On Successive Lumping of Large Scale Systems

By Laurens C. Smit

A dissertation submitted to the

Graduate School-Newark

Rutgers, The State University of New Jersey

in partial fulfillment of requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

Written under the direction of

Dr. M.N. Katehakis

and approved by

Newark, New Jersey

May, 2014

ABSTRACT OF THE DISSERTATION

On Successive Lumping of Large Scale Systems

By Laurens C. Smit

Dissertation Director:

M.N.Katehakis

The general area of research of this dissertation concerns large systems with random aspects to their behavior that can be modeled and studied in terms of the stationary distribution of Markov chains. As the state spaces of such systems become large, their behavior gets hard to analyze, either via mathematical theory, computational algorithms or via computer simulation.

In this dissertation a class of Markov chains that we call successively lumpable is specified for which we show that the stationary probabilities can be obtained by successively computing the stationary probabilities of a propitiously constructed sequence of Markov chains. Each of the latter chains has a (typically much) smaller state space and this yields significant computational improvements.

In a successively lumpable Markov chain, we denote the states by tuples of the form

(m, i), where $m \in \mathbb{Z}$ represents the "current" level of the state and $i \in \mathbb{Z}^+$ the current phase of the state. A Markov process is called *quasi skip free* (QSF) when its transition probability matrix does not permit one step transitions to states that are two or more levels away from the current state in one direction of the level variable m.

We study the class of QSF processes for which in all levels m transitions from level m can only go "down" to a single state in level $m-1$ while "upward" transitions are not restricted. Furthermore, we study the class of QSF processes for which in all level m transitions from level m can go "down" to any state in level $m-1$ while "upward" transitions go only to one state in the highest level. We derive explicit solutions and bounds for the steady state probabilities for both classes of processes, when the process is ergodic.

These two classes of QSF processes have applications in many areas of applied probability comprising computer science, queueing theory, inventory theory, reliability and the theory of branching processes. To elaborate the applicability of the method we present explicit solutions for well known queueing models. In addition we will give examples of inventory models and restart models that also fit in the framework of successively lumpable QSF processes.

Preface

This Ph.D. dissertation entitled “On Successive Lumping of Large Scale Systems” has been prepared by Laurens C. Smit during the period November 2011 to May 2014 at the department of Management Science and Information Systems at Rutgers University, Newark and New Brunswick.

The Ph.D. project has been completed under the supervision of my advisor Professor Michael N. Katehakis. The dissertation is submitted as a partial fulfillment of the requirement for obtaining the Ph.D. degree at the Rutgers University. The project was supported by a Teaching Assistantship and an appointment as a Part Time Lecturer at Rutgers Business School, Rutgers University.

First of all, I would like to express my most sincere gratitude to my advisor Professor Michael Katehakis. I have truly appreciated the enormous amount of help he gave me and the advise to pursue a Ph.D. degree. He learned me that there is always room for improvement when completing an article, even when I got impatient: “every dot has to have a reason to be where it is”. Second, I want to thank Floske Spieksma for advising me from the Netherlands, both during my periods abroad and in Leiden. Even though the teaching situation in Leiden was very hectic and time was limited, she gave many critical remarks to the articles we produced, which was highly appreciated.

I also want to give my special thanks to the other dissertation committee members: Panagiotis Karras, Lee Papayanopoulos, Andrzej Ruszczyński and Jian Yang of Rutgers University and Ohad Perry of Northwestern University. Thank you for taking the time to read this dissertation and for giving feedback.

My doctoral program has been far from standard. Therefore I express my gratitude to everybody who helped to construct a program that fitted my irregular requests. A special thanks to Gonçalo Filipe, Monnique Desilva, Jerome Williams and Luz Kosar, who all spent a lot of time taking care of this.

I am also grateful to my family and friends (both in Newark and in Leiden), who advised me where necessary and accepted my frequent relocations.

Finally, I would like to thank Geesje, who has been incredible patient with me during my long stays abroad. She has always been my emotional backbone and supported me in finishing this degree. Without this support it would have been impossible to complete this dissertation.

Laurens Smit

Newark, NJ, March 2014

Contents

List of Figures **vii**

Introduction **1**

1 Successive Lumping **4**

1.1 Introduction to Chapter 1 4
1.2 Successively Lumpable Markov Chains 5
1.2.1 Definitions and Proofs for the First Lumping Stage 6
1.2.2 Definitions and Proofs for Successive Lumping 12
1.2.3 The Algorithm and an Example 17
1.3 Multiple Success. Lump. Markov Chains 22
1.3.1 Definitions and Proofs for Multiple Successive Lumping 23
1.3.2 The Algorithm and an Example 29
1.4 Extention to Semi Markov and Cont. Time Processes. 30

2 QSF Processes **34**

2.1 Introduction to Chapter 2 34
2.2 Definitions and Basic Notation 38
2.3 Explicit Solutions for DES QSF processes 40
2.3.1 State Space Truncations 52
2.4 Explicit solutions for RES QSF processes 55
2.4.1 State Space Truncations 60
2.5 A special case of QSF processes: QBD processes 62

3 Applications **66**

3.1 Introduction to Chapter 3 66
3.2 Two Classic Queueing Models 67
3.2.1 The $M/Er/n$ Model with Batch Arrivals. 67
3.2.2 The $Er/M/n$ Model. 72
3.3 An Inventory Model with Random Yield 75
3.4 A Restart system 76

References **80**

Curriculum Vitae **84**

List of Figures

1.1 Transition diagram of a successively lumpable Markov chain $X(t)$, arrows represent possible transitions under $\underline{\underline{\text{P}}}$. 19
1.2 First iteration. 20
a Transition diagram of $\underline{\underline{\text{U}}}_{\Delta_0}$. 20
b Transition diagram of $\underline{\underline{\text{U}}}_{\Delta_1}$. 20
1.3 Second iteration. 21
a Graphical representation of $\underline{\underline{\text{U}}}_{\Delta_2}$. 21
b Transition diagram of $\underline{\underline{\text{U}}}_{\Delta_3}$. 21
1.4 Transition diagram of $X(t)$, of Figure 1, with state space $\mathcal{X}$ partitioned by $\mathcal{D}' = \{D'_0, D'_1, D'_2, D'_3, D'_4\}$. 21
1.5 Transition diagram of a process with maximum number of positive probability transitions that is successively lumpable, cf., Remark 1.4. 22
1.6 Transition diagram of a multiple successively lumpable Markov chain. 29

2.1 Graphical representation of a DES QSF process 42
2.2 Graphical representation of a RES QSF process 57
2.3 Graphical representation of a QBD process 62

3.1 An $M/Er/n$ queueing process with batch arrivals 70
3.2 The left figure displays the transition rate diagram of the $\hat{Q}$ matrix of the $Er/M/n$ queue; the right figure is the diagram for the $M/Er/n$ queue. 75
a The $Er/M/n$ process 75
b The $M/Er/n$ process 75
3.3 An Inventory Model with random yield. 76
3.4 Transition diagram for parallel system with $M = 3$ servers. 78

Introduction

The general area of research of this dissertation concerns large systems with random aspects to their behavior that can be modeled and studied in terms of the stationary distribution of Markov chains. As the state spaces of such systems become large, their behavior gets impossible to analyze, either via mathematical theory, computational algorithms, or via computer simulation. This is an instance of the so called big state space challenge or the 'curse of dimensionality'. This phenomenon is known to confound one's ability to draw valid inference for problems that are important in many areas of science, e.g. computer science, economics, building systems, statistics, etc. Traditionally, model size reduction for Markov chains has been achieved by a form of a static (one time) lumping (or aggregation) of states.

Our current research has been devoted to steady state analysis of Markov chains. In recent work we have discovered a method of successively (dynamically) lumping the states for a class of Markov chains. This new approach can lead to big computational benefits when the method is applicable and it appears to be the most efficient computational procedure available to-date to compute the stationary probabilities for large classes of problems.

This dissertation consists of three main chapters. Chapter 1 contains the explanation of the successive lumping procedure. It is thoroughly described in this chapter, proven and clarified with examples. Also the possibility of having multiple successive lumpable structures within one chain is handled. It is important to note that in this dissertation we will not discuss how to find the partition over which the Markov chain is successively lumpable: we assume that this partition is given, or, like in many applications, the model will provide it. For ease of exposition the procedure is described in discrete time space. We will show that the successive lumping procedure also works for continuous time Markov chains and for semi-Markov chains.

Chapter 2 contains a usage of successive lumping that is introduced in the context of QSF processes; these are two dimensional Markov processes in which one step transitions in one direction can only go one level higher or lower. A general solution procedure other than a 'brute force method' is not known for these processes.

In line with the well known matrix analytic method we are trying to compute a rate matrix set, used to express the steady state distribution of a level in terms of the steady state distributions of lower (or higher) levels. We will show how successive lumping can be used to find an exact rate matrix set, something which is a hard problem in general. We derive solutions for two different types of QSF processes that are successively lumpable. As far as we know, there are hardly any algorithms known to find an exact representation of the rate matrix set, only indirect iterative solution procedures that require to find the solution to a quadratic matrix equation.

The subset of QSF processes are the well studied QBD processes, in which transitions are only allowed one level up and down from any state. We will show in Chapter 2.5 how our derived results simplify for QBD processes.

Chapter 3 deals with several applications of successively lumpable processes. Many of these applications are in queueing, we can for example find an exact solution (the probability that a certain number of customers is in the system) for a queuing system where customers are served according to a Erlang process. The number of servers can vary, and various numbers of servers can be compared using the successive lumping procedure. We will also provide solutions to notoriously hard problems in inventory management and reliability theory. The usage of successive lumping in these models allows us to calculate the steady state probabilities and establish various performance measures efficiently.

To summarize, Chapter 1 deals with successive lumpable Markov processes, and is based on Katehakis and Smit (2012b). Chapter 2 deals with QSF processes and is based on Katehakis et al. (2013). A specialization of QSF processes to QBD processes can be found in Section 2.5 and shows how all proofs simplify in this case. Chapter 3 deals with applications of successive lumping, both in the QSF setting as in other frameworks. Most of these applications are described in Katehakis et al. (2013), but we added some to this dissertation for extra clarification.

CHAPTER 1

Successive Lumping

1.1 Introduction to Chapter 1

In this chapter we identify a class of Markov chains that we call successively lumpable, for which it is shown that the stationary probabilities can be obtained by successively computing the stationary probabilities of a propitiously constructed sequence of Markov chains. Each of the latter chains has a, typically much, smaller state space and a successive method of solution becomes possible with significant computational improvements. Lumping of states was first discussed in Kemeny and Snell (1960). Methods and benefits of aggregation/disaggregation are thoroughly described in Schweitzer et al. (1984), Miranker and Pan (1980) and Yap (2009). In our construction the new key idea is to identify conditions (cf., Definition 1.2) on the transition matrix of the Markov chain under which it is *successively lumpable*. A necessary condition for a chain to be successively lumpable is the existence of "entrance states" cf., Definition 1.2. These states are called "input states" by Feinberg and Chui (1987) and they are a special case of the "mandatory states" which have been studied in Kim and Smith (1989) and Kim and Smith (1990).

This chapter is organized as follows. In Section 1.2.1 after some preliminaries we provide the basic framework for the first lumping stage. The successively lumpable class of Markov chains is defined in Section 1.2.2 and their main properties are given in Theorems 1.2 and 1.3. These theorems are the main results of this chapter. In Section 1.3 another class of Markov chains is introduced for which, using our results of Section 1.2, we construct a multiple successive lumping procedure. In Section 1.4, we discuss the ramifications of the work in Sections 1.2 and 1.3 to the case of semi-Markov processes and continuous time Markov processes.

Later in this dissertation, in Chapter 3, we study applications of successively lumpable Markov chains to certain classical reliability/queueing problems. Versions of these reliability/queueing models have been studied before in Derman et al. (1980), Frostig (1999), Hooghiemstra and Koole (2000), Katehakis and Derman (1989) as well as in Katehakis and Melolidakis (1995), Righter (1996), Zhang et al. (2007) and references therein.

1.2 Successively Lumpable Markov Chains

Let $X(t)$ denote an irreducible and positive recurrent Markov chain on a finite or countable state space $\mathcal{X}$. Clearly $\mathcal{X}$ can be partitioned into a (possibly infinite) sequence of mutually *exclusive* and *exhaustive* sets $\mathcal{D} := \{D_0, D_1, \ldots, D_M\}$, with $M \leq \infty$, $\cup_{m=0}^{M} D_m = \mathcal{X}$, and $D_m \cap D_{m'} = \emptyset$, when $m \neq m'$. For notational convenience, the elements of each set D_m will be denoted (relabelled) as $\{(m,1),(m,2),\ldots,(m,\ell_m)\}$, for some fixed constants $\ell_m \leq \infty$. The transition matrix of $X(t)$ will be denoted by $\underline{\underline{P}} = [p(m,j\,|\,m',j')]$, where its $((m',j'),(m,j))-$element is

$$p(m,j\,|\,m',j') = \mathbf{Pr}[X(t+1)=(m,j)\,|\,X(t)=(m',j')].$$

In the sequel we will denote the stationary probabilities with the elements $\pi(m,j) = \lim_{t\to\infty}\mathbf{Pr}[X(t)=(m,j)]$. These probabilities exist, because the Markov chain $X(t)$

is irreducible and positive recurrent. We will use the notation

$$\underline{\pi} = \big(\pi(0,1),\ldots,\pi(0,\ell_0),\pi(1,1),\ldots,\pi(1,\ell_1),\ldots,\pi(M,1),\ldots,\pi(M,\ell_M)\big).$$

In the sequel, to avoid trivial cases we assume that $M \geq 2$, i.e., the partition $\mathcal{D}$ has at least two subsets. Note also, that we will use the symbol $\underline{\underline{\mathrm{A}}}$ to denote a matrix where $a(i,j)$ will denote its (i,j)-th element and $\underline{a}(i)$ (respectively $\underline{a}'(j)$) will denote its i-th row (respectively j-th column) vector.

1.2.1 Definitions and Proofs for the First Lumping Stage

We start with the definition of the **entrance state** of a subset D_m of a partition $\mathcal{D}$ of the state space $\mathcal{X}$.

Definition 1.1. A subset D_m of $\mathcal{D}$ has an **entrance state** $(m, \boldsymbol{\varepsilon}_m(\mathcal{D})) \in D_m$ if and only if

$$p(m,j\,|\,m',j') = 0, \text{ for all } m' \neq m \text{ with } j \neq \boldsymbol{\varepsilon}_m(\mathcal{D}), \text{ and all } j' \in D_{m'}.$$

Remark 1.1. i) Note that from the positive recurrence assumption it follows that if $D_m \in \mathcal{D}$ has an entrance state, there exists some $(m',j') \in D_{m'}$ with $m' \neq m$ such that

$$p(m,\boldsymbol{\varepsilon}_m(\mathcal{D})\,|\,m',j') > 0.$$

ii) An entrance state of a set D_m is the *only* state via which the set D_m can be entered by the chain $X(t)$ from a state in $\mathcal{X}\backslash D_m$, where given two sets A and B, $A\backslash B$ denotes the elements of A that do not belong to B.

iii) Note also that in the familiar one dimensional notation for the states, a subset D of $\mathcal{X}$ has an **entrance state** $\boldsymbol{\varepsilon} \in D$ if

$$p(j|j') = 0 \text{ for all } j \neq \boldsymbol{\varepsilon},\ j \in D \text{ and all } j' \notin D.$$

Given a partition $\mathcal{D}$ with an entrance state $(0, \varepsilon_0(\mathcal{D})) \in D_0$ we construct the following Markov chains.

a) A Markov chain $Z_0(t)$ on state space D_0 with transition matrix $\underline{\underline{U}}_{D_0}$ which elements are as follows:

$$u_{D_0}(0,j\,|\,0,i) = \begin{cases} p(0,\varepsilon_0(\mathcal{D})\,|\,0,i) + \sum\limits_{(k,j')\notin D_0} p(k,j'\,|\,0,i), & \text{if } j = \varepsilon_0(\mathcal{D}), \\ p(0,j\,|\,0,i), & \text{otherwise.} \end{cases} \tag{1.1}$$

b) A Markov chain $X_1(t)$ with state space $\mathcal{X}_1 = \{(1,0)\} \cup D_1 \cup \ldots \cup D_M$ and transition matrix $\underline{\underline{P}}_1$ where its $((k,j),(k',j'))$-th element is defined by Eq. (1.2) below if $(k,j) = (k',j') = (1,0)$ and by Eq. (1.3), otherwise.

$$p_1(1,0\,|\,1,0) = \sum_{(0,i'),\,(0,i)\in D_0} p(0,i'\,|\,0,i)\upsilon_{D_0}(0,i), \tag{1.2}$$

$$p_1(k',j'\,|\,k,j) = \begin{cases} \sum\limits_{(0,i)\in D_0} p(k',j'\,|\,0,i)\upsilon_{D_0}(0,i), & \text{if } (k,j) = (1,0), \\ \sum\limits_{(0,i)\in D_0} p(0,i\,|\,k,j), & \text{if } (k',j') = (1,0), \\ p(k',j'\,|\,k,j), & \text{otherwise.} \end{cases} \tag{1.3}$$

It is easy to see that both chains $Z_0(t)$ and $X_1(t)$ are irreducible and positive recurrent, because $X(t)$ has these properties as well. The steady state probabilities of Markov chain $Z_0(t)$ will be denoted by $\upsilon_{D_0}(0,i) = \lim_{t\to\infty} \mathbf{Pr}[Z_0(t) = (0,i)]$. The vector of the steady state probabilities of the Markov chain $X_1(t)$, will be denoted by:

$$\underline{\pi}_1 = \big(\pi_1(1,0); \pi_1(1,1), \ldots, \pi_1(1,\ell_1), \ldots, \pi_1(M,1), \ldots, \pi_1(M,\ell_M)\big).$$

Note that in the above construction of the new process $X_1(t)$, we have introduced an artificial state we denote as $(1,0)$. This state $(1,0)$ essentially represents the "lumped states" of the set D_0 of the initial process $X(t)$; we have used a semicolon in the above notation for $\underline{\pi}_1$ to emphasize this fact.

We will use the notation $\underline{\underline{\mathrm{U}}}_{D_0} = [\underline{u}'_{D_0}(0,1), \ldots, \underline{u}'_{D_0}(0,\ell_0)]$, where $\underline{u}'_{D_0}(0,j)$ denotes the j-th column of the transition matrix $\underline{\underline{\mathrm{U}}}_{D_0}$. Similarly,

$$\underline{\underline{\mathrm{P}}} = [\underline{p}'(0,1), \ldots, \underline{p}'(0,\ell_0), \ldots, \underline{p}'(M,1), \ \ldots, \underline{p}'(M,\ell_M)].$$

It is well known that $\underline{\pi}$ is the solution to the following system of equations: $\underline{\pi}\,\underline{\underline{\mathrm{P}}} = \underline{\pi}$ and $\underline{\pi}\,\underline{1}' = 1$. Here $\underline{1}$ will always denote a vector of ones of the same dimension as $\underline{\pi}$.

We will next state and prove the following proposition and theorem.

Proposition 1.1. *If D_0 has an entrance state $(0,\varepsilon_0(\mathcal{D}))$, then the following is true for all $(0,i) \in D_0$:*

$$\upsilon_{D_0}(0,i) = \frac{\pi(0,i)}{\sum_{(0,j)\in D_0} \pi(0,j)}. \tag{1.4}$$

Proof. Let $\underline{\upsilon}_{D_0} = \big(\upsilon_{D_0}(0,1), \ldots, \upsilon_{D_0}(0,\ell_0)\big)$. It is clear that for $\underline{\upsilon}_{D_0}$, defined by Eq. (1.4), the statement $\underline{\upsilon}_{D_0}\underline{1}' = 1$ holds. To prove that this choice of $\underline{\upsilon}_{D_0}$ is the solution, we will show that it also satisfies:

$$\underline{\upsilon}_{D_0}\underline{\underline{\mathrm{U}}}_{D_0} = \underline{\upsilon}_{D_0}. \tag{1.5}$$

By uniqueness of solutions to Eq. (1.5) (with $\underline{\upsilon}_{D_0}\underline{1}' = 1$) it then follows that $\underline{\upsilon}_{D_0}$ is indeed the steady state vector.

To show that Eq. (1.5) holds, we distinguish two cases: the entrance state $(0,\varepsilon_0(\mathcal{D}))$ or any of the other states.

We will use the following derivation:

$$\begin{aligned}
\sum_{(0,j)\in D_0} p(0,\boldsymbol{\varepsilon}_0(\mathcal{D})\,|\,0,j)\pi(0,j) =&\pi(0,\boldsymbol{\varepsilon}_0(\mathcal{D})) - \sum_{(k,i')\notin D_0} p(0,\boldsymbol{\varepsilon}_0(\mathcal{D})\,|\,k,i')\pi(k,i') \\
=&\pi(0,\boldsymbol{\varepsilon}_0(\mathcal{D})) - \sum_{(k,i')\notin D_0} \Big(1 - \sum_{(k',i'')\notin D_0} p(k',i''\,|\,k,i')\Big)\pi(k,i') \\
=&\pi(0,\boldsymbol{\varepsilon}_0(\mathcal{D})) - \sum_{(k,i')\notin D_0} \pi(k,i') \\
&+ \sum_{(k',i''),(k,i')\notin D_0} p(k',i''\,|\,k,i')\pi(k,i') \\
=&\pi(0,\boldsymbol{\varepsilon}_0(\mathcal{D})) - \sum_{(k,i')\notin D_0} \pi(k,i') \\
&+ \sum_{(k',i'')\notin D_0} (\,\pi(k',i'') - \sum_{(0,j)\in D_0} p(k',i''\,|\,0,j)\pi(0,j)) \\
=&\pi(0,\boldsymbol{\varepsilon}_0(\mathcal{D})) - \sum_{(0,j)\in D_0}\sum_{(k,i')\notin D_0} p(k,i'\,|\,0,j)\pi(0,j).
\end{aligned}$$

Now, using this equality we obtain for $(0,i) = (0,\boldsymbol{\varepsilon}_0(\mathcal{D}))$:

$$\begin{aligned}
&\underline{\upsilon}_{D_0}\,\underline{u}'_{D_0}(0,\boldsymbol{\varepsilon}_0(\mathcal{D})) = \\
&= \sum_{(0,j)\in D_0} \upsilon_{D_0}(0,j)u_{D_0}(0,\boldsymbol{\varepsilon}_0(\mathcal{D})\,|\,0,j) \\
&= \frac{\sum_{(0,j)\in D_0}\pi(0,j)\Big(p(0,\boldsymbol{\varepsilon}_0(\mathcal{D})\,|\,0,j) + \sum_{(k,i')\notin D_0} p(k,i'\,|\,0,i)\Big)}{\sum_{(0,i')\in D_0}\pi(0,i')} \\
&= \frac{\pi(0,\boldsymbol{\varepsilon}_0(\mathcal{D})) - \sum_{(0,j)\in D_0}\pi(0,j)\sum_{(k,i')\notin D_0}(p(k,i'\,|\,0,j) - p(k,i'\,|\,0,j))}{\sum_{(0,i')\in D_0}\pi(0,i')} \\
&= \frac{\pi(0,\boldsymbol{\varepsilon}_0(\mathcal{D}))}{\sum_{(0,j)\in D_0}\pi(0,j)} \\
&= \upsilon_{D_0}(\epsilon_0(\mathcal{D})).
\end{aligned}$$

Similarly, for $(0,i) \neq (0, \boldsymbol{\varepsilon}_0(\mathcal{D}))$:

$$\begin{aligned}
\underline{\upsilon}_{D_0} \underline{u}'_{D_0}(0,i) &= \sum_{(0,j)\in D_0} \upsilon_{D_0}(0,j) u_{D_0}(0,i \,|\, 0,j) \\
&= \frac{1}{\sum_{(0,i')\in D_0} \pi(0,i')} \sum_{(0,j)\in D_0} \pi(0,j) p(0,i \,|\, 0,j) \\
&= \frac{\pi(0,i)}{\sum_{(0,i')\in D_0} \pi(0,i')} \\
&= \upsilon_{D_0}(0,i).
\end{aligned}$$

Thus, $\underline{\upsilon}_{D_0} \underline{u}'_{D_0}(0,i) = \upsilon_{D_0}(0,i)$ for all $(0,i) \in D_0$ and the proof is complete. □

For the chain $X_1(t)$ we have the following main result concerning the steady state distribution.

Theorem 1.1. *If D_0 has an entrance state $(0, \boldsymbol{\varepsilon}_0(\mathcal{D}))$, then the following are true regarding the Markov chains $X(t)$ and $X_1(t)$.*

i) If $(k,j) \neq (1,0)$, then

$$\pi_1(k,j) = \pi(k,j),$$

ii) If $(k,j) = (1,0)$, then

$$\pi_1(k,j) = \sum_{(0,i)\in D_0} \pi(0,i).$$

Proof. We need to show that the above choice of $\underline{\pi}_1$ satisfies the steady state equations of the $X_1(t)$ process, i.e., it is the unique solution of the linear system

$$\underline{\pi}_1 \underline{\underline{\mathrm{P}}}_1 = \underline{\pi}_1,$$

together with $\underline{\pi}_1 \, \underline{1}' = 1$.

The latter equality is easy to see.

Next, for each state $(k,j) \neq (1,0)$ we have:

$$\begin{aligned}
\underline{\pi}_1\underline{p}'_1(k,j) &= \sum_{(k',j')\in\mathcal{X}_1} p_1(k,j\,|\,k',j')\pi_1(k',j') \\
&= \sum_{(k',j')\in\mathcal{X}_1\setminus\{(1,0)\}} p(k,j\,|\,k',j')\pi(k',j') \\
&\quad + \sum_{(0,i)\in D_0} p(k,j\,|\,0,i)\upsilon_{D_0}(0,i) \sum_{(0,i')\in D_0} \pi(0,i') \\
&= \sum_{(k',j')\in\mathcal{X}_1\setminus\{(1,0)\}} \pi(k',j')p(k,j\,|\,k',j') + \sum_{(0,i)\in D_0} p(k,j\,|\,0,i)\pi(0,i) \\
&= \pi(k,j) = \pi_1(k,j),
\end{aligned}$$

and thus $\pi_1(k,j) = \pi(k,j)$ satisfies the steady state equations of the $X_1(t)$ process for all $(k,j) \neq (1,0)$.

Finally, for $(k,j) = (1,0)$, we have:

$$\begin{aligned}
\underline{\pi}_1\,\underline{p}'_1(1,0) &= \sum_{(k',j')\in\mathcal{X}_1} p_1(1,0\,|\,k',j')\pi_1(k',j') \\
&= \sum_{(k',j')\in\mathcal{X}_1\setminus\{(1,0)\}} \sum_{(0,i)\in D_0} p(0,i\,|\,k',j')\pi(k',j') \\
&\quad + \sum_{(0,i),(0,i')\in D_0} p(0,i\,|\,0,i')\upsilon_{D_0}(0,i') \sum_{(0,i'')\in D_0} \pi(0,i'') \\
&= \sum_{(0,i)\in D_0} \sum_{(k',j')\in\mathcal{X}_1\setminus\{(1,0)\}} p(0,i\,|\,k',j')\pi(k',j') + \sum_{(0,i),(0,i')\in D_0} p(0,i\,|\,0,i')\pi(0,i') \\
&= \sum_{(0,i)\in D_0} \sum_{(k',j')\in\mathcal{X}} p(0,i\,|\,k',j')\pi(k',j') \\
&= \sum_{(0,i)\in D_0} \pi(0,i) = \pi_1(1,0).
\end{aligned}$$

So indeed the choice of $\pi_1(1,0) = \sum_{(0,i)\in D_0} \pi(0,i)$ satisfies the steady state equation of $X_1(t)$, that corresponds to state $(1,0)$.

The proof is now complete. □

In section 1.2.2 below we provide conditions under which it is possible to successively (or sequentially) use the lumping procedure of Section 1.2.1 over the sets $D_1,\ D_2, \ldots, D_m$. In section 1.2.3 we present the algorithm and a computational example.

1.2.2 Definitions and Proofs for Successive Lumping

We start with the following extended notation and definitions. For a Markov chain $X(t)$, with state space $\mathcal{X}$, transition matrix $\underline{\underline{P}}$ and a partition $\mathcal{D} = \{D_0, \ldots, D_M\}$, we define $\Delta_0 = D_0$, $\Delta_m = \{(m, 0)\} \cup D_m$, with $m = 1, 2, \ldots M$, where $(m, 0)$ is an artificial state, representing the lumped states: $\bigcup_{k=0}^{m-1} D_k$.

We further define the partitions $\mathcal{D}_m = \{\Delta_m, D_{m+1}, \ldots, D_M\}$ and the state spaces: $\mathcal{X}_m = \Delta_m \cup D_{m+1} \cup \ldots \cup D_M$, for $m = 0, \ldots, M$.

For notational consistency, we will use the notation: $X_0(t) = X(t)$, $\mathcal{X}_0 = \mathcal{X}$, $\mathcal{D}_0 = \mathcal{D}$, $\underline{\underline{P}}_0 = \underline{\underline{P}}$, and $\underline{\pi}_0 = \underline{\pi}$. Furthermore without loss of generality we will denote the entrance state of Δ_m in $\mathcal{D}_m$ with $(m, \epsilon_m(\mathcal{D}_m))$, if it exists.

We next state the following definition.

Definition 1.2. A Markov chain $X(t)$ is called **successively lumpable** with respect to partition $\mathcal{D} = \{D_0, \ldots, D_M\}$ if and only if the set $D_0 \cup D_1 \cup \ldots \cup D_i$ has an entrance state for all $i = 0, \ldots, M$.

The above definition means that there exists only one state in $\{D_0 \cup \ldots \cup D_{m'}\}$ that can be entered from a state in D_m when $m > m' > 0$. Note also that the definition implies that transitions out of states in $D_{m'}$ can only lead to states in D_m with $m \geq m' \geq 0$ or to the entrance state of the set $\Delta_{m'-1}$.

In the sequel, the state (m, η_m) will denote an arbitrary but fixed state in Δ_m.

Given the partition $\mathcal{D}_m$ we successively construct the following Markov chains.

a) A Markov chain $Z_m(t)$ with state space Δ_m and transition matrix $\underline{\underline{U}}_{\Delta_m}$ with

$$u_{\Delta_m}(m, j \,|\, m, i) = \begin{cases} p_m(m, j \,|\, m, i) + \sum_{(k,i') \notin \Delta_m} p_m(k, i' \,|\, m, i), & \text{if } (m, j) = (m, \eta_m), \\ p_m(m, j \,|\, m, i), & \text{otherwise.} \end{cases} \tag{1.6}$$

Suppose that the chains $Z_m(t)$ is irreducible and positive recurrent. Then its steady state probabilities will be denoted by $\upsilon_{\Delta_m}(m, i)$, i.e.,

$$\upsilon_{\Delta_m}(m, i) = \lim_{t \to \infty} \mathbf{Pr}[Z_m(t) = (m, i)].$$

b) A Markov chain $X_{m+1}(t)$ with state space $\mathcal{X}_{m+1} = \Delta_{m+1} \cup D_{m+2} \cup \ldots \cup D_M$ and transition matrix $\underline{\underline{P}}_{m+1}$, with elements $((k, j), (k', j'))$ defined by Eq. (1.7) below if $(k, j) = (k', j') = (m + 1, 0)$ and by Eq. (1.8) otherwise.

$$p_{m+1}(m + 1, 0 \,|\, m + 1, 0) = \sum_{(m,i'),\, (m,i) \in \Delta_m} p_m(m, i' \,|\, m, i) \upsilon_{\Delta_m}(m, i), \tag{1.7}$$

$$p_{m+1}(k', j' \,|\, k, j) = \begin{cases} \sum_{(m,i) \in \Delta_m} p_m(k', j' \,|\, m, i) \upsilon_{\Delta_m}(m, i), & \text{if } (k, j) = (m + 1, 0), \\ \sum_{(m,i) \in \Delta_m} p_m(m, i \,|\, k, j), & \text{if } (k', j') = (m + 1, 0), \\ p_m(k', j' \,|\, k, j), & \text{otherwise.} \end{cases} \tag{1.8}$$

Note that in order to compute $p_{m+1}(\cdot \,|\, \cdot)$ we first need to compute $\upsilon_{\Delta_m}(\cdot)$. The vector of the steady state probabilities of the chain X_{m+1} will be denoted by:

$$\underline{\pi}_{m+1}=\big(\pi_{m+1}(m,0);\pi_{m+1}(m,1),\ldots,\pi_{m+1}(m,\ell_m),\ldots,\pi_{m+1}(M,1),\ldots,\pi_{m+1}(M,\ell_M)\big).$$

We will use the notation:

$$\underline{\underline{\mathrm{U}}}_{\Delta_m} = [\underline{u}'_{\Delta_m}(m,1),\ldots,\underline{u}'_{\Delta_m}(m,\ell_m)],$$

and

$$\underline{\underline{\mathrm{P}}}_m = [\underline{p}'_m(m,0);\underline{p}'_m(m,1),\ldots,\underline{p}'_m(m,\ell_m),\ldots,\underline{p}'_m(M,1),\ldots,\underline{p}'_m(M,\ell_M)].$$

Remark 1.2. Every Markov chain is successively lumpable with respect to a partition $\mathcal{D} = \{D_0, D_1\}$ when $D_0 = \{(0,\boldsymbol{\varepsilon}_0(\mathcal{D}))\}$ is any single state and D_1 contains the remaining states.

We can now state the following proposition regarding successively lumpable Markov chains.

Proposition 1.2. *If Markov chain $X_0(t)$ is successively lumpable with respect to partition $\mathcal{D}_0$, then $X_m(t)$ is successively lumpable with respect to partition $\mathcal{D}_m$, for all $m = 1,\ldots,M$.*

Proof. To complete an induction proof we need to show that if $X_m(t)$ is successively lumpable with respect to partition $\mathcal{D}_m$, then $X_{m+1}(t)$ is successively lumpable with respect to partition $\mathcal{D}_{m+1}$.

For $m = 0$, Definition 1.2 holds by assumption on $\underline{\underline{\mathrm{P}}}_0 (= \underline{\underline{\mathrm{P}}})$. We assume the induction holds for $k = 0,\ldots,m$ and we show it holds for $m+1$. We have defined $\underline{\underline{\mathrm{P}}}_{m+1}$ in Eqs. (1.7)-(1.8).

To prove that $X_{m+1}(t)$ is successively lumpable with respect to $\mathcal{D}_{m+1}$ we first show that Δ_{m+1} has an entrance state $(m+1, \epsilon_{m+1}(\mathcal{D}_{m+1}))$.

By induction we know that $\Delta_m \cup D_{m+1}$ has an entrance state in $X_m(t)$: either $(m, \epsilon_m(\mathcal{D}_m))$ or $(m+1, i_1)$, a state in D_{m+1}. Furthermore, we know by Eq. (1.8) that for $i \neq 0$, with $k > m+1$:

$$p_{m+1}(m+1, i \,|\, k, j) = p_m(m+1, i \,|\, k, j), \tag{1.9}$$

and for $i = 0$:

$$p_{m+1}(m+1, 0 \,|\, k, j) = \sum_{(m,i') \in \Delta_m} p_m(m, i' \,|\, k, j). \tag{1.10}$$

Now, if $(m, \epsilon_m(\mathcal{D}_m))$ is the entrance state of $\Delta_m \cup D_{m+1}$ in $X_m(t)$ we get by Eq. (1.9) that $p_{m+1}(m+1, i \,|\, k, j) = 0$ for all $i > 0, k > m+1$ and thus that $(m+1, 0)$ is the entrance state of Δ_{m+1} in $X_{m+1}(t)$.

If $(m+1, i_1)$ is the entrance state of $\Delta_m \cup D_{m+1}$ in $X_m(t)$, we know by Eq.(1.9) that $p_{m+1}(m+1, i \,|\, k, j) = 0$ for all i except i_1 and by Eq.(1.10) that $p_{m+1}(m+1, 0 \,|\, k, j) = 0$. Thus $(m+1, i_1)$ is the entrance state of Δ_{m+1} in $X_{m+1}(t)$.

With a similar argument we can prove that $\Delta_{m+1} \cup \ldots \cup D_i$ has an entrance state in $X_{m+1}(t)$ for all i. Thus $X_{m+1}(t)$ is successively lumpable with respect to $\mathcal{D}_{m+1}$ when $X_m(t)$ is successively lumpable with respect to $\mathcal{D}_m$. □

Remark 1.3. Because of Proposition 1.2 we know that Δ_m has an entrance state in $X_m(t)$ for all $m \leq M$. In the construction of $Z_m(t)$, (m, η_m) was chosen arbitrarily. From now on we choose $(m, \eta_m) = (m, \epsilon_m(\mathcal{D}_m))$. Then $Z_m(t)$ is irreducible and positive recurrent as can be easily seen from a graphical representation.

We can now state the following.

Theorem 1.2. *Under the assumption of Proposition 1.2 the following are true:*

i)

$$\upsilon_{\Delta_m}(m,i) = \frac{\pi_m(m,i)}{\sum_{(m,i')\in\Delta_m}\pi_m(m,i')}. \tag{1.11}$$

ii)

$$\pi_{m+1}(k,j) = \begin{cases} \sum\limits_{(m,i')\in\Delta_m} \pi_m(m,i'), & \textit{if } (k,j)=(m+1,0),\\ \pi_m(k,j), & \textit{otherwise.}\end{cases} \tag{1.12}$$

Proof. The proof is easy to complete by induction using a similar derivation as in Proposition 1.1 and Theorem 1.1, combined with the induction result of Proposition 1.2. □

The previous results imply that the following theorem holds.

Theorem 1.3. *If $X_0(t)$ is successively lumpable with $|\mathcal{X}_0| < \infty$ the following is true:*

$$\pi_0(m,j) = \upsilon_{\Delta_m}(m,j)\prod_{k=m+1}^{M}\upsilon_{\Delta_k}(k,0),\ \ \forall(m,j)\in\mathcal{X}_0.$$

Proof. The proof follows by induction on decreasing values of $n = M, M-1, \ldots, 0$ for fixed M; note that $|\mathcal{X}_0| < \infty$ implies that M is finite.

For $n = M$, we need to show that

$$\pi_0(M,j) = \upsilon_{\Delta_M}(M,j), \forall(M,j)\in D_M.$$

Indeed, by Theorem 1.2, we have $\upsilon_{\Delta_M}(M,j) = \pi_M(M,j)/1$, where the denominator is 1 because Δ_M contains all states of $\mathcal{X}_M$. Since $j \neq 0$, (i.e. (M,j) has never been lumped by our lumping procedure) by using Theorem 1.2 repeatedly we obtain $\pi_M(M,j) = \pi_{M-1}(M,j) = \ldots = \pi_0(M,j)$, and the proof is complete for $n = M$.

We next show that the claim is true for $n = M - 1$, assuming it is true for $n = M$. So we will show that:

$$\pi_0(M-1,j) = \upsilon_{\Delta_{M-1}}(M-1,j)\prod_{k=M}^{M}\upsilon_{\Delta_k}(k,0).$$

The right hand side of the above is

$$\begin{aligned}
\upsilon_{\Delta_{M-1}}(M-1,j)\upsilon_{\Delta_M}(M,0) &= \frac{\pi_{M-1}(M-1,j)}{\sum_{(M-1,j')\in\Delta_{M-1}}\pi_{M-1}(M-1,j')}\upsilon_{\Delta_M}(M,0)\\
&= \frac{\pi_{M-1}(M-1,j)}{\sum_{(M-1,j')\in\Delta_{M-1}}\pi_{M-1}(M-1,j')}\frac{\pi_M(M,0)}{\sum_{(M,j')\in\Delta_M}\pi_M(M,j')}\\
&= \pi_{M-1}(M-1,j),
\end{aligned}$$

where the first two equalities follow from Theorem 1.2, Eq. (1.11). The last equality uses Eq. (1.12) and the fact that $\sum_{(M,\ell)\in\Delta_M}\pi_M(M,\ell) = 1$, as before. The proof for $n = M - 1$ is complete when we observe that $\pi_{M-1}(M-1,j) = \pi_{M-2}(M-1,j) = \ldots = \pi_0(M-1,j)$ since $j \neq 0$, as in the case when $n = M$.

The induction step from n to $n-1$ is easy to complete using similar algebra with albeit more cluttered equations. □

1.2.3 The Algorithm and an Example

Using the construction and results of Theorem 1.3 of the previous section, we can now state an algorithm for computing the stationary probability vector $\underline{\pi}$, of a successively lumpable Markov chain with respect to partition $\mathcal{D}$ as below.

Algorithm 1.1. [SL]

1 Construct $\underline{\underline{U}}_{D_0}$, cf., Eq. (1.1).

2 Calculate $\underline{u}_{D_0}$.

3 Lump D_0 to $(1, 0)$ and let $\Delta_1 = \{(1, 0)\} \cup D_1$.

Set $m = 1$.

While $m \leq M$

4.1 Construct $\underline{\underline{U}}_{\Delta_m}$ cf., Eq. (1.6).

4.2 Calculate $\underline{u}'_{\Delta_m}$.

4.3 Lump Δ_m to $(m + 1, 0)$ and let $\Delta_{m+1} = (m + 1, 0) \cup D_m$.

$m = m + 1$

End

5 Calculate $\underline{\pi}$, cf., Theorem 1.3.

We next clarify the previous results with a small example.

Example 1. For clarity we will number the state space according to the notation introduced in Section 1.2, so we take:

$\mathcal{X} = \{(0, 1), (0, 2), (1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2), (3, 3)\}$,

with a partition $\mathcal{D} = \{D_0, \ldots, D_3\}$ where $D_0 = \{(0, 1), (0, 2)\}, D_1 = \{(1, 1), (1, 2)\}$, $D_2 = \{(2, 1), (2, 2)\}$ $D_3 = \{(3, 1), (3, 2), (3, 3)\}$ and transition matrix $\underline{\underline{P}}$:

$\underline{\underline{P}} =$

	(0, 1)	(0, 2)	(1, 1)	(1, 2)	(2, 1)	(2, 2)	(3, 1)	(3, 2)	(3, 3)
(0, 1)	0	1/3	5/9	0	0	0	0	1/9	0
(0, 2)	0	0	1/3	2/3	0	0	0	0	0
(1, 1)	0	0	0	1/6	2/3	0	1/6	0	0
(1, 2)	0	0	0	0	1/6	3/4	0	0	1/12
(2, 1)	0	0	0	0	0	1	0	0	0
(2, 2)	0	0	0	0	0	0	0	0	1
(3, 1)	1/2	0	0	0	0	0	0	1/2	0
(3, 2)	1/2	0	0	0	0	0	0	0	1/2
(3, 3)	1/2	0	0	0	0	0	1/2	0	0

The transition diagram of the corresponding Markov chain $X(t)$ is given in Figure

1.1. It is easy to see that $X(t)$ is successively lumpable with respect to the partition $\mathcal{D}$. The first steps of the algorithm are:

1 $\underline{\underline{U}}_{\Delta_0} = \begin{bmatrix} 2/3 & 1/3 \\ 1 & 0 \end{bmatrix}$.

2 $\underline{\upsilon}_{\Delta_0} = [3/4, 1/4]$.

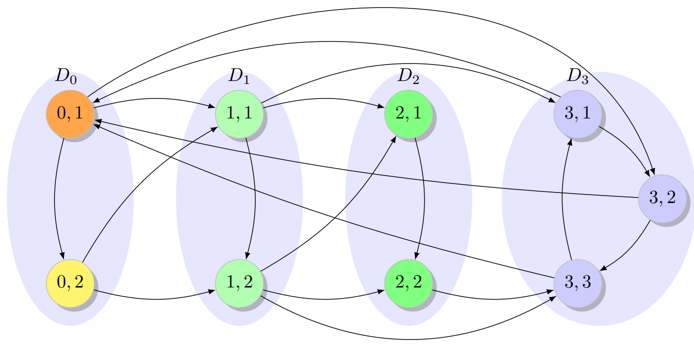


Figure 1.1: Transition diagram of a successively lumpable Markov chain $X(t)$, arrows represent possible transitions under $\underline{\underline{P}}$.

Next, we continue with $D_1 \neq \emptyset$, and note that for every positive $p(k', j' \,|\, k, j)$ with $(k', j') \in D_1$ we have $(k, j) \in D_0 \cup D_1$.

3 Lump $\{(0,1),(0,2)\}$ to $(1,0)$ and let $\Delta_1 = \{(1,0),(1,1),(1,2)\}$.

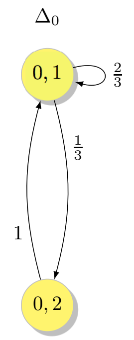
4.1 $\underline{\underline{U}}_{\Delta_1} = \begin{bmatrix} 1/3 & 1/2 & 1/6 \\ 5/6 & 0 & 1/6 \\ 1 & 0 & 0 \end{bmatrix}$.

4.2 $\underline{\upsilon}_{\Delta_1} = [4/7, 2/7, 1/7]$.

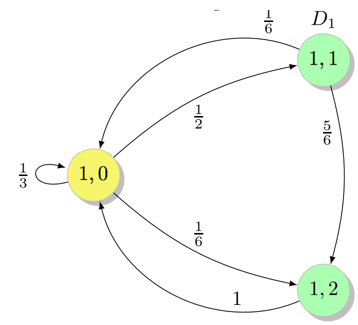
Figure 1.2b illustrates the transition diagram of this $Z_1(t)$ chain.

4.3 Lump $\{(1,0),(1,1),(1,2)\}$ to $(2,0)$ and let $\Delta_2 = \{(2,0),(2,1),(2,2)\}$. Note that since we know $\underline{\upsilon}_{\Delta_1}$ we can construct transition probabilities of a set Δ_2 without knowledge of $\underline{\pi}$ with the use of the previous states $\{(1,0),(1,1),(1,2)\}$.

4.1 $\underline{\underline{U}}_{\Delta_2} = \begin{bmatrix} 19/28 & 3/14 & 3/28 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.



(a) Transition diagram of $\underline{\underline{U}}_{\Delta_0}$.



(b) Transition diagram of $\underline{\underline{U}}_{\Delta_1}$.

Figure 1.2: First iteration.

4.2 $\underline{\upsilon}_{\Delta_2} = [28/43, 6/43, 9/43]$.

Next we look at subset D_3 and repeat the previous.

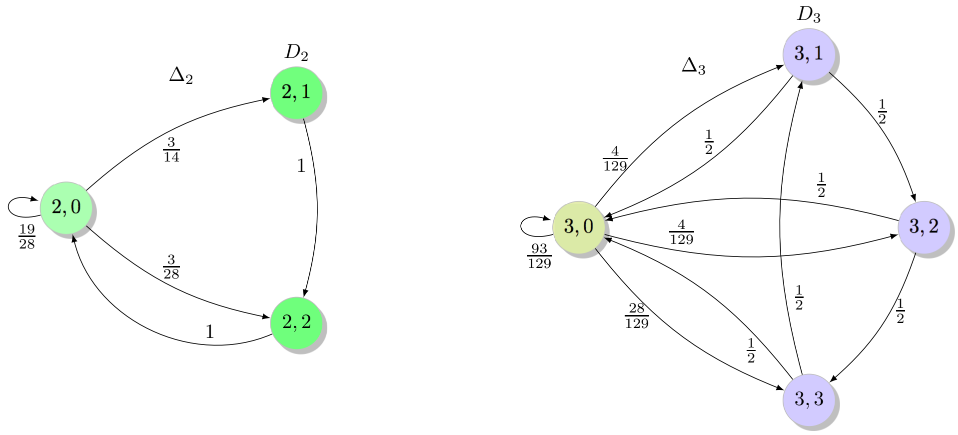
4.3 Lump $\{(2,0),(2,1),(2,2)\}$ to $(3,0)$ and let $\Delta_3 = \{(3,0),(3,1),(3,2),(3,3)\}$.

4.1 $\underline{\underline{U}}_{\Delta_3} = \begin{bmatrix} 93/129 & 4/129 & 4/129 & 4/129 \\ 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}$.

4.2 $\underline{\upsilon}_{\Delta_3} = [43/67, 142/1407, 104/1407, 248/1407]$.

We advance to step 5 and calculate $\underline{\pi}$:

$$\begin{aligned}
\pi(0,1) &= \upsilon_{\Delta_3}(3,0)\upsilon_{\Delta_2}(2,0)\upsilon_{\Delta_1}(1,0)\upsilon_{\Delta_0}(0,1) &&= \frac{43}{67}\frac{28}{43}\frac{4}{7}\frac{3}{4} &&= \frac{12}{67}\\
\pi(0,2) &= \upsilon_{\Delta_3}(3,0)\upsilon_{\Delta_2}(2,0)\upsilon_{\Delta_1}(1,0)\upsilon_{\Delta_0}(0,2) &&= \frac{43}{67}\frac{28}{43}\frac{4}{7}\frac{1}{4} &&= \frac{4}{67}\\
\pi(1,1) &= \upsilon_{\Delta_3}(3,0)\upsilon_{\Delta_2}(2,0)\upsilon_{\Delta_1}(1,1) &&= \frac{43}{67}\frac{28}{43}\frac{2}{7} &&= \frac{8}{67}\\
\pi(1,2) &= \upsilon_{\Delta_3}(3,0)\upsilon_{\Delta_2}(2,0)\upsilon_{\Delta_1}(1,2) &&= \frac{43}{67}\frac{28}{43}\frac{1}{7} &&= \frac{4}{67}\\
\pi(2,1) &= \upsilon_{\Delta_3}(3,0)\upsilon_{\Delta_2}(2,1) &&= \frac{43}{67}\frac{6}{43} &&= \frac{6}{67}\\
\pi(2,2) &= \upsilon_{\Delta_3}(3,0)\upsilon_{\Delta_2}(2,2) &&= \frac{43}{67}\frac{9}{43} &&= \frac{9}{67}\\
\pi(3,1) &= \upsilon_{\Delta_3}(3,1) &&= \frac{142}{1407}\\
\pi(3,2) &= \upsilon_{\Delta_3}(3,2) &&= \frac{104}{1407}\\
\pi(3,3) &= \upsilon_{\Delta_3}(3,3) &&= \frac{248}{1407}.
\end{aligned}$$



(a) Graphical representation of $\underline{\underline{U}}_{\Delta_2}$. (b) Transition diagram of $\underline{\underline{U}}_{\Delta_3}$.

Figure 1.3: Second iteration.

Remark 1.4.

i) To illustrate the fact that a Markov chain can be successively lumped with respect to different partitions, Figure 1.4 shows its transition diagram of the chain of example 1, where highlighted areas represent the sets of a different partition $\mathcal{D}'$ and it is easy to see that the chain is also successively lumpable with respect to partition $\mathcal{D}'$.

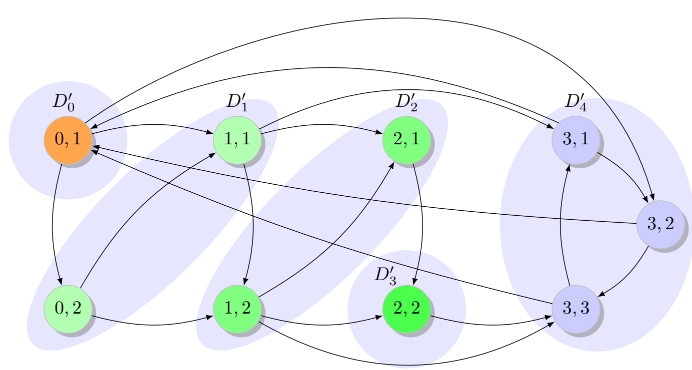


Figure 1.4: Transition diagram of $X(t)$, of Figure 1, with state space $\mathcal{X}$ partitioned by $\mathcal{D}' = \{D'_0, D'_1, D'_2, D'_3, D'_4\}$.

ii) Figure 1.5 illustrates the transition diagram of a Markov chain. An arrow from a state (m, j) to a state (m', j') is present only if the corresponding transition probability $p(m', j' \mid m, j)$ is positive; where we ignore "loop" transitions with $p(m, j \mid m, j) > 0$ that do not play a role in determining successive lumpability. The Markov chain corresponding to this transition diagram is successively lumpable with respect to partition $\mathcal{D}$ that consists of the four highlighted sets of states. This picture is interesting since it is easy to see that "adding" any additional (non-loop) arrows (i.e., transition(s), with positive probability) will result in a transition diagram of a chain for which the successive lumpability property does not hold.

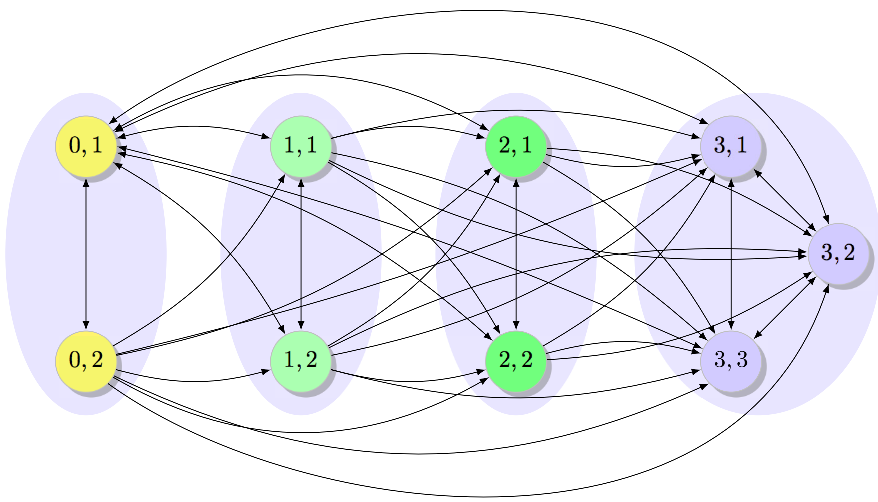


Figure 1.5: Transition diagram of a process with maximum number of positive probability transitions that is successively lumpable, cf., Remark 1.4.

1.3 Multiple Successively Lumpable Markov Chains

The main result of Section 1.2 is that when a Markov chain is successively lumpable its steady state probability vector can be calculated using successive lumping. In Section 1.3.1 we will show that it is possible to have multiple lumpable structures in one Markov chain. In this case it is possible to calculate the steady state vector using multiple times successive lumping. We also establish a product form expression, for finite state spaces. We present the algorithm and an example in Section 1.3.2.

1.3.1 Definitions and Proofs for Multiple Successive Lumping

Let $X(t)$ be a Markov chain on a finite (or countable) state space $\mathcal{X}$ with transition matrix $\underline{\underline{P}}$. We will assume that the state space $\mathcal{X}$ is composed of $N \leq \infty$ mutually exclusive and exhaustive sets, i.e.

$$\mathcal{X} = \bigcup_{n=1}^{N} \mathcal{X}^n,$$

where each subset $\mathcal{X}^n$ can be partitioned into a (possibly infinite) sequence of

$$\mathcal{D}^n = \{D_0^n, \ldots, D_{M_n}^n\}.$$

Alternatively the partition

$$\mathcal{D} = \{D_0^1, \ldots, D_{M_1}^1, \ldots, D_0^N, \ldots, D_{M_N}^N\}$$

is a sequence of $N \leq \infty$ subpartitions of $\mathcal{X}$. For notational convenience, the elements of each set D_m^n will be relabelled to a triple-notation as $\{(n, m, 1), (n, m, 2), \ldots, (n, m, \ell_{(n,m)})\}$, for given constants $\ell_{(n,m)} \leq \infty$. After this state relabeling, the transition matrix of $X(t)$ will be denoted by $\underline{\underline{P}} = [p(n', m', j' \,|\, n, m, j)]$, where the $((n, m, j)\,, (n', m', j'))$ element is given by

$$p(n', m', j' \,|\, n, m, j) = \mathbf{Pr}[X(t+1) = (n', m', j') \,|\, X(t) = (n, m, j)].$$

In the sequel, to avoid trivial cases we assume that $N \geq 2$, i.e., the partition $\mathcal{D}$ has at least two subsets.

The definition of an **entrance state** in this triple index notation is as follows.

Definition 1.3. A subset D^n_m of $\mathcal{D}$ has an **entrance state** $(n, m, \boldsymbol{\varepsilon}_{n,m}(\mathcal{D})) \in D^n_m$ iff for all $n' \neq n$, $m' \neq m$ and $j \neq \boldsymbol{\varepsilon}_{n,m}(\mathcal{D})$,

$$p(n, m, j \,|\, n', m', j') = 0.$$

Given a Markov chain $X(t)$ with partition $\mathcal{D}$ for which every subset $\mathcal{X}^n$ has an entrance state $(n, m, \epsilon_{n,m}(\mathcal{D}))$, we next study the Markov chains $X^n(t)$ on state space $\mathcal{X}^n$. We state the following definition about their transition matrix $\underline{\underline{\mathrm{P}}}^{\,n}$.

Definition 1.4. We call $X^n(t)$ the $\boldsymbol{n^{th}}$**- component Markov chain** of $X(t)$ when the

$(n, m', j' \,|\, n, m, j)$-element of $\underline{\underline{P}}^n$ is defined as follows:

a) If $(n, m', j') = (n, m', \boldsymbol{\varepsilon}_{n,m'}(\mathcal{D}))$, then

$$p^n(n, m', j' \,|\, n, m, j) = p(n, m', j' \,|\, n, m, j) + \sum_{(n', m'', j'') \notin \mathcal{X}^n} p(n', m'', j'' \,|\, n, m, j).$$

b) Otherwise,

$$p^n(n, m', j' \,|\, n, m, j) = p(n, m', j' \,|\, n, m, j).$$

We can now state the following.

Definition 1.5. A Markov chain $X(t)$ is called **multiple successively lumpable** with respect to partition $\mathcal{D} = \{\mathcal{D}^1, \ldots, \mathcal{D}^N\}$ if and only if the following conditions hold.

a) $D^n_0 \cup D^n_1 \cup \ldots \cup D^n_{M_n}$ has an entrance state in $X(t)$ for all $n = 1, \ldots, N$.

b) $D^n_0 \cup D^n_1 \cup \ldots \cup D^n_{m_n}$ has an entrance state in $X^n(t)$ for all $n = 1, \ldots, N$ and for all $m_n = 0, \ldots, M_n$.

Note that condition (a) makes the assertion that any state (n', m', j') in $D^{n'}_{m'}$ can not be entered from a state (n, m, j) in D^n_m, except when $(n', m', j') = (n', m', \varepsilon_{n',m'}(\mathcal{D}))$. Condition (b) asserts that a state (n, m', j') in $D^n_{m'}$ can not be entered from a state (n, m, j) in D^n_m when $m' < m$, except when (n, m', j') is the entrance state of $D^n_0 \cup \ldots \cup D^n_{m'}$ in $X^n(t)$.

We can now state the following lemma which shows that a multiple successively lumpable Markov chain is indeed several times successively lumpable.

Lemma 1.1. *When $X(t)$ is multiple successively lumpable with respect to $\mathcal{D}$, the n^{th}-component-Markov chain $X^n(t)$, is successively lumpable with respect to $\mathcal{D}^n$ for all $n \leq N$.*

Proof. Using their construction (cf. Definition 1.4 (a) and (b)) the transition probabilities $p^n(n, m', j' \,|\, n, m, j)$ can be shown to satisfy the conditions of Definition 1.5. □

We next introduce some notation, extending that of Section 1.2.

1. On the n^{th}- component-Markov chain $X^n(t)$ of $X(t)$ we define $\Delta^n_0 = D^n_0$, $\Delta^n_1 = \{(n, 1, 0)\} \cup D^n_1$, $\Delta^n_m = \{(n, m, 0)\} \cup D^n_m$, where states $(n, m, 0)$ are lumped states representing $\bigcup_{k=0}^{m-1} D^n_k$.

   For notational consistency we will use the notation: $X^n_0(t) = X^n(t)$, $\mathcal{X}^n_0 = \mathcal{X}^n$, $\mathcal{D}^n_0 = \mathcal{D}^n$, and $\underline{\underline{P}}{}^n_0 = \underline{\underline{P}}$. Further, we consider $X^n_m(t)$ on $\mathcal{X}^n_m$ with transitions $\underline{\underline{P}}{}^n_m$.

2. Analogously to the chains $Z_m(t)$ defined in Section 1.2, we define Markov chains $Z^n_m(t)$ with state space Δ^n_m and transition matrix $\underline{\underline{U}}{}^n_{\Delta^m_n}$.

3. Also, we define:

$$\pi(n,m,j) = \lim_{t\to\infty} \mathbf{Pr}[X(t) = (n,m,j)],$$

$$\pi^n(n,m,j) = \lim_{t\to\infty} \mathbf{Pr}[X^n(t) = (n,m,j)],$$

$$\upsilon^n_{\Delta^m_n}(j) = \lim_{t\to\infty} \mathbf{Pr}[Z^n_m(t) = (j)].$$

4. Similarly, we define $\underline{\pi}$, $\underline{\pi}^n$, $\underline{\upsilon}^n_{\Delta^m_n}$ to be the corresponding probability vectors; with dimensions $\prod_{n=1}^N \prod_{m=0}^{M_n} \ell^n_m$, $\prod_{m=0}^{M_n} \ell^n_m$, $\ell^n_m + \delta(m)$ respectively, where in the last expression the term "$\delta(m)$" is equal to one if $m > 0$, and equal to zero when $m = 0$ (note that no artificial state in Δ^n_m is used when $m = 0$).

   The elements of $\underline{\underline{\mathrm{U}}}^n_{\Delta^m_n}$ and $\underline{\underline{\mathrm{P}}}^n_{m+1}$ are akin to the elements of $\underline{\underline{\mathrm{U}}}_{\Delta_m}$ and $\underline{\underline{\mathrm{P}}}_{m+1}$, cf., Eqs. (1.6)-(1.8).

5. Finally, we define a chain $Y(t)$ with state space $\mathcal{E} = \{1,\ldots,N\}$ and transition matrix $\underline{\underline{\mathrm{Q}}}$ with its (n,n') element being equal to:

$$q(n' \,|\, n) = \sum_{(n',m',j')\in\mathcal{X}^{n'}} \sum_{(n,m,j)\in\mathcal{X}^n} \pi^n(n,m,j)p(n',m',j' \,|\, n,m,j). \tag{1.13}$$

   Note that the chain $Y(t)$ can be viewed as a chain between the different "lumped" successively lumpable chains. We will use the notation $\sigma(n)$ for the steady state probabilities of the above chain, i.e., $\sigma(n) = \lim_{t\to\infty} \mathbf{Pr}[\,Y(t) = n\,]$.

We will next show the following:

Lemma 1.2. *Assuming that $X(t)$ is a multiple successively lumpable Markov chain with its n^{th}- component Markov chain $X^n(t)$ defined as above, the following is true:*

$$\pi^n(n,m,j) = \frac{\pi(n,m,j)}{\sum_{(n,m',j')\in\mathcal{X}^n} \pi(n,m',j')}.$$

Proof. It is clear that $\underline{\pi}^n \, \underline{1}' = 1$. Now from Definition 1.5 we see that $\mathcal{X}^n$ has an entrance state $(n, m, \boldsymbol{\varepsilon}_{n,m}(\mathcal{D}))$ and therefore we can use a similar derivation as is used in Proposition 1.1 to complete the proof. □

Proposition 1.3. *For a multiple successively lumpable Markov chain $X(t)$ and with $Y(t)$ defined as above, the following is true:*

$$\sigma(n) = \sum_{(n,m,j)\in\mathcal{X}^n} \pi(n,m,j).$$

Proof. It is clear that $\underline{\sigma}\,\underline{1}' = 1$. It suffices to prove that the above choice of $\underline{\sigma}$ is the solution of the steady state equations, of the $Y(t)$ process, below:

$$\sigma(n') = \sum_{n=1}^{N} \sigma(n) q(n'|n) \quad \text{for } n' = 1, 2, \ldots, N.$$

Indeed:

$$\begin{aligned}
\sum_{n=1}^{N} \sigma(n) q(n'|n) &= \sum_{n=1}^{N} \sigma(n) \sum_{(n',m',j')\in\mathcal{X}^{n'}} \sum_{(n,m,j)\in\mathcal{X}^n} \pi^n(n,m,j) p(n',m',j' \,|\, n,m,j) \\
&= \sum_{n=1}^{N} \sum_{(n,m,j)\in\mathcal{X}^n} \pi(n,m,j) \sum_{(n',m',j')\in\mathcal{X}^{n'}} \sum_{(n,m,j)\in\mathcal{X}^n} \frac{\pi(n,m,j) p(n',m',j' \,|\, n,m,j)}{\sum_{(n,m',j')\in\mathcal{X}^n} \pi(n,m',j')} \\
&= \sum_{n=1}^{N} \sum_{(n',m',j')\in\mathcal{X}^{n'}} \sum_{(n,m,j)\in\mathcal{X}^n} \pi(n,m,j) p(n',m',j' \,|\, n,m,j) \qquad (1.14) \\
&= \sum_{(n',m',j')\in\mathcal{X}^{n'}} \sum_{n=1}^{N} \sum_{(n,m,j)\in\mathcal{X}^n} \pi(n,m,j) p(n',m',j' \,|\, n,m,j) \qquad (1.15) \\
&= \sum_{(n',m',j')\in\mathcal{X}^{n'}} \sum_{(n,m,j)\in\mathcal{X}} \pi(n,m,j) p(n',m',j' \,|\, n,m,j) \\
&= \sum_{(n',m',j')\in\mathcal{X}^{n'}} \pi(n',m',j') \\
&= \sigma(n').
\end{aligned}$$

The second equality above follows from Lemma 1.2. It is clear that the summations in Eqs. (1.14) and (1.15) can be interchanged freely. □

The main result of this section is the next theorem, for a multiple successively lumpable Markov chain $X(t)$ with respect to a partition $\mathcal{D}$ and with $|\mathcal{X}| < \infty$.

Theorem 1.4. *If $X(t)$ is multiple successively lumpable with respect to partition $\mathcal{D}$ and $|\mathcal{X}| < \infty$ then:*

$$\pi(n,m,j) = \sigma(n)\upsilon^n_{\Delta^m_n}(n,m,j) \prod_{k=m+1}^{M} \upsilon^n_{\Delta^k_n}(n,k,0) \quad \textit{for all } (n,m,j) \in \mathcal{X}.$$

Proof. Since by Lemma 1.1, $X^n(t)$ is a successively lumpable Markov chain with respect to partition $\mathcal{D}^n$ we know by Theorem 1.3 that for all n

$$\pi^n(n,m,j) = \upsilon^n_{\Delta^m_n}(n,m,j) \prod_{k=m+1}^{M} \upsilon^n_{\Delta^k_n}(n,k,0).$$

The proof is easy to complete using Lemma 1.2 and Proposition 1.3. □

Remark 1.5. When $M_n = 1$ for all n, then Theorem 1.4 becomes the main result in Feinberg and Chui (1987).

Remark 1.6. For a multiple successively lumpable Markov chain we can solve the $\prod_{n=1}^{N} M_n$ Markov chains of sizes $\ell^n_{m_n} + \delta(m_n)$ each, instead of one big system of size:

$$\prod_{n=1}^{N}\prod_{m=0}^{M_n} \ell^n_m.$$

For example, if $N = 10^4$, $M_n = 10^2$ for all n and $l^n_{m_n} = 10^4$ for all n, m, we need to solve 10^6 systems of size 10^4 instead of 1 of size 10^{10}.

1.3.2 The Algorithm and an Example

Similarly to algorithm SL for a successively lumpable Markov chain presented in Section 1.2.3, we state an algorithm for a Markov chain that is multiple successively lumpable with respect to a partition $\mathcal{D} = \{\mathcal{D}^1, \ldots, \mathcal{D}^N\}$. Again, this algorithm does not require a proof, it is a direct result of Theorem 1.4.

Algorithm 1.2. [MSL]

For $n = 1, \ldots N$

1.1 Construct $X^n(t)$ with Def. 1.4.

1.2 Call Algorithm SL and solve $X^n(t)$.

End

2 Construct $\underline{\underline{Q}}$, cf., Eq. (1.13).

3 Calculate $\underline{\sigma}$ with Proposition 1.3.

4 Calculate $\underline{\pi}$, cf., Theorem 1.4.

To clarify the algorithm, Figure 1.6 shows a multiple successively lumpable Markov chain, with $N = 2, M_1 = 2, M_2 = 2, \ell^1_{1_1} = 2, \ell^2_{1_2} = 3, \ell^1_{2_1} = 2, \ell^2_{2_2} = 3$.

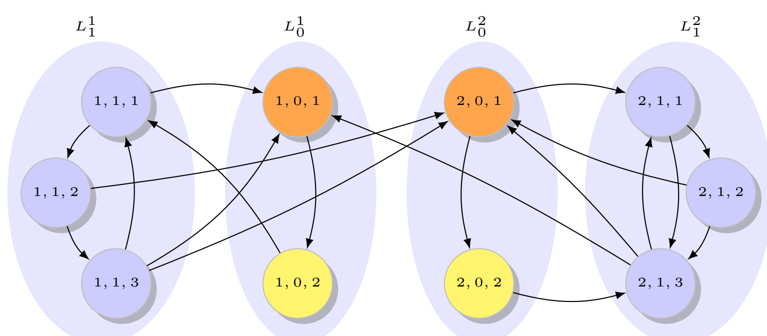


Figure 1.6: Transition diagram of a multiple successively lumpable Markov chain.

It is easy to see that in this example both $(1, 0, 1)$ and $(2, 0, 1)$ are the *entrance states* of D^1_0 and D^2_0 respectively. The procedure will solve $\mathcal{D}^1$ and $\mathcal{D}^2$ separately as successive lumpable Markov chains and conclude with solving a chain $Y(t)$ between two states representing $\mathcal{D}^1$ and $\mathcal{D}^2$. To do so, the arrows from $(1,1,2)$ and $(1,1,3)$ to $(2,0,1)$ will be "redirected" to $(1,0,1)$ and the arrow from $(2,1,3)$ to $(1,0,1)$ redirected to $(2,0,1)$ as described above.

1.4 Extension to Semi-Markov and Continuous Time Processes

It is possible to extend the successive lumpable and the multiple successive lumpable theory to semi-Markov or a continuous time Markov chains either directly or using the following construction and notation, cf., Ross (1996). A process $\{\mathcal{Z}(t),\ t \geq 0\}$ is a semi-Markov chain, on a state space $\mathcal{X}$, if it can be constructed as follows.

i) Transitions from state to state are generated from an "embedded" discrete time chain $\{X_n\,,\ n = 0, 1, \ldots\}$ with state space $\mathcal{X}$ and transition matrix $\underline{\underline{\mathrm{P}}} = [p_{ij}]$.

ii) The sojourn times (durations of a visit) for any state i are i.i.d. non-negative random variables T_i^m, $m \geq 1$, having an arbitrary distribution F_i, i.e. they are distributed as a random variable T_i with $F_i(t) = \mathbf{Pr}\,(T_i \leq t \,|\, \mathcal{Z}(0) = i)$.

A notable special case is that of a continuous time Markov process, in which case the F_i's are all exponential distributions with parameters λ_i. To avoid trivial cases with instantaneous states, it is assumed that the expected durations $\mu_i = \mathbf{E}T_i$ are finite positive constants.

Let $\mathcal{O}_i^n$ denote the time spent in states other than state i between the n^{th} and $n+1^{st}$ visits to state i. The above assumptions imply that $\mathcal{O}_i^n$, $n = 1, 2, \ldots$ are i.i.d. random variables, i.e., they are distributed as a random variable $\mathcal{O}_i$, having a distribution $G_i(t)$. The main results concerning the long run steady state probabilities ϖ_i, and π_i of $\mathcal{Z}(t)$ and X_n, respectively, are summarized in the next Theorem, cf. Ross (1996).

Theorem 1.5. *If* $\underline{\underline{P}}$ *is irreducible and if the distribution of* $T_i + \mathcal{O}_i$ *is nonlattice then the limit* $\lim_{t\to\infty} \mathbf{Pr}(\mathcal{Z}(t) = i|\mathcal{Z}(0) = j)$ *exists, it is independent of the initial state* $\mathcal{Z}(0) = j$ *and it is equal to:*

$$\varpi_i = \frac{\mathbf{E}T_i}{\mathbf{E}T_i + \mathbf{E}\mathcal{O}_i} \tag{1.16}$$

$$= \frac{\pi_i \mathbf{E}T_i}{\sum_{j\in\mathcal{X}} \pi_j \mathbf{E}T_j}. \tag{1.17}$$

Equation (1.17) provides a method for computing the steady states probabilities ϖ_i, from π_i and the expected sojourn times $\mathbf{E}T_i$. It also makes possible a similar successive construction for continuous time or Semi-Markov chains using the expected sojourn times for all states.

For completeness, we next describe the successively lumpable process for a continuous time Markov chain $X(t)$ on a state space $\mathcal{X}$ with a partition $\mathcal{D}$, of size M.

We revert back to the double index state notation of Section 1.2. The transition rates of $X(t)$ will be denoted by $\mu(k', j' \,|\, k, j)$ where:

$$\mu(k, i \,|\, k, i) = - \sum_{(k',i')\neq(k,i)} \mu(k', i' \,|\, k, i).$$

It is easy to see that $X(t)$ is successively lumpable with respect to $\mathcal{D}$ if the conditions of Definition 1.2 are valid with $p(\cdot \,|\, \cdot)$ replaced by $\mu(\cdot \,|\, \cdot)$.

We provide below the rates of the corresponding processes $Z_m(t)$ and $X_m(t)$.

For the chain $Z_0(t)$ on $D_0(=\Delta_0)$, corresponding to the process described in Eq. (1.1), its $((0,i),(0,j))-$element can be shown to be as follows:

$$\lambda_{\Delta_0}(0,j\,|\,0,i) = \begin{cases} \mu(0,j\,|\,0,i), & \text{if } i \neq \varepsilon_0(\mathcal{D}), j \neq \varepsilon_0(\mathcal{D}), \\ \sum\limits_{(k,j')\notin D_0} \mu(k,j'\,|\,0,i) + \mu(0,\varepsilon_0(\mathcal{D})\,|\,0,i), & \text{if } i \neq \varepsilon_0(\mathcal{D}), j = \varepsilon_0(\mathcal{D}), \\ \mu(0,j\,|\,0,\varepsilon_0(\mathcal{D})), & \text{if } i = \varepsilon_0(\mathcal{D}), j \neq \varepsilon_0(\mathcal{D}), \\ -\sum\limits_{(0,j')\in D_0} \mu(0,j'\,|\,0,\varepsilon_0(\mathcal{D})), & \text{if } i = j = \varepsilon_0(\mathcal{D}). \end{cases}$$

The transition rate matrix of the above rates is denoted as $\underline{\underline{\Lambda}}_{\Delta_0}$ and the steady state equations are:

$$\underline{\upsilon}_{\Delta_0}\underline{\underline{\Lambda}}_{\Delta_0} = 0,$$

and

$$\sum_{(0,i)\in D_0} \upsilon_{\Delta_0}(0,i) = 1,$$

where $\underline{\upsilon}_{\Delta_0}$ as before denotes the steady state probability vector.

As in Section 1.2 we can construct successively a sequence of processes $X_m(t)$ on $\mathcal{X}_m = \Delta_m \cup D_{m+1} \cup \ldots \cup D_M$ and $Z_m(t)$ on $\Delta_m = (m,0) \cup D_m$ (with steady state vector $\underline{\upsilon}_{\Delta_m}$) as follows for $m = 1, 2 \ldots M$.

i) For the Markov process $X_m(t)$, the transition rates $\mu_m(k',j'\,|\,k,j)$ are defined as follows.

a) If $(k,j) = (m,0)$ and $(k',j') \neq (m,0)$:

$$\mu_m(k',j'\,|\,m,0) = \sum_{(m-1,i)\in\Delta_{m-1}} \mu_{m-1}(k',j'\,|\,m-1,i)\upsilon_{\Delta_{m-1}}(m-1,i).$$

b) If $(k,j) \neq (m,0)$:

$$\mu_m(k',j'\,|\,k,j) = \begin{cases} \mu_{m-1}(m-1,\boldsymbol{\varepsilon}_{m-1}(\mathcal{D})\,|\,k,j), & \text{if } (k',j')=(m,0), \\ \mu_{m-1}(k',j'\,|\,k,j), & \text{if } (k',j')\neq(m,0). \end{cases}$$

c) And if $(k,j)=(k',j')=(m,0)$:

$$\mu_m(m,0\,|\,m,0) = -\sum_{(m,j)\in D_m} \mu_m(m,j|m,0).$$

ii) For $Z_m(t)$:

$$\lambda_{\Delta_m}(m,j\,|\,m,i) = \begin{cases} \mu_m(m,j\,|\,m,i), & \text{if } i\neq\boldsymbol{\varepsilon}_m(\mathcal{D}), j\neq\boldsymbol{\varepsilon}_m(\mathcal{D}), \\ \mu_m(m,j\,|\,m,\boldsymbol{\varepsilon}_m(\mathcal{D})), & \text{if } i=\boldsymbol{\varepsilon}_m(\mathcal{D}), j\neq\boldsymbol{\varepsilon}_m(\mathcal{D}), \\ -\sum_{(m,j')\in\Delta_m\setminus(m,\boldsymbol{\varepsilon}_m(\mathcal{D}))} \mu_m(m,j'\,|\,m,\boldsymbol{\varepsilon}_m(\mathcal{D})), & \text{if } i=j=\boldsymbol{\varepsilon}_m(\mathcal{D}). \end{cases}$$

and otherwise:

$$\lambda_{\Delta_m}(m,j\,|\,m,i) = \sum_{(k,j')\notin\Delta_m} \mu_m(k,j'\,|\,m,i) + \mu_m(m,\boldsymbol{\varepsilon}_m(\mathcal{D})\,|\,m,i).$$

CHAPTER 2

QSF Processes

2.1 Introduction to Chapter 2

In this chapter we study the class of *quasi skip free* processes, a subclass of Markov processes, for which the states can be specified by tuples of the form (m, i), where $m \in \mathbb{Z}$ represents the "current" level of the state and $i \in \mathbb{Z}^+$ the current phase of the state. The process is called *quasi skip free* (QSF) *to the left* (down) when its transition rate matrix Q (cf. Eq. (2.1)) does not permit one step transitions to states that are two or more levels away from the current state in the downwards direction of the level variable m. QSF processes generalize the well studied class of *quasi birth and death* processes (QBD) for which one step transitions to states that are two or more levels away from the current state in either direction of the level variable m are not allowed. *QSF to the right* (up) process can be defined with an apparent modification of the definition of the transition rate matrix Q. In the sequel for simplicity we will refer to a QSF process when the skip free direction is apparent (and without loss of generality taken to be the "down" direction).

Definition 2.1. The QSF processes under consideration are level dependent and ergodic processes with a transition rate matrix Q that satisfies one of the two following properties:

i) the *down entrance state* (DES) property: The structure of the nonzero elements of Q is such that *one step "down" transitions from a level m can only reach a single state in level $m-1$, for all levels m.*

ii) the *restart entrance state* (RES) property: The structure of the nonzero elements of Q is such that *one step "up" transitions from a level m can only reach a single state in level M_2 (M_2 is the highest level) for all levels m.*

The main results in this chapter are as follows. First, we derive explicit solutions, cf. Eqs. (2.14) and (2.17), for the steady state probabilities of level dependent QSFs that satisfy the DES property. Second, we use state truncation to derive tight bounds for the steady state probabilities. Then we derive explicit solutions, cf. Eqs. (2.29) and (2.32), for the steady state probabilities of level dependent QSFs that satisfy the RES property. For these type of processes we also use state truncation to derive tight bounds for the steady state probabilities.

In Chapter 3 we show that the DES and the RES property are satisfied by many models that arise in practice and we obtain explicit solutions for the well-known open problems of the $M/Er/n$ and the $Er/M/n$ queues with batch arrivals. We note that there are no explicit expressions in the literature for the rate matrix sets (and hence for the invariant measures) for any of these models. This is due to the QSF structure of the $M/Er/n$ model with batch arrivals and to the infinite to the 'down' direction state space of the $Er/M/n$ model. For either of these cases little is known outside this present research. We also demonstrate the applicability of the method with an inventory model with random yield. Regarding the RES property we show that a well known reliability (restart) problem can be easily analyzed using the proposed method. Even though the approaches used in this chapter use the successive lumping procedure described in Chapter 1 all the results presented in this chapter are useful by itself, since they provide solutions to notoriously hard to analyze processes.

Although our methodology applies to infinite values for the number of levels in the process in the downward and in the upward direction as well as for the number of phases, we take up explicitly only the case of finite values so that we can employ finite matrices in the analysis. In this way the basic features of the theory will not be obscured by additional formalism. However, we want to emphasize that all the results generalize to infinite values in a natural way using truncation, cf. Section 2.3.1 herein and Vere-Jones (1967), Tweedie (1973) and Seneta (1980). We will allow the lowest level value to be negative in order to have a natural state description for some models, for example, in the $Er/M/1$ queueing system this is necessary. We also note that the results hold for the case that either "boundary-side" of the state space is a transient class of states. However, for simplicity we will not consider this case explicitly.

The smaller class of level homogenous QBD (LHQBD) processes cf. Neuts (1981), has been used to model systems in many areas including queueing theory, cf. Riska and Smirni (2002), retrial queues, cf. Artalejo et al. (2010). For algorithmic usages of QBDs we refer to the anti-plagiarism scanner software of Viper. We note that most of the early literature devoted to level homogeneous QBDs typically follows the approach presented in Chapter 6, p. 129, in Latouche and Ramaswami (1999), whereby the computation of the steady state distribution is based on computing a rate matrix "R". This matrix is specified as one of the solutions of the, not easy to solve, matrix quadratic equation: $R^2D + RW + U = 0$, where we use our current (cf. Eq. (2.1)) notation: D, W, U for the matrices A_2 A_1, A_0 in Latouche and Ramaswami (1999). Numerical methods for computing R involve cyclic reduction Bini and Meini (1996) and logarithmic reduction Pérez and van Houdt (2011), Latouche and Ramaswami (1993). For instance R is expressed in terms of a matrix G (such that $R = U(I - W - UG)^{-1}$) which is the solution of matrix quadratic equation: $UG^2 + WG + D = 0$. For recent work on numerical methods for computing the matrix G for QBD processes of special structure we refer to van Houdt and van Leeuwaarden (2011), Etessami et al. (2010), Bini et al. (2005) and references therein. A general approach to compute G, exists only in special cases when the down transition matrix has the form $c \cdot r$ with c a column vector and r a row vector normalized to one. We note that the DES property is implied by the $c \cdot r$ structure but in our present setting we deal with the much more general class of QSF processes.

For level dependent QBD (LDQBD) processes we refer to and Chapters 8 and 12 of Latouche and Ramaswami (1999) and to Kharoufeh (2011). In Bright and Taylor (1995) recursive algorithms are given to compute the rate matrices. We note that when the LDQBD model described in Bright and Taylor (1995) satisfies the DES property then we can provide explicit solutions for the rate matrix set.

In Latouche and Ramaswami (1999), Chapter 13, p. 268-270, a method is given to analyze level homogeneous QSF (LHQSF) processes by considering them as embedded processes in suitably defined QBDs. However, as stated therein, the success of this approach has been limited. We note that the matrix analytic method has been used in Ramaswami (1988) to derive a recursive solution for the $M/G/1$ queue which is a QSF processes, using matrix 'G' described above.

The much wider class of QSF processes has the capacity to model much more general problems. Indeed in Chapter 3 we obtain explicit solutions for two well known queueing problems. In addition, QSF processes can be used to represent restart systems, cf. Katehakis and Veinott Jr. (1987), Tong et al. (2006) and Sonin (2011), chains that represent inventory systems with random yield or lead times, reliability, cf. Kapodistria (2011), and computer science and the theory of branching processes, cf. Brown et al. (2010). And as far as we know, explicit formulas for the rate matrix set of a QSF process have not been derived before.

The rest of this chapter is organized as follows. In Section 2.2 we introduce notation and formally define a QSF process. In Section 2.3 we show that the DES property implies the successive lumpable property of a QSF process. Then, we provide a recursive relation for the steady state probabilities between a level of the state space and its sub-levels. In Section 2.3.1 we show how the state space can be truncated, both in the downwards as in the upward directions. We follow the same line of thought in Section 2.4 and 2.4.1 when a QSF processes possesses the RES property. In Section 2.5 we show how our results when specialized to QBD processes provide simpler proofs and generalizations to well known theorems of the QBD literature. We will use the results of this chapter in the chapter on applications of this dissertation. In Chapter 3 we show how this methodology can be applied to the $M/Er/n$ queue with batch arrivals, to the $Er/M/n$ queue and to an inventory model with random yield.

2.2 Definitions and Basic Notation

A QSF to the left (or "down") process is a continuous time Markov process $X(t)$ on state space $\mathcal{X}$ that can be expressed as $\mathcal{X} = \bigcup_{m=M_1}^{M_2} \{(m,1),(m,2),\ldots,(m,\ell_m)\}$, where ℓ_m, M_1, M_2, are some fixed finite integers, with $1 \le \ell_m < \infty$, $-\infty < M_1 < M_2 < \infty$, and $M_1 \le m \le M_2$. A state (m,i) specifies its "current" level m and its within the level state i, with $i = 1,\ldots,\ell_m$.

Remark 2.1. The notation in this chapter differs from the notation introduced in Chapter 1, since it handles a slightly different kind of structure on the Markov chain.

For a QSF process, the transition rate matrix has the form:

$$Q = \begin{bmatrix} W^{M_1} & U^{M_1,M_1+1} & \cdots & U^{M_1,m} & U^{M_1,m+1} & \cdots & U^{M_1,M_2-1} & U^{M_1,M_2} \\ D^{M_1-1} & W^{M_1+1} & \cdots & U^{M_1+1,m} & U^{M_1+1,m+1} & \cdots & U^{M_1+1,M_2-1} & U^{M_1+1,M_2} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & W^m & U^{m,m+1} & \cdots & U^{m-1,M_2-1} & U^{m-1,M_2} \\ 0 & 0 & \cdots & D^{m+1} & W^{m+1} & \cdots & U^{m,M_2-1} & U^{m,M_2} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & W^{M_2-1} & U^{M_2-1,M_2} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & D^{M_2} & W^{M_2} \end{bmatrix}, \quad (2.1)$$

where in the above specification of Q we use the notation D^m, W^m, and $U^{m,k}$ to describe respectively the "down" (to level $m-1$), "within" (level m), and "up" (to level $k = m+1, m+2, \ldots, M_2$) transition rate sub-matrices in relation to the current level m of a state (m, i). The dimensions of these matrices are respectively $\ell_m \times \ell_{m-1}$, $\ell_m \times \ell_m$ and $\ell_m \times \ell_k$; the 0 sub-matrices of Q have position dependent dimensions so that Q is a well defined transition rate matrix.

Note that in the special case that there exist matrices D, W, U^k such that $D^m = D$, $W^m = W$ and $U^{m,k} = U^k$, for all m, the process is called (level) homogeneous. When $U^{m,m+k}$ are all 0 matrices for all $k \geq 2$ and all m, the QSF process reduces to the well studied QBD process, cf. Artalejo and Gómez-Corral (2008).

In the sequel we will assume that the QSF processes discussed are ergodic. Explicit sufficient conditions for the elements of Q can be derived to support this claim, cf. Remark 2.5; such conditions on ergodicity for multi dimensional Markov Chains can be found in Szpankowski (1988) and various criteria in Tweedie (1981), in Hordijk and Spieksma (1992) and in Spieksma and Tweedie (1994).

We consider Markov process $X(t)$, introduced above. Clearly the state space $\mathcal{X}$ of this process can be partitioned into a (possibly infinite) sequence of mutually exclusive and exhaustive *level sets*: $L_m = \{(m,1),(m,2),\ldots,(m,\ell_m)\}$.

Definition 2.2. For any fixed m we define the ***sub-level set*** of L_m to be the set of states $\underset{\sim}{L}_m = \cup_{k=M_1}^{m} L_k$ while the set $\widetilde{L}_m = \cup_{k=m}^{M_2} L_k$ is the ***super-level set*** of L_m.

We will express the steady state distribution of states in L_m in that of $\underset{\sim}{L}_m$. The presence of entrance states guarantee that sojourn times in $\underset{\sim}{L}_m$ are independent of the rate structure in $\widetilde{L}_m$.

We let $\pi(m,i)$ denote the steady state probability of state (m,i). The vectors $\pi^m := [\pi(m,1),\ldots,\pi(m,\ell_m)]$ and $\underset{\sim}{\pi}^m := [\pi^{M_1},\ldots,\pi^m]$, will denote respectively the steady state probabilities of states in level L_m and sub-level $\underset{\sim}{L}_m$. The vector of the steady state probabilities over all states will be denoted by $\pi := [\pi^{M_1},\ldots,\pi^{M_2}] = \underset{\sim}{\pi}^{M_2}$.

Using the QSF structure of Eq. (2.1) of the rate matrix Q, the (potentially non-zero) elements of the matrices D^m, W^m, $U^{m,k}$ will be denoted respectively by $d(m-1,j\,|\,m,i)$, (a "down" rate), $w(m,j\,|\,m,i)$ (a "within" rate) and $u(k,j\,|\,m,i)$, (an "up" rate) for $k>m$. Note that the diagonal elements of W^m are the negative sum of all other elements in that row of Q.

2.3 Explicit Solutions for DES QSF processes

The following lemmas show that the simple algebraic characterization of the DES property is a sufficient condition for a QSF process to be successively lumpable. Following Chapter 1, a state is an *entrance state* of a subset $\mathcal{X}_0$ of the state space $\mathcal{X}$ if all one step transitions from outside this set $\mathcal{X}_0$ into $\mathcal{X}_0$ can only occur via a transition to the entrance state.

We start with the following Lemma which characterizes an entrance state of a sub-level set L_m of a QSF process in terms of an algebraic property of the "down" transition sub-matrix D^m of its transition rate matrix Q.

Lemma 2.1. *For a QSF process $X(t)$ and for a fixed $m \in \{M_1, \dots, M_2\}$, a state $(m, \boldsymbol{\varepsilon}(L_m)) \in L_m$ is an* **entrance state** *for $\underset{\sim}{L}_m$ if and only if the following is true for all $(m+1, i) \in L_{m+1}$:*

$$d(m, j \mid m+1, i) = 0, \quad \textit{if } (m, j) \neq (m, \boldsymbol{\varepsilon}(L_m)). \tag{2.2}$$

Proof. The structure of the rate matrix Q implies that "down" transitions leaving the set $\widetilde{L}_{m+1} = L_{m+1} \cup L_{m+2} \cup \dots$ can only come from states in L_{m+1}. Further, by Eq. (2.2) the latter type of transitions are possible only when they lead into the same state $(m, \boldsymbol{\varepsilon}(L_m)) \in L_m$. □

It is easy to see that Eq. (2.2) of Lemma 2.1 is equivalent to the statement that the $\ell_m \times \ell_{m-1}$ matrix D^m has a single nonzero column.

For any fixed $n \in \{M_1, \dots, M_2\}$, let $\mathcal{D}_n$ denote the partition $\{\underset{\sim}{L}_n, L_{n+1}, \dots, L_{M_2}\}$ of $\mathcal{X}$. For a fixed n, the next lemma establishes that when D^m has a single non-zero column for all $m \geq n+1$, i.e., Q has the DES property, then the QSF process is successively lumpable with respect to the partition $\mathcal{D}_n$.

Lemma 2.2. *A QSF process is successively lumpable with respect to a partition $\mathcal{D}_{M_1}$ if D^m contains a single non-zero column vector for all $m = M_1 + 1, \dots, M_2$.*

Proof. It is a direct consequence of Lemma 2.1, that for a QSF process, a sub-level set $\underset{\sim}{L}_m$ has an entrance state $(m, \boldsymbol{\varepsilon}(L_m))$ if D^{m+1} contains a single non-zero column vector. This is true for all $m \geq M_1$. Since $\underset{\sim}{L}_m = \underset{\sim}{L}_{m-1} \cup L_m$ it follows from the definition that the chain is successively lumpable: in the notation of Chapter 1, "D_0" corresponds to $\underset{\sim}{L}_{M_1}$ and for $m > M_1$: "D_m" corresponds to L_{m-M_1}.

□

Note that when a QSF process is successively lumpable with respect to a partition

$\mathcal{D}_n$ then it is successively lumpable with respect to partition $\mathcal{D}_m$ for all $m > M_1$. A graphical representation of the transitions that are allowed in a successively lumpable QSF process can be found in Figure 2.1.

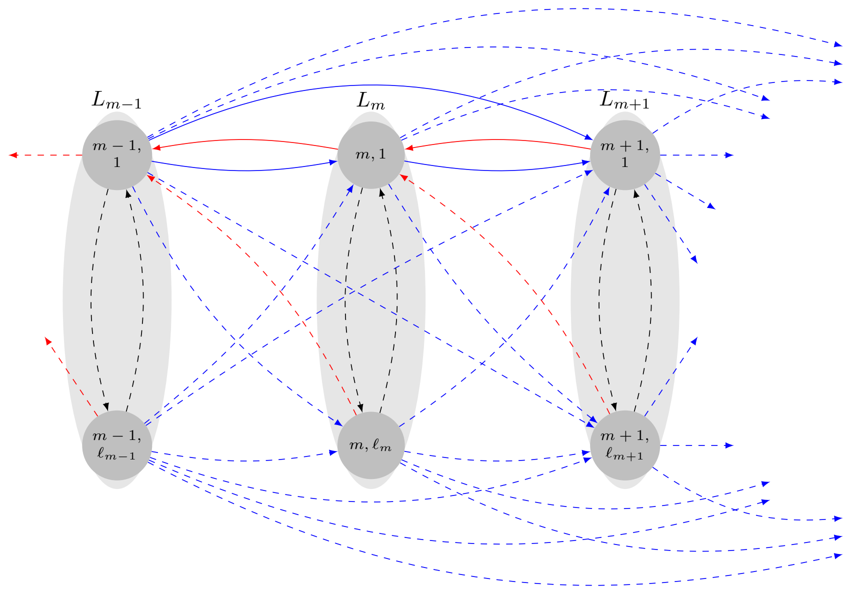


Figure 2.1: Graphical representation of a DES QSF process

We now state the following assumption that will be used in the sequel, where for notational simplicity we let the entrance state of a set $\underset{\sim}{L}_m$ be state $(m, 1)$, for all m without loss of generality.

Assumption 2.1. In this section the QSF process has a transition rate matrix Q with the following properties:

A1. The QSF process is ergodic (irreducible).

A2. For all $m \in \{M_1 + 1, \ldots, M_2\}$, only the *first column* of sub-matrix D^m contains non-zero elements, i.e., $d(m-1, 1 \,|\, m, i) > 0$ for at least one $(m, i) \in L_m$ and all other columns of D^m are equal to zero.

A3. The QSF process has bounded rates.

Remark 2.2. Part (A3) of Assumption 2.1 is given to make the proofs applicable for an infinite state space and is not strictly necessary. It can be relaxed without further conditions; it has been added to the assumption to make some of the proofs easier to state.

We will use the following notation. We first define the scalar $\underset{\sim}{\ell}_m := \sum_{k=M_1}^{m} \ell_k$ and use the symbols I_m and $\underset{\sim}{I}_m$ for the identity matrices of dimension $\ell_m \times \ell_m$ and $\underset{\sim}{\ell}_m \times \underset{\sim}{\ell}_m$, respectively. Second we define the (row)vectors of dimension ℓ_m, 0_m, 1_m and δ_m to represent a vector identically equal to 0, a vector identically equal to 1, a vector with 1 as its first coordinate and 0 elsewhere respectively. Next we define the (row)vectors of dimension $\underset{\sim}{\ell}_m$: $\underset{\sim}{0}_m$ and $\underset{\sim}{1}_m$ to be the vectors will all its coordinates equal to 0 and 1, respectively. Finally, we define the matrix $\widetilde{U}^{m,n}$ of dimension $\ell_m \times \ell_m$ by:

$$\widetilde{U}^{m,n} = \sum_{k=n}^{M_2} U^{m,k} 1'_k \delta_m, \tag{2.3}$$

where the elements of $\widetilde{U}^{m,n}$ will be denoted by $\tilde{u}(m,i)$, thus,

$$\tilde{u}(m,i) := \sum_{k=n}^{M_2} \sum_{j=1}^{\ell_k} u(k,j \,|\, m,j).$$

We next define the rate sub-matrices:

$$A_m = \begin{bmatrix} \widetilde{U}^{M_1,m+1} + U^{M_1,m} \\ \vdots \\ \widetilde{U}^{m-1,m+1} + U^{m-1,m} \end{bmatrix}, \tag{2.4}$$

$$B_m = \widetilde{U}^{m,m+1} + W^m. \tag{2.5}$$

and

$$\Gamma_m := \begin{bmatrix} A_m \\ B_m \end{bmatrix}.$$

Note that:

i) The matrix A_m contains all rates of Q corresponding to transitions from states in $\underset{\sim}{L}_{m-1}$, *into* states of L_m plus rates corresponding to transitions under Q from states in $\underset{\sim}{L}_{m-1}$ into states of $\widetilde{L}_{m+1}$, which under A_m have been re-directed to transitions into the entrance state $(m, 1)$.

ii) The $\ell_m \times \ell_m$ matrix B_m contains all rates of Q corresponding to transitions from states in L_m, *into* states of L_m plus rates corresponding to transitions under Q from states in L_m into states of $\widetilde{L}_{m+1}$, which under B_m have been re-directed to transitions into the entrance state $(m, 1)$. Thus, for $m > M_1$ since the construction of B_m excludes all down transitions it a transient transition rate matrix. However, by the definition of $\widetilde{U}^{M_1,M_1+1}$ and by its construction the matrix B_{M_1} is an $\ell_{M_1} \times \ell_{M_1}$ conservative transition rate matrix.

We next state and prove a proposition regarding basic properties of B_m.

Proposition 2.1. *The following are true:*

i) The matrix B_{M_1} is irreducible.

ii) The matrices B_m are non-singular, for all $m > M_1$.

iii) All elements of the inverse of B_m are non positive, for all $m > M_1$.

Proof. To prove i), we will show that every state (M_1, j) in level L_{M_1} that is not the entrance state $(M_1, 1)$ communicates with state $(M_1, 1)$. First, recall that in the construction of B_{M_1}, "up"-transitions are redirected to the entrance state. If there are no "up"-transitions from the communicating class containing (M_1, j), this class would be a closed class in Q which would not contain state $(M_1, 1)$, a contradiction to the irreducibility assumption of Q. So the entrance state is reachable from (M_1, j) under B_{M_1}. Second, we show that (M_1, j) is reachable from state $(M_1, 1)$ under B_{M_1}. Indeed, the only way to reach (M_1, j) from a state in $\widetilde{L}_{M_1+1}$ is via the entrance state, and such a path has to exist by irreducibility of Q. Since the "within" L_{M_1} transition rates under Q are all preserved under B_{M_1}, state (M_1, j) is reachable from state $(M_1, 1)$ under B_{M_1}. Thus every state communicates with the entrance state, and we conclude that B_{M_1} is irreducible.

For ii), we will call a matrix *diagonally dominant* if the absolute value of a diagonal elements are greater or equal than the sum of the absolute values of the off diagonal elements in that row, and strict inequality holds for at least one row. An irreducible diagonally dominant matrix is non singular by the well-known Levy-Desplanques theorem, cf. Varga (1963) (p. 85) or Varga (1976).

The construction of B_m excludes all down transitions and that makes it a diagonally dominant matrix. So when B_m is irreducible, the claim of the lemma is true.

However, in general it is possible that the matrix B_m is not irreducible (even though the matrix Q is irreducible). In this case we will show that the construction of B_m with the irreducibility of Q implies the non-singular property of B_m. Indeed, suppose B_m contains two or more communicating classes of states, say $C_1, \ldots, C_{k_m}$, where C_e contains the entrance state $(m, 1)$. We will relabel the states such that states in the same class have adjacent indices and such that if a state in a class C_i has a transition to a state in a class C_j then $i < j$. It is clear that this relabeling is feasible and we can write B_m as:

$$B_m = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & \ldots \\ 0 & Z_{22} & Z_{23} & \ldots \\ 0 & 0 & Z_{33} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

where Z_{ii} is a matrix of size $|C_i| \times |C_i|$ containing transition rates within C_i and the matrix Z_{ij} is of size $|C_i| \times |C_j|$ containing transition rates from C_i to C_j (and is possible to have some nonzero elements). We next show that the determinants of Z_{ii} are all non zero. This is sufficient to show that B_m is non-singular, since its determinant is the product of the determinants of the Z_{ii}'s.

First, at least one state in C_e has to have a transition "down" or to another class C_j within B_m, since otherwise C_e would be part of a closed absorbing class under Q. This class does not contain any states in $\underset{\sim}{L}_m$ and this is a contradiction to the irreducibility of Q. Thus Z_{ee} is diagonally dominant and therefore non-singular.

Furthermore, any other class C_j ($j \neq e$) has to have a state that has a transition rate leaving this class (i.e., down, up or to another class $C_{j'}$ ($j' \neq j$)) otherwise C_j would be a closed class under Q. Therefore, the sum of the off diagonal elements of Z_{jj} in at least one of its row is strictly less than the absolute value of the corresponding diagonal element, since at least one transition leads to a state out of C_j. Thus Z_{jj} is diagonally dominant for all j and the proof of part ii) is complete.

For iii), let τ_i denote the maximum of the absolute values of the diagonal elements of Z_{ii} and let

$$\Gamma_i = \tau_i I + Z_{ii}.$$

Since τ_i is finite and positive and Γ_i is non-negative with row sum less or equal to τ_i, (and strictly less than τ_i for at least one row), the row sum of $\tau_i^{-1}\Gamma_i$ is smaller or equal to one for each row. This implies that $\tau_i^{-1}\Gamma_i$ is a transient transition matrix with spectral radius smaller than 1, and thus all elements of $\sum_{j=0}^{\infty}(\tau_i^{-1}\Gamma_i)^j$ are non-negative and finite, see for example Seneta (1981), Theorem 4.3. Note also that

$$(Z_{ii})^{-1} = (\tau_i(\tau_i^{-1}\Gamma_i - I))^{-1} = \tau_i^{-1}(\tau_i^{-1}\Gamma_i - I)^{-1} = \tau_i^{-1}\sum_{j=0}^{\infty} -(\tau_i^{-1}\Gamma_i)^j.$$

The above implies that all elements of $(Z_{ii})^{-1}$ are non-positive.

By Woodbury's identity, cf. Woodbury (1950) we know:

$$\begin{bmatrix} Z_{11} & Z_{12} \\ 0 & Z_{22} \end{bmatrix}^{-1} = \begin{bmatrix} Z_{11}^{-1} & -Z_{11}^{-1} Z_{12} Z_{22}^{-1} \\ 0 & Z_{22}^{-1} \end{bmatrix},$$

which is a non-positive matrix by the above; an induction argument can establish the same for $(B_m)^{-1}$. □

We emphasize the following notational change with Chapter 1.

Remark 2.3. Let $\Delta_0 = \underset{\sim}{L}_m = \{(M_1, 1), \ldots, (M_1, \ell_{M_1}), \ldots, (m, 1), \ldots, (m, \ell_m)\}$. The lumped process on Δ_0 has a rate matrix "U_{Δ_0}" (defined in Chapter 1) of size $\underset{\sim}{\ell}_m \times \underset{\sim}{\ell}_m$ that can be written as:

$$[\Lambda_m \,|\, \Gamma_m]$$

where Λ_m contains the rates of transitions into states of the set $\underset{\sim}{L}_{m-1}$ (i.e., it is a matrix of dimension $\underset{\sim}{\ell}_m \times \underset{\sim}{\ell}_{m-1}$ and the construction of the $\underset{\sim}{\ell}_m \times \ell_m$, matrix Γ_m is done following Chapter 1). Note that we do not need to explicitly define the elements of the matrices Λ_m as they are not explicitly used in the sequel.

We next state the following theorem for successively lumpable Markov chains within the context and notation of the present chapter. This theorem is a consequence of Theorem 1.2, as is shown in the proof.

Theorem 2.1. *Under Assumption 2.1, the following equality is true for the steady state probabilities $\underset{\sim}{\pi}^m$ of $X(t)$ for every m :*

$$\underset{\sim}{\pi}^m \Gamma_m = 0_m. \tag{2.6}$$

Proof. For any fixed m, we consider the partition $\mathcal{D} = \{\underset{\sim}{L}_m, L_{m+1}, \ldots, L_{M_2}\}$ of the state space $\mathcal{X}$ of the chain. We note that the sets Δ_m of Chapter 1 are, within the present context, given by: $\Delta_k = \{(k\,0)\} \cup L_k$; where $(k\,0)$ represents the "lumped state".

By Lemma 2.2 we know that $X(t)$ is successively lumpable with respect to $\mathcal{D}$. Let $v_{\underset{\sim}{L}_m}$ denote the steady state probability vector of the lumped process on $\Delta_0 = \underset{\sim}{L}_m$, cf. Remark 2.3. By Proposition 1.1 (with $(k,i) \in \underset{\sim}{L}_m$ in place of $(0,i) \in \Delta_0$) we know that for all $k \leq m$:

$$\pi(k,i) = \sum_{(k',j)\in \underset{\sim}{L}_m} \pi(k' j) v_{\underset{\sim}{L}_m}(k,i). \tag{2.7}$$

Further, since $v_{\underset{\sim}{L}_m}$ is a steady state probability vector of the lumped process on $\underset{\sim}{L}_m$(="Δ_0"), it is the normalized to 1 solution of the equation below (see Remark 2.3):

$$v_{\underset{\sim}{L}_m}[\Lambda_{m-1} \,|\, \Gamma_m] = \underset{\sim}{0}_m. \tag{2.8}$$

If we let $c = \sum_{(k' j)\in \underset{\sim}{L}_m} \pi(k',j) \geq 0$, then from Eq. (2.7) we know $\underset{\sim}{\pi}^m = c\, v_{\underset{\sim}{L}_m}$. Eq. (2.9) below then follows by multiplying both sides of Eq. (2.8) by c:

$$\underset{\sim}{\pi}^m[\Lambda_{m-1} \,|\, \Gamma_m] = [\,\underset{\sim}{\pi}^m\Lambda_{m-1} \,|\, \underset{\sim}{\pi}^m\Gamma_m] = \underset{\sim}{0}_m = [\,\underset{\sim}{0}_{m-1} \,|\, 0_m], \tag{2.9}$$

and the proof is complete. □

Note that Theorem 2.1 implies, since $\Gamma_{M_1} = B_{M_1}$ and since $\underset{\sim}{\pi}^{M_1-1}$ does not exist:

$$\pi^{M_1} B_{M_1} = 0_{M_1}. \tag{2.10}$$

We next introduce the idea of a *rate matrix set* for Q as a sequence of matrices $\mathcal{R} = \{\mathcal{R}_m^k\}_{m,k}$ such that $\mathcal{R}_m^k$ satisfy Eq. (2.11), for all $k = 1, \ldots, m - M_1$ and $m = M_1 + 1, \ldots, M_2$; cf. Latouche and Ramaswami (1999) and references therein.

$$\pi^m = \underset{\sim}{\pi}^{m-k}\mathcal{R}_m^k. \tag{2.11}$$

Note that there are multiple rate matrix sets, for a given Q. To see this note that for any fixed k, m and known vectors $\pi^m = [\pi(m,1),\dots,\pi(m,\ell_m)]$ and $\underset{\sim}{\pi}^{m-k} = [\pi^{M_1},\dots,\pi^{m-k}]$ Eq. (2.11) is essentially a system of ℓ_m equations with $\underset{\sim}{\ell}_{m-k} \times \ell_m$ unknowns, the elements of the matrix $\mathcal{R}^k_m$. These equations have many solutions.

In Theorem 2.2 we show that the specific set $\mathcal{R}_0 := \{R^k_m\}_{m,k}$ obtained recursively using Eq. (2.12) starting with Eq. (2.13), is a rate matrix set for Q. For all $m = M_1+1,\dots M_2$ with $k = 2,\dots m - M_1$ we define:

$$R^k_m := \left[\underset{\sim}{I}_{m-k} \,|\, R^1_{m-(k-1)}\right] R^{k-1}_m, \tag{2.12}$$

where:

$$R^1_m := -A_m(B_m)^{-1}. \tag{2.13}$$

By virtue of Proposition 2.1 ii) we know that B_m is non sinular.

Theorem 2.2. *The set $\mathcal{R}_0$ defined by Eqs. (2.12) and (2.13) above is a rate matrix set for Q.*

Proof. The proof is by induction. For $k = 1$ we know by Theorem 2.1 that

$$\underset{\sim}{\pi}^m \Gamma_m = 0_m.$$

We can rewrite this as:

$$[\underset{\sim}{\pi}^{m-1} | \pi^m] \begin{bmatrix} A_m \\ B_m \end{bmatrix} = 0_m,$$

and thus:

$$\pi^m = -\underset{\sim}{\pi}^{m-1} A_m (B_m)^{-1} = \pi^{m-1} R^1_m.$$

Suppose the statement is true for any m and for $k-1$. We next show that the statement holds for k:

$$\begin{aligned}\pi^m &= \underset{\sim}{\pi}^{m-(k-1)} R_m^{k-1} \\ &= [\underset{\sim}{\pi}^{m-k} | \pi^{m-(k-1)}] R_m^{k-1} \\ &= [\underset{\sim}{\pi}^{m-k} | \underset{\sim}{\pi}^{m-k} R^1_{m-(k-1)}] R_m^{k-1} \\ &= \underset{\sim}{\pi}^{m-k} \left[\underset{\sim}{I}_{m-k} | R^1_{m-(k-1)}\right] R_m^{k-1} \\ &= \underset{\sim}{\pi}^{m-k} R_m^k.\end{aligned}$$

Thus the statement is true for $k = 1, \ldots, m - M_1$ and therefore:

$$\pi^m = \underset{\sim}{\pi}^{m-k} R_m^k. \qquad \square$$

Note that the above implies that we can express all vectors π^m in terms of the steady state distribution of level M_1, since M_1 is finite, By the irreducibility assumption all vectors are strictly larger than 0. Therefore we state:

$$\pi^m = \pi^{M_1} R_m^{m-M_1} > 0_m, \tag{2.14}$$

For any $m_1, m_2 \in \{M_1, \ldots, M_2\}$, with $m_1 < m_2$, we define the column vector $S_{m_1}^{m_2}$ of length ℓ_{m_1} by Eq. (2.15) below.

$$S_{m_1}^{m_2} = \left[1'_{m_1} + \sum_{m=m_1+1}^{m_2} R_m^{m-m_1} 1'_m\right]. \tag{2.15}$$

Remark 2.4.

The elements of R_m^k are non-negative $\forall\, k, m$ where $M_1+1 \leq m \leq M_2$, $1 \leq k \leq m-M_1$. To check this claim for R_m^1, it suffices to note that $A_m \geq 0$ by definition and $-B_m^{-1} \geq 0$, by Proposition 2.1. Alternatively, the (i, j)-th element of R_m^1 can be given an expected first passage time interpretation as is described for QBD processes in Latouche and Ramaswami (1999), Chapter 6. The claim for R_m^k, with $k \geq 2$, follows using Eq. (2.12).

The lemma below establishes the relation between π^{M_1} and $S_{M_1}^{M_2}$.

Lemma 2.3. *The following relation holds for π^{M_1} and $S_{M_1}^{M_2}$:*

$$\pi^{M_1} S_{M_1}^{M_2} = 1. \tag{2.16}$$

Proof. Since the chain is ergodic we have:

$$\pi^{M_1} 1'_{M_1} + \sum_{m=M_1+1}^{M_2} \pi^m 1'_m = 1,$$

thus Eq. (2.14) implies:

$$\pi^{M_1} \left[1'_{M_1} + \sum_{m=M_1+1}^{M_2} R_m^{m-M_1} 1'_m \right] = 1.$$

Substituting $\left[1'_{M_1} + \sum_{m=M_1+1}^{M_2} R_m^{m-M_1} 1'_m\right]$ by $S_{M_1}^{M_2}$ in the above gives Eq. (2.16).

□

We now state and prove the following theorem.

Theorem 2.3. *Under Assumption 2.1, the following is true:*

$$\pi^{M_1} = \delta_{M_1} \left[S_{M_1}^{M_2} \delta_{M_1} - B_{M_1} \right]^{-1}. \tag{2.17}$$

Proof. Since B_{M_1} is an irreducible rate matrix (Proposition 2.1 i)), it has rank $(\ell_{M_1} - 1)$ by basic linear algebra theory, see for example Seneta (1981). Furthermore, we know that $\pi^{M_1} B_{M_1} = 0_{M_1}$ and $\pi^{M_1} S_{M_1}^{M_2} = 1$, thus that the vector $S_{M_1}^{M_2}$ is not an element of the linear space spanned by the columns of B_{M_1}. Therefore $[S_{M_1}^{M_2} \delta_{M_1} - B_{M_1}]$ has full rank and is invertible.

We use Lemma 2.3 to state

$$\left[\pi^{M_1} S_{M_1}^{M_2}\right] \delta_{M_1} = \delta_{M_1},$$

and via

$$\pi^{M_1}\left[S_{M_1}^{M_2}\delta_{M_1} - B_{M_1}\right] = \delta_{M_1} - 0_{M_1} = \delta_{M_1},$$

we conclude

$$\pi^{M_1} = \delta_{M_1}\left[S_{M_1}^{M_2}\delta_{M_1} - B_{M_1}\right]^{-1}.$$

□

The results above justify the following algorithm to find the steady state distribution of a DES-QSF process.

Algorithm 2.1. [DES-QSF]

- Calculate R_m^1 with Eq. (2.13) for all $m = M_1 + 1, \ldots, M_2$.
- Compute R_m^k recursively via Eq. (2.12) for $m = M_1 + 1 \ldots, M_2$ and $k = 2, \ldots, m - M_1$.
- Calculate $S_{M_1}^{M_2}$ via Eq. (2.15).
- Calculate π^{M_1} via Eq. (2.17).
- Calculate π^m via Eq. (2.14) for all $m = M_1 + 1, \ldots, M_2$.

2.3.1 State Space Truncations

In this section we show how to truncate the state space in the upward direction, in order to obtain upper bounds for the steady state probabilities $\pi(m, i)$ of states in

$\underset{\sim}{L}_{m_2}$ where $m_2 \in \{M_1, M_1+1, \ldots, M_2-1\}$. To this end we first define a process $X_{m_2}(t)$ with truncated state space $\mathcal{X}_{m_2} = \underset{\sim}{L}_{m_2}$ and transition rate matrix:

$$Q_{X_{m_2}} = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots \\ \cdots & D^{m_2-2} & W^{m_2-2} & U^{m_2-2,m_2-1} & U^{m_2-2,m_2} + \widetilde{U}^{m_2-2,m_2} \\ \cdots & 0 & D^{m_2-1} & W^{m_2-1} & U^{m_2-1,m_2} + \widetilde{U}^{m_2-1,m_2} \\ \cdots & 0 & 0 & D^{m_2} & W^{m_2} + \widetilde{U}^{m_2,m_2} \end{bmatrix}, \quad (2.18)$$

where the elements of the last column are given by Eq. (2.3). We denote the steady state distribution of this process as the row vector $\pi_{X_{m_2}} = [\pi^{M_1}_{X_{m_2}}, \ldots, \pi^{m_2}_{X_{m_2}}]$ of size: $\sum_{m=M_1}^{m_2} \ell_m$, where its m^{th} component contains the steady state probabilities for level m of the truncated process.

We next state the following. We emphasize that this proposition clearly holds for $M_2 = \infty$ under the ergodicity assumption.

Proposition 2.2. *For all finite $m_2 \geq M_1$, and any level $m = M_1, M_1+1, \ldots, m_2$, the following are true:*

i)
$$\pi^{m}_{X_{m_2}} = \pi^{M_1}_{X_{m_2}} R_m^{m-M_1}, \quad (2.19)$$
$$\pi^{M_1}_{X_{m_2}} = \delta_{M_1} \left[S^{m_2}_{M_1} \delta_{M_1} - B_{M_1} \right]^{-1}. \quad (2.20)$$

ii)
$$\pi(m,i) < \pi_{X_{m_2}}(m,i).$$

iii) For all states (m,i), $\pi_{X_\nu}(m,i)$ is a strict decreasing function in $\nu = m_2, m_2+1, \ldots$

Proof. By its construction, the process $X_{m_2}(t)$ is a QSF process which satisfies Assumption $A1$. Further, by its definition the matrix $Q_{X_{m_2}}$ gives rise to the same rate matrices R_m^k as the matrix Q of the original process $X(t)$. This follows from the fact that this specific truncation ensures that the matrices $A_{m,X_{m_2}}$, $B_{m,X_{m_2}}$ of the truncated process corresponding to the matrices A_m B_m of the original process are identical and this proves Eq. (2.19). The proof of Eq. (2.20) follows as the proof of Theorem 2.3, if we replace M_2 with m_2.

For the proof of part ii), using Proposition 1.1, (where $\pi_{X_{m_2}}(m,i) = v_{\underset{\sim}{L}_m}(m,i)$) we obtain that Eq. (2.21) below is valid for all $m \le \nu$:

$$\pi_{X_\nu}(m,i) = \frac{\pi(m,i)}{\sum_{(k,j)\in \underset{\sim}{L}_\nu} \pi(k,j)} \; for\ all\ \nu = m_2, m_2+1, \ldots . \tag{2.21}$$

Since $\sum_{(k,j)\in \underset{\sim}{L}_\nu} \pi(k,j) < 1$, for all finite ν, it follows from the above that $\pi(m,i) < \pi_{X_{m_2}}(m,i)$.

For the proof of part iii), note that since $\underset{\sim}{L}_\nu \subset \underset{\sim}{L}_{\nu+1}$, we have that $\sum_{(k,j)\in \underset{\sim}{L}_\nu} \pi(k,j) < \sum_{(k,j)\in \underset{\sim}{L}_{\nu+1}} \pi(k,j)$. Thus, we conclude that $\pi_{X_{\nu+1}}(m,i) < \pi_{X_\nu}(m,i)$. We can repeat this argument for $\underset{\sim}{L}_{\nu+2}, \underset{\sim}{L}_{\nu+3}, \ldots$ and the proof is complete.

□

Note that Proposition 2.2 is closely related the results of Bright and Taylor (1995) (pp. 499-500), derived for LDQBDs. Specifically, Eq. (2.19)-(2.20) are the QSF process extensions of Eqs. (1.7)-(1.8) of that paper, with k and m reversed and the change of notation K^*, x_k, R_{k+1}, R_0 in place of our m_2, π^m, R_m R_1.

Note that the matrix $R_m^{m-M_1}$ is finite even when the the QSF process is not ergodic; such a non-ergodic case exists for instance when there is a drift to "up" direction. We can however state the following:

Remark 2.5. The successively lumpable QSF process is ergodic if $\sum_{m=M_1}^{M_2} R_m^{m-M_1} < \infty$, since then Theorem 2.3 and Eq. (2.14) show that there exist positive steady state probabilities for all states. Similarly, it follows that for QSF processes to be ergodic it is sufficient that $S_{M_1}^{M_2} < \infty$.

Remark 2.6. One can also construct truncations with respect to M_1 or to any ℓ_m separately. This is especially important when some of these constants are infinite.

There are various truncations methods possible to truncate the matrix Q of infinite size, some are described in Vere-Jones (1967), Seneta (1980) and Tweedie (1973). Most of these truncations will preserve the successively lumpable property. Such as truncation to $m_1 \geq M_1 = -\infty$ we provide below, following Seneta (1980).

Define a process $X_{m_1}(t)$ with state space $\mathcal{X}_{m_1} = \widetilde{L}_{m_1}$ and transition rate matrix:

$$Q_{X_{m_1}} = \begin{bmatrix} \overline{W}^{m_1} & U^{m_1,m_1+1} & U^{m_1,m_1+2} & U^{m_1,m_1+3} & \cdots \\ D^{m_1+1} & W^{m_1+1} & U^{m_1+1,m_1+2} & U^{m_1+1,m+3} & \cdots \\ 0 & D^{m_1+2} & W^{m_1+2} & U^{m_1+2,m_1+3} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}, \tag{2.22}$$

where $\overline{w}(m_1, j \mid m_1, i) = w(m_1, j \mid m_1, i)$ if $j \neq 1$ and $\overline{w}(m_1, 1 \mid m_1, i) = w(m_1, 1 \mid m_1, i)$ $+d(m_1 - 1, 1 \mid m_1, i)$ otherwise. Let $\pi_{X_{m_1}}$ denote the steady state distribution of $X_{m_1}(t)$. For this truncation it is shown in Seneta (1980) (Theorem on p. 262) that for all (m, i):

$$\pi(m, i) = \lim_{m_1 \to -\infty} \pi_{X_{m_1}}(m, i).$$

2.4 Explicit solutions for RES QSF processes

In this section we consider QSF processes of the form described in Definition 2.1 ii). Because the current section has the same structure as Section 2.3, most of the theorems and lemmas can be given without proof. We will refer to the corresponding statement in the previous section and add some clarification if necessary.

We start with the following lemmas that show that the simple algebraic characterization of the RES property is a sufficient condition for a QSF process to be successively lumpable.

We start with the following lemma that characterizes an entrance state of the super-level set $\widetilde{L}_m$ of a QSF process in terms of an algebraic property of the "up" transition sub-matrices $U^{m,k}$ of its transition rate matrix Q.

Lemma 2.4. *For a QSF process $X(t)$ and for all $m \in \{M_1, \ldots, M_2\}$, the state $(M_2, \boldsymbol{\varepsilon}(L_m)) \in L_{M_2}$ is an* **entrance state** *for $\widetilde{L}_m$ if the following is true for all states $(n, i) \in \underset{\sim}{L}_{m-1}$:*

$$u(k, j \,|\, n, i) = 0, \quad \text{if } (k, j) \neq (M_2, \boldsymbol{\varepsilon}(L_m)). \tag{2.23}$$

Proof. Directly from the definition of an entrance state (cf. Definition 1.1). □

It is easy to see that Eq. (2.23) of Lemma 2.4 is equivalent to the statement that the matrices U^{m,M_2} have a single nonzero column and that $U^{m,k} = 0$ for all $k \in \{m+1, \ldots, M_2 - 1\}$. This is equivalent to the RES property.

For any fixed $n \in \{M_1, \ldots, M_2\}$, let $\mathcal{D}_n$ denote the partition $\{L_{M_1}, \ldots, L_{n-1}, \widetilde{L}_n\}$ of $\mathcal{X}$. For a fixed n, the next lemma establishes that when Q has the RES property (cf. 2.1 ii)), then the QSF process is successively lumpable with respect to the partition $\mathcal{D}_n$.

Lemma 2.5. *A QSF process is successively lumpable with respect to a partition $\mathcal{D}_{M_2}$ if the matrices U^{m,M_2} have a single nonzero column and that $U^{m,k} = 0$ for all $k \in \{m+1, \ldots, M_2 - 1\}$.*

Proof. Similar to the proof of Lemma 2.2 in the previous section. □

A graphical representation of the transitions that are allowed in a RES QSF process can be found in Figure 2.2.

We now state the following assumption that will be used in the sequel of this section, where for notational simplicity we let the entrance state of a set $\widetilde{L}_m$ be state $(M_2, 1)$, for all m without loss of generality.

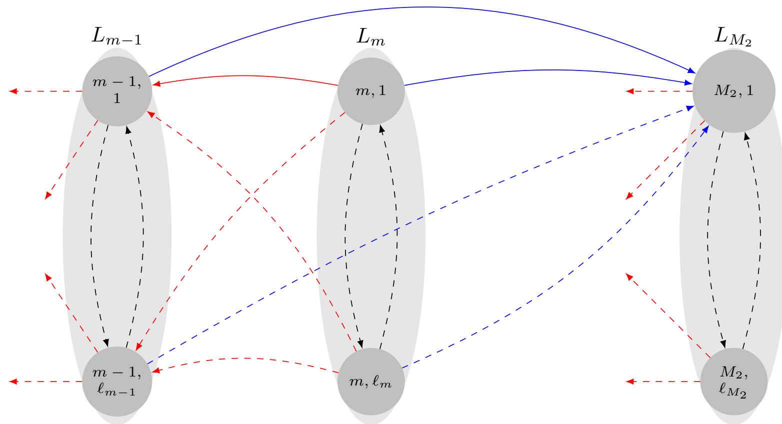


Figure 2.2: Graphical representation of a RES QSF process

Assumption 2.2. The QSF process under consideration has a transition rate matrix Q with the following properties:

A1. The QSF process is ergodic (irreducible);

A2. For all $m \in \{M_1, \ldots, M_2 - 1\}$, only the *first column* of sub-matrix U^{m,M_2} can contain non-zero elements. Since the process is positive recurrent, there has to be a transition back to the entrance state from a state in $\underset{\sim}{L}_m$, i.e. $u(M_2, 1 \,|\, n, i) > 0$ for at least one $(n,i) \in \underset{\sim}{L}_m$ and all other columns of U^{m,M_2} are equal to zero. In addition, $U^{m,k} = 0$ for all $k \in \{m, \ldots, M_2 - 1$;

A3. The QSF process has bounded rates.

We will use the notation introduced in the previous section. In addition we define the matrix $\widetilde{D}^m$ of dimension $\ell_m \times \ell_m$ by:

$$\widetilde{D}^m = D^m 1_k' \delta_m. \tag{2.24}$$

We next state and prove a proposition regarding basic properties of W_m and $\widetilde{D}^{M_2}$.

Proposition 2.3. *The following are true:*

i) The matrix $W^{M_2} + \widetilde{D}^{M_2}$ *is irreducible.*

ii) The matrices W^m are non-singular, for all $m \in \{M_1, \ldots, M_2\}$.

iii) The elements of the inverse of W^m are non positive, for all $m < M_2$.

Proof. The proves are analogous to the one of Proposition 2.1 in the previous section, that states similar results for a DES QSF process. □

We can now state the following theorem for successively lumpable Markov chains, that satisfy the RES property, within the context and notation of the present section.

Theorem 2.4. *Under Assumption 2.2, the following equality is true for the steady state probabilities π^m of $X(t)$ for every $m \in \{M_1, \ldots, M_2 - 1\}$:*

$$\pi^m W^m + \pi^{m+1} D^{m+1} = 0_m. \tag{2.25}$$

$$\pi^{M_2}(W^{M_2} + \widetilde{D}^{M_2}) = 0_{M_2}. \tag{2.26}$$

Proof. The proof is along the same lines as Theorem 2.1 and is a consequence of Theorem 1.2, since the QSF process is successively lumpable. We can complete the proof by only considering the states that have possibly positive transitions out of set $\widetilde{L}_{m+1}$. □

For a RES QSF process, we define a *rate matrix set* for Q as a sequence of matrices $\mathcal{R} = \{\mathcal{R}_m\}_m$ such that $\mathcal{R}_m$ satisfy Eq. (2.27), for all $m = M_1, \ldots, M_2 - 1$. Note that this is a slightly different definition than the ones introduced for DES QSF processes.

$$\pi^m = \pi^{m+1} \mathcal{R}_m. \tag{2.27}$$

In Theorem 2.5 we show that the specific set $\mathcal{R}_0 := \{R_m\}_m$ obtained Eq. (2.28), is a rate matrix set for Q. For all $m = M_1, \ldots, M_2 - 1$ we define:

$$R_m := -D^{m+1}(W^m)^{-1}. \tag{2.28}$$

By virtue of Proposition 2.3 ii) we know that W^m is non sinular.

Theorem 2.5. *The set $\mathcal{R}_0$ defined by (2.28) above is a rate matrix set for Q.*

Proof. Follows directly from Theorem 2.4. □

Note that the above implies that we can express all vectors π^m in terms of the steady state distribution of level M_2, since M_2 is finite. By the irreducibility assumption all vectors are strictly larger than 0. Therefore we state:

$$\pi^m = \pi^{M_2} \prod_{k=0}^{M_2-1-m} R_{M_2-1-k} > 0_m, \tag{2.29}$$

For any $m_1 \in \{M_1, \ldots, M_2\}$, with $m_1 < M_2$, we define the column vector $T_{m_1}^{M_2}$ of length ℓ_{M_2} by Eq. (2.30) below.

$$T_{m_1}^{M_2} = \left[1'_{M_2} + \sum_{m=m_1}^{M_2-1} \prod_{k=0}^{M_2-1-m} R_{M_2-1-k} 1'_m \right]. \tag{2.30}$$

Note that R_m is non negative for all m.

The lemma below establishes the relation between π^{M_2} and $T_{M_1}^{M_2}$.

Lemma 2.6. *The following relation holds for π^{M_2} and $T_{M_1}^{M_2}$:*

$$\pi^{M_2} T_{M_1}^{M_2} = 1. \tag{2.31}$$

Proof. Analogous to Lemma 2.6. □

We now state and prove the following theorem.

Theorem 2.6. *Under Assumption 2.2, the following is true:*

$$\pi^{M_2} = \delta_{M_2} \left[T^{M_2}_{M_1} \delta_{M_2} - W^{M_2} - \widetilde{D}^{M_2} \right]^{-1}. \tag{2.32}$$

Proof. Since $W^{M_2} - \widetilde{D}^{M_2}$ is an irreducible rate matrix (Proposition 2.3 i)), it has rank $(\ell_{M_2} - 1)$ by basic linear algebra theory, see for example Seneta (1981). Furthermore, we know that $\pi^{M_2}(W^{M_2} - \widetilde{D}^{M_2}) = 0_{M_2}$ and $\pi^{M_2} T^{M_2}_{M_1} = 1$, thus that the vector $T^{M_2}_{M_1}$ is not an element of the linear space spanned by the columns of $W^{M_2} - \widetilde{D}^{M_2}$. Therefore $[T^{M_2}_{M_1} \delta_{M_1} - W^{M_2} - \widetilde{D}^{M_2}]$ has full rank and is invertible. The remainder of the proof is similar to the proof of Theorem 2.3. □

The results above justify the following algorithm to find the steady state distribution of a RES QSF process.

Algorithm 2.2. [RES-QSF]

- Compute R_m via Eq. (2.28) for $m = M_1 \ldots, M_2 - 1$.
- Calculate $T^{M_2}_{M_1}$ via Eq. (2.30).
- Calculate π^{M_2} via Eq. (2.32).
- Calculate π^m via Eq. (2.29) for all $m = M_1, \ldots, M_2 - 1$.

2.4.1 State Space Truncations

In this section we show how to truncate the state space of a RES QSF process in the downward direction in order to obtain upper bounds for the steady state probabilities $\pi(m, i)$ of states in $\widetilde{L}_{m_1}$ where $m_1 \in \{M_1 + 1, M_1 + 2, \ldots, M_2\}$. To this end we define a process $X_{m_1}(t)$ with truncated state space $\mathcal{X}_{m_1} = \widetilde{L}_{m_1}$ and transition rate matrix:

$$Q_{X_{m_1}} = \begin{bmatrix} W^{m_1} & 0 & \cdots & 0 & 0 & 0 & \widetilde{D}^{m_1} + U^{m_1,M_2} \\ D^{m_1+1} & W^{m_1+1} & \cdots & 0 & 0 & 0 & U^{m_1+1,M_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & D^{M_2-2} & W^{M_2-2} & 0 & U^{M_2-2,M_2} \\ 0 & 0 & \cdots & 0 & D^{M_2-1} & W^{M_2-1} & U^{M_2-1,M_2} \\ 0 & 0 & \cdots & 0 & 0 & D^{M_2} & W^{M_2} \end{bmatrix}. \quad (2.33)$$

We denote the steady state distribution of this process as the row vector $\pi_{X_{m_1}} = [\pi^{m_1}_{X_{m_1}}, \ldots, \pi^{M_2}_{X_{m_1}}]$ of size: $\sum_{m=m_1}^{M_2} \ell_m$, where its m^{th} component contains the steady state probabilities for level m of the truncated process.

We next state the following. We emphasize that this proposition clearly holds for $M_1 = \infty$ under the ergodicity assumption.

Proposition 2.4. *For all finite $m_1 \leq M_2$, and any level $m = m_1, m_1+1, \ldots, M_2$, the following are true:*

i)
$$\pi^m_{X_{m_1}} = \pi^{M_2}_{X_{m_1}} \prod_{k=0}^{M_2-1-m} R_{M_2-1-k} \quad (2.34)$$
$$\pi^{M_2}_{X_{m_1}} = \delta_{M_2} \left[T^{M_2}_{m_1} \delta_{M_2} - W^{M_2} - \widetilde{D}^{M_2} \right]^{-1}. \quad (2.35)$$

ii) $\pi(m,i) < \pi_{X_{m_1}}(m,i)$.

iii) For all states (m,i), $\pi_{X_\nu}(m,i)$ is a strict decreasing function in $\nu = m_1, m_1-1, \ldots$

Proof. Similar to the proof of Proposition 2.2: the RES property remains intact, the rate matrices do not change. The entrance state will never be removed from the state space. □

Remark 2.7. It is not useful to truncate the process in the upward direction. Since we consider a ergodic process with a restart entrance state, returns will go to the highest level. Removing this set would effect the structure of the process too much to give any bounds of interest.

2.5 A special case of QSF processes: QBD processes

A Quasi Birth and Death process is a special case of a QSF process, where $U^{m,k} = 0$ for $k \geq m+2$. Therefore we rename in this section $U^{m,m+1}$ to U^m. All the proofs of the previous section for a QSF process also hold for a QBD process, but the algebra simplifies. In the sequel we will assume that $M_1 = 0$ and that Q is irreducible. The QBD process $X(t)$ has the transition rate matrix of Eq. (2.36) below.

$$Q = \begin{bmatrix} W^{M_1} & U^{M_1} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ D^{M_1-1} & W^{M_1+1} & U^{M_1+1} & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & W^m & U^m & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & D^{m+1} & W^{m+1} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & D^{M_2-1} & W^{M_2-1} & U^{M_2-1} \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & D^{M_2} & W^{M_2} \end{bmatrix}. \tag{2.36}$$

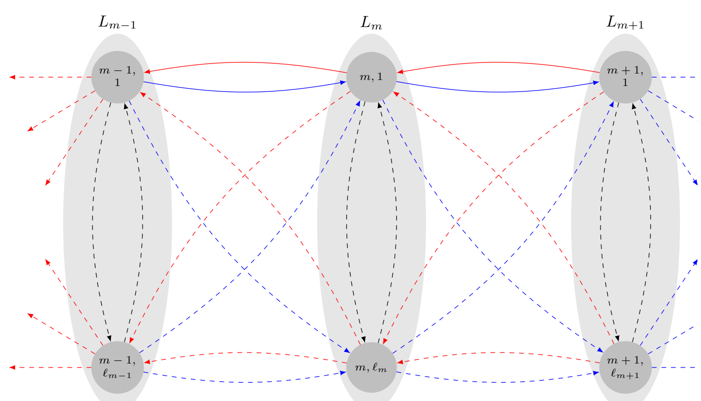


Figure 2.3: Graphical representation of a QBD process

In a QBD process $\widetilde{U}^m$ is as follows:

$$\widetilde{U}^m = U^m 1'_{m-1}\delta_m.$$

In a successively lumpable QBD process, the rate matrix set $\{\mathcal{R}_m\}_m$ satisfies the equation below, since there are only transitions in the upward direction from one level lower.

$$\pi^m = \pi^{m-1}\mathcal{R}_m\,, \qquad \text{for } m = 1, \ldots, M_2. \tag{2.37}$$

Remark 2.8. Note that Eq. (2.37) is Eq. (1.5) in Bright and Taylor (1995) (with k and m reversed and the change of notation their x_k denotes π^m and R_{k+1} denotes our $\mathcal{R}_m$. To find the rate matrix set they provided iterative algorithms. In contrast, Eq. (2.38) below provides explicit formulas for the rate matrix set $\{\mathcal{R}_m\}_m$; this equation is made possible by the assumption of the successively lumpable property.

Again, let $\mathcal{R}_0 := \{R_m\}_m$ be as follows (the QBD equivalent of Eq. (2.12)):

$$R_m = -U^{m-1}(\widetilde{U}^m + W^m)^{-1}. \tag{2.38}$$

We now state and prove the following simplification of Theorem 2.2 for a QBD process.

Theorem 2.7. *The set $\mathcal{R}_0$ is a rate matrix set for Q.*

Proof. Note that in the present QBD case the matrices A_m, B_m (defined in Eq. (2.4) and (2.5)) simplify to:

$$A_m = \begin{bmatrix} 0_{m-2,m} \\ U^{m-1} \end{bmatrix},$$

$$B_m = \widetilde{U}^m + W^m,$$

where $0_{m-2,m}$ is a matrix of size $\underset{\sim}{\ell}_{m-2} \times \ell_m$ with a 0 at every entry. By Theorem 2.1 we know:

$$\pi^m = -\underset{\sim}{\pi}^{m-1} A_m (B_m)^{-1}, \tag{2.39}$$

and because of the structure of A_m we write:

$$\underset{\sim}{\pi}^{m-1} A_m = \pi^{m-1} U^{m-1}.$$

And thus:

$$\pi^m = -\underset{\sim}{\pi}^{m-1} A_m (B_m)^{-1} = -\pi^{m-1} U^{m-1} (\widetilde{U}^m + W^m)^{-1}$$

and the proof is complete. □

Remark 2.9.

i) Note that Eq. (2.38) implies that the following recursive relation holds for all $\nu = 0, \ldots, m-1$:

$$\pi^m = \pi^\nu \prod_{k=\nu+1}^{m} R_k. \tag{2.40}$$

ii) It is easy to see that the above defined π^m and R_m satisfy the non-linear Eq. (12.2) of Latouche and Ramaswami (1999). The matrices R_m are solutions to Eq. (12.11) of the same book, given there without explicit solution.

Remark 2.10. For each $m = 1, 2, \ldots, M_2$ the matrices R_m are easy to compute; the computation only involves inversion of the $\ell_m \times \ell_m$ matrix $\widetilde{U}^m + W^m$ and pre-multiplication of the inverse by the $\ell_{m-1} \times \ell_m$ matrix U^{m-1}.

Theorem 2.8. *The following are true for the successively lumpable QBD process $X(t)$:*

$$\pi^0 = \delta_0 \left[S_0^{M_2} \delta_0 - (\widetilde{U}^0 + W^0) \right]^{-1} \tag{2.41}$$

where

$$S_0^{M_2} = (1_0' + \sum_{m=1}^{M_2} \prod_{k=1}^{m} R_k 1_m'). \tag{2.42}$$

Proof. Direct consequence of the QBD structure and Theorems 2.2 and 2.3. □

Remark 2.11. Note that a RES QSF process does not have a simplified QBD simplification. Transition need to go up to the restart state from the lowest level, otherwise the Markov process is not positively recurrent. Therefore the process is not a QBD process.

CHAPTER 3

Applications

3.1 Introduction to Chapter 3

To illustrate the application of the results we provide explicit solutions and approximations to well known open problems of queueing, cf. Adan et al. (2013), and to a stochastic inventory theory problem, cf. Veinott (1965). The queueing models under consideration are the $M/Er/n$ queueing model with batch arrivals and the $Er/M/n$ queueing model. In the two subsequent sections, we take the number of phases to be constant, i.e. $\ell_m = \ell$ for all m, this is done solely for presentation simplicity. The analysis is easy to extend when the number of phases of the corresponding distribution is a function of "m" - the state of the queue, the number of customers in line. The steady state distribution of the $M/Er/n$ is known for the case of Poisson arrivals, as for example discussed in Latouche and Ramaswami (1999). As far as we known this is the first time the steady state distribution of the $M/Er/n$ model with batch arrivals is obtained. We note that the same book gives an exact solution procedure for QBD processes only when M_1 is finite. Below we show that our direct method works for the $Er/M/n$ queueing system, i.e., we provide explicit formulas for the rate matrix set, even when $M_1 = -\infty$. For the inventory model we show that is has the same structure as the $M/Er/n$ and it can be handled similarly.

The construction in the following remark can be used to extend the applicability of the methods described in the previous section to models that are QSF and successively lumpable in the 'upward' direction. The 'QBD' version of this remark is used in Section 3.2.2.

Remark 3.1. Consider a process with a transition rate matrix Q that has the block form shown in Eq. (3.1), where its elements are labeled by $(m, i) \in \mathcal{X}$, the states of the underlying process.

$$Q = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\ \cdots & D^{m-1,m-2} & W^{m-1} & U^{m-1} & 0 & 0 & \cdots \\ \cdots & D^{m,m-2} & D^{m,m-1} & W^{m} & U^{m} & 0 & \cdots \\ \cdots & D^{m+1,m-2} & D^{m+1,m-1} & D^{m+1,m} & W^{m+1} & U^{m+1} & \cdots \\ \iddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{3.1}$$

Then, we can construct a transition rate matrix $\hat{Q}$ of the form of Eq. (2.1) by relabeling the states so that a new state $(-m, i)$ corresponds to the original $(m, i) \in \mathcal{X}$ by redefining the down, within and up sub-matrices of $\hat{Q}$ as follows: $\hat{D}^{-m} = U^m$, $\hat{U}^{-m,-k} = D^{m,k}$, and $\hat{W}^{-m} = W^m$. The steady state probabilities of the Q-process can be readily obtained from those of the $\hat{Q}$-process.

In this Chapter we will use the notation introduced in Chapter 2.

3.2 Two Classic Queueing Models

3.2.1 The $M/Er/n$ Model with Batch Arrivals.

In a $M/Er/n$ queueing system with batch arrivals the service of a customer occurs in ℓ phases, each exponentially distributed with parameter μ_i for the i-th phase of the service. For notational simplicity of the exposition we describe in detail the case in which a batch may contain either 1 or 2 customers. Batches with a single customer arrive according to a Poisson process with rate $p\lambda_{m,i}$ when there are m customers in the system and the served customer has gone through the first i phases of services. Similarly, batches of 2 customers arrive with rate $(1-p)\lambda_{m,i}$ with $p \in [0,1]$. The service of a customer has to be completed before another customer can start his first phase.

In order to have state notation that is consistent with that of Section 2.2, we use the following state description. For $i < \ell$, state (m,i) denotes the event that there are m customers in the waiting line of the system and a customer in service that has gone through i phases of the service. State (m,ℓ) denotes the event that there are m customers in the waiting line and a service completion has just occurred, so that one of the waiting customers is starting service. Note that with this awkward but convenient notation the empty state of the system is state $(0,\ell)$. Then, it is easy to see that this system can be modeled as a QSF process. Its state space is $\mathcal{X} = \{L_0, L_1, \ldots, L_{M_2}\}$ with $L_m = \{(m,1),\ldots,(m,\ell)\}$ and $M_2 \leq \infty$. The Q matrix is defined by Eq. (2.1), with U^m, W^m and D^m (all of size $\ell \times \ell$) as given below.

$$W^0 = \begin{bmatrix} -\lambda_{0,1}-\mu_1 & \mu_1 & 0 & \cdots & 0 \\ 0 & -\lambda_{0,2}-\mu_2 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\lambda_{0,\ell-1}-\mu_{\ell-1} & \mu_{\ell-1} \\ 0 & \cdots & 0 & 0 & -\lambda_{0,\ell} \end{bmatrix},$$

and for $m \geq 1$:

$$W^m = \begin{bmatrix} -\lambda_{m,1}-\mu_1 & \mu_1 & 0 & \cdots & 0 \\ 0 & -\lambda_{m,2}-\mu_2 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\lambda_{m,\ell-1}-\mu_{\ell-1} & \mu_{\ell-1} \\ 0 & \cdots & 0 & 0 & -\lambda_{m,\ell}-\mu_\ell \end{bmatrix},$$

$$D^m = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_\ell & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

For $m = 0, 1, \ldots :$

$$U^{m,m+1} = p \begin{bmatrix} \lambda_{m,1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{m,2} & 0 & \cdots & 0 \\ 0 & 0 & \lambda_{m,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{m,\ell} \end{bmatrix},$$

$$U^{m,m+2} = (1-p) \begin{bmatrix} \lambda_{m,1} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_{m,2} & 0 & \cdots & 0 \\ 0 & 0 & \lambda_{m,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{m,\ell} \\ 0, & & & & \end{bmatrix}.$$

Note that $(m, 1)$ is the entrance state of the set $\underset{\sim}{L}_m$, because D_m has a single nonzero

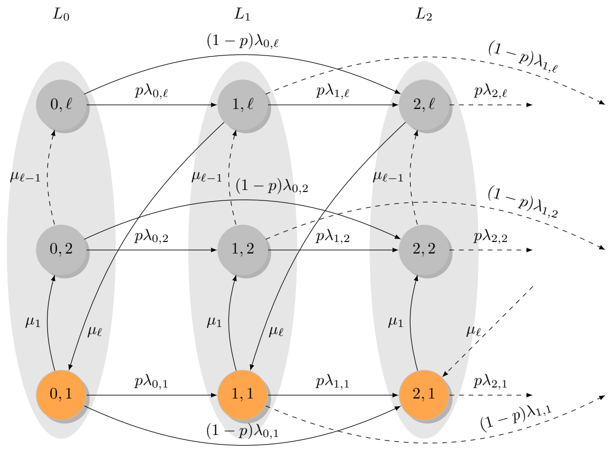


Figure 3.1: An $M/Er/n$ queueing process with batch arrivals

column. The matrices A_m and B_m, described in Section 2.3, are:

$$A_m = \left[\begin{array}{ccccc} \multicolumn{5}{c}{0_{m-3,m}} \\ \hline (1-p)\lambda_{m-2,1} & 0 & 0 & \dots & 0 \\ 0 & (1-p)\lambda_{m-2,2} & 0 & \dots & 0 \\ 0 & 0 & (1-p)\lambda_{m-2,3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & (1-p)\lambda_{m-2,\ell} \\ \lambda_{m-1,1} & 0 & 0 & \dots & 0 \\ (1-p)\lambda_{m-1,2} & p\lambda_{m-1,2} & 0 & \dots & 0 \\ (1-p)\lambda_{m-1,3} & 0 & p\lambda_{m-1,3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (1-p)\lambda_{m-1,\ell} & 0 & 0 & \dots & p\lambda_{m-1,\ell} \end{array}\right]$$

where $0_{m-3,m}$ is a matrix of size $\underset{\sim}{\ell}_{m-3} \times \ell_m$ with a 0 at every entry. For $m = 1, 2, \ldots$:

$$B_m = \begin{bmatrix} -\mu_1 & \mu_1 & 0 & \cdots & 0 \\ \lambda_{m,2} & -\lambda_{m,2} - \mu_2 & \mu_2 & \cdots & 0 \\ \lambda_{m,3} & 0 & -\lambda_{m,3} - \mu_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{m,\ell} & 0 & 0 & \cdots & -\lambda_{m,\ell} - \mu_\ell \end{bmatrix}$$

and

$$B_0 = \begin{bmatrix} -\mu_1 & \mu_2 & 0 & \cdots & 0 \\ \lambda_{0,2} & -\lambda_{0,2}-\mu_2 & \mu_2 & \cdots & 0 \\ \lambda_{0,3} & 0 & -\lambda_{0,3}-\mu_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \lambda_{0,\ell} & 0 & 0 & \cdots & -\lambda_{0,\ell} \end{bmatrix}.$$

Now we can calculate R^1_m using Eq. (2.13): $R^1_m = -A_m(B_m)^{-1}$, where $\pi^m = \underset{\sim}{\pi}^{m-1}R^1_m$.

Since the first $\underset{\sim}{\ell}_{m-3}$ rows of R^1_m are zero (due to multiplication of $(B_m)^{-1}$ with the $0_{m-3,m}$ sub-matrix of A_m) this expression reduces to the following

$$\pi^m = [\pi^{m-2}|\pi^{m-1}]R^{*1}_m,$$

where R^{*1}_m denotes the nonzero rows of R^1_m.

We can construct the sequence of rate matrices R^k_m using Eq. (2.12) and the notation R^{*k}_m for the sub-matrix of the nonzero rows of R^k_m we obtain:

$$\pi^m = [\pi^{m-k-1}|\pi^{m-k}]R^{*k}_m$$

and $\pi^m = \pi^0 R^{*m}_m$.

When M_2 is finite (i.e. there is a finite buffer for the number of customers allowed in the system), then Theorem 2.3 readily provides the solution: $\pi^0 = \delta_0\left[S_0^{M_2}\delta_0 - B_0\right]^{-1}$, $\pi^m = \pi^0 R^{*m}_m$.

When M_2 is infinite, using Proposition 2.2, we can construct upper bounds for $\pi(m,i)$ via the process $X_{m_2}(t)$ described therein. This result is stated in the next theorem.

Theorem 3.1. *The following is true for the $M/Er/n$ model with batch arrivals:*

$$\pi^0_{X_{m_2}} = \delta_0 \left[S_0^{m_2} \delta_0 - B_0 \right]^{-1} \tag{3.2}$$

where $S_0^{m_2} = [1'_0 + \sum_{m=1}^{m_2} R_m^{*m} 1'_m]$ *and*

$$\pi^m \leq \pi^m_{X_{m_2}} = \pi^0_{X_{m_2}} R_m^{*m}.$$

Proof. Directly from Theorem 2.3 (for Eq. (3.2)) and Section 2.3 (Proposition 2.2) for the second claim.

□

3.2.2 The $Er/M/n$ Model.

In this section we derive limit solutions for the $Er/M/n$ queueing system. Specifically we consider a system with n identical servers, where the service time of a customer is exponentially distributed with parameter μ. The inter-arrival times are modeled as a sum of ℓ ($\ell < \infty$) distinguishable exponentially distributed 'phases', where the rate of the i-th phase may be a function of the number m of customers in line and it is denoted by $\lambda_{m,i}$. We can model the $Er/M/n$ queueing process as a QBD process on the state space $\mathcal{X} = \{L_0, L_1, \ldots\}$ with $L_m = \{(m,1), \ldots, (m,\ell)\}$. We use the state notation (m,i) to denote the event that there are m customers in the system and the customer inter-arrival time is at its i-th phase. In this QBD process, U^m, W^m and

D^m are of size $\ell \times \ell$ and are given below. For $m = 0, 1, \ldots$:

$$D^m = \mu_m I_\ell, \qquad U^m = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{m,\ell} & 0 & \cdots & 0 & 0 \end{bmatrix},$$

$$W^0 = \begin{bmatrix} -\lambda_{0,1} & \lambda_{0,1} & 0 & \cdots & 0 \\ 0 & -\lambda_{0,2} & \lambda_{0,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\lambda_{0,\ell-1} & \lambda_{0,\ell-1} \\ 0 & 0 & \cdots & 0 & -\lambda_{0,\ell} \end{bmatrix},$$

and for $m = 1, 2, \ldots$:

$$W^m = \begin{bmatrix} -(\lambda_{m,1}+\mu_m) & \lambda_{m,1} & 0 & \cdots & 0 \\ 0 & -(\lambda_{m,2}+\mu_m) & \lambda_{m,2} & \cdots & 0 \\ \vdots & \vdots & \vdots\ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -(\lambda_{m,\ell}+\mu_m) & \lambda_{m,\ell} \\ 0 & 0 & \cdots & 0 & -(\lambda_{m,\ell}+\mu_m) \end{bmatrix},$$

where $\mu_m = m\mu$ for $m \leq n-1$, $\mu_m = n\mu$ for $m \geq n$. We can now use the relabeling specified in Remark 3.1. In this case $\hat{D}^{-m} = U^m, \hat{U}^{-m} = D^m$, and $\hat{W}^{-m} = W^m$. The relabeled process satisfies Assumption 2.1 and has a QBD structure with $M_1 = -\infty$. The state space can be truncated at level m_1 as in Remark 2.6 (where the $\overline{\hat{w}}$'s are defined). Then, Eq. (2.38) and Theorem 2.8 from Appendix B can be used to compute limiting approximations $\hat{\pi}_{X_{m_1}}(-m, i)$ for the steady state probabilities $\pi(m, i)$ of the

$Er/M/n$ model as described below.

$$\hat{R}_{-m} = \mu_{m+1} \begin{bmatrix} \lambda_{m,1} & -\lambda_{m,1} & 0 & \cdots & 0 \\ -\mu_m & \lambda_{m,2}+\mu_m & -\lambda_{m,2} & \cdots & 0 \\ -\mu_m & 0 & \lambda_{m,3}+\mu_m & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mu_m & 0 & \cdots & 0 & \lambda_{m,\ell}+\mu_m \end{bmatrix}^{-1},$$

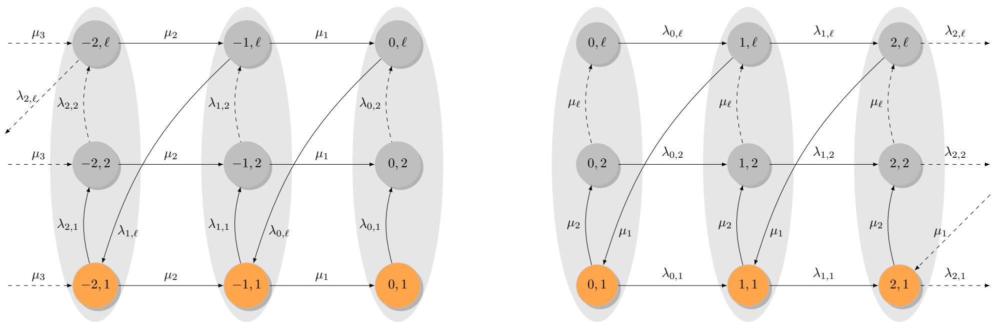
$$\hat{R}_0 = \mu_1 \begin{bmatrix} \lambda_{0\,1} & -\lambda_{0\,1} & 0 & \cdots & 0 \\ 0 & \lambda_{0\,2} & -\lambda_{0\,2} & \cdots & 0 \\ 0 & 0 & \lambda_{0\,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_{0\,\ell} \end{bmatrix}^{-1}.$$

The following are true for the steady state probabilities of the $\hat{Q}$-process are: $\hat{\pi}^{m_1}_{X_{m_1}} = \delta_{m_1}\left[S^0_{m_1}\delta_{m_1} - \widetilde{\hat{U}}^{\,m_1} - \overline{\hat{W}}^{\,m_1}\right]^{-1}$, $\hat{\pi}^{m}_{X_{m_1}} = \hat{\pi}^{m_1}_{X_{m_1}} \prod_{k=m_1}^{m-1} \hat{R}_k$ where $S^0_{m_1} = 1'_{m_1} + \sum_{m=m_1}^{0} \prod_{k=m_1}^{m-1} \hat{R}_k 1'_m$.

In addition, by Remark 3.1 we have:

$$\pi(m,i) = \hat{\pi}(-m,i) = \lim_{m_1 \to -\infty} \hat{\pi}_{X_{m_1}}(-m,i).$$

The homogeneous $M/Er/n$ and the $Er/M/n$ queueing systems have a very similar structure when the number of phases ℓ is equal. When considering the relabeled level process of $Er/M/n$ process and change the role of λ and μ, the $Er/M/n$ process is exactly the negative extension of the $M/Er/n$ queue as it is shown in Figure 3.2a.



(a) The $Er/M/n$ process (b) The $M/Er/n$ process

Figure 3.2: The left figure displays the transition rate diagram of the $\hat{Q}$ matrix of the $Er/M/n$ queue; the right figure is the diagram for the $M/Er/n$ queue.

3.3 An Inventory Model with Random Yield

In this section we consider an inventory model with random yield. Specifically, we investigate a system where customers arrive with rate λ and a batch of products arrives according to an exponential distribution with rate μ. Random yield is possible in this model, i.e. the size of the batch is $n\ell$ with probability p_n, where ℓ is a fixed positive constant, cf. Veinott (1965). We model this inventory model process as a QSF process $X(t)$ on state space $\mathcal{X} = \{L_0, L_1, \ldots\}$ with $L_m = \{(m,1), \ldots, (m,\ell)\}$. In state (m,i) there are $m\ell + i$ products in stock. Figure 3.3 displays the transition diagram of the described model with $\ell = 3$ and where the size of the batch is 3 with probability p and 6 with probability $1-p$. This model is a successively lumpable QSF process, where states $(m, \ell-1)$ are the entrance states of sub-sets $\underset{\sim}{L}_m$. In this QSF process, $U^{m,k}$, W^m and D^m are of size $\ell \times \ell$. It has the same structure as the queueing model described in Section 3.2.1 and it can be solved analogously.

We note that we can easily obtain explicit formulas for the steady state probabilities even in the case that both λ and μ may depend on the state. For example, when there is too much (little) inventory a discount (premium) price may be used for the product and this may change the arrival rate of the customers, i.e., $\lambda = \lambda(m,i)$. Also, when there is a high level of inventory one may decide not to order. This can be easily incorporated in product arrival rate, i.e., $\mu = \mu(m, i)$. Finally we note that just as easily one can handle the extension where the batch size $n\ell$ is replaced by $n\ell_m$ to represent dependency on the inventory level m. We omitted all these dependencies in this exposition only to simplify the notation.

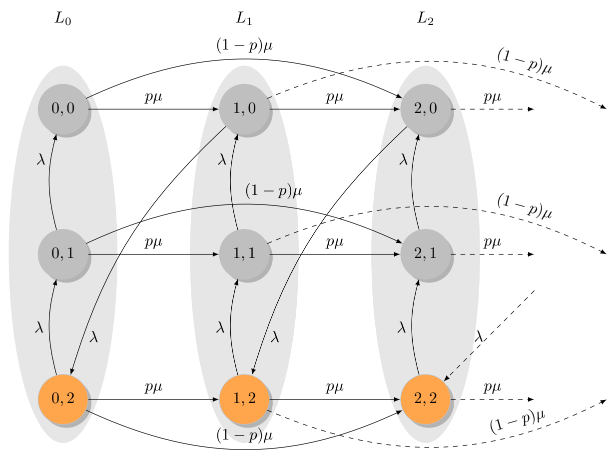


Figure 3.3: An Inventory Model with random yield.

3.4 A Restart system

Consider the classical restart (reliability) problem where a system of known structure is composed of N components and it operates continuously. The time to failure of component $i = 1, \ldots, N$ is exponentially distributed with rate μ_i and it is independent of the state of the other components. This type of systems has been studied in Derman et al. (1980), Katehakis and Derman (1984), Frostig (1999) and Hooghiemstra and Koole (2000) as well as Katehakis and Melolidakis (1995), Righter (1996), Koole and Spieksma (2001) and Ungureanu et al. (2006).

In this section we assume that when the system fails it is restored (or replaced) to a state "as good as new" and the time it takes for this restoration is exponentially distributed with rate λ.

These assumptions imply that at any point in time the state of the system can be identified by a boolean M-vector $x = (x_1, \ldots, x_M)$, with $x_i = 1$ if the $i-$th component is working, else $x_i = 0$. Hence $\mathcal{X} = \{0,1\}^M$ is the set of all possible states. Under these conditions the time evolution of the state of the system can be described by a continuous time Markov chain. The structure of the system is specified by a binary function ϕ defined on $\mathcal{X}$. Let $G = \{x \,:\, \phi(x) = 1\}$ denote the set of all operational (good) states of the system and let $B = \{x \,:\, \phi(x) = 0\}$ denote all failed states of the system. For such a system it is important to compute measures of performance such as the availability of the system defined as $\alpha_\phi = \sum_{x \in G} \pi(x)$. Regardless of the choice of the structure ϕ it is easy to see that the corresponding chain is successively lumpable.

For example, for the parallel system we have $B = \{(0, \ldots, 0)\}$. Figure 3.4 illustrates the transition diagram of the corresponding Markov process for the parallel system when $M = 3$. It is clear that this process is a RES QSF process with respect to partition $\mathcal{D} = \{L_0, \ldots, L_3\}$ of size $M + 1$, with $\forall x \in \mathcal{X}$:

$$x \in L_m, \text{ if } \sum_i x_i = m.$$

In this example of a restart process, when modeled as a QSF process, $M_1 = 0$ and $M_2 = 3$. The states are ordered as is shown in Figure 3.4, i.e. for example $(0,0,1)$ is the first state of level 1. We derive that D^m, W^m and U^{m,M_2} have the from given below:

$$W^0 = -\lambda, \quad U^{0,3} = \lambda,$$

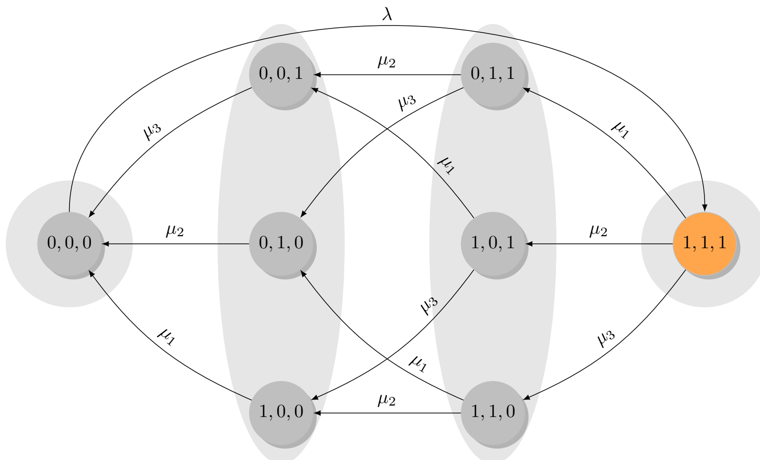


Figure 3.4: Transition diagram for parallel system with $M = 3$ servers.

$$D^1 = \begin{bmatrix} \mu_3 \\ \mu_2 \\ \mu_1 \end{bmatrix}, \quad W^1 = \begin{bmatrix} -\mu_3 & 0 & 0 \\ 0 & -\mu_2 & 0 \\ 0 & 0 & -\mu_1 \end{bmatrix}, \quad U^{1,3} = 0'_1$$

$$D^2 = \begin{bmatrix} \mu_2 & \mu_3 & 0 \\ \mu_1 & 0 & \mu_3 \\ 0 & \mu_1 & \mu_2 \end{bmatrix}, \quad W^2 = \begin{bmatrix} -\mu_2-\mu_3 & 0 & 0 \\ 0 & -\mu_1-\mu_3 & 0 \\ 0 & 0 & -\mu_1-\mu_2 \end{bmatrix}, \quad U^{2,3} = 0'_2$$

and

$$D^3 = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix}, \quad W^3 = -(\mu_1+\mu_2+\mu_3).$$

Using the first step in Algorithm 2.2 we derive the following rate matrices:

$$R_0 = -D^1(W^0)^{-1} = 1/\lambda \begin{bmatrix} \mu_3 \\ \mu_2 \\ \mu_1 \end{bmatrix}, \; R_1 = -D^2(W^1)^{-1} = \begin{bmatrix} \mu_2/\mu_3 & \mu_3/\mu_2 & 0 \\ \mu_1/\mu_3 & 0 & \mu_3/\mu_1 \\ 0 & \mu_1/\mu_2 & \mu_2/\mu_1 \end{bmatrix},$$

$$R_2 = -D^3(W^2)^{-1} = \begin{bmatrix} \mu_1/(\mu_2+\mu_3) & \mu_2/(\mu_1+\mu_3) & \mu_3/(\mu_1+\mu_2) \end{bmatrix}.$$

We can find the steady state distribution with the remaining steps of Algorithm 2.2. Note that in this case T_0^3 is a scalar and $\delta_3 = 1$.

- $T_0^3 = 1 + \sum_{m=0}^{2} \prod_{k=0}^{2-m} R_{2-k} 1'_m = 1 + R_2 R_1 R_0 1'_0 + R_2 R_1 1'_1 + R_2 1'_2$,

- $\pi^3 = \pi(1,1,1) = \delta_3 [T_0^3 \delta_3 - W^3 - \widetilde{D}^3]^{-1} = [T_0^3]^{-1}$,

- $\pi^2 = \pi^3 R_2$, $\pi^1 = \pi^2 R_1$, $\pi^0 = \pi^1 R_0$.

It is important to note that the successively lumpable property of the process results in the following computational gains for large M: instead of solving a system of size 2^M we only need to solve M systems the largest of which is of size $\binom{M}{\lfloor M/2 \rfloor} + 1$.

References

I.J.B.F. Adan, S. Kapodistria, and J.S.H. van Leeuwaarden. Erlang arrivals joining the shorter queue. *Queueing Systems*, 74(2-3):273–302, 2013.

J.R. Artalejo and A. Gómez-Corral. *Retrial queueing systems: a computational approach*. Springer Verlag, 2008.

J.R. Artalejo, A. Economou, and M.J. Lopez-Herrero. The maximum number of infected individuals in SIS epidemic models: Computational techniques and quasi-stationary distributions. *Journal of computational and applied mathematics*, 233 (10):2563–2574, 2010.

D. Bini and B. Meini. On the solution of a nonlinear matrix equation arising in queueing problems. *SIAM Journal on Matrix Analysis and Applications*, 17(4): 906–926, 1996.

D. Bini, G. Latouche, and B. Meini. *Numerical methods for structured Markov chains*. Oxford University Press, 2005.

L. Bright and P.G. Taylor. Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes. *Stochastic Models*, 11(3):497–525, 1995.

M. Brown, E.A. Peköz, and S.M. Ross. Some results for skip-free random walk. *Probability in the Engineering and Informational Sciences*, 24(04):491–507, 2010.

C. Derman, G.J. Lieberman, and S.M. Ross. On the optimal assignment of servers and a repairman. *Journal of Applied Probability*, pages 577–581, 1980.

K. Etessami, D. Wojtczak, and M. Yannakakis. Quasi-birth-death processes, tree-like qbds, probabilistic 1-counter automata, and pushdown systems. *Performance Evaluation*, 67(9):837–857, 2010.

B.N. Feinberg and S.S. Chui. A method to calculate steady state distributions of large Markov chains. *Operations Research*, 35(2):282–290, 1987.

E. Frostig. Jointly optimal allocation of a repairman and optimal control of service rate for machine repairman problem. *European Journal of Operational Research*, 116(2):274–280, 1999.

G. Hooghiemstra and G. Koole. On the convergence of the power series algorithm. *Performance Evaluation*, 42(1):21–39, 2000.

A. Hordijk and F. Spieksma. On ergodicity and recurrence properties of a Markov chain with an application to an open Jackson network. *Advances in Applied Probability*, 24(2):343–376, 1992.

S. Kapodistria. The M/M/1 queue with synchronized abandonments. *Queueing Systems*, 68(1):79–109, 2011.

M.N. Katehakis and C. Derman. Optimal repair allocation in a series system. *Mathematics of Operations Research*, 9(4):615–623, 1984.

M.N. Katehakis and C. Derman. On the maintenance of systems composed of highly reliable components. *Management Science*, pages 551–560, 1989.

M.N. Katehakis and A.F. Veinott Jr. The multi-armed bandit problem: decomposition and computation. *Mathematics of Operations Research*, 12(2):262–268, 1987.

M.N. Katehakis and C. Melolidakis. On the optimal maintenance of systems and control of arrivals in queues. *Stochastic Analysis and Applications*, 13(2):137–164, 1995.

M.N. Katehakis and L.C. Smit. On computing optimal (Q, r) replenishment policies under quantity discounts. *Annals of Operations Research*, 200(1):279–298, 2012a.

M.N. Katehakis and L.C. Smit. A successive lumping procedure for a class of Markov chains. *Probability in the Engineering and Informational Sciences*, 26(4):483–508, 2012b.

M.N. Katehakis, L.C. Smit, and F.M. Spieksma. DES and RES processes and their explicit solutions. *Submitted to Operations Research*, 2013.

J.G. Kemeny and J.L. Snell. *Finite Markov Chains*. D. van Nostrand Company, inc., Princeton, N.J., 1960.

J.P. Kharoufeh. *Level-Dependent Quasi-Birth-and-Death Processes*. Wiley Encyclopedia of Operations Research and Management Science., 2011.

D.S. Kim and R.L. Smith. An exact aggregation algorithm for a special class of Markov chains. *Technical Report*, 1989.

D.S. Kim and R.L. Smith. An exact aggregation-disaggregation algortihm for mandatory set decomposable Markov chains. *Numerical Solution of Markov Chains*, pages 89–104, 1990.

G.M. Koole and F.M. Spieksma. On deviation matrices for birth–death processes. *Probability in the Engineering and Informational Sciences*, 15(2):239–258, 2001.

G Latouche and V. Ramaswami. A logarithmic reduction algorithm for quasi-birth-death processes. *Journal of Applied Probability*, pages 650–674, 1993.

G. Latouche and V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling*, volume 5. ASA-SIAM Series on Statistics and Applied Probability. SIAM, Philadelphia, PA., 1999.

W. L. Miranker and V. Ya. Pan. Methods of aggregation. *Linear Algebra and its Application*, 29:231–257, 1980.

M.F. Neuts. *Matrix-Geometric Solutions in Stochastic Models - An Algorithmic Approach*. Dover Publications, Inc., New York, 1981.

J.F. Pérez and B. van Houdt. Quasi-birth-and-death processes with restricted transitions and its applications. *Performance Evaluation*, 68(2):126–141, 2011.

V. Ramaswami. A stable recursion for the steady state vector in Markov chains of M/G/1 type. *Communications in Statistics. Stochastic Models*, 4:183–188, 1988.

R. Righter. Optimal policies for scheduling repairs and allocating heterogeneous servers. *Journal of applied probability*, pages 436–547, 1996.

A. Riska and E. Smirni. M/G/1-type Markov processes: A tutorial. *Performance Evaluation of Complex Systems: Techniques and Tools*, 2459:36–63, 2002.

S.M. Ross. *Stochastic Processes*. John Wiley and Sons, NY., 1996.

P.J. Schweitzer, M.L. Puterman, and W.L. Kindle. Iterative aggregation-disaggregation procedures for discounted semi-Markov reward processes. *Operations Research*, 33(3):589–605, 1984.

E. Seneta. Computing the stationary distribution for infinite Markov chains. *Linear Algebra and Its Applications*, 34:259–267, 1980.

E. Seneta. *Non-negative matrices and Markov chains*. Springer, New York, 1981.

I.M. Sonin. Optimal stopping of Markov chains and three abstract optimization problems. *An International Journal of Probability and Stochastic Processes*, 83 (4-6):405–414, 2011.

F.M. Spieksma and R.L. Tweedie. Strengthening ergodicity to geometric ergodicity for Markov chains. *Stochastic Models*, 10(1):45–74, 1994.

W. Szpankowski. Stability conditions for multidimensional queueing systems with computer applications. *Operations Research*, 36(6):944–957, 1988.

H. Tong, C. Faloutsos, and J.Y. Pan. Fast random walk with restart and its applications. *Data Mining, ICDM '06: Sixth International Conference on Data Mining*, pages 613–622, 2006.

R.L. Tweedie. The calculation of limit probabilities for denumerable Markov processes from infinitesimal properties. *Journal of Applied Probability*, pages 84–99, 1973.

R.L. Tweedie. Criteria for ergodicity, exponential ergodicity and strong ergodicity of Markov processes. *Journal for Applied Probability*, 18(1):122–130, 1981.

V. Ungureanu, B. Melamed, M. Katehakis, and P.G. Bradford. Deferred assignment scheduling in cluster-based servers. *Cluster Computing*, 9(2):57–65, 2006.

B. van Houdt and J.S.H. van Leeuwaarden. Triangular M/G/1-type and tree-like quasi-birth-death Markov chains. *INFORMS Journal on Computing*, 23(1):165–171, 2011.

R.S. Varga. *Matrix iterative analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1963.

R.S. Varga. On recurring theorems on diagonal dominance. *Linear Algebra and its Applications*, 13(1):1–9, 1976.

A.F. Veinott. The optimal inventory policy for batch ordering. *Operations Research*, 13:424–432, 1965.

D. Vere-Jones. Ergodic properties of nonnegative matrices I. *Pacific Journal of Mathematics*, 22(2):361–386, 1967.

M.A. Woodbury. Inverting modified matrices. *Memorandum Report, Statistical Research Group*, 42, 1950.

V.B. Yap. Similar states in continuous time Markov chains. *Journal of Applied Probability*, 46(2):497–506, 2009.

Q. Zhang, N. Mi, A. Riska, and E. Smirni. Performance-guided load (un) balancing under autocorrelated flows. *IEEE Transactions on Parallel and Distributed Systems*, pages 652–665, 2007.

Curriculum Vitae

Laurens Christiaan Smit

1986	March 20, born in Leiden, Netherlands
2004-2007	Bachelor Mathematics at Leiden University, Netherlands
2008-2011	Master Applied Mathematics at Leiden University
2011	Spring Semester at Rutgers University, New Jersey, USA
2011-current	Ph.D program at Leiden University in Applied Probability
2012-2014	Ph.D program at Business School, Rutgers University, Newark & New Brunswick in Management Science

Articles

2012	M.N. Katehakis and L.C. Smit. On computing optimal (Q, r) replenishment policies under quantity discounts, *Annals of Operations Research*, 200 (1) 279–298.
2012	M.N. Katehakis and L.C. Smit. A successive lumping procedure for a class of Markov chains, *Probability in the Engineering and Informational Sciences*, 26(4):483–508.
2013	M.N. Katehakis, L.C. Smit and F.M. Spieksma. Explicit Solutions for a class of Quasi Skip Free Processes, *Submitted*.
2014	M.N. Katehakis, L.C. Smit and F.M. Spieksma. A Comparative Analysis of Successive Lumping and Lattice Path Counting, *Work in Progress*
2014	D.Ertiningsih, M.N. Katehakis, L.C. Smit and F.M. Spieksma. Level product form Quasi-Skip Free Processes: an application to the $E_k/M/1$-queue, *Work in Progress.*

Honors

2012	Finalist at the New Jersey Chapter Contest, Research competition for Operation Research Students from New Jersey
2013	Dean's Award for Ph.D. Student Research, Rutgers Business School, Newark & New Brunswick
2013	Winner of the New Jersey Chapter Contest, Research competition for Operation Research Students from New Jersey