d_i)/b_i \\ &= \lambda_i \frac{\pi^*(k-d_i)}{\pi^*(k)} \end{aligned}$$

and

$$\begin{aligned} E^{(k)}X_i^2 &= \sum_{x:f(x)=k} x_i^2 \frac{\pi(x)}{\pi^*(k)} \\ &= \frac{1}{\pi^*(k)} \sum_{x:f(x)=k} x_i \lambda_i \pi(x-\delta_i), \qquad \text{as in (4.2.4)} \\ &= \frac{\lambda_i}{\pi^*(k)} \sum_{x:f(x)=k-d_i} (x_i+1)\pi(x) \\ &= \frac{\lambda_i}{\pi^*(k)} \sum_{x:f(x)=k-d_i} x_i \pi(x) \quad + \quad \frac{\lambda_i}{\pi^*(k)} \sum_{x:f(x)=k-d_i} \pi(x) \\ &= \lambda_i^2 \frac{\pi^*(k-2d_i)}{\pi^*(k)} \quad + \quad \lambda_i \frac{\pi^*(k-d_i)}{\pi^*(k)}, \qquad \text{as in (4.2.4).} \end{aligned}$$

■

**Lemma 4.2.6** *(Iscoe, McDonald and Qian (1992b) Lemma 3.6)*

*For sufficiently large $k$, $\pi^*(k-1)/\pi^*(k) - \pi^*(k)/\pi^*(k+1)$ is well defined and bounded. Hence $Var^{(k)}(X_i)/E^{(k)}(X_i)$ is uniformly bounded in $i$ and $k$.*

**Proof:** In the case that all the $d_i$'s coincide (and equal 1), the conclusion follows from explicit computation. As a matter of fact, $\pi^*$ is then a Poisson density and the

ratio $\pi^*(k-1)/\pi^*(k)$ is a linear function of $k$.

Assume for the remainder of the proof, that the $d_i$'s are aperiodic. Then, for sufficiently large $k$, $k \in S^*$. Thus $\pi^*(k) > 0$ and therefore the ratio $g(k) := \pi^*(k-1)/\pi^*(k)$ is well defined. We shall show that

$$g(k) = ck^p + \mathcal{O}(1), \text{ as } k \to \infty; \quad c > 0,\ p = 1/d. \tag{4.2.5}$$

The first part of the lemma then follows immediately from (4.2.5), for

$$\begin{aligned} g(k+1) - g(k) &= c[(k+1)^p - k^p] + \mathcal{O}(1) \\ &= ck^p[(1+\frac{1}{k})^p - 1] + \mathcal{O}(1) \\ &= ck^p[\frac{p}{k} + \mathcal{O}(k^{-2})] + \mathcal{O}(1) \\ &= cpk^{p-1} + \mathcal{O}(1) \\ &= o(1) + \mathcal{O}(1) = \mathcal{O}(1). \end{aligned}$$

The second part of the lemma then follows (for large $k$) from Lemma 4.2.4 and the factorization:

$$\frac{\pi^*(k-2d_i)}{\pi^*(k)} - \left[\frac{\pi^*(k-d_i)}{\pi^*(k)}\right]^2 = \left[\frac{\pi^*(k-2d_i)}{\pi^*(k-d_i)} - \frac{\pi^*(k-d_i)}{\pi^*(k)}\right]\frac{\pi^*(k_i-d)}{\pi^*(k)}.$$

A somewhat weaker form of (4.2.5) was established in Lemma 4.1.2: namely that $g(k) \sim ck^p$. The analysis here is similar, so we shall only outline the extra work involved in obtaining the stronger result (4.2.5). The starting point was the asymptotic expansion

$$\pi^*(k) = \frac{1}{\sqrt{2\pi}} \frac{s(k)^{-k}\exp[\sum_{j=1}^n d_j^2\lambda_j(s(k)^{d_j}-1)]}{\left[\sum_{j=1}^n d_j^2\lambda_j s(k)^{d_j}\right]^{1/2}} \left\{1 + \frac{c_1}{k} + \frac{c_2}{k^2} + \cdots\right\}$$

where $s \equiv s(k)$ is the positive solution of $k = \sum_{j=1}^{n} d_j^2 \lambda_j s^{d_j}$; so that $s(k) \sim (k/\lambda d)^{1/d}$, as $k \to \infty$. Therefore

$$g(k) = \left[\frac{s(k)}{s(k-1)}\right]^k \left[\exp\left(\sum_{j=1}^{n} \lambda_j [s(k-1)^{d_j} - s(k)^{d_j}]\right)\right] \left[\frac{\sum_{j=1}^{n} d_j^2 \lambda_j s(k)^{d_j}}{\sum_{j=1}^{n} d_j^2 \lambda_j s(k-1)^{d_j}}\right]^{\frac{1}{2}} \times$$

$$s(k-1)\,[1 + \mathcal{O}(k^{-2})], \text{ as } k \to \infty, \tag{4.2.6}$$

since $[1 + c_1/(k-1) + \mathcal{O}(k^{-2}]/[1 + c_1/k + \mathcal{O}(k^{-2})] = 1 + \mathcal{O}(k^{-2})$, as $k \to \infty$.

Now by the expression of $s(k)$ given in the proof of Lemma 4.1.2 we obtain

$$\frac{s(k)}{s(k-1)} = 1 + \frac{1}{d^2 \lambda s(k)^d}[1 + \mathcal{O}(s(k)^{-1})], \text{ as } k \to \infty. \tag{4.2.7}$$

Also, from the defining relation for $s(k)$, it follows that

$$s(k) = \left(\frac{k}{\lambda d}\right)^{1/d} [1 + \mathcal{O}(s(k)^{-1})], \text{ as } k \to \infty. \tag{4.2.8}$$

Therefore

$$\frac{s(k)}{s(k-1)} = 1 + \frac{(1/d)}{k}\,[1 + \mathcal{O}(s(k)^{-1})], \text{ as } k \to \infty \tag{4.2.9}$$

and

$$\begin{aligned}\left[\frac{s(k)}{s(k-1)}\right]^k &= \exp(k \log[1 + \frac{(1/d)}{k}\,[1 + \mathcal{O}(s(k)^{-1})]) \\ &= \exp[\frac{1}{d} + \mathcal{O}(s(k)^{-1})] \\ &= e^{1/d}\,[1 + \mathcal{O}(s(k)^{-1})], \text{ as } k \to \infty. \end{aligned} \tag{4.2.10}$$

Next

$$\exp\left(\sum_{j=1}^{n} \lambda_j [s(k-1)^{d_j} - s(k)^{d_j}]\right)$$

$$= \exp\left(\sum_{j=1}^{n} \lambda_j s(k)^{d_j} \left[\left(\frac{s(k-1)}{s(k)}\right)^{d_j} - 1\right]\right)$$

$$= \exp\left(\sum_{j=1}^{n} \lambda_j s(k)^{d_j}[-\frac{d_j}{dk}[1+\mathcal{O}(s(k)^{-1})] + o(k^{-1})]\right), \text{ by (4.2.9):}$$

$$= \exp\left(-\frac{1}{d}\frac{\lambda d\, s(k)^d}{k}[1+\mathcal{O}(s(k)^{-1})]\right)$$

$$= \exp(-\frac{1}{d}[1+\mathcal{O}(s(k)^{-1})]), \text{ by (4.2.8):}$$

$$= e^{-1/d}\,[1+\mathcal{O}(s(k)^{-1})], \text{ as } k \to \infty. \tag{4.2.11}$$

Next,

$$\left[\frac{\sum_{j=1}^n d_j^2 \lambda_j s(k)^{d_j}}{\sum_{j=1}^n d_j^2 \lambda_j s(k-1)^{d_j}}\right]^{1/2} = \left[\frac{s(k)}{s(k-1)}\right]^{d/2}\left[\frac{1+\mathcal{O}(s(k)^{-1})}{1+\mathcal{O}(s(k-1)^{-1})}\right]^{1/2}$$

$$= [1+\mathcal{O}(k^{-1})]^{d/2}[1+\mathcal{O}(s(k)^{-1})]^{1/2}, \text{ by (4.2.8):}$$

$$= 1+\mathcal{O}(s(k)^{-1}), \text{ as } k \to \infty. \tag{4.2.12}$$

Finally, from (4.2.9) and then (4.2.8)

$$s(k-1) = s(k)[1+\mathcal{O}(k^{-1})]$$

$$= \left(\frac{k}{d\lambda}\right)^{1/d}[1+\mathcal{O}(s(k)^{-1})][1+\mathcal{O}(k^{-1})]$$

$$= \left(\frac{k}{d\lambda}\right)^{1/d}[1+\mathcal{O}(s(k)^{-1})], \text{ as } k \to \infty. \tag{4.2.13}$$

Therefore, substituting (4.2.10),(4.2.11),(4.2.12), and (4.2.13) into (4.2.6), we obtain

$$g(k) = \left(\frac{k}{d\lambda}\right)^{1/d}[1+\mathcal{O}(s(k)^{-1})]$$

$$= ck^p + \mathcal{O}(1), \text{ by (4.2.8).}$$

where $p = 1/d$ and $c = (d\lambda)^{-1/d}$; which establishes (4.2.5). ■

The main tool used in the proof of Proposition 4.2.8 is the following variational estimate which is a special case, sufficient for our purposes, of a result due to Temple and Kato (see Kato (1949)).

**Theorem 4.2.7** *Let $\mathcal{L}$ be a self-adjoint operator on a Hilbert space $\mathcal{H}$, having a discrete spectrum of eigenvalues, bounded below: $\lambda_0 < \lambda_1 \leq \lambda_2 \cdots$. Fix $0 \neq v \in \mathcal{H}$ and denote by $\lambda$, the Rayleigh quotient:*

$$\lambda = (\mathcal{L}v, v)/||v||^2$$

*If $\lambda < \lambda_1$ then for any $\Lambda^* \in (\lambda_0, \lambda_1)$ such that $\lambda < \Lambda^*$:*

$$\lambda - \frac{\epsilon^2}{\Lambda^* - \lambda} \leq \lambda_0 \leq \lambda$$

*where $\epsilon^2 \equiv \epsilon(v)^2 = ||\mathcal{L}v||^2/||v||^2 - \lambda^2$.*

**Proposition 4.2.8** *(Iscoe, McDonald and Qian (1992b) Proposition 3.8)*

*Let $\lambda_0(B) \equiv \lambda_0(\ell)$, $\lambda_0^*(B^*) \equiv \lambda_0^*(\ell)$ be the principal eigenvalues of $L^B$ and $(L^B)^* \equiv L^{*B^*}$. Then $\lambda_0(\ell) \sim \lambda_0^*(\ell)$, as $\ell \to \infty$.*

**Proof:** We apply the Temple-Kato result, Theorem 4.2.7, with test function $\varphi \equiv \tilde{\rho} = \rho^* \circ f$ where $\rho^*$ is the principal (positive) eigenvector associated with $\lambda_0^*(\ell)$, normalized such that $||\rho^*||_{\pi^*} = 1$; and $f(x) = \sum_{i=1}^n d_i x_i$. Thus

$$\tilde{\rho}(x) = \rho^*(k), \text{ if } \sum_{i=1}^n d_i x_i = k.$$

As such, $||\tilde{\rho}||_\pi = 1$ and the Rayleigh-Ritz quotient, viz.

$$(L^B\tilde{\rho}, \tilde{\rho})_\pi = ((L^B)^*\rho^*, \rho^*)_{\pi^*} = \lambda_0^*(\ell).$$

Taking $v = \varphi$ in Theorem 4.2.7 we have

$$\lambda_0^*(\ell) - \frac{\epsilon^2}{\Lambda^* - \lambda_0(\ell)} \leq \lambda_0(\ell) \leq \lambda_0^*(\ell),$$

where $\epsilon^2 = \|L^B\tilde{\rho}\|_\pi^2/\|\tilde{\rho}\|_\pi^2 - (\lambda_0^*(\ell))^2$ and $\Lambda^*$ is chosen fixed in $(\lambda_0(\ell), Gap(L))$. We need to show that $\epsilon^2/\lambda_0^*(\ell) \to 0$, as $\ell \to \infty$ and this is equivalent to

$$\frac{\|L^B\tilde{\rho} - \lambda_0^*(\ell)\tilde{\rho}\|_\pi^2}{\lambda_0^*(\ell)} \to 0, \text{ as } \ell \to \infty.$$

Now, if $f(x) = k$,

$$\begin{aligned} -L^B\tilde{\rho}(x) &= \sum_{i=1}^n a_i[\tilde{\rho}(x+\delta_i) - \tilde{\rho}(x)] + \sum_{i=1}^n b_i x_i[\tilde{\rho}(x-\delta_i) - \tilde{\rho}(x)] \\ &= \sum_{i=1}^n a_i|\rho^*(k+d_i) - \rho^*(k)| + \sum_{i=1}^n b_i x_i|\rho^*(k-d_i) - \rho^*(k)| \end{aligned}$$

and

$$\begin{aligned} -\lambda_0^*(\ell)\tilde{\rho}(x) &= -\lambda_0^*(\ell)\rho^*(k) = [-L^B]^*\rho^*(k) \\ &= \sum_{i=1}^n a_i[\rho^*(k+d_i) - \rho^*(k)] \\ &+ \sum_{i=1}^n [\sum_{x:f(x)=k} b_i x_i \frac{\pi(x)}{\pi^*(k)}]|[\rho^*(k-d_i) - \rho^*(k)]. \end{aligned}$$

Therefore

$$\begin{aligned} \|L^B\tilde{\rho} - \lambda_0^*(\ell)\tilde{\rho}\|_\pi^2 &= \sum_{x\in B}[L^B\tilde{\rho}(x) - \lambda_0^*(\ell)\tilde{\rho}(x)]^2\pi(x) \\ &= \sum_{k=0}^{\ell-1}\sum_{x:f(x)=k}[L^B\tilde{\rho}(x) - \lambda_0^*(\ell)\tilde{\rho}(x)]^2\pi(x) \\ &= \sum_{k=0}^{\ell-1}\sum_{x:f(x)=k}\left(\sum_{i=1}^n|b_i x_i - \sum_{x:f(x)=k} b_i x_i\frac{\pi(x)}{\pi^*(k)}|\,|\rho^*(k-d_i) - \rho^*(k)|\right)^2\pi(x) \\ &\le \sum_{k=0}^{\ell-1}\sum_{x:f(x)=k} n\sum_{i=1}^n|b_i x_i - \sum_{x:f(x)=k} b_i x_i\frac{\pi(x)}{\pi^*(k)}|^2\,|\rho^*(k-d_i) - \rho^*(k)|^2\pi(x) \\ &= n\sum_{i=1}^n\sum_{k=0}^{\ell-1}\sum_{x:f(x)=k}|b_i x_i - \sum_{x:f(x)=k} b_i x_i\frac{\pi(x)}{\pi^*(k)}|^2\,|\rho^*(k-d_i) - \rho^*(k)|^2\pi^*(k) \\ &= n\sum_{i=1}^n\sum_{k=0}^{\ell-1} Var^{(k)}(b_i X_i)|\rho^*(k-d_i) - \rho^*(k)|^2\pi^*(k) \end{aligned}$$

where $(X_i : 1 \le i \le n)$ are random variables as in Lemma 4.2.5. For the remainder of the calculation, $C$ denotes a generic constant which may vary from line to line. By Lemma 4.2.6, the last step may be estimated by

$$\begin{aligned} ||L^B\tilde{\rho} - \lambda_0^*(\ell)\tilde{\rho}||_\pi^2 &\le C\sum_{i=1}^{n}\sum_{k=d_i}^{\ell-1} b_i^2 E^{(k)}(X_i)\,[\rho^*(k-d_i)-\rho^*(k)]^2\pi^*(k) \\ &= C\sum_{i=1}^{n}\sum_{k=d_i}^{\ell-1} a_i b_i [\rho^*(k-d_i)-\rho^*(k)]^2\pi^*(k-d_i) \\ &= C\sum_{i=1}^{n}\sum_{k=0}^{\ell-d_i-1} a_i\,[\rho^*(k+d_i)-\rho^*(k)]^2\pi^*(k) \\ &= o(\lambda_0^*(\ell)), \quad \text{as } \ell \to \infty \end{aligned}$$

by Corollary 4.2.4. ■

**Corollary 4.2.9** *(Iscoe, McDonald and Qian (1992b) Corollary 3.9)*

$$E_\pi\tau \sim \lambda_0(\ell)^{-1} \sim \lambda_0^*(\ell) \sim [a\pi^*(\ell-d)]^{-1} \sim \overline{\kappa}^{-1}, \text{ as } \ell \to \infty.$$

**Proof:** This is an immediate consequence of Theorem 2.3.5, Proposition 4.2.8, Proposition 4.2.3 and the fact that $\overline{\kappa} = \overline{\kappa}^*$. ■

# Chapter 5

# Steady State Analysis of an ATM Buffer

This chapter is devoted to the steady state distribution of the content of the common buffer shared by a number of categories of multiple sources in an ATM switch. In particular, we will concentrate on the probability of the buffer content exceeding some limit value. This is a very important issue in telecommunication applications. Here we adopt the so-called *fluid model* in which the content can take any value in $[0, \infty)$. This is an approximation to the real model used in an ATM switch, where the buffer content is a multiple of the number of information packages (cells). But according to Daigle and Langford (1986), this approximation is very good. We start by recalling some landmark works and some of the models studied. The notation we use for the models in this chapter is different from that of the ATM multiplexor model in the previous chapters. We basically follow the notation used by the authors whose works we frequently reference.

## 5.1 Physical and Mathematical Models

A data handling switch receives messages from a group of $N$ identical mutually independent bursty sources, which can each be *on* or *off*. The changes from on to off and from off to on take place according to Poisson processes with densities $\lambda$

and $\mu$, respectively. Each source produces information at a rate of $y$ units per unit time while it is on. The output channel of the switch can handle information at a maximum rate of $c$ units per unit time. If there are $r$ sources on simultaneously and if $ry > c$, the information not handled immediately is stored in a buffer at a rate of $ry - c$. If however $ry < c$, the buffer is depleted at a rate of $c - ry$, until it is empty. We assume that the buffer size is infinite and the system is stationary. The problem is to determine the probability $G(x)$ of the buffer content exceeding some value $x$, or if the buffer has a maximum size, the probability of overflow. We use a parameter vector $(N, \lambda, \mu, y)$ to describe the characteristics of the above system which we regard as the one-group-finite-source(OGFS) model.

Three other models which are closely related to OGFS model are one-group-infinite-source (OGIS) model, multi-group-finite-source(MGFS) model and multi-group-infinite-source(MGIS) model. Since the names are quite self-explanatory, we only give the differences . The OGIS model is the model when we have an infinite number of identical and mutually independent sources in the group (i.e., $N = \infty$ in OGFM model). If a system consists of $m(> 1)$ groups of identical sources, each group having a parameter vector $(N_i, \lambda_i, \mu_i, y_i)$ $(N_i < \infty, i = 1, 2, \cdots, m)$, and a single buffer shared by all sources, then it is a MGFS model. If $N_i = \infty$ for all $i$, we get the MGIS model.

The OGFS model was thoroughly discussed by Anick, Mitra and Sondhi (1982), further referred to as AMS[8]. In this model a complete and elegant solution for the probability of overflow is possible. In an earlier paper, Kosten (1974) treated the same problem for OGIS model which can be considered as the limit case, as $N \to \infty$, $\lambda \to 0$, and $N\lambda$ remains the fixed, of the OGFS model. This was pointed out by AMS[8]. The OGFS model was also discussed in Daigle and Langford (1986) as a model for analyzing packet voice communications systems. There it was called

the uniform arrival and service model and was compared with two other models.

The MGFS model was studied by Kosten (1984). The idea was to divide the total system into several OGFS systems. A complete solution is achievable by trial and error and by solving eigenvalue problem for those subsystems. This procedure is not practical in most cases, as was pointed out by the author.

In the ATM multiplexor problem studied in the early chapters, the MGIS model was used. There we calculated the probability of the total rate at which information produced exceeded maximum handling rate $c$ of the multiplexor during a certain time interval. Also, the mean time $E\tau$ for such a rare event to happen was investigated. The exact solution for the above problems is difficult to obtain for a relatively large system. We employed the technique of induced Dirichlet forms for reversible Markov processes to reduce the dimension of such problems.

In the present chapter, the stationary probability of overflow of the buffer of an ATM switch modeled by the MGFS model is considered. Results from Chapter 3 and 4 show that we should use the induced Dirichlet forms technique to get good approximations. In the rest of this section we present the mathematical representations for the OGFS and MGFS models with a brief review of some of the relevant methods and results given by AMS[8] and Kosten (1984). In the next section we introduce our approximation methods. We compare the approximations with the exact solutions by numerical examples in Section 3. Conclusion and comments are given in the last section.

For simplicity of analysis, the following assumption is made throughout this chapter:

*The maximum rate $c$ at which information units can be handled by the switch is a noninteger positive number.*

### 5.1.1 OGFS Model

Let $(N, \lambda, \mu, y)$ be the parameter vector for the OGFS system. We consider the case when $y = 1$ which is the same as in AMS[8] where $\mu = 1$ also. Denote by $N_t$ and $Q_t$ the number of sources which are on and the buffer content at time $t$, respectively. We say $Q_t$ is driven by $N_t$ in a sense that can be precisely described by the following relation

$$Q_t = \int_0^t (N_s - c)\, 1_{\{Q_{s-}>0\}} ds + Q_0$$

where $1_A$ is the indicator function of set $A$ and $Q_0 (\geq 0)$ is the initial content of the buffer. The infinitesimal generator matrix $\mathbf{G} = (g_{ij})$ for $N_t$ is defined by

$$g_{ij} = \begin{cases} (N-i)\lambda & \text{if } j = i+1 \\ i\mu & \text{if } j = i-1 \\ -((N-i)\lambda + i\mu)) & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

Also, $(N_t, Q_t)$ is a Markov process on the space $S \times R^+$, where $S = \{0, 1, \cdots, N\}$. We assume that $N\lambda/(\lambda + \mu) < c$ so that the system is stationary. For $i \in S$, define $F_i(x)$ as the equilibrium probability that $i$ sources are on and the buffer content does not exceed $x$. Then (see AMS[8] for details),

$$\begin{aligned} (i-c)\frac{dF_i(x)}{dx} &= (N-i+1)\lambda F_{i-1}(x) - \{(N-i)\lambda + i\mu\} F_i(x) \\ &+ (i+1)\mu F_{i+1}(x) \end{aligned} \tag{5.1.1}$$

for all $i \in S$, where $F_i(x) = 0$ if $i \notin S$. Define

$$\mathbf{D} = diag\{-c, 1-c, 2-c, \cdots, N-c\}$$

and

$$\mathbf{F}(x) = |\, F_0(x), F_1(x), \cdots, F_N(x)\, |^T,$$

then equation (5.1.1) can be written in a matrix form

$$\mathbf{D}\frac{d}{dx}\mathbf{F}(x) = \mathbf{G}^T\mathbf{F}(x), \quad x \geq 0. \tag{5.1.2}$$

To solve equation (5.1.2) we need to know the initial conditions $\mathbf{F}_0 := \mathbf{F}(0)$ and eigenvalues, together with left and right eigenvectors of the matrix $\mathbf{D}^{-1}\mathbf{G}^T$ (Notice that the diagonal matrix $\mathbf{D}$ is invertible because of our general assumption on c). Using the method of generating functions for eigenvectors, AMS[8] constructs $N+1$ quadratic equations whose roots are all the eigenvalues of $\mathbf{D}^{-1}\mathbf{G}^T$. It is shown that all the eigenvalues are real and different, and among them there are $[c]$ positive and $N-[c]$ negative ones and a zero. Especially, the largest negative eigenvalue $z^0$ is simply given by :

$$z^0 = -\frac{1+\lambda/\mu - N\lambda/c\mu}{1-c/N}. \tag{5.1.3}$$

Explicit formulas are also given for obtaining the right and left eigenvectors while the initial conditions follows from the following argument:

First, an empty buffer is inconsistent with more than $[c]$ sources being on, so we should have:

$$F_r(0) = 0, \quad r = [c]+1, [c]+2, \cdots, N. \tag{5.1.4}$$

This means we have only $[c]+1$ unknowns left for the initial vector $\mathbf{F}_0$. Secondly, the solution to the differential equations in (5.1.2) with $\mathbf{F}(0) = \mathbf{F}_0$ can be written as

$$\mathbf{F}(x) = \sum_{i=0}^{N} a_i \phi_i e^{z_i x} \tag{5.1.5}$$

where $\phi_i$ ($\psi_i^T\mathbf{D}$) is the right (left) eigenvector of $\mathbf{D}^{-1}\mathbf{G}^T$ corresponding to the eigenvalue $z_i$ and

$$a_i = \frac{\psi_i^T\mathbf{D}\mathbf{F}_0}{\psi_i^T\mathbf{D}\phi_i}.$$

Clearly, positive eigenvalues are related to components in the solution of (5.1.2) that grow to infinity according to $exp(zx)$. This is inconsistent with our assumption of the system being stationary. We delete those components in the solution by setting the corresponding coefficients $a_i$ equal to 0. This provides $|c|$ linear equations for $\mathbf{F_0}$. The last equation comes from normalizing the solution so that

$$\mathbf{1}^T \mathbf{F}(\infty) = 1. \tag{5.1.6}$$

since $F_i(\infty)$ is the probability that $i$ out of $N$ sources are on simultaneously. The solution for $\mathbf{F}(x)$ can then be expressed by

$$\mathbf{F}(x) = \mathbf{F}(\infty) + \sum_{z_i<0} a_i \phi_i e^{z_i x}. \tag{5.1.7}$$

In general, we have to go through the whole procedure mentioned above to find $\mathbf{F_0}$ and then $a_i$. But AMS[8] contains simple formulas for calculating $a_i$ without first solving equations for $\mathbf{F_0}$. The probability of overflow $G(x)$ is then given by

$$\begin{aligned} G(x) &= 1 - \mathbf{1}^T \mathbf{F}(x) \\ &= -\sum_{i:z_i<0} (a_i \mathbf{1}^T \phi_i) e^{z_i x} \end{aligned} \tag{5.1.8}$$

where $\mathbf{1} = [1, 1, \cdots, 1]^T$ with suitable dimension. Suppose that $z_0$ is the largest one among all negative eigenvalues, then asymptotically $G(x)$ behaves like its *dominant term* $-(a_0 \mathbf{1}^T \phi_0) e^{z_0 x}$.

### 5.1.2 MGFS Model

In the MGFS model we have $m(>1)$ groups of identical sources. Let $(N_i, \lambda_i, \mu_i, y_i)$ be the parameter vector of group $i$, $i = 1, 2, \cdots, m$. We denote by $N_i(i)$ the number of sources which are on in group $i$. Then information units will be stored in the

common buffer when $\sum_{i=1}^{m} y_i N_t(i) > c$. The following condition is assumed for the system to be stationary:

$$\sum_{i=1}^{m} \frac{y_i N_i \lambda_i}{\lambda_i + \mu_i} < c.$$

Define

$$N_t = [\, N_t(1), N_t(2), \cdots, N_t(m) \,]^T$$

and

$$Q_t = \int_0^t \left( \sum_{i=1}^{m} y_i N_s(i) - c \right) 1_{\{Q_s > 0\}} ds + Q_0$$

Then $Q_t$ is the buffer content at time $t$ with initial value $Q_0 (\geq 0)$. Now $N_t$ is a Markov process on the state space $S := S_1 \times S_2 \times \cdots \times S_m$, where $S_i = \{0, 1, \cdots, N_i\}$. So we have a total of $N := \prod_{i=1}^{m} (N_i + 1)$ states for $N_t$. Suppose an order is taken, say lexicographical, for the states so that $S = \{\vec{n}_1, \vec{n}_2, \cdots, \vec{n}_N\}$. According to this order, the infinitesimal generator $\mathbf{G} = (g_{ij})$ is defined as: for $\vec{n}_i = (n_{i1}, n_{i2}, \cdots, n_{im}) \in S$

$$g_{ij} = \begin{cases} (N_k - n_{ik})\lambda_k & \text{if } \vec{n}_j = \vec{n}_i + \delta_k \\ n_{ik}\mu_k & \text{if } \vec{n}_j = \vec{n}_i - \delta_k \\ -\sum_{k=1}^{m} \{(N_k - n_{ik})\lambda_k + n_{ik}\mu_k)\} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

where $\delta_k$ is the $k^{th}$ basis vector in $R^N$. We also define the stationary probability distributions $\pi = [\, \pi(\vec{n}_1), \cdots, \pi(\vec{n}_N) \,]^T$ for $N_t$ and $\mathbf{F}(x) = [\, F(\vec{n}_1, x), \cdots, F(\vec{n}_N, x) \,]^T$ for $(N_t(1), \cdots, N_t(m), Q_t)$. It is known that

$$\pi(\vec{n}_k) = \prod_{i=1}^{m} \pi_i(n_{ki}) := \prod_{i=1}^{m} \binom{N_i}{n_{ki}} \rho_i^{n_{ki}} (1 - \rho_i)^{N_i - n_{ki}} \tag{5.1.9}$$

where

$$\rho_i = \frac{\lambda_i}{\lambda_i + \mu_i}.$$

Also, $F(\vec{n}_k, x)$ satisfies the following differential equation (see Kosten (1984) for details):

$$(\sum_{i=1}^{m} y_i n_{ki} - c)\frac{dF(\vec{n}_k, x)}{dx} = \sum_{i=1}^{m}[\,(N_i - n_{ki} + 1)\lambda_i F(\vec{n}_k - \delta_i, x)$$

$$(5.1.10) \qquad - ((N_i - n_{ki})\lambda_i + n_{ki}\mu_i)F(\vec{n}_k, x) + (n_{ki} + 1)\mu_i F(\vec{n}_k + \delta_i, x)\,]$$

for $\vec{n}_k = (n_{k1}, \cdots, n_{km}) \in S$ and $x \geq 0$. If we define $\mathbf{D} = diag\{d_1, \cdots, d_N\}$ such that $d_k = \sum_{i=1}^{m} y_i n_{ki} - c$ corresponding to $\vec{n}_k = (n_{k1}, \cdots, n_{km})$, then (5.1.10) has the matrix form

$$(5.1.11) \qquad \mathbf{D}\frac{d\mathbf{F}(x)}{dx} = \mathbf{G}^T\mathbf{F}(x), \quad \text{for } x \geq 0.$$

This is formally the same equation as (5.1.2) for the OGFS model. So basically we should carry on the same procedure to solve it, *i.e.*, we need to find the eigenvalues and eigenvectors of the matrix $\mathbf{D}^{-1}\mathbf{G}^T$ and also the initial conditions $\mathbf{F}_0 := \mathbf{F}(0)$. Apparently, things become much more complicated since we are dealing with a huge $N \times N$ matrix, where $N := \prod_{i=1}^{m}(N_i + 1)$. Kosten (1984) uses the multiple generating functions for eigenvectors and shows that the eigenvalue problem of the total MGFS system can be decomposed into that of $m$ OGFS subsystems. It turns out that $z$ is an eigenvalue of the total MGFS system if a set of values $c_i$ $(i = 1, 2, \cdots, m)$ can be found such that:

1. $\sum_{i=1}^{m} y_i c_i = c$;
2. for any $i \in \{1, \cdots, m\}$, $z_i = zy_i/\mu_i$ is an eigenvalue of the OGFS system with parameter vector $(N_i, \lambda_i, \mu_i, 1)$ and deleting rate $c_i$.

It is noted that $c_i$ can take negative values. In fact, the eigenvalue problem of the matrix is irrelevant to what $c_i$ means in the physical model. Kosten (1984) then constructs exactly $N$ values of $z$ obeying the above criterion and claimed the following results.

**Theorem 5.1.1** *Matrix* $\mathbf{D}^{-1}\mathbf{G}^T$ *has the following eigenvalue properties:*

*(i) There are exactly* $N$ *eigenvalues which are all real:*

*(ii) There is one zero eigenvalue:*

*(iii) There are* $K-1$ *positive eigenvalues and* $N-K$ *negative ones if*

$$\sum_{i=1}^{m} \frac{y_i N_i \lambda_i}{\lambda_i + \mu_i} < c, \tag{5.1.12}$$

*where* $K$ *is the number of states* $\vec{n} \in S$ *such that* $\sum_{i=1}^{m} y_i n_i < c$. *When the above inequality is reversed, we have one more positive and one less negative eigenvalue.*

Theoretically, we can construct the eigenvalues in the way described by Kosten (1984) and then find the corresponding eigenvectors which can be composed from the eigenvectors of the subsystems. But the amount of calculation is substantial. In addition, there is no easy way as for the OGFS model case to get the initial conditions $\mathbf{F}_0$. So, a complete solution for the stationary distribution $F(x)$ for the MGFS model remains intractable in most cases. On the other hand, the asymptotic behavior of the probability of overflow is mainly dependent on the largest negative eigenvalue $z_0$, *i.e.*,

$$G(x) \sim \text{constant} * e^{z_0 x}, \text{ as } x \to \infty.$$

Kosten (1984) suggests using simulation to get $G(x)$ up to some value and then making the continuation by the above asymptotics. According to the construction of the eigenvalues, $z_0$ can be found by first solving the following polynomial equations for $c_1, \cdots, c_m$

$$\sum_{i=1}^{m} y_i c_i = c; \tag{5.1.13}$$

$$\frac{\mu_i}{y_i} \frac{1 + \lambda_i/m\mu_i - N_i\lambda_i/c_i\mu_i}{1 - c_i/N_i} = \frac{\mu_1}{y_1} \frac{1 + \lambda_1/\mu_1 - N_1\lambda_1/c_1\mu_1}{1 - c_1/N_1}, \text{ for } i = 2, \cdots, m$$

and then calculate

$$z_0 = -\frac{\mu_1}{y_1}\frac{1+\lambda_1/\mu_1 - N_1\lambda_1/c_1\mu_1}{1-c_1/N_1}. \tag{5.1.14}$$

In general, we get multiple solutions to (5.1.13) and (5.1.14). But if the condition for the system being stationary is satisfied, the largest negative eigenvalue $z_0$ is contained in those solutions and is the one corresponding to the solution to (5.1.13) such that

$$\frac{N_i\lambda_i}{\lambda_i+\mu_i} < c_i < N_i, \text{ for } i = 1, \cdots, m.$$

In the next section, an alternative approach to get approximations to the probability of overflow for MGFS models is presented. We first use the induced Dirichlet forms technique to create a so-called induced OGFS model and discuss the probability of overflow problem for it. We show that this probability is asymptotically a lower bound for that of the MGFS model. In the end, we point out that a better approximation of this probability for any MGFS model is achievable by combining the results from the induced OGFS model with the asymptotic behavior of the probability for the original MGFS model.

## 5.2 Approximation Methods

### 5.2.1 Induced OGFS Models

We use the same definitions and notation as in the previous sections, especially those for the MGFS models. For simplicity of analysis, we make the following assumption in the rest of this chapter: *$y_i$ are distinct positive integers and $y_1$ is equal to one.*

Recall that $N_t$ is a multidimensional Markov process on the state space $S = S_1 \times S_2 \times \cdots \times S_m$, where $S_i = \{0, 1, \cdots, N_i\}$, with stationary distribution $\pi$ given by (5.1.9). The states in $S$ are ordered and the k*th* state is denoted as $\vec{n}_k =$

$(n_{k1}, \cdots, n_{km})$. Notice that $N_t$ is a reversible Markov process with respect to $\pi$. The Dirichlet form associated with $N_t$ is defined as follows:

$$\mathcal{E}(u,u) = \sum_{\vec{n}_k \in S} \sum_{i=1}^{m} (u(\vec{n}_k + \delta_i) - u(\vec{n}_k))^2 (N_i - n_{ki}) \lambda_i \pi(\vec{n}_k). \tag{5.2.15}$$

Define $S^* = \{0, 1, \cdots, N^*\}$, where $N^* := \sum_{i=1}^{m} y_i N_i$ is an integer by the assumption on $y_i$, and define a map $f : S \mapsto S^*$ such that $f(\vec{n}_k) = \sum_{i=1}^{m} y_i n_{ki}$. Then $f$ induces a probability distribution $\pi^*$ defined by

$$\pi^*(r) := \sum_{\vec{n}_k \in S : f(\vec{n}_k) = r} \pi(\vec{n}_k).$$

It also induces a regular Dirichlet form $\mathcal{E}^*$ on $R^+$. For any $h$ defined on $S^*$,

$$\begin{aligned} \mathcal{E}^*(h,h) &= \mathcal{E}(h \circ f, h \circ f) \\ &= \sum_{r \in S^*} \sum_{\vec{n}_k \in S : f(\vec{n}_k) = r} \sum_{i=1}^{m} [\, (h(r + y_i) - h(r))^2 (N_i - n_{ki}) \lambda_i \,] \pi(\vec{n}_k) \\ &= \sum_{r=0}^{N^*} \sum_{i=1}^{m} (h(r + y_i) - h(r))^2 \left( \sum_{\vec{n}_k \in S : f(\vec{n}_k) = r} \frac{(N_i - n_{ki}) \lambda_i \pi(\vec{n}_k)}{\pi^*(r)} \right) \pi^*(r) \end{aligned} \tag{5.2.16}$$

The form $\mathcal{E}^*$ is associated with a one-dimensional reversible Markov jump process $N_t^*$ on $S^*$ having stationary measure $\pi^*$ and infinitesimal generator $\mathbf{G}^*$, which jumps to the right from $r$ to $r + y_i \in S^*$ with intensity

$$g^*_{r, r+y_i} = \sum_{\vec{n}_k \in S : f(\vec{n}_k) = r} \frac{(N_i - n_{ki}) \lambda_i \pi(\vec{n}_k)}{\pi^*(r)}. \tag{5.2.17}$$

Define

$$Q_t^* = \int_0^t (N_s^* - c)\, 1_{\{Q_{s-}^* > 0\}} ds + Q_0^*$$

then $(N_t^*, Q_t^*)$ is a Markov process. $Q_t^*$ is different from $Q_t$ since the driving process $\sum_{i=1}^{m} y_i N_t(i)$ has been substituted by $N_t^*$ (Notice that the former is not usually Markovian). If we think of $Q_t^*$ as the content of a buffer in a new model, which we

refer to as the induced OGFS model, we then have the same problem of determining the probability of overflow of the buffer as we had for the original MGFS model. It turns out that we have, in a suitable notation, the same differential equations as (5.1.11) which control the stationary distribution of $(N_t^*, Q_t^*)$. To be precise, define

$$\mathbf{D}^* = diag\{-c, 1-c, \cdots, N^* - c\}$$

and

$$\mathbf{F}^*(x) = [F_0(x), F_1(x), \cdots, F_{N^*}(x)]^T$$

where $F_i(x)$ is the stationary probability of having $i$ sources on and the buffer content not exceeding $x$ for the process $(N_t^*, Q_t^*)$. Then we get the following differential equations by a discussion analogous to that for the OGFS model:

$$\mathbf{D}^* \frac{d\mathbf{F}^*(x)}{dx} = \mathbf{G}^{*T}\mathbf{F}^*(x). \tag{5.2.18}$$

Solving these equations, we can find the probability of overflow for the induced OGFS model:

$$G^*(x) := 1 - \mathbf{1}^T\mathbf{F}^*(x).$$

Our idea is to use $G^*(x)$ and its variants as approximations to $G(x)$, the probability of overflow of the original MGFS model. Since $N^* + 1$, the total number of states in $S^*$, is now a linear function of $N_1, N_2, \cdots, N_m$, the dimension of the buffer content problem is hopefully remarkably reduced.

To investigate the properties of matrices $\mathbf{D}^{-1}\mathbf{G}^T$ and $\mathbf{D}^{*-1}\mathbf{G}^{*T}$, we need the following immediate results. Define

$$\Pi^{1/2} := diag\left\{\sqrt{\pi(\vec{n}_1)}, \sqrt{\pi(\vec{n}_2)}, \cdots, \sqrt{\pi(\vec{n}_N)}\right\} \tag{5.2.19}$$

$$\Pi^{*1/2} := diag\left\{\sqrt{\pi^*(0)}, \cdots, \sqrt{\pi^*(N^*)}\right\}, \tag{5.2.20}$$

then, since $N_t$ and $N_t^*$ are reversible Markov processes with respect to $\pi$ and $\pi^*$ respectively,

$$\mathbf{A} := \Pi^{1/2}\mathbf{G}\Pi^{-1/2}, \quad \mathbf{A}^* := \Pi^{*1/2}\mathbf{G}^*\Pi^{*-1/2}$$

are symmetric matrices. Note that, since all the eigenvalues of $-\mathbf{G}$ and $-\mathbf{G}^*$ are nonnegative, $-\mathbf{A}$ and $-\mathbf{A}^*$ are nonnegative definite matrices. It is also easy to see that

$$\Pi^{1/2}\mathbf{D}\Pi^{-1/2} = \mathbf{D}, \quad \Pi^{*1/2}\mathbf{D}^*\Pi^{*-1/2} = \mathbf{D}^*.$$

Part of the following lemma showing a relationship between $\mathbf{A}$ and $\mathbf{A}^*$ provides a matrix representation of the procedure of obtaining the induced Markov process $N_t^*$ from the multidimensional process $N_t$. The proof of this lemma is given in the appendix at the end of this chapter.

**Lemma 5.2.1** *Define matrix* $\mathbf{P}^T := (p_{ij})$ *as follows:*

$$p_{ij} = \sqrt{\frac{\pi(\vec{n}_j)}{\pi^*(i)}}\, 1_{\{f(\vec{n}_j)=i\}}$$

*for* $i \in S^*, \vec{n}_j \in S$. *Then* $\mathbf{P}$ *is a column orthonormal matrix, i.e.,* $\mathbf{P}^T\mathbf{P} = \mathbf{I}$ *where* $\mathbf{I}$ *is the unitary matrix of size* $N^* + 1$, *and*

$$(a) \quad \mathbf{A}^* = \mathbf{P}^T\mathbf{A}\mathbf{P};$$

$$(b) \quad \mathbf{D}^* = \mathbf{P}^T\mathbf{D}\mathbf{P}; \quad \mathbf{D}^{*-1} = \mathbf{P}^T\mathbf{D}^{-1}\mathbf{P}.$$

We only use this representation in theoretical analysis. In practice, we use simple recursion formulas to calculate $\pi^*$ and $\mathbf{G}^*$. This enables us to avoid creating all the states in $S$.

**Theorem 5.2.2** *For* $k = 1, 2, \cdots, m$, *define*

$$\gamma_k(r) := \sum_{\vec{n}_i \in S: f(\vec{n}_i)=r} n_{ik}\pi(\vec{n}_i),$$

*then we have the following recursion formulas:*

$$\gamma_k(r) = \frac{N_k\lambda_k}{\mu_k}\pi^*(r-y_k) - \frac{\lambda_k}{\mu_k}\gamma_k(r-y_k) \tag{5.2.21}$$

$$\pi^*(r) = \frac{1}{r}\sum_{k=1}^{m} y_k\gamma_k(r) \tag{5.2.22}$$

with initial conditions given by: $\gamma_k(0) = 0$ and

$$\pi^*(0) = \prod_{k=1}^{m}\left(\frac{\mu_k}{\lambda_k+\mu_k}\right)^{N_k}.$$

And the generator $\mathbf{G}^*$ can be obtained as follows:

$$g^*_{r,r+y_k} = \lambda_k\left(N_k - \frac{\gamma_k(r)}{\pi^*(r)}\right)$$

$$g^*_{r,r-y_k} = \frac{\pi^*(r-y_k)}{\pi^*(r)}g^*_{r-y_k,r}. \tag{5.2.23}$$

**Proof:** Let $(X_1, X_2, \cdots, X_m)$ be a random vector defined on $S$ with probability distribution $\pi$. For $i = 1, 2, \cdots, m$, define the conditional expectations as follows:

$$E^{(r)}X_i := E(X_i|\sum_{i=1}^{m} y_iX_i = r) = \sum_{\vec{n}:f(\vec{n})=r} n_i\pi(\vec{n})/\pi^*(r).$$

Then

$$r = E(\sum_{i=1}^{m} y_iX_i|\sum_{i=1}^{m} y_iX_i = r) = \sum_{i=1}^{m} y_iE^{(r)}X_i.$$

So, by the definition of $\gamma_i(r)$,

$$r\pi^*(r) = \sum_{i=1}^{m} y_iE^{(r)}X_i\pi^*(r) = \sum_{i=1}^{m} y_i\gamma_i(r).$$

By the reversibility of $\pi_i$ in the sense that

$$(N_i - x + 1)\frac{\lambda_i}{\mu_i}\pi_i(x-1) = x\pi_i(x),$$

we have

$$\begin{aligned}
\gamma_i(r) &= \sum_{\vec{n}:f(\vec{n})=r} n_i\pi(\vec{n}) = \sum_{\vec{n}:f(\vec{n})=r} (N_i - n_i + 1)\frac{\lambda_i}{\mu_i}\pi_i(n_i-1)\prod_{j\neq i}\pi_j(n_j) \\
&= \sum_{\vec{n}:f(\vec{n})=r} N_i\frac{\lambda_i}{\mu_i}\pi_i(n_i-1)\prod_{j\neq i}\pi_j(n_j) - \sum_{\vec{n}:f(\vec{n})=r}(n_i-1)\frac{\lambda_i}{\mu_i}\pi_i(n_i-1)\prod_{j\neq i}\pi_j(n_j) \\
&= N_i\frac{\lambda_i}{\mu_i}\sum_{\vec{n}:f(\vec{n}+\delta_i)=r}\pi(\vec{n}) - \frac{\lambda_i}{\mu_i}\sum_{\vec{n}:f(\vec{n}+\delta_i)=r} n_i\pi_i(n_i)\prod_{j\neq i}\pi_j(n_j) \\
&= N_i\frac{\lambda_i}{\mu_i}\pi^*(r-y_i) - \frac{\lambda_i}{\mu_i}\pi^*(r-y_i)E^{(r-y_i)}N_i \\
&= N_i\frac{\lambda_i}{\mu_i}\pi^*(r-y_i) - \frac{\lambda_i}{\mu_i}\gamma_i(r-y_i).
\end{aligned}$$

This proves the recursion formulas. The initial conditions are obvious. For the generator $G^*$ we have

$$\begin{aligned}
g^*_{r,r+y_i} &= \sum_{\vec{n}:f(\vec{n})=r}\frac{\lambda_i(N_i-n_i)\pi(\vec{n})}{\pi^*(r)} \\
&= \lambda_i N_i - \frac{\lambda_i}{\pi^*(r)}\sum_{\vec{n}:f(\vec{n})=r} n_i\pi(\vec{n}) \\
&= \lambda_i(N_i - \frac{\gamma_i(r)}{\pi^*(r)}).
\end{aligned}$$

Then by the reversibility again,

$$g^*_{r,r-y_i} = \frac{\pi^*(r-y_i)}{\pi^*(r)}g^*_{r-y_i,r}.$$

### 5.2.2 Eigenvalue Properties

We present the eigenvalue properties of matrices $\mathbf{D}^{*-1}\mathbf{G}^{*T}$ and $\mathbf{D}^{-1}\mathbf{G}^T$ and numerical comparisons of these eigenvalues. The following theorem is analogous to Theorem 5.1.1.

**Theorem 5.2.3** *Matrix* $\mathbf{D}^{*-1}\mathbf{G}^{*T}$ *has the following eigenvalue properties:*

*(i) There are exactly* $N^*+1$ *eigenvalues which are all real;*

*(ii) There is one zero eigenvalue;*

*(iii) There are* $|c|$ *positive eigenvalues and* $N^*-|c|$ *negative ones if*

$$\sum_{i=0}^{N^*} i\pi^*(i) \;<\; c. \tag{5.2.24}$$

*When the above inequality is reversed, we have one more positive and one less negative eigenvalue. Moreover, the above inequality is equivalent to*

$$\sum_{i=1}^{m} \frac{y_i N_i \lambda_i}{\lambda_i + \mu_i} < c.$$

Here we provide a rigorous proof for Theorem 5.1.1 which also gives the proof of Theorem 5.2.3. Denote the spectrum of matrix $M$ by $\Lambda(M)$, then we have the following observation which plays the role throughout the present chapter:

$$\begin{aligned}
\Lambda(\mathbf{D}^{-1}\mathbf{G}^T) &= \Lambda(\mathbf{G}\mathbf{D}^{-1}) \\
&= \Lambda(\Pi^{1/2}\mathbf{G}\mathbf{D}^{-1}\Pi^{-1/2}) \\
&= \Lambda(\Pi^{1/2}\mathbf{G}\Pi^{-1/2}\Pi^{1/2}\mathbf{D}^{-1}\Pi^{-1/2}) \\
&= \Lambda(\mathbf{A}\mathbf{D}^{-1}) \qquad (5.2.25)
\end{aligned}$$

**Proof of Theorem 5.1.1:** First, *(ii)* is easy to see from the fact that $\mathbf{G}$ is the generator matrix of an irreducible finite Markov chain and $\mathbf{D}$ is invertible.

In the rest of the proof, we discuss the equivalent problem for $\mathbf{A}\mathbf{D}^{-1}$. Define

$$\mathbf{A}(\epsilon) := \epsilon\mathbf{I} - \mathbf{A},$$

for $\epsilon > 0$. Then $\mathbf{A}(\epsilon)$ is a positive definite matrix since $-\mathbf{A}$ is a nonnegative definite matrix. The eigenvalues of $\mathbf{A}(\epsilon)$ are continuous functions of $\epsilon$, so when $\epsilon \to 0$,

$$\Lambda(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) \;\longrightarrow\; \Lambda(-\mathbf{A}\mathbf{D}^{-1}) = -\Lambda(\mathbf{A}\mathbf{D}^{-1}) \tag{5.2.26}$$

On the other hand, we have

$$\begin{aligned} \Lambda(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) &= \Lambda^{-1}(\mathbf{D}\mathbf{A}^{-1}(\epsilon)) \\ &= \Lambda^{-1}(\mathbf{A}^{-1}(\epsilon)\mathbf{D}) \end{aligned} \tag{5.2.27}$$

where $\Lambda^{-1}(M)$ stands for the set of reciprocals of eigenvalues of any invertible matrix $M$. Since $\mathbf{A}(\epsilon)$ is positive definite and $\mathbf{D}$ is symmetric, there exists a nonsingular matrix $\mathbf{X} = [\,\mathbf{x}_1, \cdots, \mathbf{x}_N\,]$ such that

$$\mathbf{X}^T\mathbf{A}(\epsilon)\mathbf{X} = diag\,\{a_1(\epsilon), \cdots, a_N(\epsilon)\}$$

$$\mathbf{X}^T\mathbf{D}\mathbf{X} = diag\,\{b_1(\epsilon), \cdots, b_N(\epsilon)\}$$

Moreover,

$$\mathbf{A}(\epsilon)\mathbf{x}_i = \lambda_i\mathbf{D}\mathbf{x}_i, \;\; \lambda_i = \frac{a_i(\epsilon)}{b_i(\epsilon)}$$

where $a_i(\epsilon) > 0, b_i(\epsilon) \neq 0$ are real for all $i$. This and (5.2.26) give the proof of $(i)$.

We also see that the number of negative $\lambda_i$ is equal to the number of negative $d_i$, which is $K$, by the law of inertia. According to $(ii)$, there is one $\lambda_i$ which tends to 0 when $\epsilon \to 0$. So, we will lose exclusively either a positive $\lambda_i$ or a negative one. Let $\lambda^-_{max}(M)$ be the largest negative eigenvalue and $\lambda^+_{min}(M)$ be the smallest positive one of any matrix $M$. Also, denote by $\lambda_{max}(M)$ and $\lambda_{min}(M)$ the maximum and minimum eigenvalues. Clearly, we have two possibilities

$$\lim_{\epsilon \to 0} \lambda^-_{max}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) = 0$$

or

$$\lim_{\epsilon \to 0} \lambda^+_{min}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) = 0.$$

But, according to (5.2.27) we have

$$\lambda^-_{max}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) = \frac{1}{\lambda^-_{min}(\mathbf{D}\mathbf{A}^{-1}(\epsilon))} = \frac{1}{\lambda_{min}(\mathbf{D}\mathbf{A}^{-1}(\epsilon))} \tag{5.2.28}$$

$$\lambda^+_{min}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) = \frac{1}{\lambda^+_{max}(\mathbf{D}\mathbf{A}^{-1}(\epsilon))} = \frac{1}{\lambda_{max}(\mathbf{D}\mathbf{A}^{-1}(\epsilon))} \tag{5.2.29}$$

Using Rayleigh quotients for the positive-symmetric pencil $(\mathbf{A}(\epsilon), \mathbf{D})$, we obtain

$$\begin{aligned} \lambda^{-}_{max}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) &= [\lambda_{min}(\mathbf{A}^{-1}(\epsilon)\mathbf{D})]^{-1} \\ &= \left[\min_{x} \frac{x^T\mathbf{D}x}{x^T\mathbf{A}(\epsilon)x}\right]^{-1} \end{aligned} \tag{5.2.30}$$

and

$$\lambda^{+}_{min}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) = \left[\max_{x} \frac{x^T\mathbf{D}x}{x^T\mathbf{A}(\epsilon)x}\right]^{-1}. \tag{5.2.31}$$

The necessary and sufficient condition for one $\lambda_i$ tends to 0 is that one of the following two expressions is true:

$$(a) \quad \min_{x} \frac{x^T\mathbf{D}x}{x^T\mathbf{A}(\epsilon)x} \longrightarrow -\infty;$$

$$(b) \quad \max_{x} \frac{x^T\mathbf{D}x}{x^T\mathbf{A}(\epsilon)x} \longrightarrow +\infty.$$

as $\epsilon \rightarrow 0$. On the other hand, there exists a unique (except for a constant factor) $x_0 := \Pi^{1/2}\mathbf{1} \in R^N$ such that

$$\begin{aligned} \lim_{\epsilon \to 0} x_0^T\mathbf{A}(\epsilon)x_0 &= -\mathbf{1}^T(\Pi^{1/2})^T\mathbf{A}\Pi^{1/2}\mathbf{1} \\ &= -\pi^T\mathbf{G}\mathbf{1} \\ &= 0. \end{aligned} \tag{5.2.32}$$

We conclude that (a) (or (b)) is equivalent to $x_0^T\mathbf{D}x_0$ less (or greater) than zero.

$$\begin{aligned} x_0^T\mathbf{D}x_0 &= \sum_{i=1}^{N} d_i\pi(\vec{n}_i) = \sum_{i=1}^{N}\left(\sum_{j=1}^{m} y_j n_{ij} - c\right)\pi(\vec{n}_i) \\ &= \sum_{j=1}^{m} y_j \sum_{i=1}^{N} n_{ij}\pi(\vec{n}_i) - c = \sum_{j=1}^{m} y_j N_j \rho_j - c. \end{aligned} \tag{5.2.33}$$

where $\rho_j = \lambda_j/(\lambda_j+\mu_j)$. Noticing that the number of positive eigenvalues of $-\mathbf{A}\mathbf{D}^{-1}$ is the number of negative eigenvalues of $\mathbf{A}\mathbf{D}^{-1}$, we achieved the proof for *(iii)* ■

**Proof of Theorem 5.2.3:** Notice that $|c| + 1$ is the number of negative diagonal elements of $\mathbf{D}^*$. An analogous discussion as in the proof of Theorem 5.1.1 gives the proof of this theorem except for the equivalence of the inequalities, which is proved as follows: Let $X_i$ be binomial random variables with parameters $N_i$ and $\rho_i$, where $\rho_i = \lambda_i/(\lambda_i + \mu_i)$ and $i = 1, 2, \cdots, m$. Define $X := \sum_{i=1}^{m} y_i X_i$. Then $X$ is of probability distribution $\pi^*$. So,

$$\begin{aligned}\sum_{i=1}^{m} \frac{N_i y_i \lambda_i}{\lambda_i + \mu_i} &= \sum_{i=1}^{m} y_i E(X_i) \\ &= E(X) \\ &= \sum_{i=0}^{N^*} i\pi^*(i) \blacksquare\end{aligned}$$

Recall that $\lambda^-_{max}(M)$ (or $\lambda^+_{min}(M)$) denotes the largest negative (or the smallest positive) eigenvalue of any matrix $M$. Also, denote by $\lambda_{max}(M)$ and $\lambda_{min}(M)$ the maximum and minimum eigenvalues. The following theorem shows an important relationship between eigenvalues of matrix $\mathbf{D}^{*-1}\mathbf{G}^{*T}$ and matrix $\mathbf{D}^{-1}\mathbf{G}^T$.

**Theorem 5.2.4** *Suppose that the condition for the MGFS system to be stationary is satisfied, i.e.,*

$$\sum_{i=1}^{m} \frac{y_i N_i \lambda_i}{\lambda_i + \mu_i} < c. \tag{5.2.34}$$

*Then we have the following inequalities:*

$$(i) \quad \lambda^-_{max}(\mathbf{D}^{*-1}\mathbf{G}^{*T}) \leq \lambda^-_{max}(\mathbf{D}^{-1}\mathbf{G}^T)$$

$$(ii) \quad \lambda^+_{min}(\mathbf{D}^{*-1}\mathbf{G}^{*T}) \geq \lambda^+_{min}(\mathbf{D}^{-1}\mathbf{G}^T)$$

$$(iii) \quad \lambda_{max}(\mathbf{D}^{*-1}\mathbf{G}^{*T}) \leq \lambda_{max}(\mathbf{D}^{-1}\mathbf{G}^T)$$

$$(iv) \quad \lambda_{min}(\mathbf{D}^{*-1}\mathbf{G}^{*T}) \geq \lambda_{min}(\mathbf{D}^{-1}\mathbf{G}^T).$$

**Proof:** (*i*) The proof is given for the equivalent problem for matrices $\mathbf{A}\mathbf{D}^{-1}$ and $\mathbf{A}^*\mathbf{D}^{*-1}$. For $\epsilon > 0$, since $-\mathbf{A}$ is nonnegative definite, the following defines a positive definite matrix:

$$\mathbf{A}(\epsilon) = \epsilon\mathbf{I} - \mathbf{A}.$$

Then, according to Lemma 5.2.1,

$$\mathbf{P}^T\mathbf{A}(\epsilon)\mathbf{P} = \mathbf{A}(\epsilon)^* := \epsilon\mathbf{I} - \mathbf{A}^*,$$

which is also a positive definite matrix. For any eigenvalue $\lambda(\mathbf{A}(\epsilon)\mathbf{D}^{-1})$, we have

$$\begin{aligned}\lambda(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) &= \frac{1}{\lambda(\mathbf{D}\mathbf{A}^{-1}(\epsilon))} \\ &= \frac{1}{\lambda(\mathbf{A}^{-1}(\epsilon)\mathbf{D})} \qquad (5.2.35)\end{aligned}$$

and

$$\lambda^+_{min}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) = \frac{1}{\lambda_{max}(\mathbf{A}^{-1}(\epsilon)\mathbf{D})}. \qquad (5.2.36)$$

Since $\mathbf{A}(\epsilon)$ is positive definite and $\mathbf{D}$ is symmetric, Rayleigh quotients for relative eigenvalues for the positive-symmetric pencil can be applied.

$$\begin{aligned}\lambda_{max}(\mathbf{A}^{-1}(\epsilon)\mathbf{D}) &= \max_x \frac{x^T\mathbf{D}x}{x^T\mathbf{A}(\epsilon)x} \geq \max_{x=\mathbf{P}y} \frac{x^T\mathbf{D}x}{x^T\mathbf{A}(\epsilon)x} \\ &= \max_y \frac{y^T\mathbf{P}^T\mathbf{D}\mathbf{P}y}{y^T\mathbf{P}^T\mathbf{A}(\epsilon)\mathbf{P}y} = \max_y \frac{y^T\mathbf{D}^*y}{y^T\mathbf{A}(\epsilon)^*y} \\ &= \lambda_{max}(\mathbf{A}(\epsilon)^{*-1}\mathbf{D}^*) \qquad (5.2.37)\end{aligned}$$

Using (5.2.36) again, we obtain

$$\lambda^+_{min}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) \leq \frac{1}{\lambda_{max}(\mathbf{A}(\epsilon)^{*-1}\mathbf{D}^*)}$$

$$= \lambda^{+}_{min}(\mathbf{A}(\epsilon)^{\bullet}\mathbf{D}^{\bullet^{-1}}) \tag{5.2.38}$$

Since (5.2.34) is assumed, we know (see the proof of Theorem 5.1.1) that

$$\lambda^{+}_{min}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}) \not\to 0$$

as $\epsilon \to 0$. Then, by letting $\epsilon \to 0$, we get

$$\lambda^{+}_{min}(-\mathbf{A}\mathbf{D}^{-1}) \leq \lambda^{+}_{min}(-\mathbf{A}^{\bullet}\mathbf{D}^{\bullet^{-1}}).$$

This is equivalent to

$$\lambda^{-}_{max}(\mathbf{A}\mathbf{D}^{-1}) \geq \lambda^{-}_{max}(\mathbf{A}^{\bullet}\mathbf{D}^{\bullet^{-1}}),$$

since

$$\lambda^{+}_{min}(-\mathbf{A}\mathbf{D}^{-1}) = -\lambda^{-}_{max}(\mathbf{A}\mathbf{D}^{-1}).$$

*(ii)* Clearly,

$$\begin{aligned} \lambda^{+}_{min}(\mathbf{D}^{\bullet^{-1}}\mathbf{G}^{\bullet^{T}}) &= -\lambda^{-}_{max}(-\mathbf{D}^{\bullet^{-1}}\mathbf{G}^{\bullet^{T}}) \\ &\geq -\lambda^{-}_{max}(-\mathbf{D}^{-1}\mathbf{G}^{T}) \\ &= \lambda^{+}_{min}(\mathbf{D}^{-1}\mathbf{G}^{T}). \end{aligned} \tag{5.2.39}$$

Here we have used the result of *(i)*.

*(iii)* We use the same notation as in the proof of *(i)*. Clearly,

$$\lambda_{min}(\mathbf{A}(\epsilon)^{\bullet}\mathbf{D}^{\bullet^{-1}}) < 0.$$

Since for any matrices $\mathbf{X}$ and $\mathbf{Y}$, $\mathbf{XY}$ and $\mathbf{YX}$ have the same non-zero eigenvalues, we have

$$\begin{aligned} \lambda_{min}(\mathbf{A}(\epsilon)^{\bullet}\mathbf{D}^{\bullet^{-1}}) &= \lambda_{min}(\mathbf{P}^{T}\mathbf{A}(\epsilon)\mathbf{P}\mathbf{P}^{T}\mathbf{D}^{-1}\mathbf{P}) \\ &= \lambda_{min}(\mathbf{A}(\epsilon)\mathbf{P}\mathbf{P}^{T}\mathbf{D}^{-1}\mathbf{P}\mathbf{P}^{T}). \end{aligned}$$

Now, $\mathbf{A}(\epsilon)$ is positive definite and $\mathbf{PP}^T\mathbf{D}^{-1}\mathbf{PP}^T$ is symmetric, so

$$\lambda_{min}(\mathbf{A}(\epsilon)^*\mathbf{D}^{*^{-1}}) = \min_y \frac{y^T\mathbf{PP}^T\mathbf{D}^{-1}\mathbf{PP}^T y}{y^T\mathbf{A}^{-1}(\epsilon)y} \leq \min_{y=\mathbf{P}y_1} \frac{y^T\mathbf{PP}^T\mathbf{D}^{-1}\mathbf{PP}^T y}{y^T\mathbf{A}^{-1}(\epsilon)y}$$

$$= \min_{y_1} \frac{y_1^T\mathbf{P}^T\mathbf{D}^{-1}\mathbf{P}y_1}{y_1^T\mathbf{P}^T\mathbf{A}^{-1}(\epsilon)\mathbf{P}y_1} = \min_{y_1} \frac{y_1^T\mathbf{D}^{*^{-1}}y_1}{y_1^T\mathbf{A}(\epsilon)^{*^{-1}}y_1}$$

$$= \lambda_{min}(\mathbf{A}(\epsilon)^*\mathbf{D}^{*^{-1}}).$$

So, the above inequality is actually an equality. This means

$$\lambda_{min}(\mathbf{A}(\epsilon)^*\mathbf{D}^{*^{-1}}) = \min_{y_1} \frac{y_1^T\mathbf{P}^T\mathbf{D}^{-1}\mathbf{P}y_1}{y_1^T\mathbf{P}^T\mathbf{A}^{-1}(\epsilon)\mathbf{P}y_1} = \min_{x=\mathbf{P}y_1} \frac{x^T\mathbf{D}^{-1}x}{x^T\mathbf{A}^{-1}(\epsilon)x}$$

$$\geq \min_x \frac{x^T\mathbf{D}^{-1}x}{x^T\mathbf{A}^{-1}(\epsilon)x} = \lambda_{min}(\mathbf{A}(\epsilon)\mathbf{D}^{-1}).$$

Let $\epsilon \to 0$ on both sides of the above inequality, we get

$$\lambda_{min}(-\mathbf{AD}^{-1}) \leq \lambda_{min}(-\mathbf{A}^*\mathbf{D}^{*^{-1}}),$$

which is equivalent to (*iii*) since

$$\lambda_{max}(\mathbf{A}) = -\lambda_{min}(-\mathbf{A})$$

is true for any matrix $\mathbf{A}$.

(*iv*) An analogous discussion as in the proof of (*iii*). ■

Theorem 5.2.4 shows that the positive and negative eigenvalues of $\mathbf{D}^{*^{-1}}\mathbf{G}^{*^T}$ are convex combinations of the positive and negative eigenvalues of matrix $\mathbf{D}^{-1}\mathbf{G}^T$. We see from various numerical examples that they are actually interlaced with each other very well. The curves of eigenvalues of $\mathbf{D}^{*^{-1}}\mathbf{G}^{*^T}$ and $\mathbf{D}^{-1}\mathbf{G}^T$ are of the same shape.

## 5.3 Numerical Results

As was mentioned earlier, we use the probability $G^*(x) := 1 - \mathbf{1}^T\mathbf{F}^*(x)$ of overflow of the induced OGFS model and its variants to estimate the probability $G(x)$ of overflow of the original MGFS model. The procedure for obtaining $G^*(x)$ is the same as that for the OGFS model, which is described in the previous section. In other words, we look for solution

$$\begin{aligned} G^*(x) &= 1 - \mathbf{1}^T\mathbf{F}^*(x) \\ &= -\sum_{i:z_i^*<0} e^{z_i^* x} a_i^* \mathbf{1}^T \phi_i^* \end{aligned} \tag{5.3.40}$$

where $\mathbf{1} = [\,1, 1, \cdots, 1\,]^T$ with suitable dimension. Since $G(x)$ and $G^*(x)$ behave asymptotically like $-a_0\mathbf{1}^T\phi_0 e^{z_0 x}$ and $-a_0^*\mathbf{1}^T\phi_0^* e^{z_0^* x}$ respectively, and $z_0 > z_0^*$ by Theorem 5.2.4, we have

**Corollary 5.3.1** *The probability of overflow of the induced OGFS model is asymptotically a lower bound for that of the original MGFS model, i.e., for $x$ large enough,*

$$G^*(x) < G(x).$$

In Figure 5.1 through Figure 5.3 we compare $\log G(x)$ with $\log G^*(x)$ for three two-dimensional models in which $N_1 = 10, N_2 = 15$ are fixed and the depleting rates $c$ are set to be *traffic intensity* $\rho$ plus 1.85. Table 5.1 shows the comparison of some interesting parameters corresponding to those models. The parameters are:

1. the probabilities $G(0)$ and $G^*(0)$ of non-empty buffers,

2. the mean buffer contents *mean* and *mean*$^*$,

3. the principal eigenvalues $z_0$ and $z_0^*$ and

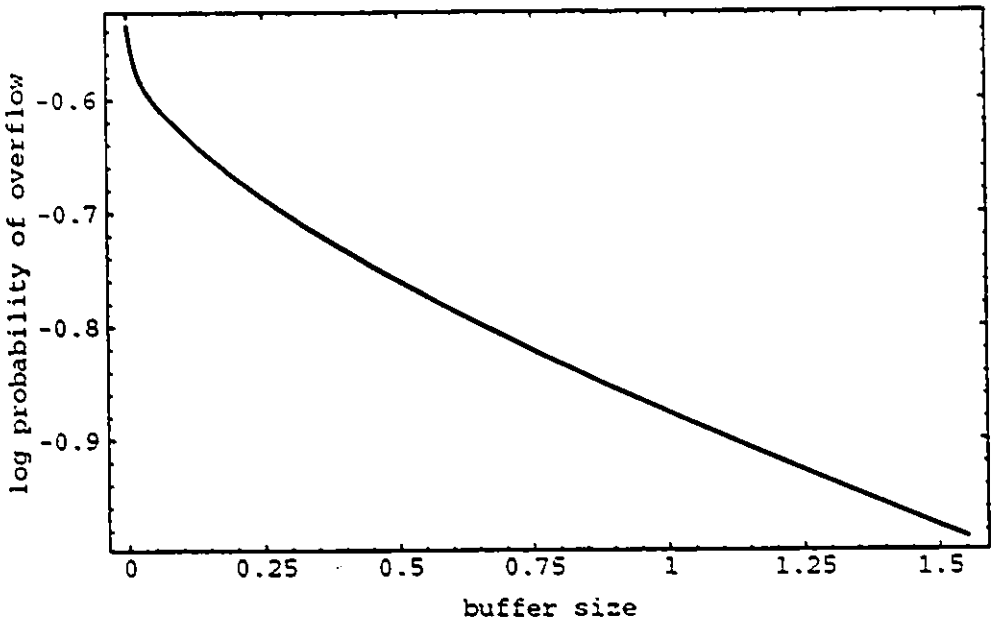


Figure 5.1: $\log G^*(x)$ vs $\log G(x)$ for traffic intensity $\rho = 21.0$.

4. the coefficients $-a_0 \mathbf{1}^T \phi_0$ and $-a_0^* \mathbf{1}^T \phi_0^*$ of the dominant terms in $G(x)$ and $G^*(x)$.

The lower bounds are good for various situations, especially when the values of $x$ are not too large. This motivates us to combine our method with the asymptotics given by Kosten (1984). Precisely, we solve the polynomial equations (5.1.13) and

| *No.* | $\rho$ | $G(0)$ | $G^*(0)$ | *mean* | *mean*$^*$ | $-z_0$ | $-z_0^*$ | $-a_0 \mathbf{1}^T \phi_0$ | $-a_0^* \mathbf{1}^T \phi_0^*$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 21.0 | .58562 | .58564 | 3.0425 | 3.0424 | .14957 | .14958 | .43386 | .43398 |
| 2 | 15.5 | .57889 | .57652 | 2.3872 | 1.9052 | .18322 | .23704 | .40906 | .42863 |
| 3 | 6.6 | .46063 | .45875 | 1.7760 | 1.4271 | .16257 | .21611 | .22968 | .26053 |

Table 5.1: Comparison of some important parameters.

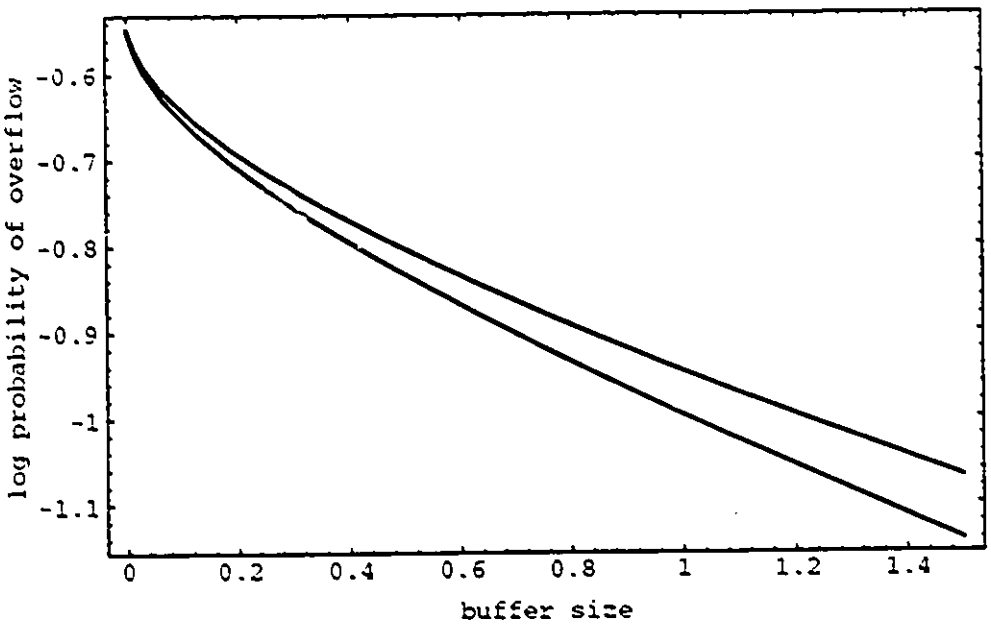


Figure 5.2: $\log G^*(x)$ vs $\log G(x)$ for traffic intensity $\rho = 15.5$.

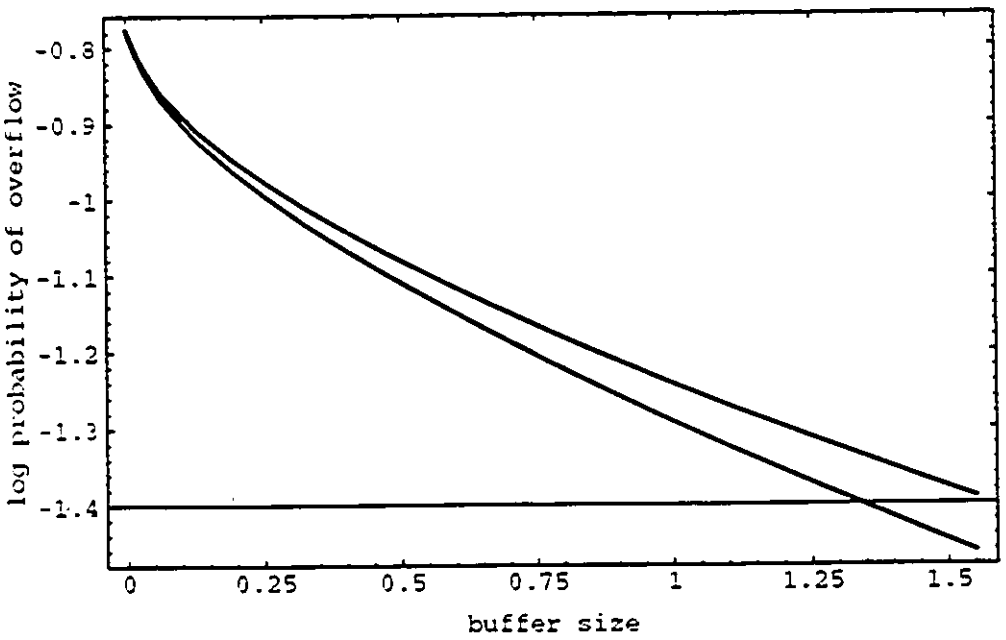


Figure 5.3: $\log G^*(x)$ vs $\log G(x)$ for traffic intensity $\rho = 6.6$.

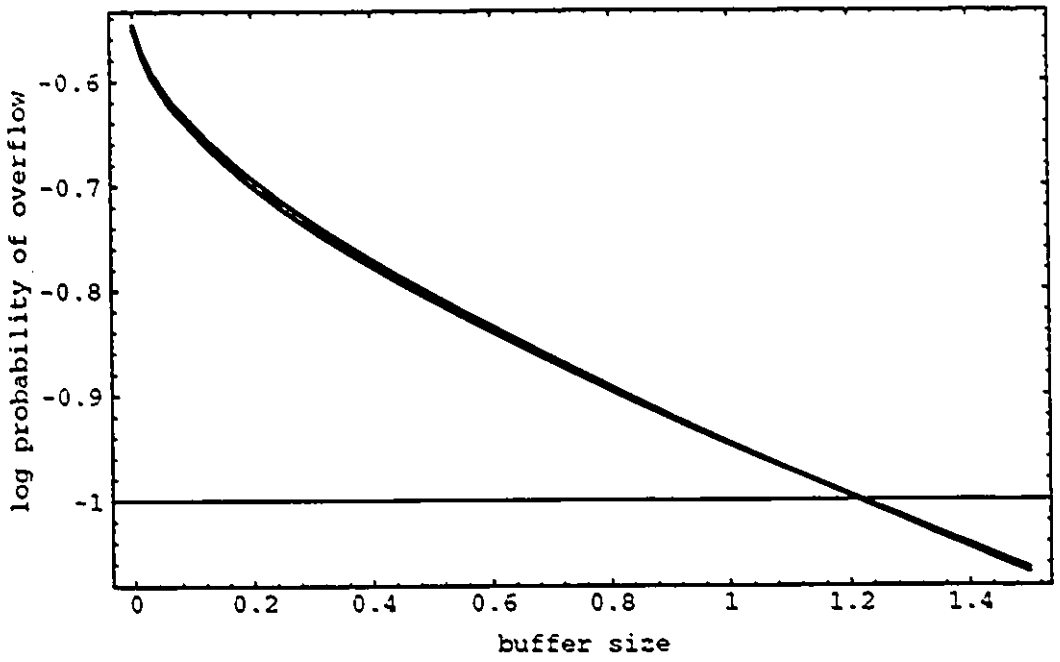


Figure 5.4: $\log G^{**}(x)$ vs $\log G(x)$ for traffic intensity $\rho = 15.5$.

use (5.1.14) to find the largest negative eigenvalue $z_0$ and then substitute $z_0^*$ in $G^*(x)$ by $z_0$. The result, denoted by $G^{**}(x)$, is a better approximation to $G(x)$ as we can see from Figure 5.4 which is to be compared with Figure 5.2. The approximation of the mean buffer content is also improved by this combined method.

## 5.4 Comments

We present the induced Dirichlet form method to reduce the dimension of computing the probability of overflow of the buffer content in a MGFS model. What we get are lower bounds and approximations rather than the exact solutions. Since the number of states for the induced process $N_t^*$ is $\sum_{i=1}^m y_i N_i + 1$, which can be much smaller than $\prod_{i=1}^m (N_i + 1)$, the number of states of the original multidimensional process $N_t$, this method can provide good approximations to some interesting parameters for MGFS models, while the exact solution is intractable due to the huge size of the state space. This is especially true when $y_i$'s are comparatively small. In our analysis we assumed that $y_i$'s are distinct integers and one of them is one. It is obvious that with some equal $y_i$'s things becomes even easier. So this technique requirement can be achieved by rescaling and rounding up $y_i$'s and $c$. In fact, this minimizes computation. The numerical examples given in this chapter are experimental to illustrate the method. Since we wanted to compare the approximations with the exact solutions we had to choose small models.

For the induced OGFS model, we do not have simple formulas, as those obtained for the standard OGFS models given by AMS[8], for computing the eigenvalues and eigenvectors and we do have to solve matrix equations to find the initial conditions. The generating function method fails for the induced Markov jump process with nonlinear jumping rates. Further study of this issue is needed.

Apparently, the technique introduced in this chapter can be applied to the MGIS models. The technique applies moreover to transient problems associated with these models such as the time until the buffer is overflow.

## 5.5 Appendix

We give a matrix representation for the procedure of inducing a one-dimensional Markov jump process from a general multi-dimensional birth and death process. Lemma 5.2.1 is then proved in a general form.

Let $N_t(r)$ be a stationary birth and death process on the space $S_r \subset Z^+$ with birth rate $a_i(r)$ and death rate $b_i(r)$ for $r = 1, 2, \cdots, m$ and $i \in S_r$. The generator $\mathbf{G}_r := (g_{ij})$ of $N_t(r)$ is defined as

$$g_{ij}(r) = a_i(r)(1_{\{j=i+1\}} - 1_{\{j=i\}}) + b_i(r)(1_{\{j=i-1\}} - 1_{\{j=i\}}).$$

Suppose $N_t(r), r = 1, 2, \cdots, m$ are independent and stable. Define

$$N_t = (N_t(1), \cdots, N_t(m)).$$

Then $N_t$ is a multi-dimensional birth and death process on the space

$$S = S_1 \times S_2 \times \cdots \times S_m := \{\vec{n} = (n_1, \cdots, n_m)\}.$$

Suppose an order is taken in the space such that the generator $\mathbf{G} := (g_{ij})$ is defined by

$$\begin{aligned} g_{ij} &= \sum_{r=1}^{m} a_{n_r}(r)(1_{\{\vec{n}_j = \vec{n}_i + \delta_r\}} - 1_{\{\vec{n}_j = \vec{n}_i\}}) \\ &+ \sum_{r=1}^{m} b_{n_r}(r)(1_{\{\vec{n}_j = \vec{n}_i - \delta_r\}} - 1_{\{\vec{n}_j = \vec{n}_i\}}) \end{aligned} \tag{5.5.41}$$

where $\vec{n}_i = (n_1, \cdots, n_m) \in S$. Let $\pi(\vec{n})$ be the strictly positive, stationary distribution of $N_t$. Since $N_t$ is reversible with respect to $\pi$, $\mathbf{G}$ can be symmetrized. Define $\mathbf{A} := (a_{ij})$ as follows

$$a_{ij} = \sqrt{\frac{\pi(\vec{n}_i)}{\pi(\vec{n}_j)}} g_{ij}.$$

Associate with $N_t$ we can define the Dirichlet (zero) form

$$\mathcal{E}(u,u) = \frac{1}{2}\sum_{\vec{n}_i,\vec{n}_j}(u(\vec{n}_i)-u(\vec{n}_j))^2\pi(\vec{n}_i)g_{ij} \tag{5.5.42}$$

$$= \sum_{\vec{n}\in S}\sum_{r=1}^{m}(u(\vec{n})-u(\vec{n}+\delta_r))^2 a_{n_r}(r)\pi(\vec{n}).$$

Here we have used the reversibility of $N_t$ and (5.5.41). Define $f : S \mapsto S^* \subset Z^+$ such that

$$f(\vec{n}) = \sum_{i=1}^{m} y_i n_i$$

where $y_i$ are distinct positive integers and $y_1$ is equal to one. Then $f$ induces a probability distribution $\pi^*$ defined by

$$\pi^*(k) = \sum_{\vec{n}\in S: f(\vec{n})=k}\pi(\vec{n}).$$

It also induces a regular Dirichlet form $\mathcal{E}^*$ on $R^+$.

$$\mathcal{E}^*(h,h) = \mathcal{E}(h\circ f, h\circ f)$$

$$= \sum_{k\in S^*}\sum_{\vec{n}\in S:f(\vec{n})=k}\sum_{r=1}^{m}[\,(h(k+y_r)-h(k))^2 a_{n_r}(r)\,]\pi(\vec{n})$$

$$= \sum_{k\in S^*}\sum_{r=1}^{m}(h(k+y_r)-h(k))^2\left(\sum_{\vec{n}\in S:f(\vec{n})=k}\frac{a_{n_r}(r)\pi(\vec{n})}{\pi^*(k)}\right)\pi^*(k).$$

The form $\mathcal{E}^*$ is associated with a reversible Markov jump process $N_t^*$ on $S^*$ having stationary measure $\pi^*$ and infinitesimal generator $\mathbf{G}^*$, which is given by

$$g_{ij}^* = \sum_{r=1}^{m}\sum_{\vec{n}\in S:f(\vec{n})=i}\frac{a_{n_r}(r)\pi(\vec{n})}{\pi^*(i)}(1_{\{j=i+y_r\}}-1_{\{j=i\}}) \tag{5.5.43}$$

$$+ \sum_{r=1}^{m}\sum_{\vec{n}\in S:f(\vec{n})=i}\frac{b_{n_r}(r)\pi(\vec{n})}{\pi^*(i)}(1_{\{j=i-y_r\}}-1_{\{j=i\}})$$

The symmetric version of $\mathbf{G}^*$ is $\mathbf{A}^* := (a_{ij}^*)$, where

$$a_{ij}^* = \sqrt{\frac{\pi^*(i)}{\pi^*(j)}}g_{ij}^* \tag{5.5.44}$$

$$= \frac{1}{\sqrt{\pi^*(i)\pi^*(j)}} \sum_{r=1}^{m} \sum_{\vec{n}\in S: f(\vec{n})=i} \pi(\vec{n}) [\, a_{n_r}(r)(1_{\{j=i+y_r\}} - 1_{\{j=i\}})$$

$$+ \; b_{n_r}(r)(1_{\{j=i-y_r\}} - 1_{\{j=i\}})\,]$$

**Proof of Lemma 5.2.1:** The $ij^{th}$ element of $\mathbf{P}^T\mathbf{P}$ is

$$\sum_{\vec{n}_l\in S} p_{il}p_{jl} = \sum_{\vec{n}_l\in S} \sqrt{\frac{\pi(\vec{n}_l)}{\pi^*(i)}}\sqrt{\frac{\pi(\vec{n}_l)}{\pi^*(j)}} 1_{\{f(\vec{n}_l)=i, f(\vec{n}_l)=j\}}$$

$$= 1_{\{j=i\}}$$

This proves the orthonormal property of $\mathbf{P}$.

The $ij^{th}$ element of matrix $\mathbf{P}^T\mathbf{A}\mathbf{P}$ is

$$\sum_{k,l} p_{ik}a_{kl}p_{jl} = \sum_{\vec{n}_k,\vec{n}_l\in S} \sqrt{\frac{\pi(\vec{n}_k)}{\pi^*(i)}}\sqrt{\frac{\pi(\vec{n}_l)}{\pi^*(j)}} a_{kl} 1_{\{f(\vec{n}_k)=i, f(\vec{n}_l)=j\}}$$

$$= \sum_{\vec{n}_k,\vec{n}_l\in S} \sqrt{\frac{\pi(\vec{n}_k)}{\pi^*(i)}}\sqrt{\frac{\pi(\vec{n}_l)}{\pi^*(j)}}\sqrt{\frac{\pi(\vec{n}_k)}{\pi(\vec{n}_l)}} g_{kl} 1_{\{f(\vec{n}_k)=i, f(\vec{n}_l)=j\}}$$

$$= \frac{1}{\sqrt{\pi^*(i)\pi^*(j)}} \sum_{l,k} g_{kl}\pi(\vec{n}_k) 1_{\{f(\vec{n}_k)=i, f(\vec{n}_l)=j\}}$$

$$= \frac{1}{\sqrt{\pi^*(i)\pi^*(j)}} \sum_{l,k} \sum_{r=1}^{m} [\, a_{n_r}(r)(1_{\{\vec{n}_l=\vec{n}_k+\delta_r\}} - 1_{\{\vec{n}_l=\vec{n}_k\}})$$

$$+ \; b_{n_r}(r)(1_{\{\vec{n}_l=\vec{n}_k-\delta_r\}} - 1_{\{\vec{n}_l=\vec{n}_k\}})\,]\pi(\vec{n}_k) 1_{\{f(\vec{n}_k)=i, f(\vec{n}_l)=j\}}$$

$$= \frac{1}{\sqrt{\pi^*(i)\pi^*(j)}} \sum_{r=1}^{m} \sum_{k} [\, a_{n_r}(r)\pi(\vec{n}_k) \sum_{l} (1_{\{\vec{n}_l=\vec{n}_k+\delta_r\}} - 1_{\{\vec{n}_l=\vec{n}_k\}})$$

$$\times \; 1_{\{f(\vec{n}_k)=i, f(\vec{n}_l)=j\}}$$

$$+ \; b_{n_r}(r)\pi(\vec{n}_k) \sum_{l} (1_{\{\vec{n}_l=\vec{n}_k-\delta_r\}} - 1_{\{\vec{n}_l=\vec{n}_k\}}) 1_{\{f(\vec{n}_k)=i, f(\vec{n}_l)=j\}}\,]$$

$$= \frac{1}{\sqrt{\pi^*(i)\pi^*(j)}} \sum_{r=1}^{m} \sum_{\vec{n}_k\in S} \pi(\vec{n}_k) [\, a_{n_r}(r)(1_{\{j=i+y_r\}} - 1_{\{j=i\}})$$

$$+ \; b_{n_r}(r)(1_{\{j=i-y_r\}} - 1_{\{j=i\}})\,]$$

$$= a^*_{ij}.$$

The last equality is obtained by (5.5.44).

We now prove

$$(i) \quad \mathbf{D}^* = \mathbf{P}^T\mathbf{D}\mathbf{P}$$

$$(ii) \quad \mathbf{D}^{*-1} = \mathbf{P}^T\mathbf{D}^{-1}\mathbf{P}.$$

where $\mathbf{D}^*, \mathbf{D}$ are defined as in the previous sections. $(i)$ For all $i = 0, 1, \cdots, N^*, j = 1, 2, \cdots, N$, the $ij^{th}$ element of $\mathbf{P}^T\mathbf{D}\mathbf{P}$ is,

$$\begin{aligned}
\sum_{k=1}^{N} p_{ik} d_k p_{jk} &= \sum_{k=1}^{N} \sqrt{\frac{\pi(\vec{n}_k)}{\pi^*(i)}} 1_{\{f(\vec{n}_k)=i\}} \sqrt{\frac{\pi(\vec{n}_k)}{\pi^*(j)}} 1_{\{f(\vec{n}_k)=j\}} d_k \\
&= \sum_{k=1}^{N} \frac{\pi(\vec{n}_k)}{\pi^*(i)} d_k 1_{\{f(\vec{n}_k)=i=j\}} \\
&= \sum_{k=1}^{N} \frac{\pi(\vec{n}_k)}{\pi^*(i)} (i-c) 1_{\{f(\vec{n}_k)=i=j\}} \\
&= (i-c) 1_{\{i=j\}} \sum_{k=1}^{N} \frac{\pi(\vec{n}_k)}{\pi^*(i)} 1_{\{f(\vec{n}_k)=i\}} \\
&= (i-c) 1_{\{i=j\}} \\
&= d^*_{ij}
\end{aligned}$$

$(ii)$ The same as $(i)$. ■

# Bibliography

[1] Aldous, D. (1982). Markov chains with almost exponential hitting times. Stochastic Proc. Appl., 13: 305-310.

[2] Aldous, D. (1983a). On the time taken by random walks on finite groups to visit every state. Z. Wahrsch. Verw. Gebiete, 62: 3611-374.

[3] Aldous, D. (1983b). Random walks on finite groups and rapidly mixing Markov chains. In Seminaire de Probabilites XVII, pages 243-297. Springer-Verlag, New York, 1983.

[4] Aldous, D. (1989). Probability approximations via the Poisson clumping heuristic. Springer-Verlag, New York, 1989.

[5] Aldous, D., Brown, M. (1992). Inequalities for rare events in time-reversible Markov processes I. In M. Shaked and Y. L. Tong, editors, Inequalities in Statistics and Probability, Institute of Mathematical Statistics. To appear.

[6] Aldous, D., Diaconis, P (1986). Shuffling cards and stopping times. Amer. Math. Monthly, 8: 69-87.

[7] Anderson, B. D. O., Frater, M. R. (1988). Estimation of rare event statistics in data communication networks. Third Australian Teletraffic research seminar, Paper No. 4.3, 14 p.p.

[8] Anick, D., Mitra, D., Sondhi, M.M.(1982). Stochastic theory of a data-handling system with multiple sources. BSTJ 61, 1871-1894.

[9] Asmussen, S. (1987). Applied Probability and Queues. Wiley, New York, 1987.

[10] Brémaud, P. (1981). Point processes and queues: martingale dynamics. Springer-Verlag, 355 p.p.

[11] Cogburn, R. (1985). On the distribution of first passage and return times for small sets. Annals of Probability, 13: 1219-1223.

[12] Daigle, J.N., Langford, J.D.(1986). Models for Analysis of Packet Voice Communications Systems. IEEE J. Selec. Areas in Communications, Vol. SAC-4, No. 6.

[13] Daniels, H.E. (1954). Saddlepoint approximations in statistics. Ann. Math. Statist. 25, 631-650.

[14] Davies, E. B. (1980). One-parameter semigroups. Academic Press, London, New York, 230 p.p.

[15] Diaconis, P. and Stroock, D. (1991). Geometric bounds for eigenvalues of Markov chains. Ann. Appl. Probab., 1: 36-61, 1991.

[16] Dynkin, E.B. (1965). Markov Processes. Vol.1. Springer-Verlag, Berlin, 1965.

[17] Ethier, S. and Kurtz, T. (1986). Markov Processes. John Wiley & Sons, New York, 1986.

[18] Feller, W. (1957). An Introduction to Probability Theory and its Applications, Vol. 1, Wiley, N.Y.

[19] Feller, W. (1971). An Introduction to Probability Theory and its Applications. Vol. 2, Wiley, N.Y.

[20] Fukushima, M. (1980). Dirichlet forms and Markov Processes. North Holland Mathematical Library, 196 p.p.

[21] Hong, S., Perros, H.G. (1992). An approximate analysis of an ATM multiplexor with multiple heterogeneous bursty arrivals. preprint.

[22] Iscoe, I., McDonald, D.(1989). Large deviations for $l^2$-valued Ornstein-Uhlenbeck processes. Ann. Prob., Vol. 17, No. 1, 58-73.

[23] Iscoe, I., McDonald, D. (1990). Induced Dirichlet forms and capacity inequalities. Ann. Probab., Vol.18, No.3, 1195-1221.

[24] Iscoe, I., McDonald, D.(1991). Asymptotics of exit times for Markov jump processes with applications to Jackson networks. preprint.

[25] Iscoe, I., McDonald, D.(1992). Asymptotics of exit times for Markov jump processes. Submitted to Ann. Probab.

[26] Iscoe, I., McDonald, D., Qian, K. (1992a). Capacity of ATM switches. To appear in Ann. Appl. Probab.

[27] Iscoe, I., McDonald, D., Qian, K. (1992b). Asymptotics of exit distribution for Markov jump processes; application to ATM. To appear in Canadian J. of Mathematics.

[28] Karlin, S., Taylor, H. (1975). A first course in stochastic processes, second edition. Academic Press, N.Y.

[29] Kato, T (1949). On the upper and lower bounds of eigenvalues. J. Phys. Soc. Japan 4, 334-339.

[30] Keilson, J. (1966) A limit theorem for passage times in ergodic regenerative processes. The Annals of Mathematical Statistics, 37, No. 4, pp. 866-870.

[31] Keilson, J. (1979) Markov Chain models — Rarity and Exponentiality. Springer-Verlag.

[32] Kelly, F. P. (1979). Reversibility and Stochastic Networks. Wiley, 230 p.p.

[33] Kosten, L. (1974). Stochastic theory of a multi-entry buffer (I). Delft Progr. Rep., Series F: Mathematical engineering, mathematics and information engineering, 1, 10-18.

[34] Kosten, L. (1984). Stochastic theory of data handling systems with groups of multiple sources. Performance of Computer Communication Systems. 321-331.

[35] Kreyszig, E. (1978). Introductory functional analysis with applications. John Wiley & Sons, New York, 1978.

[36] Lawler, G.F., Sokal, A.D. (1988). Bounds on the $L^2$ spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. TAMS. Vol. 309, No. 2, 557-580.

[37] Li, S. (1989). Study of information loss in packet voice systems. IEEE Trans Commun., Vol. 37, No. 11, 1192-1202.

[38] Liggett,T. M. (1985). Interacting Particle Systems. Springer, New York, Heidelberg Berlin.

[39] Liggett,T. M. (1989). Exponential $L_2$ Convergence of Attractive Reversible Nearest Particle Systems. Ann. Prob., Vol. 17, No. 2, 403-432.

[40] Moser, L., Wyman, M. (1956). Asymptotic expansions. Can. J. Math., Vol. 8, 225-233.

[41] Norros, I., Roberts, J. M., Simonian, A. and Virtamo, J. T. (1991). The superposition of variable bit rate sources in an ATM multiplexer. IEEE J. Select. Areas Commun., vol. 9, No. 3, pp. 378-xxx. April 1991.

[42] Nummelin, E. (1984). General irreducible Markov chains and non-negative operators. Cambridge Univ. Press, Cambridge, England.

[43] Reed, M., Simon, B. (1978). Methods of modern mathematical physics Vol. iv: Analysis of operators. Academic Press, New York, San Francisco, London.

[44] Schweitzer, P. (1990). A survey of aggregation-disaggregation in large Markov chains. First international workshop on the numerical solution of Markov chains. Raleigh, NC.

[45] Seneta, E. (1981). Non-negative matrices and Markov chains, second edition. Springer Verlag, New York, 279 p.p.

[46] Stern, T. E. and Elwalid, A. I. (1991). Analysis of separable Markov-modulated rate models for information-handling systems. Adv. Appl. Prob. 23, 105-139.

[47] Stewart, G.W. (1971). Error bounds for approximate invariant subspaces of closed linear operators. SIAM J Numer. Anal. Vol. 8, No. 4, 796-808.

[48] Stroock, D. W. (1984). An Introduction to the Theory of Large Deviations. Springer, New York, Berlin Heidelberg, Tokyo, 196 p.p.

[49] Woodruff, G.M., Kositpaiboon, R. (1990). Multimedia traffic management principles for guaranteed ATM network performance. IEEE J. Select. Areas Commun., Vol. 8, No. 3, 437-446.