

Laboratório Nacional de Computação Científica  
Programa de Pós Graduação em Modelagem Computacional

**Aproximações para Redes Estocásticas Sinalizantes Sob  
Tráfego Pesado**

Por  
**Saul de Castro Leite**

PETRÓPOLIS, RJ - BRASIL

AGOSTO DE 2009

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

**APROXIMAÇÕES PARA REDES ESTOCÁSTICAS  
SINALIZANTES SOB TRÁFEGO PESADO**

**Saul de Castro Leite**

TESE SUBMETIDA AO CORPO DOCENTE DO LABORATÓRIO NACIONAL  
DE COMPUTAÇÃO CIENTÍFICA COMO PARTE DOS REQUISITOS NE-  
CESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS  
EM MODELAGEM COMPUTACIONAL

Aprovada por:

---

Prof. Marcelo Dutra Fragoso, Ph.D.  
(Presidente)

---

Prof. Márcio Arab Murad, D.Sc.

---

Prof. João Bosco Ribeiro do Val, Ph.D.

---

Prof. Takashi Yoneyama, Ph.D.

PETRÓPOLIS, RJ - BRASIL  
AGOSTO DE 2009

Leite, Saul de Castro

\_533a            aproximações para redes estocásticas sinalizantes sob tráfego pesado /  
Saul de Castro Leite. Petrópolis, RJ. : Laboratório Nacional de Computação  
Científica, 2009.

xviii, 237 p. : il.; 29 cm

Orientador: Marcelo Dutra Fragoso

Tese (D.Sc.) – Laboratório Nacional de Computação Científica, 2009.

1. Teoria das filas. 2. Processos Estocásticos. 3. Aproximações por  
tráfego pesado. 4. Redes estocásticas. I. Fragoso, Marcelo Dutra. II.  
LNCC/MCT. III. Título.

CDD - 519.82

Patience, patience,  
Patience dans l'azur!  
Chaque atome de silence  
Est la chance d'un fruit mûr!

*Paul Valéry (retirado de "Um Pouco Mais  
de Azul" por Hubert Reeves, 1986).*

À memória do meu avô Leite,  
quem participa das minhas mais saudosas  
lembranças de infância.

# Agradecimentos

Aos meus pais, Heloisa e José Luiz, pelo total apoio e suporte que foram imprescindíveis para enfrentar esse desafio.

À minha família, aos meus irmãos, Sarah e Samuel, e à Fernanda pelo apoio, carinho, e compreensão durante esses anos.

Ao meu orientador, Prof. Marcelo D. Fragoso, que dedicou inúmeras horas do seu tempo me escutando em seminários semanais durante alguns anos. Me sinto muito privilegiado pela oportunidade de trabalhar junto a ele.

Gostaria de agradecer também ao Prof. Kushner, que foi muito atencioso e prestativo ao responder meus emails com questões relacionadas ao seu livro. Agradeço também ao Prof. Ziviani pela ajuda na discussão sobre a aplicação do modelo em sistemas de busca na rede.

A todos envolvidos no projeto COMUT, sem o qual não seria possível ter acesso à tantos artigos essenciais para o desenvolvimento deste trabalho.

A todos envolvidos no programa de Pós-graduação do LNCC e ao CNPq pelo apoio financeiro.

Resumo da Tese apresentada ao LNCC/MCT como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

## APROXIMAÇÕES PARA REDES ESTOCÁSTICAS SINALIZANTES SOB TRÁFEGO PESADO

Saul de Castro Leite

Agosto , 2009

**Orientador:** Marcelo Dutra Fragoso, Ph.D.

Este trabalho apresenta a caracterização de limites no sentido fraco dos sistemas de filas em redes que podem enviar e receber sinais. Estes sinais podem ser usados, entre outras coisas, para que as filas se auto controlem. Mostra-se que, sob certas condições, o sistema pode ser aproximado por uma equação diferencial estocástica refletida. Os benefícios de tais aproximações são que elas descrevem a evolução transiente destes sistemas e possibilitam a introdução de controles. Em seguida, uma abordagem mais abrangente é apresentada através de redes de Petri estocásticas. Uma nova classe destas redes é introduzida como uma forma unificadora para tratar sistemas que podem ser descritos por quantidades discretas que sofrem trocas estocásticas ao longo do tempo. A classe é geral o suficiente para incluir as redes de Petri estocásticas, as redes de Jackson, e as redes de Gelenbe com sinais do tipo “cliente negativo” e do tipo “*triggers*”. O objetivo principal é obter, de maneira unificada, uma aproximação por difusão que possa ser facilmente aplicável em um número grande de problemas práticos.

Abstract of Thesis presented to LNCC/MCT as a partial fulfillment of the requirements for the degree of Doctor of Sciences (D.Sc.)

## HEAVY TRAFFIC APPROXIMATIONS FOR SIGNALING NETWORKS

Saul de Castro Leite

August, 2009

**Advisor:** Marcelo Dutra Fragoso, Ph.D.

This work is concerned with the characterization of weak-sense limits of networks of queues which can send and receive signals. These signals can be used, among other things, to auto control the queues. It is shown that, under certain conditions, the system can be approximated by a reflected stochastic differential equation. The benefits of such an approximation are that it describes the transient evolution of these systems and allows for the introduction of controls. Following that, a more general approach is presented using stochastic Petri nets. A new class of these nets is introduced to serve as an unifying approach to treat systems that can be described by discrete quantities that suffer stochastic changes in time. The class is general enough to include the stochastic Petri nets, the Jackson network, and Gelenbe's network with signals of the type "negative customer" and "triggers." The main goal in this approach is to obtain, in a unifying way, an approximation that can be readily applied in several practical problems.

# Sumário

<b>1</b>	Introdução	1
1.1	Estrutura da Tese . . . . .	4
<b>2</b>	Da Teoria Clássica à Aproximação Por Difusão	6
2.1	Preliminares . . . . .	7
2.2	Teoria Clássica . . . . .	10
2.3	O Modelo $M/GI/1$ para Filas . . . . .	14
2.4	Métodos Analíticos Matriciais ( <i>Matrix Analytic Methods</i> ) . . . . .	23
2.5	O Modelo $GI/GI/1$ e Condições de Estabilidade . . . . .	30
2.6	Aproximações de Filas Sob Tráfego Pesado . . . . .	34
2.7	Redes de Filas . . . . .	44
2.8	Aproximação por Fluido . . . . .	52
<b>3</b>	Modelos de Fila com Sinais	54
<b>4</b>	Filas com Sinais Sob Tráfego Pesado	61
4.1	Modelo para o Número de Clientes no Servidor . . . . .	63
4.2	Modelo para o Processo de Carga de Trabalho . . . . .	67
4.3	Modelo com Chegadas Moduladas por um Processo de Markov . . . . .	70
4.4	Experimentos Numéricos . . . . .	79
<b>5</b>	Redes de Filas com Sinais Sob Tráfego Pesado	84
5.1	Modelo de Filas . . . . .	89

5.2	Limite em Regime de Tráfego Pesado . . . . .	95
5.2.1	Enfraquecendo as Condições Sob as Funções $f_i^\alpha$ e $f_{ij}^\beta$ . . . . .	106
5.3	Topologias de Rede Satisfazendo a Condição 5.2.2(b) . . . . .	110
5.3.1	Duas Filas em Sequência . . . . .	110
5.3.2	Redes com Duas Camadas . . . . .	111
5.4	Experimentos Numéricos . . . . .	112
<b>6</b>	<b>Estudo de Caso: Aplicação para Sistemas de Processamento Paralelo</b>	<b>117</b>
6.1	Modelo de Filas . . . . .	120
6.2	Limite em Regime de Tráfego Pesado . . . . .	123
6.3	Aplicação: Sistemas de Busca na Internet . . . . .	129
6.3.1	Simulação Computacional . . . . .	130
6.3.2	Aplicação do Modelo Teórico . . . . .	131
6.3.3	Formulação do Problema de Controle Ótimo . . . . .	133
6.3.4	Resultados Numéricos . . . . .	133
<b>7</b>	<b>Uma Estrutura Unificante para Redes com Sinais</b>	<b>143</b>
7.1	Redes de Petri . . . . .	146
7.2	Redes de Petri Estocásticas . . . . .	148
7.2.1	Limite em Regime de Tráfego Pesado . . . . .	152
7.3	Redes de Petri Estocásticas com Transições Aleatórias . . . . .	159
7.3.1	Limite em Regime de Tráfego Pesado . . . . .	162
7.4	Experimentos Numéricos . . . . .	167
7.4.1	O Modelo de Predador-Presa de Lotka-Volterra . . . . .	167
7.4.2	Controle de Congestionamento de Rede . . . . .	169
7.4.3	Certificador de Qualidade em Linhas de Produção . . . . .	173
<b>8</b>	<b>Conclusões e Trabalhos Futuros</b>	<b>175</b>

<b>Referências Bibliográficas</b>	178
-----------------------------------	-----

## **Apêndice**

<b>A</b>	Processo de Difusão Refletido	192
A.1	Existência e Unicidade de Soluções para Equações Diferenciais Estocásticas Refletidas . . . . .	193
A.1.1	Existência e Unicidade Forte . . . . .	194
A.1.2	Existência de Solução Fraca . . . . .	207
A.2	Convergência Fraca de Processos de Reflexão Aproximados . . . . .	212
A.3	Propriedade do Termo de Reflexão nas Fronteiras . . . . .	215
<b>B</b>	Resultados Auxiliares e Definições	223
B.1	Convergência Fraca . . . . .	223
B.1.1	Espaço de Skorohod . . . . .	224
B.1.2	Critérios . . . . .	225
B.1.3	Convergência Fraca para Processos de Wiener . . . . .	226
B.2	Processos de Salto . . . . .	230

# Lista de Figuras

## Figura

3.1	Duas filas em sequência com sinais. O símbolo + (resp., -) indica a possibilidade de chegada de clientes (resp., sinais). . . . .	55
4.1	Uma linha de eventos ilustrando a discrepância gerada quando um cliente é removido em serviço. . . . .	62
4.2	Uma linha de eventos ilustrando a diferença entre o tempo entre clientes removidos e chegada de sinais. . . . .	63
4.3	Gráfico do número médio de clientes no sistema para diferentes valores de intensidade de tráfego $\rho$ . Os resultados foram obtidos usando uma simulação computacional (SIM) e através da aproximação por tráfego pesado (HTA). Para a simulação, também é mostrado o intervalo de 95% t-confiança (veja (Gross e Harris, 1998), página 392). As Figuras (a) e (b) mostram o resultado para o primeiro e segundo exemplo respectivamente . . . . .	81
4.4	Gráfico do número médio de carga de trabalho para diferentes valores de intensidade de tráfego $\rho$ . Os resultados foram obtidos usando uma simulação computacional (SIM) e através da aproximação por tráfego pesado (HTA). Para a simulação, também é mostrado o intervalo de 95% t-confiança (veja (Gross e Harris, 1998), página 392). As Figuras (a) e (b) mostram o resultado para o primeiro e segundo exemplo respectivamente. . . . .	82

4.5	Gráfico do número médio de clientes na fila para o modelo com processo de entrada modulado por uma cadeia de Markov para diferentes valores de intensidade de tráfego $\rho$ . Os resultados foram obtidos usando uma simulação computacional (SIM) e através da aproximação por tráfego pesado (HTA). Para a simulação, também é mostrado o intervalo de 95% t-confiança (veja (Gross e Harris, 1998), página 392). Figuras (a) e (b) mostram o resultado para o primeiro e segundo exemplo respectivamente. . . . .	83
5.1	Duas filas em sequência com sinais. O símbolo “+” indica a entrada de clientes, já o símbolo “-” indica a entrada de sinais. . . . .	84
5.2	Gráfico representando um problema na representação da direção de reflexão na fronteira $\partial G_2 = \{\xi \in \mathbb{R}^2 : \xi_2 = 0\}$ . Cada eixo representa o número de clientes no sistema . . . . .	86
5.3	Duas filas em sequência com sinais. O símbolo + (resp., -) indica a possibilidade de chegada de clientes (resp., sinais). . . . .	110
5.4	Uma rede de filas com duas camadas. . . . .	111
5.5	Função amostra de $x$ com $g \equiv s$ (azul) e $g \equiv 0$ (vermelho). (a) Gráfico de $x_1$ no tempo. (b) Gráfico de $x_2$ no tempo. (c) Gráfico de $x_1$ versus $x_2$ . . . . .	114
5.6	Curvas de chaveamento dadas pelo controle ótimo para diferentes valores de $c$ e $v$ . O controle é aplicado em força máxima na região denominada de <i>Active</i> e não é aplicado na região denominada de <i>Inactive</i> . . . . .	116
6.1	Diagrama de duas estações em paralelo. (a) Filas antes da chegada do quarto cliente. (b) Um sistema sem estratégia de balanceamento. (c) Um sistema com controle centralizado. (d) Um sistema com controle distribuído. . . . .	119

6.2	Gráfico da função de distribuição cumulativa exponencial e hiperexponencial . . . . .	132
6.3	(a) Gráfico do tempo de resposta do sistema simulado para taxas de entradas variadas (o intervalo de 95% $t$ -confiança para simulação também está presente no gráfico). (b) Curvas de <i>switching</i> para o controle ótimo $f_2^{c,*}$ para valores variados de $\eta \in \{1, 5, 10, 15, 20, 25\}$ (o espaço de estado mostrado na figura está escalonado por $1/\sqrt{100}$ ).	135
6.4	(a) Região do espaço de estado (escalonado) onde cada controle age. Controle 1 e 2 são referentes à $f_1^c$ e $f_2^c$ , respectivamente. As regiões são dadas aproximadamente por $R_1 = \{x \in \mathbb{R}_+^2 : x_1 + 0.5 > x_2\}$ e $R_2 = \{x \in \mathbb{R}_+^2 : x_1 - 0.5 < x_2\}$ . A linha azul atravessando o meio do espaço de estado é a linha $x_1 = x_2$ . (b) O gráfico do tempo de resposta do sistema calculado usando a simulação com valores variados de taxa de chegada para o sistema com controle e o sistema sem controle (o intervalo de 95% $t$ -confiança para simulação também está presente no gráfico) . . . . .	137
6.5	(a) Ganho relativo do sistema com controle em relação ao ( <i>w.r.t</i> ) sistema sem controle e o sistema usando o controle SIB. (b) Ganho relativo do sistema com controle em relação ao sistema sem controle variando o número de estações em paralelo com taxa de entrada fixa em 28 perguntas/segundo (o intervalo de 95% $t$ -confiança para simulação também está presente no gráfico). . . . .	139

- 6.6 (a) Região do estado de espaço (escalonado por  $1/\sqrt{n}$ ) onde cada controle ótimo age. Os controles 1 e 2 são referentes às funções  $f_1^c$  e  $f_2^c$ , respectivamente. As regiões são aproximadamente dadas por  $R_1 = \{x \in \mathbb{R}_+^2 : 1.34x_1 + 0.2 > x_2\}$  e  $R_2 = \{x \in \mathbb{R}_+^2 : 1.34x_1 - 1.6 < x_2\}$ . A linha azul que cruza o espaço de estados é a linha  $x_1 = x_2$ . Neste gráfico, a fila 2 é mais rápida que a fila 1. (b) Gráfico do tempo de resposta do sistema com diferentes valores para a taxa de chegada para o sistema com controle e sem controle. (o intervalo de 95%  $t$ -confiança para a simulação também está presente no gráfico). 140
- 6.7 (a) Ganho relativo do sistema com controle em relação ao sistema sem controle, com o controle SIB, e usando somente o controle para sistemas heterogêneos. (b) Ganho relativo do sistema com controle em relação ao sistema sem controle variando o número de servidores no sistema com taxa de entrada fixa em 28 perguntas/segundo (o intervalo de 95%  $t$ -confiança também está presente no gráfico). . . . 142
- 7.1 Representação gráfica de uma rede de Petri. (a) Os círculos em preto dentro dos nós representam item antes da execução da transição  $t_1$ . (b) Os círculos em preto indicam a marca da rede após a execução da transição  $t_1$ . No estado presente da rede, a transição  $t_1$  não pode ser disparada novamente já que não existem itens suficientes no nó  $p_2$ . 147
- 7.2 (a) Função  $x$  versus o tempo para o modelo predador-presa. (b) Gráfico da componente  $x_1$  versus a componente  $x_2$  para a mesma função amostra. . . . . 170
- 7.3 Diagrama de um sistema de filas com controle de congestionamento. 171
- 7.4 Superfície de chaveamento para o controle ótimo no espaço de estados escalonado. O controle é aplicado em força máxima na região debaixo da curva. Abaixo, tem-se o mapa de contorno desta superfície. 172

7.5 Diagrama de um sistema de linha de produção com certificador de qualidade. A linha pontilhada indica produtos que são enviados de volta para reprocessamento ou são enviados para uma terceira estação para serem restaurados. . . . . 173

# Lista de Tabelas

## Tabela

- 2.1 Tabela mostrando o número médio de clientes na rede para cada fila em regime estacionário. Para a simulação, o intervalo de 95% t-confiança está em parenteses. . . . . 51
  
- 6.1 Tabela mostrando o tempo de resposta médio obtido da simulação para um sistema com controle ótimo calculado usando diferentes valores para o parâmetro  $\eta$  na função de custo. . . . . 136

# Lista de Símbolos

- $(\Omega, \mathbb{P}, \mathcal{F}, \mathcal{F}_t)$  - base estocástica, onde  $(\Omega, \mathbb{P}, \mathcal{F})$  é um espaço de probabilidade e  $\{\mathcal{F}_t, t \geq 0\}$  é uma filtragem. Sempre será assumido que  $\mathcal{F}_t$  satisfaz as hipóteses usuais (veja a Definição 2.25 de (Karatzas e Shreve, 1988), na página 10).
- $\Delta_l^a$ : Intervalo de tempo entre a chegada do cliente  $l$  e o cliente  $l - 1$ .
- $\Delta_l^{d,n}$ : Tempo de serviço do cliente  $l$ .
- $\mathbb{N} \triangleq \{1, 2, \dots\}$ : o conjunto dos números naturais e  $\mathbb{N}_0 \triangleq \mathbb{N} \cup \{0\}$ .
- $\{S + u\} \triangleq \{a + u | a \in S\}$ , para um conjunto  $S$  e escalar  $u$ .
- $\mathbb{R}_{\geq 0} \triangleq \{\xi \in \mathbb{R} | \xi \geq 0\}$  e  $\mathbb{R}_{> 0} \triangleq \{\xi \in \mathbb{R} | \xi > 0\}$ .
- $\rho$ : intensidade de tráfego de um sistema de filas.
- $a'$ : transposto de  $a$ , para um vetor ou matriz  $a$ .
- $\text{diag}(v)$ , para  $v \in \mathbb{R}^n$ , é definido como a matriz diagonal em  $\mathbb{R}^{n \times n}$  com estradas diagonais dadas pelo vetor  $v$ .
- $\langle y, x \rangle$ : produto interno entre dois vetores  $x$  e  $y$ .
- $\partial B$ : fronteira de um subconjunto  $B$  de um espaço topológico.
- $B^0$ : interior de um subconjunto  $B$  de um espaço topológico.
- $D(\mathbb{R}^n; 0, \infty)$ : conjunto das funções *càdlàg*. Veja o Apêndice  $B$ .
- $C(\mathbb{R}^n; 0, \infty)$ : conjunto das funções contínuas de  $[0, \infty)$  à  $\mathbb{R}^n$ . Veja Apêndice  $B$ .
- $\vee$ : denota o máximo, i.e.,  $a \vee b \triangleq \max\{a, b\}$ .
- $\wedge$ : denota o mínimo, i.e.,  $a \wedge b \triangleq \min\{a, b\}$ .
- $\Rightarrow$  representa convergência fraca. Veja o Apêndice  $B$ .
- $\sigma(A)$  representa o raio espectral da matriz  $A$ .
- $\text{dist}(\xi, G) \triangleq \inf\{\|x - \xi\| : x \in G\}$  para um ponto  $\xi$  e um domínio  $G$ .
- $\text{Cone}(\{d_i\})$  representa o conjunto de vetores gerados por uma combinação convexa dos vetores  $\{d_i\}$ .

# Capítulo 1

## Introdução

No início do século passado, A. K. Erlang deu início aos primeiros estudos de teoria de filas em suas pesquisas sobre o congestionamento de tráfego de ligações telefônicas (Erlang, 1909). Desde aquela época, viu-se a importância de construir modelos matemáticos para obter quantidades, tais como: o tempo médio de espera de cada cliente e o tempo ocioso dos servidores (Gross e Harris, 1998; Kleinrock, 1975), que são fundamentais para o planejamento destes sistemas.

Atualmente, quando se pensa em sistemas de computadores, é extremamente fácil identificar processos que podem ser modelados sob o formalismo da teoria de filas. De fato, qualquer recurso computacional limitado que necessita atender a uma grande demanda de usuários pode ser reconhecido como um sistema de filas. Desta forma, com o advento dos computadores, aumentaram a necessidade e a demanda de modelos mais complexos para atender as novas aplicações que surgiam, especialmente em sistemas em redes e de comunicações. Contudo, dois grandes problemas permeiam a teoria clássica de filas e dificultam a sua aplicação em certos problemas.

Um destes problemas é a incapacidade de gerar formas fechadas ou de fácil utilização e interpretação para sistemas onde o processo de chegada ou saída não possuem uma estrutura Markoviana. Talvez um dos artigos mais representativos desta dificuldade seja o de Kingman (Kingman, 1966), no qual é apresentado uma álgebra com o propósito de unificar os resultados de trabalhos para filas em casos

gerais (i.e. não necessariamente Markovianos). Uma das soluções para este problema consiste em criar classes de processos que podem ser usados para aproximar o processo de entrada ou saída desejado através da elaboração do chamado “método de estágios” de Erlang. Nestes métodos, um processo Markoviano bidimensional dado por  $\{(N(t), J(t)); t \in \mathbb{R}_{\geq 0}\}$ , com  $N$  sendo um processo de contagem e  $J$  uma variável de fase (ou estágio), é introduzido de forma que uma estrutura markoviana ainda possa ser recuperada no processo que descreve o estado da fila (veja, por exemplo, (Neuts, 1975)). Atualmente, estes métodos de análise são comumente chamado de Métodos Analíticos Matriciais. Apesar de trazerem vários avanços na teoria, estes métodos não são facilmente estendíveis para o caso de redes de filas, que é mais comumente encontrado nos problemas atuais.

Outro problema é a dificuldade de se obter resultados que dizem respeito a evolução transiente destes sistemas. Geralmente, os resultados da teoria são obtidos a partir da hipótese de que o sistema está em regime estacionário. Os poucos resultados que tratam a dinâmica destes sistemas são de difícil uso prático para poderem responder, por exemplo, questões relacionadas a controles ótimos, mesmo para os modelos mais simples. Todas estas dificuldades tornam-se ainda maiores a medida que as filas são combinadas em redes.

Assim, estas duas dificuldades impulsionaram o estudo de aproximações. Existem dois principais tipos de aproximações encontradas na literatura: as aproximações por fluido e por difusão. Geralmente, modelos de fluido descrevem a dinâmica do valor médio do sistema por uma equação diferencial. As aproximações por difusão diferem das aproximações por fluido no fato que o efeito estocástico encontrado em sistemas de filas (devido à incertezas geradas pelos processos de chegada de clientes e de serviço) não é eliminado do modelo, aparecendo como um processo de Wiener ou como uma integral de Itô. Portanto, neste sentido, estas aproximações são mais fiéis à dinâmica do sistema. Contudo, este ganho vem acompanhado da necessidade da “hipótese do tráfego pesado” que exige que a taxa de clientes entrando em cada fila do sistema seja próxima da taxa com que os clientes

saem da fila. Este é um cenário comum em várias aplicações de interesse, como é o caso de sistemas computacionais modernos, que geralmente operam próximo de sua capacidade máxima (Kushner, 2001).

Aproximações por difusão vêm sendo estudadas desde os trabalhos pioneiros de Kingman (Kingman, 1961), Prohorov (Prohorov, 1963) e Borovkov (Borovkov, 1964, 1965), no início da década de 60. Um aspecto interessante das aproximações por difusão é que elas oferecem uma “visão macroscópica” (Whitt, 1974) da interação complexa presente nestas redes, deixando claro aquilo que é regular e o que é ruído. Além disso, a análise via tráfego pesado tem sido bastante promissora na simplificação e resolução de problemas de controle em teoria de filas e redes de comunicação (Kushner, 2001). Existe atualmente uma vasta bibliografia com aplicações destes modelos em problemas de controle em sistemas computacionais (veja, por exemplo, (Altman e Kushner, 2002; Buche e Kushner, 2002; Altman e Kushner, 1999; Kushner e Martins, 1993)).

A proposta deste trabalho é o estudo de aproximações por difusão em sistemas de fila que podem receber e enviar sinais entre elas. Estes sinais podem ser usados para remover clientes de qualquer fila no sistema ou mover clientes entre filas. Estes sistemas de filas tiveram suas origens com Gelenbe em (Gelenbe et al., 1991), e vem sendo amplamente estudados nos últimos anos (veja, por exemplo, (Artalejo, 2000) para uma revisão no assunto), motivado principalmente por uma série de aplicações. Dentre estas aplicações incluem-se o balanceamento de carga de trabalho em uma rede de computadores, e também aplicações em biologia como em redes neurais (Gelenbe, 1994; Gelenbe e Stafylopatis, 1991; Gelenbe, 1989) e modelos para regulação genética (Gelenbe, 2007; Arazi et al., 2004). Contudo, pouco foi explorado em relação à evolução transiente destes sistemas que é tão importante para tratar problemas de controle. Recentemente, alguns trabalhos foram feitos nesta direção (Guffens et al., 2006; Arazi et al., 2005) usando aproximações por fluido.

Através de redes de Petri, foi possível criar uma classe de redes que extrai

a característica essencial do problema tratado: redes que podem ser descritas por quantidades discretas que sofrem trocas estocásticas entre si ao longo do tempo. Esta nova classe de redes, que foi chamada de redes de Petri com transições probabilísticas, inclui como casos especiais as redes de filas com sinais de Gelenbe, com sinais do tipo “cliente negativo” e “*triggers*”, as redes clássicas de Jackson, e as redes de Petri estocásticas. A aproximação por difusão foi então desenvolvida para esta classe de redes tornando-a, assim, um modelo unificado para estes tipos de problemas.

## 1.1 Estrutura da Tese

Como a área de aproximações por difusão é relativamente nova e pouco difundida, encontram-se poucos trabalhos introdutórios na área. Uma excelente referência, sem dúvida, é o livro (Kushner, 2001). Os livros (Whitt, 2002) e (Chen e Yao, 2001) também são muito úteis. Apesar disso, estes livros não abordam a razão do aparecimento destas aproximações e porque elas são tão apreciadas. Para isso, foi preciso resgatar vários artigos iniciais e ganhar conhecimento mais detalhado sobre a teoria de filas. A intenção do Capítulo 2 é justamente facilitar o acesso à área para futuros leitores. Este capítulo contém uma revisão sobre teoria de filas destacando os resultados que ajudaram e motivaram o aparecimento das aproximações por tráfego pesado.

O Capítulo 3 contém uma breve revisão sobre redes de filas com sinais, destacando como estes modelos se encaixam no cenário geral da teoria das filas. Além disso, o capítulo resume alguns resultados recentes sobre aproximações por fluido em redes de filas com sinais.

Os capítulos seguintes (Cap. 4 - Cap. 7) contém os resultados obtidos durante o desenvolvimento desta tese, incluindo aplicações. O Capítulo 4 contém a primeira aplicação da aproximação por tráfego pesado em uma fila única com sinais, *os resultados contidos neste capítulo são referentes aos trabalhos* (Leite e Fragoso, 2007, 2008d). No Capítulo 5, é tratada a aproximação para redes de

filas com sinais do tipo “cliente negativo”. As taxas dos processos de chegada e de serviço são dependentes do estado, bem como a matriz de roteamento. *O conteúdo é referente ao trabalho* (Leite e Fragoso, 2008a). O Capítulo 6 contém o desenvolvimento da aproximação para filas com o sinal do tipo “*triggers*”, com a aplicação em sistemas de processamento paralelo. *Os resultados deste capítulo são referentes ao trabalho* (Leite e Fragoso, 2008b,c). Em seguida, o Capítulo 7 trata a abordagem mais geral através de redes de Petri discutida anteriormente. *Os resultados deste capítulo são referentes ao trabalho* (Leite e Fragoso, 2009).

O Apêndice A foi redigido seguindo a mesma linha de pensamento do que foi dito para o Capítulo 2. Ele traz o aspecto fundamental para entender equações diferenciais estocásticas refletidas - o processo de reflexão. O estudo deste processo foi fundamental para poder tratar um problema que surgiu no desenvolvimento desta tese. Além disso, os resultados contidos neste apêndice serão usados durante todo o trabalho.

O Apêndice B contém uma miscelânea de resultados e definições que serão usados no decorrer deste trabalho. Eles estão contidos aqui para facilitar a referência do leitor. Além disso, o apêndice contém provas de resultados menores que foram desenvolvidas para este trabalho e, portanto, não são facilmente encontradas em outras referências.

# Capítulo 2

## Da Teoria Clássica à Aproximação Por Difusão

O objetivo deste capítulo é reunir um conjunto pequeno mas representativo de resultados da teoria clássica de filas, destacando resultados motivadores para o aparecimento e desenvolvimento das aproximações por tráfego pesado. Um tema constante neste capítulo será o de estabelecer condições de estabilidade e construir analiticamente formas fechadas para as distribuições de quantidades que descrevem o sistema em regime estacionário. O capítulo está dividido em oito seções que serão descritas a seguir. Na primeira seção, será introduzido o problema de filas e algumas nomenclaturas. Em seguida, será considerado o primeiro modelo de um sistema de filas com distribuições exponenciais (chamado de modelo  $M/M/1$ ). Na Seção 2.3, será discutido o modelo de filas  $M/G/1$  e o método de análise chamando de “cadeia de Markov embutida”. Além disso, condições de estabilidade serão introduzidas sob o formalismo de cadeias de Markov. Na Seção 2.4, será sumarizado alguns resultados de modelos que utilizam os métodos analíticos matriciais (*Matrix Analytic Methods*), como são geralmente chamados na literatura. Estes modelos são relacionados aos das filas  $M/G/1$  pois são construídos de forma que uma estrutura Markoviana é preservada e pode-se ainda aplicar o método da cadeia de Markov embutida. A Seção 2.5 contém uma análise de estabilidade para sistemas de filas mais gerais. Esta análise é feita através da exploração do fato de que o tempo de espera de um cliente no sistema pode ser visto como um passeio aleatório

com uma barreira em zero. Na seção seguinte, são introduzidos alguns resultados que foram os primeiros passos direcionados à obtenção das aproximações por tráfego pesado. E na Seção 2.7, será introduzida a idéia de redes de filas e como a aproximação por tráfego pesado é utilizada nestes casos. A Seção 2.8 trata de uma forma breve as aproximações por fluido, que se assemelham às de tráfego pesado em aplicação, mas possuem algumas diferenças fundamentais.

É importante mencionar aqui que as provas presentes neste capítulo foram incluídas, dentre outras coisas, pelos seguintes motivos: 1) algumas provas são detalhamento (ou com correções) daquelas encontradas nas referências, e 2) as provas apresentadas aqui foram elaboradas no contexto da discussão e apresentação dos resultados da Tese. Além disso, acreditamos que isso facilitará o trabalho daqueles interessados em utilizar o texto em questão como um instrumento de introdução nesse tópico de pesquisa.

## **2.1 Preliminares**

Considere uma estação que provê serviço para uma série de clientes que chegam de forma aleatória. Cada cliente precisa permanecer por um certo período de tempo nesta estação até que sua tarefa termine. Durante este tempo, nenhum outro cliente pode ser servido (diz-se que a estação, ou servidor, está ocupado durante este tempo). Clientes que chegam quando o servidor está ocupado formam uma fila para esperar por serviço. Quando um cliente completa sua tarefa, ele deixa o sistema e imediatamente um outro cliente começa a ser servido. Se a fila estiver vazia, o servidor permanece ocioso a espera de um novo cliente. Este sistema simples é o tipo tratado pela teoria de filas.

Como apontado por Kendall em (Kendall, 1953), um sistema de filas pode ser descrito por três características básicas: (i) a entrada, (ii) o mecanismo de serviço e (iii) a disciplina da fila. A entrada descreve o processo que governa a demanda por serviço, ou em outras palavras, de que maneira os clientes chegam no sistema. Este processo de entrada pode ser caracterizado pelo tempo entre entradas sucessivas

de clientes. O mecanismo de serviço descreve o processo associado à necessidade de serviço de cada cliente, isto é, o tempo em que cada cliente passa recebendo serviço na estação (geralmente chamado de tempo de serviço). A disciplina da fila descreve como os clientes esperando na fila são servidos. Geralmente, quando se pensa em filas, imagina-se que o cliente que chegou primeiro será atendido primeiro. Este tipo de disciplina de fila é chamado de FIFO (*first in first out*). Contudo, existem outras formas de atender clientes esperando na fila, por exemplo, tem-se as seguintes disciplinas: LIFO (*last in first out*) o primeiro cliente que entrar na fila será o último a sair, que corresponde à um sistema de empilhamento; SJF (*shortest job first*), onde o cliente com a menor tarefa é atendido primeiro; e RR (*round robin*), onde cada cliente recebe frações de serviço e voltam para o final da fila até que o serviço completo termine. Existem vários outros exemplos além dos mencionados nesta tese. Sempre assumiremos que a disciplina de serviço é FIFO.

Para estabelecer uma nomenclatura simples para classificar diferentes modelos de filas, Kendall em (Kendall, 1953) introduziu a seguinte notação para cada tipo de distribuição de tempo de serviço e de tempo entre chegada de clientes:  $M$  denota distribuição exponencial;  $E_k$  denota a distribuição de Erlang com  $k$  estágios;  $G$  denota uma distribuição arbitrária. Quando os intervalos entre chegadas de clientes ou os tempos de serviço são independentes e identicamente distribuídos, escreve-se  $GI$ . A notação de Kendall é dividida em quatro partes  $\cdot / \cdot / \cdot / \cdot$ , a primeira denota a distribuição do intervalo entre chegada de clientes, o segundo denota a distribuição do tempo em serviço, o terceiro denota o número de estações do sistema, e o último denota a capacidade da fila (isto é, o número máximo de clientes que podem esperar na fila antes que ela transborde). Portanto,  $M/GI/c/k$  denota uma fila com intervalo entre clientes distribuídos por uma distribuição exponencial, tempo de serviço dado por qualquer distribuição mas independente e identicamente distribuídos,  $c$  estações de serviço, e uma capacidade de  $k$  clientes na fila. Quando a última parte desta nomenclatura é deixada em branco, assume-se que a fila possui uma capacidade infinita. Portanto,  $M/GI/c$  denota o mesmo

sistema descrito acima mas com espaço infinito para clientes esperarem na fila.

Se o modelo de interesse possuir o intervalo entre a chegada de clientes dado por variáveis aleatórias independentes e identicamente distribuídas, o processo que conta o número total de clientes que já entrou no sistema é um processo de renovação. Isto é, seja  $\{\Delta_i^a\}$ , para  $i \in \mathbb{N}$ , variáveis aleatórias representando o intervalo de tempo entre as chegadas de clientes, onde se assume que  $\mathbb{E}[\Delta_i^a] < \infty$ . Então, o número de clientes que entrou no sistema até o instante  $t$  é o processo de renovação  $N(t) \triangleq \max\{n \in \mathbb{N}_0 : \sum_{i=1}^n \Delta_i^a \leq t\}$  (onde a convenção  $\sum_{k=1}^0 = 0$  é usada). Um resultado bem conhecido diz que para  $N$  ser um processo de Markov,  $\Delta_i^a$  tem que ser exponencialmente distribuído. E, claramente, se  $\Delta_i^a$  for exponencialmente distribuído,  $N$  é um processo de Markov. Isso é devido à propriedade “sem memória” que a distribuição exponencial possui, que é a única distribuição contínua com esta propriedade. Isso explica também porque Kendall escolheu a letra  $M$  para denotar a distribuição exponencial.

Muitas vezes, quando um modelo estocástico está sob consideração, é possível relaxar as hipóteses para que alguns processos do modelo se tornem Markovianos, introduzindo assim, um conjunto de ferramentas matemáticas poderosas que o acompanha. Contudo, em teoria das filas, isto implica em ter que assumir que o intervalo de tempo entre a chegada de clientes é exponencialmente distribuído, que pode não ser realista para alguns sistemas. Esta troca entre a propriedade de Markov e modelos muito simplificados é uma constante fonte de problemas em teoria de filas, já que muitas vezes o problema se torna intratável quando esta propriedade não está presente. É claro que um processo não-Markoviano pode ser visto como um processo de Markov quando se aumenta o número de variáveis de estado. No caso tratado aqui, o processo  $(N, \zeta)$  é Markoviano, onde  $\zeta(t)$  denota o tempo desde a última chegada de cliente (isto é,  $\zeta(t) = t - \sum_{i=1}^{N(t)} \Delta_i^a$ ). Apesar disto, esta observação nem sempre facilita a análise destes sistemas e outras formas de lidar com este problema serão consideradas.

Na próxima seção, será introduzido o modelo clássico  $M/M/1$  e como Erlang

tratou o problema mencionado acima.

## 2.2 Teoria Clássica

Considere um modelo de filas onde o intervalo de tempo entre a chegada de clientes e o tempo de serviço são identicamente distribuídos e dados por distribuições exponenciais com parâmetros  $\lambda$  e  $\mu$ , respectivamente. Sejam  $N$  e  $S$  processos de contagem para o número de clientes que entraram e deixaram o sistema até o tempo  $t$ , respectivamente. Isto é, para cada  $t \in \mathbb{R}_{\geq 0}$ ,  $N(t)$  representa o número de clientes que entrou no sistema até o instante  $t$ , e  $S(t)$  representa o número de clientes servidos até o mesmo instante. Em teoria de filas, o processo  $X \triangleq N - S$  é de grande interesse já que ele diz quantos clientes estão presentes na fila em cada instante de tempo  $t$ . Observe que o processo  $N$  é de Poisson já que o tempo entre a chegada de clientes é suposto exponencialmente distribuído (veja, por exemplo, (Kleinrock, 1975) página 60). Além disso, o processo  $S$  pode ser visto como um processo de Poisson quando o sistema está ocupado, já que os tempos de serviço são exponencialmente distribuídos. Pode ser mostrado que  $X$  é um *processo de nascimento e morte* com taxas de transição infinitesimais dadas por:

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ \mu & -(\lambda + \mu) & \lambda & & \\ & \mu & -(\lambda + \mu) & \lambda & \\ & & \ddots & \ddots & \ddots \end{pmatrix},$$

já que da definição tem-se que, para  $\tau \in \mathbb{R}_{\geq 0}$ :

$$(1) \mathbb{P}(X(t + \tau) = k + 1 | X(t) = k) = \lambda\tau + o(\tau), \text{ para } k \geq 0$$

$$(2) \mathbb{P}(X(t + \tau) = k - 1 | X(t) = k) = \begin{cases} o(\tau) & \text{para } k = 0 \\ \mu\tau + o(\tau) & \text{para } k > 0 \end{cases}$$

$$(3) \mathbb{P}(X(t + \tau) = k | X(t) = k) = 1 - \lambda\tau - \mu\tau + o(\tau), \text{ para } k \geq 0.$$

Suponha que  $0 < \lambda, \mu < \infty$ , e seja o vetor linha  $\alpha$  dado por  $\alpha_k \triangleq \mathbb{P}(X_0 = k)$ .

Assuma ainda que o seguinte limite existe:

$$\lim_{t \rightarrow \infty} \alpha P(t) = \pi, \quad (2.1)$$

onde  $\pi$  é um vetor linha com componentes satisfazendo  $\sum_k \pi_k = 1$ , e  $P(t)$  uma matriz para cada  $t \in \mathbb{R}_{\geq 0}$  com elementos  $(P(t))_{ij} = \mathbb{P}(X(t) = j | X(0) = i)$ . O vetor linha  $\pi$  representa a distribuição do número de clientes no sistema quando a fila está em regime estacionário. Por enquanto, não haverá preocupação com as condições para que esta distribuição exista. Multiplicando a equação *forward* de Kolmogorov  $\frac{d}{dt}P(t) = P(t)Q$  por  $\alpha$  e deixando  $t \rightarrow \infty$ , tem-se que  $0 = \pi Q$ . Portanto, chega-se ao sistema de equações abaixo

$$\begin{cases} 0 = -(\lambda + \mu)\pi_k + \lambda\pi_{k-1} + \mu\pi_{k+1} \\ 0 = -\lambda\pi_0 + \mu\pi_1 \\ 1 = \sum_k \pi_k. \end{cases}$$

Em vista de  $\pi_1 = \left(\frac{\lambda}{\mu}\right)\pi_0$ , pela segunda equação do sistema acima, e resolvendo de forma recursiva, tem-se que

$$\pi_k = \left(\frac{\lambda}{\mu}\right)^k \pi_0 = \rho^k \pi_0,$$

onde  $\rho \triangleq \lambda/\mu$ . Usando a equação  $1 = \sum_k \pi_k$  tem-se que  $\sum_k \rho^k \pi_0 = 1$  e portanto  $\pi_0 = 1/\sum_k \rho^k$ . Assumindo que  $\rho < 1$ , tem-se que  $\pi_0 = 1 - \rho$ , e

$$\pi_k = (1 - \rho)\rho^k, \quad \text{para } k \geq 0. \quad (2.2)$$

Será visto mais à frente, na Seção 2.3, que a condição  $\rho < 1$  implica na existência do limite da Equação (2.1).

Com o resultado acima, pode-se computar algumas quantidades de interesse, como o *número médio de clientes na fila*, dado por  $\bar{N}$ , em regime estacionário:

$$\begin{aligned}\bar{N} &= \sum_k k\pi_k = (1-\rho) \sum_k k\rho^k = (1-\rho)\rho \frac{d}{d\rho} \left( \sum_k \rho^k \right) \\ &= (1-\rho)\rho \frac{d}{d\rho} \left( \frac{1}{1-\rho} \right) = \frac{\rho}{1-\rho}.\end{aligned}\quad (2.3)$$

Pode-se também computar o *tempo médio* que cada cliente passa no sistema,  $\bar{T}$ , usando a fórmula de Little (Little, 1961), que afirma que  $\bar{T} = \bar{N}/\lambda$ . Portanto, tem-se que

$$\bar{T} = \frac{\bar{N}}{\lambda} = \frac{\mu^{-1}}{1-\rho}.\quad (2.4)$$

Usando as Equações (2.3) e (2.4), observa-se que o sistema fica mais cheio (ou mais congestionado) quando  $\rho \uparrow 1$ . Por esta razão, o parâmetro  $\rho$  é chamado de *intensidade de tráfego*. Será visto nas seções seguintes que este parâmetro é muito importante em teoria de filas, até mesmo para modelos mais gerais.

Um modelo mais interessante pode ser construído quando se supõe que  $X$  é um processo com taxas de transições dependentes do estado, isto é, um processo de nascimento e morte com taxas infinitesimais de transição dadas por:

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & & & \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & & \\ & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Seguindo o roteiro anterior, tem-se que

$$\pi_k = \prod_{i=0}^{k-1} \left( \frac{\lambda_i}{\mu_{i+1}} \right) \pi_0 \quad \text{e} \quad \pi_0 = \frac{1}{1 + \sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}}.$$

Com este modelo, pode-se considerar, por exemplo, filas com clientes desencorajados, que considera o caso onde clientes estão menos aptos a esperar na fila se ela

estiver muito longa. Isto pode ser feito considerando o seguinte:

$$\lambda_k \triangleq \frac{\theta}{k+1} \quad \text{e} \quad \mu_k \triangleq \mu,$$

para  $0 < \theta, \mu < \infty$ . Portanto, tem-se que

$$\begin{aligned} \pi_k &= \pi_0 \prod_{i=0}^{k-1} \frac{\theta}{(i+1)\mu} \Rightarrow \pi_k = \pi_0 \left(\frac{\theta}{\mu}\right)^k \frac{1}{k!} \\ \pi_0 &= \left(1 + \sum_{k=1}^{\infty} \left(\frac{\theta}{\mu}\right)^k \frac{1}{k!}\right)^{-1} \Rightarrow \pi_0 = e^{-\theta/\mu}. \end{aligned}$$

Outros modelos podem ser considerados apenas alterando os valores dados à  $\mu_k$  e  $\lambda_k$ . Por exemplo, pode-se considerar o modelo  $M/M/\infty$  fazendo com que  $\lambda_k \triangleq \lambda$  e  $\mu_k \triangleq k\mu$ ; o modelo  $M/M/m$  definindo  $\lambda_k \triangleq \lambda$  e  $\mu_k \triangleq \min(k\mu, m\mu)$ ; e o modelo  $M/M/1/C$  definindo

$$\lambda_k \triangleq \begin{cases} \lambda & k < C \\ 0 & k \geq C \end{cases} \quad \text{e} \quad \mu_k \triangleq \mu.$$

Para maior detalhes veja, por exemplo, o livro (Kleinrock, 1975).

Por volta de 1917, A. K. Erlang observou que distribuições exponenciais nem sempre representavam de forma adequada vários problemas de filas que desejava-se modelar. Contudo, tratar casos mais gerais implicava em perder a propriedade de Markov do processo  $X$ , e o modelo se tornava muito difícil para ser tratado. Portanto, Erlang criou um método para generalizar as distribuições exponenciais de tal forma que ainda poderia ser tratável (Erlang, 1917). Esta generalização foi obtida através da seguinte estratégia: considere uma estação de serviço formada por dois estágios estatisticamente independentes. Cada cliente que entra em serviço permanece em cada estágio um tempo exponencialmente distribuído com taxa  $\mu$ . Nenhum outro cliente é permitido entrar em um estágio caso um outro cliente já esteja sendo servido em qualquer outro estágio. Portanto, a distribuição do tempo de serviço é dada pela convolução de duas distribuições exponenciais com

parâmetro  $\mu$ . De forma mais geral, pode-se considerar uma estação de serviço com  $k$  estágios com tempo de permanência exponencialmente distribuídos com taxa  $\mu$ . A função de densidade de probabilidade desta distribuição é dada por

$$b(\xi) = \frac{\mu^k}{(k-1)!} \xi^{k-1} e^{-\mu\xi}, \quad \text{para } \xi \geq 0.$$

A análise do modelo  $M/E_k/1$  é feita usando o processo  $(X, J)$  com espaço de estados  $\mathbb{E} = \mathbb{N}_0 \times \{1, \dots, k\}$  que é Markoviano, onde  $X(t)$  representa o número de clientes na fila no instante  $t$  e  $J(t)$  representa o estágio em que o cliente em serviço está presente no instante  $t$ . Este tipo de fila será considerado com mais detalhes na Seção 2.4. Pode-se também tratar este modelo usando o processo estocástico dado por  $\tilde{X}$ , que representa o número de estágios a serem percorridos por todos os clientes no sistema. Para maior detalhes sobre este modelo, veja, por exemplo, o livro (Kleinrock, 1975).

### 2.3 O Modelo $M/GI/1$ para Filas

Nesta seção será estudado o modelo de filas  $M/GI/1$  onde clientes são servidos em ordem de chegada. O sistema é descrito pelas variáveis aleatórias  $\{\Delta_i^a\}$  e  $\{\Delta_i^d\}$  onde  $\Delta_i^\alpha$  são mutuamente independentes, para  $\alpha \in \{a, d\}$  e  $i \in \mathbb{N}$ . A variável  $\Delta_i^a$  representa o tempo entre a chegada de clientes, e a variável  $\Delta_i^d$  representa o tempo em serviço. Para este modelo, um processo de Poisson alimenta o sistema, portanto  $\Delta_i^a$  são exponencialmente distribuídos. A distribuição do tempo de serviço é dada por  $B(t) \triangleq \mathbb{P}(\Delta_i^d \leq t)$ , e nenhuma hipótese mais forte é colocada sobre ela. Além disso, seja  $\bar{\Delta}^\alpha \triangleq \mathbb{E}[\Delta_i^\alpha] > 0$ , para  $\alpha \in \{a, d\}$ .

O método mais usado (e de maior sucesso) para a análise deste tipo de modelo começa pela identificação de uma *cadeia de Markov embutida* dentro do processo estocástico que descreve o número de clientes no sistema. Kendall, em (Kendall, 1953), faz uma excelente introdução à identificação de cadeias de Markov embutidas para modelos de filas. A definição abaixo segue de uma simplificação em termos de tempos de parada da definição dada em (Kendall, 1953).

**Definição 2.3.1** (Cadeia de Markov Embutida). *Seja  $X$  um processo estocástico definido em um espaço de probabilidade filtrado  $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$  tomando valores em algum espaço de estados  $\mathcal{X}$ . Seja  $\mathcal{T} \triangleq \{t_k\}$ ,  $k \in \mathbb{N}$ , uma sequência de tempos de parada  $\mathcal{F}_t$ -mensuráveis tais que  $t_{k-1} < t_k < t_{k+1}$ , para todo  $k \in \mathbb{N}$ , quase certamente. Defina o processo  $\{Y_k\}$ ,  $k \in \mathbb{N}$ , como  $Y_k \triangleq X(t_k)$ . Se  $Y$  é um processo de Markov, isto é, se satisfaz:*

$$\mathbb{P}(Y_{k+1}|Y_i, 0 \leq i \leq k) = \mathbb{P}(Y_{k+1}|Y_k) \quad \forall k \in \mathbb{N},$$

então diz-se que  $Y$  é uma cadeia de Markov embutida de  $X$ .

Para construir uma cadeia de Markov embutida para o modelo considerado, defina o processo  $X \triangleq N - S$ , onde  $N$  conta o número de clientes que entram no sistema e  $S$  o número de clientes que deixam o sistema. Defina  $t_k$  como o  $k$ -ésimo instante em que um cliente deixa o sistema, isto é:

$$\begin{aligned} t_1 &\triangleq \inf\{s \geq 0 : S(s) - S(s-) = 1\} \\ t_k &\triangleq \inf\{s > t_{k-1} : S(s) - S(s-) = 1\}, \text{ para } k > 1. \end{aligned}$$

Para provar que  $Y_k \triangleq X(t_k)$  é uma cadeia de Markov, escreve-se  $Y_k$  como:

$$\begin{aligned} Y_{k+1} &= X_{t_k} + (X_{t_{k+1}} - X_{t_k}) \\ &= Y_k + (N(t_{k+1}) - N(t_k)) - (S(t_{k+1}) - S(t_k)) \\ &= Y_k + (N(t_{k+1}) - N(t_k)) - 1, \end{aligned} \tag{2.5}$$

para  $t_k, t_{k+1} \in \mathcal{T}$ . Observe que  $\{N(t_{k+1}) - N(t_k)\}$ , para  $k \in \mathbb{N}$ , são mutuamente independentes e identicamente distribuídos para qualquer sequência  $\{t_k\} \subset \mathbb{R}_{\geq 0}$  tal que  $t_{k-1} < t_k < t_{k+1}$ . Portanto,  $Y$  é uma cadeia de Markov homogênea no tempo.

Para computar as probabilidades de transição  $p_{ij} \triangleq \mathbb{P}(Y_{k+1} = j | Y_k = i)$ ,

considere o seguinte:

$$\begin{aligned} p_{ij} &= \mathbb{P}(Y_k + (N(t_{k+1}) - N(t_k)) - 1 = j \mid Y_k = i) \\ &= \mathbb{P}(N(t_{k+1} - t_k) = j - i + 1). \end{aligned}$$

Para  $i \geq 1$ , sabe-se que pelo menos um cliente é deixado na fila quando o  $k$ -ésimo cliente deixa o sistema. Além disso, este cliente na fila inicia seu serviço imediatamente após a partida do cliente  $(k+1)$ . Portanto, tem-se que  $(t_{k+1} - t_k) = \Delta_{k+1}^d$ , ou seja,  $(t_{k+1} - t_k)$  é o tempo em serviço do cliente  $(k+1)$ . Portanto,

$$\begin{aligned} p_{ij} &= \int_0^\infty \mathbb{P}(N(t) = j - i + 1 \mid \Delta_{k+1}^d = t) dB(t) \\ &= \frac{1}{(j - i + 1)!} \int_0^\infty e^{-t/\bar{\Delta}^a} \left(\frac{t}{\bar{\Delta}^a}\right)^{j-i+1} dB(t), \quad \text{para } i \geq 1. \end{aligned}$$

Caso  $i = 0$ , tem-se que no instante que o  $k$ -ésimo cliente deixou o sistema, a fila estava vazia. Portanto, o sistema permaneceu ocioso até a chegada do próximo cliente. Seja  $s_k$  o instante em que o cliente  $(k+1)$  chega no sistema. Observe que:

$$\begin{aligned} p_{ij} &= \mathbb{P}(N(t_{k+1}) - N(t_k) = j + 1) \\ &= \mathbb{P}(N(t_{k+1}) - N(s_{k+1}) = j \mid N(s_{k+1}) - N(t_k) = 1) \\ &= \mathbb{P}(N(t_{k+1}) - N(s_{k+1}) = j) = \mathbb{P}(N(t_{k+1} - s_{k+1}) = j). \end{aligned}$$

Observando que  $(t_{k+1} - s_{k+1})$  é o tempo de serviço para o cliente  $(k+1)$ , tem-se que

$$p_{0j} = \frac{1}{j!} \int_0^\infty e^{-t/\bar{\Delta}^a} \left(\frac{t}{\bar{\Delta}^a}\right)^j dB(t). \quad (2.6)$$

Desta forma, a matriz de transição  $P = \{p_{ij}\}$  de  $Y$  é dada por

$$P = \begin{pmatrix} k_0 & k_1 & k_2 & \cdots \\ k_0 & k_1 & k_2 & \cdots \\ 0 & k_0 & k_1 & \cdots \\ 0 & 0 & k_0 & \cdots \\ \vdots & \vdots & & \ddots \end{pmatrix}$$

onde  $k_j \triangleq p_{0j}$ . Usando esta matriz de transição, observa-se que a cadeia é irredutível, isto é, qualquer estado da cadeia pode ser alcançado a partir de qualquer outro estado dado tempo suficiente. E além disso, a cadeia é aperiódica já que  $p_{ii} > 0$  para todo  $i$ .

Neste ponto, o comportamento ergódico da cadeia de Markov  $Y$  será estudada. Para isso, será usado uma série de resultados e definições provenientes da teoria de cadeias de Markov em espaços de estados enumeráveis.

**Definição 2.3.2** (Classificação de estados para cadeias de Markov). *Seja  $X$  uma cadeia de Markov a tempo discreto (homogênea) tomando valores em um espaço de estado enumerável dado por  $\mathbb{S}$ . Defina  $\rho_i \triangleq \inf\{k > 0 : X_k = i\}$  como o primeiro instante em que a cadeia atinge o estado  $i$ . Defina-se a seguinte nomenclatura para os estados  $i \in \mathbb{S}$ :*

- (1) Diz-se que o estado  $i$  é positivo recorrente se  $\mathbb{P}(\rho_i < \infty | X_0 = i) = 1$ , e  $\mathbb{E}[\rho_i | X_0 = i] < \infty$ .
- (2) Diz-se que o estado  $i$  é nulo recorrente se  $\mathbb{P}(\rho_i < \infty | X_0 = i) = 1$ , e  $\mathbb{E}[\rho_i | X_0 = i] = \infty$ .
- (3) Diz-se que o estado  $i$  é transiente quando  $\mathbb{P}(\rho_i = \infty | X_0 = i) > 0$ , e consequentemente,  $\mathbb{E}[\rho_i | X_0 = i] = \infty$ .

A interpretação da definição acima é a seguinte: um estado  $i$  é recorrente quando a cadeia retorna para o estado  $i$  em tempo finito quase certamente. O

estado é transiente quando isso não acontece. Um estado  $i$  recorrente é chamado de positivo quando o tempo médio de retorno da cadeia para o estado  $i$  é finito. Caso contrário, o estado é chamado de nulo.

Os seguintes teoremas serão usados. A prova dos seguintes teoremas podem ser encontradas em (Stroock, 2005) no capítulo 3. O critério para recorrência pode também ser encontrado em (Foster, 1953).

**Teorema 2.3.3.** *Para uma cadeia de Markov irredutível, os estados  $i \in \mathbb{S}$  são todos recorrentes positivos, recorrentes nulos, ou transientes.*

Portanto, se qualquer estado possui uma das propriedades acima, diz-se que a cadeia é recorrente positiva, recorrente nula, ou transiente, respectivamente.

**Teorema 2.3.4** (Critério para recorrência (Foster)). *Considere uma cadeia de Markov irredutível tomando valores no espaço de estados  $\mathbb{S}$  e com matriz de transição  $P = \{p_{ij}\}$ . Se para alguma função  $u : \mathbb{S} \rightarrow [0, \infty)$ , tal que  $\{k : u(k) \leq L\}$  é finito para cada  $L \in (0, \infty)$ , e algum  $j_0 \in \mathbb{S}$  tem-se que*

$$\sum_{j \in \mathbb{S}} p_{ij} u(j) \leq u(i) \quad \forall i \neq j_0,$$

então a cadeia é recorrente em  $\mathbb{S}$ .

Para a cadeia  $Y$ , a condição do teorema acima pode ser escrita como

$$\sum_{j=0}^{\infty} k_j u(j+i-1) \leq u(i)$$

para  $i \geq 1$ . Escolha a função  $u$  como a seguinte:  $u(i) = i$ . Então para todo  $i \geq 1$ , tem-se que

$$\sum_{j=0}^{\infty} k_j (j+i-1) \leq i \Rightarrow \sum_{j=0}^{\infty} j k_j + i - 1 \leq i \Rightarrow \sum_{j=0}^{\infty} j k_j \leq 1,$$

já que  $\sum_{j=0}^{\infty} k_j = 1$ . Escolhendo  $j_0$  do Teorema 2.3.4 como sendo o estado 0, tem-se a seguinte condição para  $i = 0$ :  $\sum_{j=0}^{\infty} j k_j > 0$  (que é satisfeita trivialmente). Seja

$K(z) = \sum_{j=0}^{\infty} z^j k_j$ , para  $|z| \leq 1$ , e observe que

$$\begin{aligned} K(z) &= \sum_{j=0}^{\infty} \frac{1}{j!} \int_0^{\infty} e^{-t/\bar{\Delta}^a} \left( \frac{tz}{\bar{\Delta}^a} \right)^j dB(t) = \int_0^{\infty} e^{-t/\bar{\Delta}^a} \left[ \sum_{j=0}^{\infty} \frac{1}{j!} \left( \frac{tz}{\bar{\Delta}^a} \right)^j \right] dB(t) \\ &= \int_0^{\infty} e^{-(t/\bar{\Delta}^a)(1-z)} dB(t) = B^* \left( \frac{1-z}{\bar{\Delta}^a} \right) \end{aligned}$$

onde  $B^*(s)$  é a transformada de Laplace da função de distribuição cumulativa  $B(\cdot)$  do tempo de serviço. Portanto,

$$\sum_{j=0}^{\infty} j k_j = K'(z)|_{z=1} = - \left( \frac{1}{\bar{\Delta}^a} \right) [B^*]'(1-z)|_{z=1} = \frac{\bar{\Delta}^d}{\bar{\Delta}^a} = \rho.$$

Desta forma, a condição para que a cadeia seja recorrente é  $0 < \rho \leq 1$ .

O seguinte teorema será usado para estabelecer sob quais condições a cadeia  $Y$  é recorrente positiva ou recorrente nula. A prova do teorema abaixo pode ser encontrada em (Stroock, 2005) no capítulo 3.

**Teorema 2.3.5.** *Para uma cadeia de Markov irredutível e aperiódica definida sobre o espaço de estados  $\mathbb{S}$  e com matriz de transição dada por  $P = \{p_{ij}\}$ , tem-se que*

$$\lim_{n \rightarrow \infty} p_{ij}^n \stackrel{\Delta}{=} \pi_{ij} = \frac{1}{\mathbb{E}[\rho_j | X_0 = j]}$$

e  $\sum_{j \in \mathbb{S}} \pi_{ij} \leq 1$  para todo  $i \in \mathbb{S}$ , onde  $P^n = \{p_{ij}^n\}$  é a matriz  $P$  elevada à  $n$ . (É importante destacar que o lado direito da equação acima não depende de  $i$ ).

Observe que, pela Definição 2.3.2, um estado  $j$  é recorrente positivo se e somente se  $\pi_{ij} > 0$ , para qualquer  $i \in \mathbb{S}$ . Além disso, o Teorema 2.3.3 garante que se  $\pi_{ij} > 0$ , para algum  $(i, j) \in \mathbb{S}^2$ , então  $\pi_{ik} > 0 \forall k \in \mathbb{S}$ .

Para estabelecer se  $Y$  é recorrente positiva, será usado o desenvolvimento apresentado por (Kendall, 1951). Considere o seguinte:

$$\pi_{ij} = \lim_{n \rightarrow \infty} p_{ij}^{(n+1)} = \sum_{a \in \mathbb{S}} p_{aj} \lim_{n \rightarrow \infty} p_{ia}^n = \sum_{a=0}^{\infty} \pi_{ia} p_{aj}.$$

Multiplicando ambos os lados por  $z^j$ , para  $|z| \leq 1$ , e somando em  $j$ , tem-se

$$\sum_{a=0}^{\infty} \pi_{ia} \sum_{j=0}^{\infty} z^j p_{aj} = \sum_{j=0}^{\infty} \pi_{ij} z^j.$$

Defina  $\Pi(z) = \sum_{j=0}^{\infty} \pi_{ij} z^j$  (observe que, pelo Teorema 2.3.5,  $\pi_{ij}$  não depende de  $i$ ) e  $K(z) = \sum_{j=0}^{\infty} z^j k_j$ . Portanto,

$$\begin{aligned} K(z)\pi_{i0} + \sum_{a=1}^{\infty} \pi_{ia} z^{a-1} K(z) &= \Pi(z) \\ K(z)\pi_{i0} + \frac{K(z)}{z} (\Pi(z) - \pi_{i0}) &= \Pi(z) \\ K(z)\pi_{i0} &= \left( \frac{K(z) - z}{1 - z} \right) \Pi(z) \end{aligned}$$

quando  $z \uparrow 1$  tem-se que  $K(z) \rightarrow 1$ ,  $\Pi(z) \rightarrow \Pi(1) = \sum_{j \in \mathbb{S}} \pi_{ij} \leq 1$  e

$$\lim_{z \uparrow 1} \frac{K(z) - z}{1 - z} = \lim_{z \uparrow 1} (1 - K'(z)) = 1 - \rho.$$

Desta forma, tem-se  $\pi_{i0} = \Pi(1)(1 - \rho)$ , onde  $\pi_{i0}$  e  $\Pi(1)$  são não-negativas e finitas. Portanto, se  $\rho = 1$ ,  $\pi_{i0}$  tem que ser zero, levando à conclusão que quando  $\rho = 1$  a cadeia é recorrente nula. Os resultados ergódicos estabelecidos nesta seção são resumidos no seguinte teorema:

**Teorema 2.3.6** (Comportamento ergódico da cadeia de Markov embutida). *A cadeia de Markov embutida  $Y$  para a fila  $M/GI/1$  tem as seguintes propriedades:*

- (1) *Para  $0 < \rho < 1$ , a cadeia é recorrente positiva.*
- (2) *Para  $\rho = 1$ , a cadeia é recorrente nula.*
- (3) *Para  $\rho > 1$ , a cadeia é transiente.*

Interpretando estes resultado em relação ao modelo de filas, sabe-se que a fila será instável quando  $\rho \geq 1$ , no sentido que o número de clientes no sistema vai crescer descontroladamente. Para  $\rho < 1$ , a fila é estável, ou em outras palavras, o tamanho da fila será finito quase certamente em todo instante.

Neste ponto, seria interessante obter uma expressão da distribuição estacionária de  $Y$  quando  $\rho < 1$ , como foi feito anteriormente para o modelo  $M/M/1$ . Infelizmente, derivar tal expressão diretamente se torna um problema muito complicado, e tudo que pode-se fazer é derivar uma expressão para a função geradora de probabilidade  $\Pi(z)$ . Para isso, suponha que  $r_k = N(\Delta_k^d)$ , isto é, a variável  $r_k$  denota o número de clientes que chegam durante o serviço do  $k$ -ésimo cliente. Observe que a distribuição de  $r_k$  já foi computada anteriormente e é dada por  $\mathbb{P}(r_k = i) = p_{0i}$ , dado pela Equação (2.6). Relembrando a derivação anterior, tinha-se que quando  $Y_k \neq 0$ ,  $N(t_{k+1}) - N(t_k) = N(\Delta_k^d) = r_{k+1}$ , para  $t_k, t_{k+1} \in \mathcal{T}$ . Além disso, para  $Y_k = 0$ , tinha-se que  $N(t_{k+1}) - N(t_k) = N(\Delta_k^d) + 1 = r_{k+1} + 1$ . Portanto, pode-se escrever o seguinte:

$$Y_{k+1} = Y_k + r_{k+1} - \mathbb{I}\{Y_k \neq 0\} \quad (2.7)$$

onde  $\mathbb{I}\{\cdot\}$  é a função indicadora. Portanto,

$$\mathbb{E} [z^{Y_{k+1}}] = \mathbb{E} [z^{Y_k + r_{k+1} - \mathbb{I}\{Y_k \neq 0\}}] = \mathbb{E} [z^{Y_k - \mathbb{I}\{Y_k \neq 0\}}] \mathbb{E} [z^{r_{k+1}}], \quad (2.8)$$

que segue do fato de  $r_{k+1}$  ser independente do número de clientes que já estão presentes no sistema. Sabe-se que  $\mathbb{E} [z^{r_{k+1}}] = K(z) = B^* \left( \frac{1-z}{\Delta^a} \right)$ , portanto, considere apenas o termo

$$\begin{aligned} \mathbb{E} [z^{Y_k - \mathbb{I}\{Y_k \neq 0\}}] &= \mathbb{P}(Y_k = 0) z^0 + \sum_{i=1}^{\infty} z^{i-1} \mathbb{P}(Y_k = i) \\ &= \mathbb{P}(Y_k = 0) + \frac{1}{z} \sum_{i=0}^{\infty} z^i \mathbb{P}(Y_k = i) - \frac{1}{z} \mathbb{P}(Y_k = 0) \\ &= \frac{\mathbb{E} [z^{Y_k}] - (1-z) \mathbb{P}(Y_k = 0)}{z}. \end{aligned}$$

Desta forma, depois de usar o resultado acima em combinação com o resultado da

Equação (2.8) e deixando  $k \rightarrow \infty$ , tem-se que

$$\Pi(z) = B^* \left( \frac{1-z}{\Delta^a} \right) \left( \frac{\Pi(z) - (1-z)P_0}{z} \right).$$

onde  $P_0$  denota a distribuição de probabilidade para a qual  $\mathbb{P}(Y_k = 0)$  converge. Pelo teorema de representação de Skorohod (veja por exemplo Teorema 6.7 de (Billingsley, 1999)), pode-se supor a existência do processo  $Y$  para qual  $Y_k$  converge em probabilidade e  $\mathbb{P}(Y = 0) = P_0$ . Observe que, tomando a esperança de ambos os lados da Equação (2.7), tem-se que  $E[Y_{k+1}] = E[Y_k] + E[r_{k+1}] - \mathbb{P}(Y_k \neq 0)$ , que leva à

$$\mathbb{P}(Y = 0) = 1 - \rho \tag{2.9}$$

depois que o limite é tomado e lembrando que  $\mathbb{E}[r_{k+1}] = \rho, \forall k \in \mathbb{N}$ . Finalmente, chega-se à

$$\Pi(z) = \frac{B^* \left( \frac{1-z}{\Delta^a} \right) (1-\rho)(1-z)}{B^* \left( \frac{1-z}{\Delta^a} \right) - z}, \quad \text{para } |z| \leq 1.$$

Portanto, tem-se a expressão da função geradora de probabilidade do número de clientes presentes na fila em regime estacionário nos instantes em que clientes deixam o sistema. É possível mostrar que esta é também a função geradora de probabilidade da fila estacionária para qualquer instante.

O método da cadeia de Markov embutida é um método muito bem sucedido para a análise de filas em que exista alguma estrutura Markoviana presente. Um exemplo imediato é o modelo de fila  $GI/M/1$ , que pode ser tratada usando o método da cadeia de Markov embutida. Neste caso, os tempos de parada para definir a cadeia são os instantes em que os clientes entram no sistema. Para maiores detalhes referente a este modelo veja (Kendall, 1953).

## 2.4 Métodos Analíticos Matriciais (*Matrix Analytic Methods*)

Nesta seção será feita uma breve introdução aos chamados métodos analíticos matriciais. Estes métodos generalizam o método de estágios proposto por Erlang no qual troca-se o processo de chegada (ou de serviço) por outro mais geral que ainda preserve uma estrutura Markoviana, fazendo assim com que o modelo ainda seja tratável. A primeira destas generalizações foi proposta por Neuts em (Neuts, 1975) e foi chamada de processo de renovação do tipo fase (*phase-type renewal process*). A definição deste processo começa na definição das distribuições do tipo fase (*phase-type distributions*). Seja  $X$  um processo de Markov sob um espaço de estado finito  $E = \{1, \dots, m + 1\}$  e matriz geradora dada por

$$Q = \begin{pmatrix} T & \eta \\ \mathbf{0}' & 0 \end{pmatrix},$$

onde  $T \in \mathbb{R}^{m \times m}$  é invertível,  $\mathbf{0}' = (0, \dots, 0) \in \mathbb{R}^m$  e  $\eta \in \mathbb{R}^m$ . Como as linhas de uma  $Q$ -matriz somam zero, tem-se que  $\eta_i = -\sum_{j=1}^m T_{ij}$ , para todo  $i \in \{1, \dots, m\}$ . Seja  $\tilde{\alpha} = (\alpha, \alpha_{m+1})$ , onde  $\alpha' \in \mathbb{R}^m$ , é a distribuição inicial do processo  $X$ . Observe que o estado  $m + 1$  é absorvente. Portanto, quando o processo atingir este estado, ele permanecerá nele indefinidamente. Como  $T$  é invertível, pode-se mostrar que o processo irá atingir o estado absorvente com probabilidade um. Seja  $Z = \inf\{s \in \mathbb{R}_{\geq 0} : X_s = m + 1\}$ . A distribuição de  $Z$  é chamada de *distribuição do tipo fase* com parâmetros  $(\alpha, T)$ , e escreve-se o seguinte:  $Z \sim PH(\alpha, T)$ .

**Teorema 2.4.1.** *Seja  $Z \sim PH(\alpha, T)$ , então*

$$\mathbb{P}(Z \leq t) = 1 - \alpha \exp(Tt)\mathbf{e}$$

para todo  $t \geq 0$ , onde  $\mathbf{e} = (1, \dots, 1)' \in \mathbb{R}^m$ .

*Demonstração.* Como o processo  $X$  tem matriz de probabilidade de transição dada

por  $P(t) = \exp(Qt)$  para cada  $t \in \mathbb{R}_{\geq 0}$ , tem-se que

$$\mathbb{P}(Z \leq t) = \mathbb{P}(X_t = m + 1) = \sum_{i=1}^{m+1} \mathbb{P}(X_0 = i) \exp(Qt)_{i,m+1} = (\tilde{\alpha} \exp(Qt))_{m+1}.$$

Além disso, como  $\exp(Qt) \triangleq \sum_{m=1}^{\infty} \frac{t^m}{m!} Q^m$ , pode-se escrever

$$\exp(Qt) = \begin{pmatrix} \exp(Tt) & \gamma \\ \mathbf{0}' & 1 \end{pmatrix}$$

onde  $\gamma$  é um vetor em  $\mathbb{R}^m$ . As linhas de  $\exp(Tt)$  somam um, portanto,  $\gamma_i = 1 - \sum_{j=1}^m \exp(Tt)_{ij}$ , ou em notação matricial,  $\gamma = \mathbf{e} - \exp(Tt)\mathbf{e}$ . Logo,

$$\begin{aligned} \mathbb{P}(Z \leq t) &= (\alpha \exp(Tt), \alpha\gamma + \alpha_{m+1})_{m+1} \\ &= \alpha(\mathbf{e} - \exp(Tt)\mathbf{e}) + \alpha_{m+1} = 1 - \alpha \exp(Tt)\mathbf{e}, \end{aligned}$$

que prova o teorema. □

Um *processo de renovação do tipo fase N* é simplesmente um processo de renovação onde o tempo entre os intervalos de renovação são distribuídos de acordo com uma distribuição tipo fase. Isto é,  $N(t) \triangleq \max\{n \in \mathbb{N}_0 : \sum_{i=1}^n Z_i \leq t\}$ , onde  $\{Z_i, i \in \mathbb{N}\}$  são variáveis aleatórias i.i.d. (mutualmente independentes e identicamente distribuídas) com distribuição  $PH(\alpha, T)$ .

Defina  $J$  como um processo a tempo contínuo que marca o estado em que se encontra o processo de Markov associado ao intervalo de renovação com distribuição do tipo fase. Pode-se mostrar que o processo  $X \triangleq (N, J)$ , onde  $N$  é um processo de renovação do tipo fase, é um processo de Markov sob o espaço de estados  $\mathbb{E} = \mathbb{N}_0 \times \{1, \dots, m\}$ . O processo  $J$  também é Markoviano com matriz de transições infinitesimais dada por  $Q^* = T + A$ , em que  $A = \eta\alpha \in \mathbb{R}^{m \times m}$ . Ordenando o espaço de estados lexicograficamente, pode-se escrever a matriz de transição infinitesimal

de  $X$  como

$$(G_{(x,i),(y,j)}) = (\tilde{G}_{xm+i, ym+j}) = \begin{pmatrix} T & A & 0 & \dots \\ 0 & T & A & \dots \\ 0 & 0 & T & \ddots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

O processo acima pode ser entendido de forma mais intuitiva da seguinte forma: existem dois tipos de transições, aquelas de estados  $(n, i)$  para estados  $(n, j)$ , e aquelas de estados  $(n, i)$  para estados  $(n+1, j)$ , para  $n \in \mathbb{N}_0$ , e  $(i, j) \in \{1, \dots, m\}^2$ . No primeiro caso as taxas de transição são dadas pela matriz  $T$ . O segundo caso corresponde ao fim de um período de renovação onde um novo período é iniciado no estado  $j$  com distribuição  $\alpha_j$  e tempo de permanência dado por  $(-T\mathbf{e})_j$ .

Observe que o processo de renovação tipo fase é igual ao *processo de Poisson* com taxa  $\lambda$  quando se escolhe  $m = 1$ ,  $T = -\lambda$ , e  $\alpha = 1$ . Além disso, este processo é igual a um processo de renovação com tempo de renovação dado pela *distribuição de Erlang* com  $k$  estágios quando escolhe-se

$$T = \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & \ddots & \ddots & \\ & & & -\lambda & \lambda \\ & & & & -\lambda \end{pmatrix} \quad \alpha = (1, 0, \dots, 0),$$

onde  $T \in \mathbb{R}^{k \times k}$  e  $\alpha' \in \mathbb{R}^k$ . Existem outras distribuições que estão contidas nesta classe do tipo fase. Alguns exemplos incluem a *distribuição hiperexponencial* e as *distribuições de Cox*.

Talvez o resultado de maior importância sobre distribuições do tipo fase é o dado pelo seguinte teorema.

**Teorema 2.4.2.** *A classe das distribuições tipo fase é densa (no sentido de convergência fraca) no espaço de todas as distribuições sobre  $\mathbb{R}_{\geq 0}$ .*

*Demonstração.* Esta prova é um detalhamento do curto roteiro apresentado em (Asmussen, 2000). Seja  $F$  uma função de distribuição cumulativa sobre  $\mathbb{R}_{\geq 0}$ . Como  $F$  é monotonicamente crescente, limitada e contínua à direita, sabe-se que, para cada escolha de  $\epsilon$ , existe uma função escada  $S^N$  com um número  $N$  finito de saltos nos instantes  $\{t_n, n \in \{1, \dots, N\}\}$ , onde  $0 \leq \dots < t_n < t_{n+1} < \dots < t_N$ , tal que  $\sup_t |F(t) - S^N(t)| < \epsilon$ . Sem perda de generalidade, suponha que  $S^N(t) \triangleq \sum_{i=1}^N (\tilde{F}(i) - \tilde{F}(i-1)) \mathbb{I}\{t \geq t_i\}$ , onde  $\tilde{F}(0) \triangleq 0$ ,  $\tilde{F}(N) \triangleq 1$ , e  $\tilde{F}(i) \triangleq F(t_i)$ , para  $0 < i < N$ .

Por outro lado, seja  $\nu_k^\lambda$  uma variável aleatória com distribuição de Erlang com  $k$  estágios e parâmetro  $\lambda k$ . Quando  $k \rightarrow \infty$ , a distribuição de  $\nu_k^\lambda$  converge (no sentido fraco) para a função degrau  $H : \mathbb{R}_{\geq 0} \rightarrow [0, 1]$ , dada por  $H(t) \triangleq \mathbb{I}\{t \geq 1/\lambda\}$ . De fato, com a aplicação da da inequação de Chebyshev (veja por exemplo Teorema 2.4.9 na página 88 de (Ash e Doléans-Dade, 2000)) nas variáveis aleatórias  $f(\nu_k^\lambda) = \nu_k^\lambda - 1/\lambda$  e  $f(\nu_k^\lambda) = 1/\lambda - \nu_k^\lambda$  tem-se as seguintes inequações:

$$\mathbb{P}\left(\nu_k^\lambda \geq \frac{1}{\lambda} + \delta\right) \leq \frac{1}{\delta^2 \lambda^2 k} \quad \text{and} \quad \mathbb{P}\left(\nu_k^\lambda \leq \frac{1}{\lambda} - \delta\right) \leq \frac{1}{\delta^2 \lambda^2 k}.$$

Seja então as variáveis  $\{\nu_k^{1/t_i}, i \in \{1, \dots, N\}\}$ . Defina a variável aleatória  $\zeta$ , independente de  $\{\nu_k^{1/t_i}, i \in \{1, \dots, N\}\}$ , tomando valores em  $\{1, \dots, N\}$  com probabilidade  $\mathbb{P}(\zeta = i) = (\tilde{F}(i) - \tilde{F}(i-1))$ . Então, tem-se que

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^N \mathbb{I}\{\zeta = i\} \nu_k^{1/t_i} \leq t\right) &= \sum_{j=1}^N \mathbb{P}\left(\sum_{i=1}^N \mathbb{I}\{\zeta = i\} \nu_k^{1/t_i} \leq t \mid \zeta = j\right) \mathbb{P}(\zeta = j) \\ &= \sum_{j=1}^N \mathbb{P}\left(\nu_k^{1/t_j} \leq t\right) \mathbb{P}(\zeta = j) = \sum_{j=1}^N \mathbb{P}\left(\nu_k^{1/t_j} \leq t\right) (\tilde{F}(j) - \tilde{F}(j-1)) \end{aligned}$$

Como a classe das distribuições tipo fase é fechada sob misturas finitas (veja por exemplo Teorema 2.2.4 na página 53 de (Neuts, 1981)), sabe-se que a variável  $\beta_k^N \triangleq \sum_{i=1}^N \mathbb{I}\{\zeta = i\} \nu_k^{1/t_i}$  tem distribuição tipo fase. Além disso, quando tomamos  $k \rightarrow \infty$ , a distribuição de  $\beta_k^N$  converge fracamente para a função degrau  $S^N$ , que por sua vez satisfaz  $\sup_t |F(t) - S^N(t)| < \epsilon$ .  $\square$

Portanto, quando este modelo é aplicado em um problema real, pode-se aproximar a distribuição desejada com qualquer nível de precisão. Contudo, em termos práticos, isso implica em ter que lidar com uma matriz  $T$  muito grande. Por causa desta propriedade, processos de renovação do tipo fase são muito usados em problemas de fila, dando a origem ao modelo  $PH/GI/1$ , onde o processo de chegada é dado por uma distribuição tipo fase. Estes modelos ainda podem ser analisados usando o método da cadeia de Markov embutida, que é o método mais utilizado para tratar esse tipo de modelo. Isso somente é possível porque as distribuições tipo fase mantêm uma estrutura Markoviana. De forma análoga ao que foi feito na seção anterior, pode-se construir uma cadeia de Markov embutida para o modelo  $PH/GI/1$  discretizando o tempo nos instantes em que os clientes deixam o sistema. Seguindo os passos do que foi feito para o modelo  $M/GI/1$ , assumamos que  $t_{n-1}, t_n$  são variáveis aleatórias denotando o instante em que os clientes  $(n-1)$  e  $n$  deixam o sistema, respectivamente. Seja  $X$  o processo que conta o número de clientes no sistema, e considere o seguinte

$$\begin{aligned}
& \mathbb{P}((X(t_n), J(t_n)) = (x, j) | (X(t_{n-1}), J(t_{n-1})) = (y, i)) = \\
& = \mathbb{P}(X(t_{n-1}) + N(t_n) - N(t_{n-1}) - 1 = x, J(t_n) = j | (X(t_{n-1}), J(t_{n-1})) = (y, i)) \\
& = \mathbb{P}(N(t_n) - N(t_{n-1}) = x - y + 1, J(t_n) = j | J(t_{n-1}) = i).
\end{aligned}$$

Se  $y \geq 1$ , tem-se que  $(t_n - t_{n-1})$  é o tempo de serviço para o cliente  $n$ . Portanto, pode-se escrever o seguinte

$$\begin{aligned}
& \mathbb{P}((X(t_n), J(t_n)) = (x, j) | (X(t_{n-1}), J(t_{n-1})) = (y, i)) = \\
& = \int_0^\infty \mathbb{P}((N(t), J(t)) = (x - y + 1, j) | (N(0), J(0)) = (0, i)) dB(t) \\
& = \int_0^\infty p_{ij}(x - y + 1, t) dB(t),
\end{aligned}$$

onde  $B$  é a função de distribuição cumulativa do tempo em serviço, e  $p_{ij}(n, t) \triangleq \mathbb{P}((N(t), J(t)) = (n, j) | (N(0), J(0)) = (0, i))$ . Portanto, define-se  $A_n \in \mathbb{R}^{m \times m}$ ,

para  $n \in \mathbb{N}_0$ , como  $(A_n)_{ij} \triangleq \int_0^\infty p_{ij}(n, t) dB(t)$ . Caso  $y = 0$ , a fila permanecerá vazia até que um novo cliente chegue no sistema. Seja  $s_n$  o instante em que o cliente  $n$  chega no sistema. Então tem-se que

$$\begin{aligned}
& \mathbb{P}((X(t_n), J(t_n)) = (x, j) | (X(t_{n-1}), J(t_{n-1})) = (0, i)) = \\
&= \sum_{k=1}^m \mathbb{P}(N(t_n) - N(t_{n-1}) = x + 1, J(t_n) = j, J(s_n) = k | J(t_{n-1}) = i) \\
&= \sum_{k=1}^m \mathbb{P}(N(t_n) - N(s_n) = x, J(t_n) = j, J(s_n) = k | J(t_{n-1}) = i, N(s_n) - N(t_{n-1}) = 1) \\
&= \int_0^\infty \int_0^\infty \sum_{k=1}^m p_{kj}(x, t) dB(t) (e^{-T_s \eta})_i \alpha_k ds.
\end{aligned}$$

Defina  $B_n \in \mathbb{R}^{m \times m}$ , para  $n \in \mathbb{N}_0$ , como  $(B_n)_{ij} = \int_0^\infty (e^{-Ty} \eta \alpha A_n)_{ij} dy$ . Portanto, a matriz de probabilidade de transição da cadeia de Markov embutida é dada por

$$P_{(x,i),(y,j)} = \tilde{P}_{xm+i, ym+j} = \begin{pmatrix} B_0 & B_1 & B_2 & \cdots \\ A_0 & A_1 & A_2 & \cdots \\ 0 & A_0 & A_1 & \cdots \\ 0 & 0 & A_0 & \cdots \\ \vdots & \vdots & & \ddots \end{pmatrix},$$

que é similar à dada anteriormente na seção passada. Com esta matriz, pode-se mostrar que o Teorema 2.3.6 também aplica-se neste caso e também pode-se computar a função geradora de probabilidade para o sistema em regime estacionário.

Várias generalizações das distribuições tipo fase existem, todas tendo em comum um processo de Markov bidimensional  $(N, J)$ , onde  $N$  é usado como um componente aditivo que conta o número de chegadas, e  $J$  é usado para representar o estado atual da fase do processo. Uma destas generalizações é introduzida em (Neuts, 1979) e é chamada de “o versátil processo pontual Markoviano” (*the versatile Markovian point process*). A notação empregada na descrição deste processo para o caso de chegada de um cliente por vez foi simplificada por (Lucantoni et al., 1990) e o processo resultante foi chamado de “processos de chegada Markoviano”

(*Markovian arrival process*) (MAP). O MAP é definido por uma distribuição inicial  $\alpha \in \mathbb{R}^m$  e duas matrizes  $C, D \in \mathbb{R}^{m \times m}$ , que são respectivamente uma matriz estável e uma matriz não-negativa cuja a soma é a matriz  $Q$ , que é a matriz de taxas de transição infinitesimal do processo de Markov  $J$ , que geralmente é assumido ser irredutível. A matriz  $D$  possui a taxa de mudança de estados com chegada de clientes, e a matriz  $C$  possui as taxas de mudança de estado sem a chegada de clientes (para maior detalhe neste assunto veja (Asmussen, 2000; Neuts, 1979; Lucantoni et al., 1990)).

O processo acima foi generalizado por Lucantoni para considerar a possibilidade de múltiplas chegadas no mesmo instante (conhecido como chegada em bloco), dando origem ao processo Markoviano de chegada em bloco (*batch Markovian arrival process*) (BMAP). Foi mostrado que BMAP e o versátil processo pontual Markoviano do Neuts são equivalentes (Lucantoni, 1991). Este processo de chegada é descrito pelas matrizes  $D_k \in \mathbb{R}^{m \times m}$ ,  $k \in \mathbb{N}_0$ , onde  $D_0$  é uma matriz estável, e  $D_k$  são matrizes não-negativas tais que  $\sum_{k=0}^{\infty} D_k$  é uma matriz de taxas de transição infinitesimais do processo de Markov  $J$ . Portanto, a matriz  $D_0$  governa as transições que correspondem à nenhuma chegada, e  $D_j$  corresponde as transições correspondentes à chegada de  $j$  clientes ao mesmo tempo. A fila *BMAP/GI/1* é analisada em (Ramaswami, 1980) (sob o nome de *N/G/1*), onde o método da cadeia de Markov novamente é utilizado. Uma revisão mais detalhada do assunto pode ser encontrada em (Lucantoni, 1993).

É interessante incluir aqui, para finalizar esta seção, a definição da classe de processos (como definida em (Pacheco e Prabhu, 1995)) que extrai a característica principal dos processos de entrada Markovianos:

**Definição 2.4.3** (Processos Aditivos Markovianos). *Um processo  $\mathcal{X} \triangleq \{(N(t), J(t)) : t \in \mathbb{R}_{\geq 0}\}$  sob o espaço de estados  $\mathbb{R}^r \times \mathcal{E}$ , onde  $r > 1$  e  $\mathcal{E}$  é um conjunto contável, é chamado de processo Markoviano aditivo se:*

(i)  $\mathcal{X}$  é um processo Markoviano.

(ii) Para  $s, t \in \mathbb{R}_{\geq 0}$ , onde  $s < t$ , a distribuição condicional de  $(X(s+t) -$

$X(s), J(s+t)$  dado  $(X(s), J(s))$  somente depende de  $J(s)$ .

## 2.5 O Modelo $GI/GI/1$ e Condições de Estabilidade

Neste capítulo será considerado o modelo de fila  $GI/GI/1$ , onde o tempo entre chegadas consecutivas de clientes é dado por  $\{\Delta_i^a\}$  e o tempo de serviço é dado por  $\{\Delta_i^d\}$ , para  $i \in \mathbb{N}$ . As variáveis aleatórias  $\{\Delta_i^a, \Delta_i^d; i \in \mathbb{N}\}$  são mutualmente independentes. A função de distribuição cumulativa de  $\Delta_i^a$  é dada por  $A$ , e a de  $\Delta_i^d$  é dada por  $B$ , com média dada por  $\bar{\Delta}^a \triangleq \mathbb{E}[\Delta_i^a]$  e  $\bar{\Delta}^d \triangleq \mathbb{E}[\Delta_i^d]$  que são assumidas ser finitas e não-negativas. O principal propósito desta seção é estabelecer um resultado similar ao dado pelo Teorema 2.3.6 para o modelo geral que será tratado aqui. O primeiro passo fundamental para chegar a este objetivo é derivar a equação de Lindley (Lindley, 1952). O conteúdo desta seção é baseado em (Sigman, 2004). Uma outra excelente fonte de referência neste assunto é o livro do Prabhu (Prabhu, 1980).

Seja  $w_i$  o tempo total em que o cliente  $i$  permanece esperando na fila (ou seja, o tempo em que o cliente terá que permanecer na fila antes de ser atendido pelo servidor). A variável  $w_i$  pode ser também interpretada como a carga de trabalho do servidor no instante anterior a entrada do cliente  $i$ , onde a carga de trabalho é definida como o tempo total em que o servidor tem que trabalhar para completar todas as tarefas no sistema. Por simplicidade, assume-se que o sistema está vazio no instante em que o primeiro cliente entra na fila, isto é,  $w_0 = 0$ . Observe que, se o cliente  $i$  encontra o sistema ocupado no instante da sua chegada, ele deverá esperar o mesmo tempo de espera que resta para o cliente em sua frente (i.e.,  $w_{i-1} - \Delta_i^a$ ) mais o tempo de serviço deste cliente (i.e.,  $\Delta_{i-1}^d$ ). Logo,  $w_i = (w_{i-1} - \Delta_i^a) + \Delta_{i-1}^d$ . Caso contrário, se o sistema estiver vazio, o cliente deve esperar  $w_i = 0$ . Observe que este último caso corresponde ao evento  $\Delta_i^a \geq w_{i-1} + \Delta_{i-1}^d$ . A partir destas observações, chega-se à equação de Lindley, dada abaixo:

$$w_{i+1} = (w_i + y_i)^+ \quad i \geq 1 \quad (2.10)$$

onde  $(x)^+ \triangleq \max(0, x)$ , para  $x \in \mathbb{R}$ , e  $y_i \triangleq \Delta_i^d - \Delta_{i+1}^a$ .

Seja o passeio aleatório dado por  $S_k = \sum_{j=0}^{k-1} y_j$ ,  $k > 0$ , com  $S_0 = 0$ . Além disso, considere o passeio aleatório com incrementos em ordem invertida dado por  $S_{k,n} = \sum_{j=0}^{k-1} y_{n-j-1}$ ,  $1 \leq k \leq n$ , com  $S_{0,n} = 0$ . Defina ainda  $\tilde{M}_n$  como o máximo entre os  $n$  passos deste passeio aleatório, isto é,

$$\begin{aligned} \tilde{M}_n &\triangleq \max\{0, y_{n-1}, y_{n-1} + y_{n-2}, \dots, y_{n-1} + \dots + y_0\} \\ &= \max_{0 \leq k \leq n} S_{k,n}. \end{aligned}$$

Então tem-se os seguintes resultados:

**Lema 2.5.1.** *Para cada  $n \in \mathbb{N}$  fixo,  $w_n = \tilde{M}_n$ .*

*Demonstração.* A prova é feita com uma indução em  $n$ . Para  $n = 1$ , tem-se

$$w_1 = (w_0 + y_0)^+ = \max(0, y_0) = \max(S_{0,1}, S_{1,1}) = \tilde{M}_1.$$

Agora assumamos que  $w_n = \tilde{M}_n$  para algum  $n \in \mathbb{N}$ . Então,

$$\begin{aligned} w_{n+1} &= (w_n + y_n)^+ = (\tilde{M}_n + y_n)^+ = \max\left(0, \max_{0 \leq k \leq n} S_{k,n} + y_n\right) \\ &= \max_{0 \leq k \leq n+1} S_{k,n+1} = \tilde{M}_{n+1}. \end{aligned}$$

□

**Teorema 2.5.2.** *Para cada  $n \in \mathbb{N}$  fixo,*

$$w_n \stackrel{d}{=} M_n \triangleq \max_{0 \leq k \leq n} S_k,$$

onde o símbolo  $\stackrel{d}{=}$  denota igualdade em distribuição. Além disso, quando  $n \rightarrow \infty$ ,  $\{w_n\}$  converge em distribuição para  $w$  que satisfaz

$$w \stackrel{d}{=} M \triangleq \max_{k \geq 0} S_k.$$

*Demonstração.* A primeira parte do teorema segue de forma imediata após observar que a troca de ordem dos incrementos não altera a distribuição do passeio aleatório, já que os incrementos são identicamente distribuídos. Portanto, para todo  $n \geq 1$

$$\{S_{k,n} : 0 \leq k \leq n\} \stackrel{d}{=} \{S_k : 0 \leq k \leq n\}$$

que implica que  $\tilde{M}_n \stackrel{d}{=} M_n$  e o resultado segue do Lema 2.5.1.

A segunda parte do teorema pode ser provada a partir da observação que  $\{M_n\}$  é uma sequência não-decrescente convergindo para  $M$ . Logo, para todo  $x \in \mathbb{R}_{\geq 0}$ ,

$$\mathbb{P}(w_n \leq x) = \mathbb{P}(M_n \leq x) \rightarrow P(M \leq x),$$

que segue da continuidade das medidas (veja Teorema 1.2.7 página 9 de (Ash e Doléans-Dade, 2000)). Se  $w$  denota uma variável aleatória tal que  $P(w \leq x) = P(M \leq x)$ , então  $\{w_n\}$  converge fracamente para  $w$ .  $\square$

Com estes resultados pode-se mostrar o teorema de estabilidade para este sistema.

**Teorema 2.5.3** (Condição de Estabilidade para a Fila GI/GI/1). *Uma fila GI/GI/1 tem as seguintes condições de estabilidade:*

- (i) Se  $\rho < 1$ , então  $\mathbb{P}(w < \infty) = 1$  e a fila é chamada de estável.
- (ii) Se  $\rho > 1$ , então  $\mathbb{P}(w = \infty) = 1$  e a fila é chamada de instável.
- (iii) Se  $\rho = 1$  e  $\text{var}(y_n) > 0$ , então  $\mathbb{P}(w = \infty) = 1$  e a fila é chamada de instável.

Observe que  $\rho = \bar{\Delta}^d / \bar{\Delta}^a$ .

*Demonstração.* Observe que  $\rho < 1$  implica que  $\mathbb{E}[y_n] < 0$ . Com a aplicação da lei forte dos grandes números (i.e., Teorema 6.2.5 na página 242 de (Ash e Doléans-

Dade, 2000)) tem-se que

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mathbb{E}[y_n] < 0 \quad \text{q.c..}$$

Portanto,  $\lim_{n \rightarrow \infty} S_n = -\infty$  em quase toda parte. Consequentemente, o passeio aleatório atinge um máximo finito e segue para  $-\infty$  eventualmente. Logo,  $\mathbb{P}(M < \infty) = \mathbb{P}(w < \infty) = 1$ .

Para  $\rho > 1$ , tem-se que  $\mathbb{E}[y_n] > 0$ . A mesma análise acima leva a conclusão que  $\lim_{n \rightarrow \infty} S_n = \infty$ , q.c.. Logo  $M_n \rightarrow \infty$ , q.c., que implica que  $\mathbb{P}(w = \infty) = 1$ .

No caso de  $\rho = 0$ , uma aplicação da lei do logaritmo iterado (veja Teorema 13.25 na página 291 de (Breiman, 1993)) leva à

$$\begin{aligned} \limsup_n \frac{S_n}{\sqrt{2\text{var}(y_n) \log(\log(n))}} &= 1 \\ \liminf_n \frac{S_n}{\sqrt{2\text{var}(y_n) \log(\log(n))}} &= -1 \end{aligned}$$

quase certamente. Logo, tem-se que  $\limsup_n S_n = \infty$  e  $\liminf_n S_n = -\infty$ , q.c.. Portanto,  $M_n \rightarrow \infty$ , q.c., e  $\mathbb{P}(w = \infty) = 1$ .  $\square$

Neste momento, a equação integral de Lindley será derivada. A solução desta equação é a distribuição estacionária do tempo de espera do sistema.

**Teorema 2.5.4.** *Quando  $\rho < 1$ , o tempo de espera do sistema, dado por  $w$ , em regime estacionário satisfaz*

$$w \stackrel{d}{=} (w + y)^+,$$

onde  $y$  denota uma variável aleatória com a mesma distribuição que  $y_n$ .

*Demonstração.* Observe que  $w_n + y_n$  converge fracamente para  $w + y$  pelo Teorema 2.5.2. Como  $(x)^+ = \frac{1}{2}(x + |x|)$  é uma função contínua, então  $(w_n + y_n)^+$  converge para  $(w + y)^+$  em distribuição. Logo, o resultado segue tomando o limite em ambos os lados da Equação (2.10).  $\square$

Seja  $W(x) = \mathbb{P}(w \leq x)$  e  $F(x) = \mathbb{P}(y \leq x)$ , então pelo Teorema 2.5.4 tem-se que, para  $x \geq 0$ ,

$$W(x) = \mathbb{P}((w + y)^+ \leq x) = \mathbb{P}(w + y \leq x) = \int_{\mathbb{R}} \mathbb{P}(w \leq x - \xi) dF(\xi).$$

E portanto, chega-se à chamada “equação integral de Lindley”, dada abaixo:

$$\begin{cases} W(x) = \int_{\mathbb{R}} W(x - \xi) dF(\xi) & x \geq 0 \\ W(x) = 0 & x < 0 \end{cases}.$$

Em (Lindley, 1952), foi observado que existe somente uma solução para esta equação, já que a distribuição estacionária de passeios aleatórios (recorrente positivos) são únicas. Lindley também encontra a solução desta equação para a fila  $M/G/1$  (além de outros casos). O resultado é o seguinte:

$$W^*(s) = \frac{s(1 - \rho)}{s - (\bar{\Delta}^a)^{-1}(1 - B^*(s))},$$

onde  $G^*$  denota a transformada de Laplace de qualquer função  $G$ . Esta equação é chamada de equação de Pollaczek-Khinchin. A equação integral de Lindley também foi resolvida por Smith em (Smith, 1952), para casos onde uma fatoração específica do termo  $A^*(s)B^*(s) - 1$  é possível. Uma construção passo a passo da solução da equação integral de Lindley pode ser encontrada em (Kleinrock, 1975).

## 2.6 Aproximações de Filas Sob Tráfego Pesado

Até este ponto, nas seções anteriores, foram destacados três pontos principais, são eles:

- (i) A condição de estabilidade dependente do parâmetro  $\rho$ .
- (ii) A construção de formas fechadas para a distribuição estacionária de alguma quantidade de interesse (i.e., o número de clientes no sistema nas Seções 2.2, 2.3, e 2.4, e o tempo de espera na Seção 2.5). Observe, que exceto no

caso mais simples, referente ao modelo  $M/M/1$ , estas expressões sempre aparecem sob a forma de transformadas  $z$  ou de Laplace.

- (iii) A dificuldade de trabalhar com modelos onde uma estrutura Markoviana não está presente.

No início da década de 60, muitos se preocupavam com a forma das expressões obtidas na teoria. Para ter uso prático, as expressões requeriam muita informação sobre o sistema (i.e., é necessário conhecimento suficiente para construir uma distribuição, para em seguida computar sua transformada, aplicar na formula, e inverter a transformada). Muitas vezes não se conhecia o sistema de forma detalhada e não se deseja resultados exatos, mas sim uma idéia geral do comportamento do fenômeno. Uma frase marcante de Kendall nesta época diz que a teoria da fila ainda se mantinha escondida atrás de cortinas Laplacianas (Kendall, 1964).

A seguir serão apresentados alguns resultados desenvolvido por Kingman em uma série de artigos (Kingman, 1962, 1963, 1970) que ajudam a entender o problema com relativamente poucos parâmetros do sistema (outras inequações, além das que serão apresentadas aqui, podem também ser encontradas nestes artigos). Nesta seção, será usado as mesmas hipóteses da seção anterior. O seguinte lema será usado no Teorema 2.6.2.

**Lema 2.6.1.** *Seja  $\gamma : (0, \infty) \rightarrow [0, 1]$  uma função não-crescente tal que*

$$\int_{-\infty}^x \gamma(x - \xi) dF(\xi) + 1 - F(x) \leq \gamma(x) \quad \forall x > 0.$$

onde  $F(x) \triangleq \mathbb{P}(y \leq x)$ . Então, tem-se que

$$\mathbb{P}(w \geq x) \leq \gamma(x) \quad \forall x > 0.$$

*Demonstração.* Esta prova é feita por indução. Observe que  $\mathbb{P}(w_0 \geq x) \leq \gamma(x)$  já

que  $w_0 = 0$ . Agora suponha que  $\mathbb{P}(w_n \geq x) \leq \gamma(x)$  para algum  $n \in \mathbb{N}$ . Então,

$$\begin{aligned} \mathbb{P}(w_{n+1} \geq x) &= \mathbb{P}(w_n + y_n \geq x) = \int_{-\infty}^{\infty} \mathbb{P}(w_n \geq x - \xi) dF(\xi) \\ &= \int_{-\infty}^x \mathbb{P}(w_n \geq x - \xi) dF(\xi) + \int_x^{\infty} dF(\xi) \\ &\leq \int_{-\infty}^{\infty} \gamma(x - \xi) dF(\xi) + 1 - F(x) \leq \gamma(x). \end{aligned}$$

Portanto, o resultado segue deixando  $n \rightarrow \infty$ . □

**Teorema 2.6.2.** *Suponha que a distribuição do tempo de serviço possua uma “calda leve”, isto é, existe um  $\epsilon > 0$  tal que  $\mathbb{E}[e^{\epsilon \Delta_i^d}] < \infty$ , e que  $\rho < 1$ . Então,*

$$\phi(\theta) \triangleq \mathbb{E}[e^{\theta y}] < 1$$

para alguns valores de  $\theta$  positivos e, além disso, tem-se que  $\nu = \sup\{\theta > 0; \phi(\theta) < 1\}$  é positivo e finito (exceto para o caso trivial  $\mathbb{P}(y \leq 0) = 1$ , que implica que  $\mathbb{P}(w = 0) = 1$ ). Além disso, tem-se que

$$\mathbb{P}(w \geq x) \leq e^{-\nu x} \quad \forall x > 0. \quad (2.11)$$

*Demonstração.* Se  $\mathbb{P}(y \leq 0) = \mathbb{P}(w = 0) = 1$ , o teorema é trivial, portanto, assumamos que este não é o caso. Observe que a função  $\phi(\theta)$  é a função geradora de momentos e, portanto, possui uma expansão em série de Taylor em volta do ponto  $\theta = 0$  com um raio positivo de convergência. Além disso, como  $\rho < 1 \Rightarrow \mathbb{E}[y] < 0$ ,  $\phi'(0) = \mathbb{E}[y] < 0$  e portanto a função é estritamente decrescente em volta do zero. Desta forma, o fato de  $\phi(0) = 1$  implica em  $\phi(\theta) < 1$  para valores pequenos e positivos de  $\theta$ . Com este mesmo argumento, pode-se estabelecer a existência de  $\nu$  finito e positivo.

O limite superior dado pela Equação (2.11) é consequência do Lema (2.6.1).

De fato, fazendo  $\gamma(x) \equiv e^{-\nu x}$  tem-se que

$$\begin{aligned}
\int_{-\infty}^x e^{-\nu(x-\xi)} dF(\xi) + 1 - F(x) &= e^{-\nu x} \int_{-\infty}^x e^{\nu\xi} dF(\xi) + \int_x^{\infty} dF(\xi) \\
&= e^{-\nu x} \left( \int_{-\infty}^x e^{\nu\xi} dF(\xi) + \int_x^{\infty} e^{\nu x} dF(\xi) \right) \\
&\leq e^{-\nu x} \underbrace{\left( \int_{-\infty}^{\infty} e^{\nu\xi} dF(\xi) \right)}_{\phi(\nu) < 1} \leq e^{-\nu x},
\end{aligned}$$

que completa a prova. □

Um outro teorema, chamado de Teorema da Inequação do Tempo Médio, pode ser computado a partir da equação de Lindley, dada pela Equação (2.10).

**Teorema 2.6.3.** *Suponha que  $\rho < 1$  e que  $\mathbb{E}[y^2] < \infty$ , então tem-se que*

$$\frac{\mathbb{E}[(y^+)^2]}{2|\mathbb{E}[y]|} \leq \mathbb{E}[w] \leq \frac{\text{var}(y)}{2|\mathbb{E}[y]|}.$$

*Demonstração.* Suponha primeiramente que  $\mathbb{E}[w^2] < \infty$ . Sabe-se do Teorema 2.5.4 que

$$w \stackrel{d}{=} (w + y)^+. \quad (2.12)$$

Observe que é possível escrever o seguinte:  $(w + y) = (w + y)^+ - (w + y)^-$ , onde  $(x)^- = -\min(0, x)$ . Tomando o valor esperado de ambos os lados da equação acima, tem-se  $\mathbb{E}[w] + \mathbb{E}[y] = \mathbb{E}[(w + y)^+] - \mathbb{E}[(w + y)^-]$ . Portanto, a Equação (2.12) implica que  $\mathbb{E}[(w + y)^-] = -\mathbb{E}[y] = |\mathbb{E}[y]|$ , já que  $\mathbb{E}[y] < 0$  (i.e.,  $\rho < 1$ ).

Para qualquer  $x \in \mathbb{R}$ , tem-se que  $x^2 = (x^+)^2 + (x^-)^2$  já que  $-2x^+x^- = 0$ . Portanto,

$$\begin{aligned}
\text{var}(x) &= \mathbb{E}[x^2] - (\mathbb{E}[x])^2 = \mathbb{E}[x^2] - (\mathbb{E}[x^+] - \mathbb{E}[x^-])^2 \\
&= \mathbb{E}[(x^+)^2] - \mathbb{E}[x^+]^2 + \mathbb{E}[(x^-)^2] - \mathbb{E}[(x^-)^2] + 2\mathbb{E}[x^+] \mathbb{E}[x^-] \\
&= \text{var}(x^+) + \text{var}(x^-) + 2\mathbb{E}[x^+] \mathbb{E}[x^-].
\end{aligned}$$

Usando a equação acima, tem-se que

$$\begin{aligned}\text{var}(w + y)^+ + \text{var}(w + y)^- &= \text{var}(w + y) - 2\mathbb{E}[(w + y)^+] \mathbb{E}[(w + y)^-] \\ &= \text{var}(w) + \text{var}(y) - 2\mathbb{E}[w] |\mathbb{E}[y]|.\end{aligned}$$

Desta forma,

$$\mathbb{E}[w] = \frac{\text{var}(y) - \text{var}(w + y)^-}{2|\mathbb{E}[y]|}, \quad (2.13)$$

que gera o limite superior quando se usa o fato que  $\text{var}(w + y)^- \geq 0$ . Para chegar ao limite inferior, considere o seguinte:

$$((w + y)^-)^2 \leq (y^-)^2,$$

já que  $w \geq 0$ . Então,

$$\begin{aligned}\text{var}(w + y)^- &\leq \mathbb{E}[(w + y)^-]^2 - \mathbb{E}[y]^2 \leq \mathbb{E}[(y^-)^2] - \mathbb{E}[y]^2 \\ &= \mathbb{E}[y^2 - (y^+)^2] - \mathbb{E}[y]^2 = \text{var}(y) - \mathbb{E}[(y^+)^2],\end{aligned}$$

que leva ao resultado quando aplicado na Equação (2.13).

O caso geral, quando não se assume que  $\mathbb{E}[w^2] < \infty$ , pode ser tratado através da técnica de truncagem. Considere o sistema truncado onde  $y^N \triangleq \min\{y, N\}$ , para algum  $N \in \mathbb{N}$ . Logo, pelo Teorema 2.6.2 tem-se que  $w^N$  tem momentos finitos. Pelo Teorema da convergência monótona, tem-se que

$$\begin{aligned}\mathbb{E}[w^N] &\rightarrow \mathbb{E}[w], & \mathbb{E}[y^N] &\rightarrow \mathbb{E}[y] \\ \mathbb{E}[(y^N)^2] &\rightarrow \mathbb{E}[y^2], & \mathbb{E}[(y^N)^+]^2 &\rightarrow \mathbb{E}[(y^+)^2],\end{aligned}$$

e portanto segue o resultado. □

Kingman já havia observado que quando o sistema estava em regime de tráfego pesado, isto é, quando  $1 - \epsilon < \rho < 1$ , para  $\epsilon$  muito pequeno, o limite superior do Teorema 2.6.3 era bastante preciso. Isto é explicado porque o erro no limite superior é introduzido quando a variância de  $(w + y)^-$  é desprezada, e sistemas em regime de tráfego pesado possuem  $(w + y)^- = 0$  com alta probabilidade. Observe que o Teorema 2.6.2 combinado com o limite superior do Teorema 2.6.3 sugere que talvez

$$\mathbb{P}(w \leq x) \approx 1 - e^{x(2\mathbb{E}[y]/\text{var}(y))}.$$

De fato, Kingman mostrou (Kingman, 1962) que é possível aproximar a distribuição de  $w$  por uma exponencial quando  $\rho \approx 1$ . Para obter tal resultado, foi estabelecido uma sequência de filas  $GI/GI/1$  indexadas por um parâmetro  $n$ . A medida que  $n$  tendia a  $\infty$ , tinha-se que  $\mathbb{E}[y^n] = \bar{\Delta}^{d,n} - \bar{\Delta}^{a,n} \uparrow 0$ , ou seja  $\rho^n \triangleq \bar{\Delta}^{d,n}/\bar{\Delta}^{a,n} \uparrow 1$ . Supondo algumas condições de regularidade gerais (como, por exemplo, a de que  $y^n$  converge fracamente para  $y$  e que  $0 < \text{var}(y) < \infty$ ) tinha-se que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{2|\mathbb{E}[y]|}{\text{var}(y)} w^n \leq x\right) = 1 - e^{-x} \quad (2.14)$$

para  $x \geq 0$ . Este resultado é considerado o primeiro resultado de aproximação de filas em regime de tráfego pesado.

Como foi observado por Whitt em (Whitt, 1972), um outro resultado já conhecido na época (i.e., veja (Erdős e Kac, 1946)) era o seguinte: se  $\mathbb{E}[y] \equiv 0$  (que é equivalente a dizer que  $\rho = 1$ ), então

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{1}{\text{var}(y)\sqrt{n}} \max_{0 \leq k \leq n} S_n \leq x\right) = \left(\frac{2}{\pi}\right)^{1/2} \int_0^x e^{-\xi^2/2} d\xi \quad \text{para } x \geq 0, \quad (2.15)$$

onde  $S_n = \sum_{j=0}^{n-1} y_j$ , para  $n > 0$ , e  $S_0 \triangleq 0$  (veja, por exemplo, Teorema 7.33 de (Chung, 2001) na página 232). É claro que do resultado de Lindley, tem-se que

$$\mathbb{P}(w_n \leq x) = \mathbb{P}\left(\max_{0 \leq k \leq n} S_n \leq x\right)$$

e, portanto, a distribuição de  $w_n/\text{var}(y)\sqrt{n}$  converge para o mesmo limite dado na Equação (2.15).

Este tipo de escalonamento e aplicação de convergência no sentido fraco é a base de aproximações de tráfego pesado. Como dito por Whitt em (Whitt, 1974), “a idéia é evitar complexidade desnecessária e tentar obter representações simples que capturam características essenciais do problema suficientes para serem úteis”. O foco se torna a visão macroscópica do problema de onde alguma regularidade estatística pode ser encontrada e que não é aparente do ponto de vista microscópico.

Os dois resultados acima são reflexo do resultado que será mostrado a seguir. Este resultado parte da seguinte observação:

$$w_n = S_n - \min_{0 \leq k \leq n} S_k \quad n \geq 0, \quad (2.16)$$

que pode ser facilmente mostrado com o uso de um argumento indutivo. De fato, considere o seguinte:

$$w_1 = (w_0 + y_0)^+ = \max(0, y_0) = \max(S_0, S_1) = S_1 - \min(S_0, S_1).$$

Assuma que a Equação (2.16) seja válida para um  $n \in \mathbb{N}$  e defina  $m_n \triangleq \min_{0 \leq k \leq n} S_k$ , então tem-se que

$$\begin{aligned} w_{n+1} &= (w_n + y_n)^+ = \max(0, S_n - m_n + y_n) \\ &= \max(0, S_{n+1} - m_n) = (S_{n+1} - m_n) - \min(0, S_{n+1} - m_n) \\ &= S_{n+1} - \min(S_{n+1}, m_n) = S_{n+1} - \min_{0 \leq k \leq n+1} S_k. \end{aligned}$$

Com a expressão acima chega-se à:

**Teorema 2.6.4.** *Suponha uma sequência de filas GI/GI/1 indexadas pelo parâmetro  $n$ , onde  $y^n \Rightarrow y$  e  $0 < \text{var}(y) < \infty$ . Assuma ainda que*

$$\lim_{n \rightarrow \infty} \mathbb{E}[y^n] \sqrt{n} = c \in (-\infty, \infty),$$

e seja  $\sigma^2 \triangleq \text{var}(y)$ . Seja o processo  $Z^n$  definido como

$$Z^n(t) \triangleq \frac{w_{nt}^n}{\sqrt{n}},$$

para  $t \in \mathbb{R}_{\geq 0}$  (onde aqui e no decorrer deste trabalho,  $nt$  denota o maior inteiro igual ou menor que o número real  $nt$ , quando colocado no papel de um número inteiro). Então tem-se que

$$Z^n \Longrightarrow Z_{\sigma,c}$$

onde, para cada  $t \in \mathbb{R}_{\geq 0}$ ,  $Z_{\sigma,c}(t) \triangleq \sigma W(t) + ct + 0 \vee \sup_{s \leq t} (-\sigma W(s) - cs)$ , onde  $W$  é um processo de Wiener padrão.

*Demonstração.* Usando a Equação (2.16), pode-se escrever o seguinte

$$\frac{w_{nt}^n}{\sqrt{n}} = \frac{S_{nt}^n}{\sqrt{n}} - \min_{0 \leq k \leq nt} \left( \frac{S_k^n}{\sqrt{n}} \right).$$

Por outro lado, uma aplicação do Teorema B.1.9 (que é uma extensão do conhecido Teorema de Donsker (Teorema B.1.8) ou Teorema do Limite Funcional Central) implica que o processo tomando valores

$$\frac{1}{\sqrt{n}} (S_{nt}^n - nt\mathbb{E}[y^n])$$

converge no sentido fraco para um processo de Wiener com variância  $\sigma^2$ . Como  $\sqrt{n}\mathbb{E}[y^n] \rightarrow c$ , tem-se que o processo tomando valores  $S_{nt}^n/\sqrt{n}$  converge fracamente para o processo com valores  $\sigma W(t) + ct$ , para  $t \in \mathbb{R}_{\geq 0}$ .

Seja  $\Psi$  definido como

$$\Psi(x)(t) = x(t) + 0 \vee \sup_{s \leq t} (-x(s)),$$

para qualquer função contínua  $x$ , tal que  $x(0) = 0$ , e  $t \geq 0$ . Esta função  $\Psi$  é geralmente chamada de “Função de Reflexão” e é tratada com detalhes no Apêndice A. Observe que a Inequação (A.3), dada pelo Teorema A.1.5, afirma que esta função  $\Psi$  é contínua sob o espaço das funções contínuas. Como as funções amostras de um processo de Wiener são contínuas quase certamente, o Teorema do mapeamento contínuo pode ser aplicado (e.g., Teorema 2.7 na página 21 de (Billingsley, 1999)) para chegar ao resultado.  $\square$

A relação deste resultado com os dados pelas Equações (2.14) e (2.15) é clara. A distribuição do processo  $Z_{\sigma,c}$  é dada por

$$\mathbb{P}(Z_{\sigma,c}(t) \leq x) = \Theta\left(\frac{x-ct}{\sigma\sqrt{t}}\right) - e^{-2|c|/\sigma^2} \Theta\left(\frac{-x-ct}{\sigma\sqrt{t}}\right), \quad (2.17)$$

para  $x \geq 0$  e onde  $\Theta$  é a função de densidade cumulativa de uma normal padrão (veja, por exemplo, Equação 4.5 de (Whitt, 1974)). Claramente, tem-se a exponencial da Equação (2.14) quando  $t \rightarrow \infty$ . Além disso, se for fixado  $t \equiv 1$ , tem-se o resultado dado pela Equação (2.15).

É importante mostrar também que o número de clientes na fila, dado pelo processo  $X \triangleq N(t) - S(t)$ , pode também ser aproximado por um movimento Browniano refletido. Novamente, suponha a existência de uma sequência de filas  $GI/GI/1$  indexadas pelo parâmetro  $n$ , e assuma as seguintes condições:

**Condição 2.6.5.**

- (a) Seja  $\xi_i^{\alpha,n} \triangleq 1 - \Delta_i^{\alpha,n}/\bar{\Delta}^{\alpha,n}$ , para  $\alpha \in \{a, d\}$ , e suponha que  $\xi_i^{\alpha,n}$  satisfaz as condições do Teorema B.1.9.
- (b) Suponha que  $\sqrt{n} [(\bar{\Delta}^{a,n})^{-1} - (\bar{\Delta}^{d,n})^{-1}] \triangleq c^n \rightarrow c$  (a medida que  $n \rightarrow \infty$ ), que é a chamada “condição do tráfego pesado”, i.e.  $\rho \uparrow 1$ .

Sob estas condições e usando o Teorema B.1.9 sabe-se que

$$W^{\alpha,n}(t) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{nt} \xi_i^{\alpha,n} \quad (2.18)$$

converge fracamente para um processo de Wiener com variância dada pelo limite ( $n \rightarrow \infty$ ) de

$$\text{var}(\xi_i^{\alpha,n}) = \frac{\text{var}(\Delta^{\alpha,n})}{(\bar{\Delta}^{\alpha,n})^2} \triangleq (\sigma_n^\alpha)^2.$$

Além disso, pode-se mostrar (e.g., Teorema B.2.6), que  $N^n(nt)/n$  converge fracamente para o processo tomando valores  $t/\bar{\Delta}^a$ , onde  $\bar{\Delta}^a$  é o limite de  $\bar{\Delta}^{\alpha,n}$  quando  $n \rightarrow \infty$ . O mesmo pode ser dito para  $S^n(nt)/n$  que converge fracamente para um processo com valores  $t/\bar{\Delta}^d$ .

Defina o seguinte processo escalonado  $x^n(t) \triangleq X^n(nt)/\sqrt{n} = N^n(nt)/\sqrt{n} - S^n(nt)/\sqrt{n}$ . E observe que

$$\begin{aligned} \frac{N^n(nt)}{\sqrt{n}} &= \frac{1}{\sqrt{n}} \sum_{l=1}^{N^n(nt)} 1 = \frac{1}{\sqrt{n}} \sum_{l=1}^{N^n(nt)} \left(1 - \frac{\Delta_l^{a,n}}{\bar{\Delta}^{a,n}}\right) + \frac{1}{\sqrt{n}\bar{\Delta}^{a,n}} \sum_{l=1}^{N^n(nt)} \Delta_l^{a,n} \\ &= W^{a,n} \left( \frac{N^n(nt)}{n} \right) + \frac{1}{\sqrt{n}\bar{\Delta}^{a,n}} (nt) + \varepsilon^{a,n}(t), \\ \frac{S^n(nt)}{\sqrt{n}} &= \frac{1}{\sqrt{n}} \sum_{l=1}^{S^n(nt)} 1 = \frac{1}{\sqrt{n}} \sum_{l=1}^{S^n(nt)} \left(1 - \frac{\Delta_l^{d,n}}{\bar{\Delta}^{d,n}}\right) + \frac{1}{\sqrt{n}\bar{\Delta}^{d,n}} \sum_{l=1}^{S^n(nt)} \Delta_l^{d,n} \\ &= W^{d,n} \left( \frac{S^n(nt)}{n} \right) + \frac{1}{\sqrt{n}\bar{\Delta}^{d,n}} (nt - I^n(nt)) + \varepsilon^{d,n}(t), \end{aligned}$$

onde  $\varepsilon^{\alpha,n}$  é um erro desprezível e  $I^n(t)$  denota o tempo ocioso total do servidor até o instante  $t$ . Observe que

$$z^n(t) \triangleq \frac{1}{\sqrt{n}\bar{\Delta}^{d,n}} I^n(nt)$$

satisfaz as seguintes propriedades:  $z^n(0) = 0$ ,  $z^n$  é não decrescente, e somente aumenta nos instantes  $t$  tais que  $x^n(t) = 0$ . Portanto, usando o Teorema A.1.4, sabe-se que  $z^n$  é único e é dado por  $0 \vee \sup_{s \leq t} (-W^{a,n}(N^n(ns)) + W^{d,n}(S^n(ns)) - c^n s)$

modulo um erro desprezível. Portanto, procedendo como na prova do Teorema 2.6.4, chega-se ao limite

$$x(t) = W^a(t/\bar{\Delta}^a) - W^d(t/\bar{\Delta}^d) + ct + z(t)$$

onde  $z(t) \triangleq 0 \vee \sup_{s \leq t} (-W^a(s/\bar{\Delta}^a) + W^d(s/\bar{\Delta}^d) - cs)$ .

Portanto, a análise de tráfego pesado aproxima sistemas de filas com qualquer distribuição de entrada e de serviço em regime de tráfego pesado, isto é, sob uma carga intensa de trabalho e beirando o limite de estabilidade, por um movimento Browniano refletido. Visto a teoria clássica, onde os resultados geralmente estão sob a forma de transformadas, este resultado é de extrema importância pois descreve o comportamento do sistema de forma simples e dependente do tempo. Com este modelo pode-se empregar ferramentas poderosas disponíveis na teoria de probabilidade para auxiliar o entendimento destes sistemas em regime de tráfego pesado. Estas aproximações ganharam grande força nos anos 80 com o aparecimento de sistemas de computadores, que em várias aplicações operam em regime de tráfego pesado. Além disso, a análise é facilmente adaptada para o caso de redes de fila como será visto no capítulo a seguir.

## 2.7 Redes de Filas

Nesta seção, serão considerados alguns problemas de redes de filas de forma a destacar como a teoria das aproximações por tráfego pesado ajuda o entendimento destes sistemas.

Considere primeiramente um sistema formado por duas filas em sequência, isto é, clientes que entram no sistema juntam-se à fila 1, ao terminar seu serviço, os clientes saem da fila 1 e entram na fila 2, somente ao terminar o serviço na fila 2, os clientes podem deixar o sistema. Vamos supor que cada fila deste sistema é  $M/M/1$ , com taxas de chegada e serviço dadas por  $\lambda_1^a \triangleq 1/\bar{\Delta}_1^a$  e  $\lambda_i^d \triangleq 1/\bar{\Delta}_i^d$ , respectivamente, onde  $i \in \{1, 2\}$ . O objetivo do que segue é estabelecer a distribuição do tempo entre chegadas consecutivas na fila 2. Defina o evento

$E_l \triangleq \{\text{a fila 1 está vazia no instante em que o cliente } l \text{ termina seu serviço}\}$ . Observe que o intervalo de tempo entre os instantes em que clientes sucessivos deixam a fila 1, dado por  $d_l$ , pode ser escrito da seguinte forma:

$$d_l = \mathbb{I}_{E_{l-1}}(\Delta_l^a + \Delta_l^d) + \mathbb{I}_{E_{l-1}^c}(\Delta_l^d).$$

Portanto, para  $x \geq 0$ , pode-se escrever o seguinte:

$$\begin{aligned} \mathbb{P}(d_l \leq x) &= \mathbb{P}(\Delta_l^a + \Delta_l^d \leq x \mid E_{l-1}) \mathbb{P}(E_{l-1}) + \mathbb{P}(\Delta_l^d \leq x \mid E_{l-1}^c) \mathbb{P}(E_{l-1}^c) \\ &= \mathbb{P}(\Delta_l^a + \Delta_l^d \leq x) \mathbb{P}(E_{l-1}) + \mathbb{P}(\Delta_l^d \leq x) \mathbb{P}(E_{l-1}^c), \end{aligned}$$

onde a última passagem é possível porque  $E_{l-1}$  e  $E_{l-1}^c$  podem ser escritos completamente em termos de  $\{\Delta_k^a, \Delta_k^d; 0 \leq k \leq l-1\}$  e, portanto, são independentes de  $\Delta_l^a$  e  $\Delta_l^d$ . Defina  $\rho_1 = \lambda_1^a/\lambda_1^d$ , assumamos que a fila 1 é estável ( $\rho_1 < 1$ ) e deixemos  $l \rightarrow \infty$  para tratar o sistema em regime estacionário. Seja  $D^*(s)$  a transformada de Laplace da distribuição de  $d_n$  em regime estacionário, então tem-se que

$$D^*(s) = A_1^*(s)B_1^*(s)(1 - \rho_1) + B_1^*(s)\rho_1,$$

onde  $A_1^*(s)$  e  $B_1^*(s)$  são as transformadas de Laplace das distribuições de  $\Delta_{l,1}^a$  e  $\Delta_{l,1}^d$ , respectivamente. Também foi usado acima o fato que a probabilidade do sistema estar vazio quando um cliente deixa o sistema em regime estacionário é dado por  $(1 - \rho)$  veja as Equações (2.2) ou (2.9).

Como a transformada de Laplace de uma distribuição exponencial com parâmetro  $\alpha$  é dada por  $\alpha/(\alpha + s)$ , tem-se que

$$\begin{aligned} D^*(s) &= \left( \frac{\lambda_1^a}{\lambda_1^a + s} \right) \left( \frac{\lambda_1^d}{\lambda_1^d + s} \right) (1 - \rho_1) + \left( \frac{\lambda_1^d}{\lambda_1^d + s} \right) \rho_1 \\ &= \frac{\lambda_1^a}{\lambda_1^a + s}, \end{aligned}$$

que é um resultado surpreendente: a distribuição do intervalo de tempo entre clien-

tes sucessivos deixando o sistema é exponencial com parâmetro  $\lambda_1^a$ . Este resultado é geralmente chamado de “resultado de Burke”, dado que foi ele que primeiro verificou esta propriedade em (Burke, 1966), veja também (Kleinrock, 1975) para mais detalhes.

Considere agora um sistema composto por  $K$  filas em rede. Cada fila  $i$ , para  $i \in \{1, \dots, K\}$  recebe clientes externos com distribuição entre chegada de clientes dada por uma exponencial com taxa  $\lambda_i^a$ . O tempo de serviço de cada cliente é exponencialmente distribuído com taxa  $\lambda_i^d$ . Cada cliente pode se juntar à uma fila  $j$  presente no sistema após terminar seu serviço em uma fila  $i$  com probabilidade  $r_{ij}$ . O cliente também pode deixar o sistema com probabilidade  $1 - \sum_{j=1}^K r_{ij}$ . Suponha que a matriz de roteamento, dada por  $R \triangleq (r_{ij})$ , tenha raio espectral menor do que 1. Isto implica que, quando considerado como uma matriz de transição de uma cadeia de Markov, todos os estados devem ser transiente. Em termos de teoria de filas, esta condição significa que todo cliente que entra no sistema sairá do sistema eventualmente.

Defina  $\lambda_i$ , para  $i \in \{1, \dots, K\}$ , como o conjunto de variáveis que satisfaz as equações abaixo:

$$\lambda_i = \lambda_i^a + \sum_{j=1}^K \lambda_j r_{ji}, \quad i \in \{1, \dots, K\}, \quad (2.19)$$

ou em forma vetorial,  $\lambda = \lambda^a + R'\lambda$ , que tem solução única dada por  $\lambda = (I - R')^{-1}\lambda^a$  sob a condição no raio espectral de  $R$ . Observe que, visto o resultado de Burke,  $\lambda_i$  pode ser visto como a taxa total de entrada de clientes na fila  $i$ , isto é, a taxa com quem clientes externos e clientes roteados entram na fila. De fato, se  $r_{ii} = 0$  e  $\rho \triangleq \lambda_i/\lambda_i^d < 1$ , para todo  $i \in \{1, \dots, K\}$ , o sistema em regime estacionário pode ser visto como  $K$  filas  $M/M/1$  independentes com parâmetros  $\lambda_i$  e  $\lambda_i^d$ , denotando respectivamente a taxa de entrada e a de serviço.

Quando  $r_{ii} > 0$  para algum  $i \in \{1, \dots, K\}$ , isto é, quando existe realimentação direta de uma fila, pode-se mostrar que o intervalo entre a saída consecutiva de clientes não será exponencial (veja (Disney, 1981)). Contudo Jackson mostrou

em (Jackson, 1957, 1963) que apesar destas filas não serem filas  $M/M/1$ , elas se comportam como se fossem, no sentido que a distribuição conjunta do número de clientes em cada fila em regime estacionário é um produto das distribuições dada pela Equação (2.2). Ou seja, para  $(n_1, \dots, n_K) \in \mathbb{N}_0^K$ ,

$$\mathbb{P}(X_1 = n_1, \dots, X_K = n_K) = \prod_{i=1}^K \rho_i^{n_i} (1 - \rho_i), \quad (2.20)$$

onde  $X_i$  denota o número de clientes na fila  $i$  em regime estacionário, para  $i \in \{1, \dots, K\}$ . Portanto, este resultado pode ser interpretado como se a rede fosse composta por  $K$  filas  $M/M/1$  independentes. Este resultado é geralmente chamado de Teorema de Jackson e pode ser provado mostrando que (2.20) é a distribuição estacionária do processo de Markov  $(X_1, \dots, X_K)$ . Para detalhes neste assunto veja (Gross e Harris, 1998) Seção 4.2.

O Teorema de Jackson é de extrema importância para teoria de filas e possibilita o estudo de vários problemas práticos. Contudo, este resultado não se estende para outros casos mais gerais, como por exemplo, filas onde o tempo entre a chegada de cliente não seja exponencial. Além disso, o Teorema de Burke praticamente não existe para outros modelos de fila. De fato, Chang mostrou que dentro da classe das filas  $G/GI/1$ , em que a distribuição do tempo de serviço possui suporte não limitado, o número de distribuições do tempo entre a chegada de clientes que induzem distribuições idênticas para o tempo entre a saída de clientes é no máximo um (Chang, 1994). Em particular, para todas as filas  $G/M/1$ , somente aquela com entrada de Poisson possui tal propriedade. Este tipo de observação indica que a análise de modelos mais gerais não é fácil, até mesmo para redes compostas por filas mais simples como a  $PH/M/1$ .

É neste cenário que a análise de tráfego pesado mostra todo seu potencial. A aproximação pode ser estendida para redes de filas sem maiores dificuldades. A equação de limite toma a forma de um movimento Browniano refletido que possibilita, entre outras coisas, tratar de problemas de controle. A seguir será construído a aproximação de uma forma resumida, para um tratamento mais rigoroso veja as

seguintes referências (Kushner, 2001; Reiman, 1984).

Da mesma forma como feito anteriormente, suponha a existência de uma seqüência de redes de filas indexadas por um parâmetro  $n \in \mathbb{N}$ . Sejam  $\{\Delta_{i,l}^{a,n}, l \in \mathbb{N}\}$  os intervalos entre a chegada de clientes externos na fila  $i$  e  $\{\Delta_{i,l}^{d,n}, l \in \mathbb{N}\}$  os tempos de serviço na fila  $i$ . Sejam também  $\mathbb{I}_{ij,l}^n$  uma função indicadora do evento que o  $l$ -ésimo cliente que deixa a fila  $j$  se junta à fila  $i$ . Defina o seguinte vetor  $\mathbb{I}_{i,l}^n \triangleq (\mathbb{I}_{ij,l}^n, j \in \{1, \dots, K\})$  e as constantes  $\bar{\Delta}_{i,l}^{\alpha,n} \triangleq 1/\lambda_{i,l}^{\alpha,n} \triangleq \mathbb{E}[\Delta_{i,l}^{\alpha,n}]$ , para  $\alpha \in \{a, d\}$ .

**Condição 2.7.1.** *As seguintes condições serão usadas:*

- (a) *As variáveis aleatórias  $\Delta_{i,l}^{\alpha,n}$  e  $\mathbb{I}_{i,l}^n$ , para  $\alpha \in \{a, d\}$ , são mutuamente independentes nos índices  $i, l$ , e  $\alpha$ . Além disso, as variáveis  $\xi_{i,l}^n \triangleq 1 - \Delta_{i,l}^{\alpha,n} / \bar{\Delta}_i^{\alpha,n}$  satisfazem as condições do Teorema B.1.9.*
- (b) *Existem constantes  $r_{ij}^n$  e  $r_{ij}$  tais que  $\mathbb{P}(\mathbb{I}_{ij,l}^n = 1) = r_{ij}^n \rightarrow r_{ij}$ . Além disso, o raio espectral da matriz  $R = (r_{ij})$  é menor do que um.*
- (c) *Existem  $c_i$  tais que*

$$\sqrt{n} \left[ \frac{1}{\bar{\Delta}_i^{a,n}} + \sum_{j=1}^K \frac{r_{ji}^n}{\bar{\Delta}_j^{d,n}} - \frac{1}{\bar{\Delta}_i^{d,n}} \right] \triangleq c_i^n \rightarrow c_i.$$

Observe que a Condição 2.7.1(c) é a condição do tráfego pesado. A condição exige que a taxa de clientes entrando em cada fila do sistema se aproxime da taxa com que os clientes saem da fila, a medida que  $n \uparrow \infty$ .

Seja  $N_i^n$  o processo que conta o número de clientes exógenos que entram na fila  $i$ , e  $S_i^n$  o processo que conta o número de clientes servidos na fila  $i$ . Seja também  $S_{ij}^n$  o processo que conta o número de clientes que saíram da fila  $i$  e foram para a fila  $j$ . Então pode-se escrever o seguinte:

$$x_i^n(t) \triangleq \frac{1}{\sqrt{n}} X^n(nt) = \frac{1}{\sqrt{n}} \left[ N_i^n(nt) - S_i^n(nt) + \sum_{j=1}^K S_{ji}^n(nt) \right], \quad (2.21)$$

onde  $X_i^n(t)$  representa o número de clientes na fila  $i$ , e  $x_i^n$  é o processo escalonado correspondente.

Procedendo da mesma forma como feito na seção anterior, pode-se escrever o seguinte:

$$\begin{aligned}\frac{N_i^n(nt)}{\sqrt{n}} &= W_i^{a,n} \left( \frac{N_i^n(nt)}{n} \right) + \frac{nt}{\sqrt{n}\bar{\Delta}_i^{a,n}} + \varepsilon^{a,n}(t) \\ \frac{S_i^n(nt)}{\sqrt{n}} &= W_i^{d,n} \left( \frac{S_i^n(nt)}{n} \right) + \frac{nt - I_i^n(nt)}{\sqrt{n}\bar{\Delta}_i^{d,n}} + \varepsilon^{d,n}(t)\end{aligned}$$

onde  $W_i^{\alpha,n}$  são definidos de forma análoga à Equação (2.18), e  $I_i^n(t)$  denota o tempo total que o servidor da fila  $i$  ficou ocioso até o tempo  $t$ . Defina  $W_{ij}^{r,n}(t) \triangleq (1/\sqrt{n}) \sum_{l=1}^{nt} [\mathbb{I}_{ij,l}^n - r_{ij}^n]$  e observe que

$$\begin{aligned}\frac{S_{ji}^n(nt)}{\sqrt{n}} &= \frac{1}{\sqrt{n}} \sum_{l=1}^{S_j^n(nt)} \mathbb{I}_{ji,l}^n = W_{ij}^{r,n} \left( \frac{S_j^n(nt)}{n} \right) + \frac{r_{ji}^n}{\sqrt{n}} \sum_{l=1}^{S_j^n(nt)} 1 \\ &= W_{ji}^{r,n} \left( \frac{S_j^n(nt)}{n} \right) + r_{ji}^n W_j^{d,n} \left( \frac{S_j^n(nt)}{n} \right) + \frac{r_{ji}^n}{\sqrt{n}\bar{\Delta}_j^{d,n}} (nt - I_j^n(nt)) + \varepsilon^{r,n}(t),\end{aligned}\tag{2.22}$$

Com as Condições 2.7.1(a) e (b), e o Teorema B.1.9, tem-se que

$$\begin{aligned}W_i^{a,n} \left( \frac{N_i^n(n\cdot)}{n} \right) &\Longrightarrow \sigma_i^a W_i^a(\lambda_i^a \cdot) \\ W_i^{d,n} \left( \frac{S_i^n(n\cdot)}{n} \right) &\Longrightarrow \sigma_i^d W_i^d(\lambda_i^d \cdot) \\ \left\{ W_{ij}^{r,n} \left( \frac{S_i^n(n\cdot)}{n} \right), j \leq K \right\} &\Longrightarrow W_i^r(\lambda_i^d \cdot),\end{aligned}$$

onde  $W_i^r(\lambda_i^d \cdot)$  é um processo de Wiener  $K$ -dimensional com matriz de covariância dada por

$$\Sigma_i = \begin{pmatrix} (1 - r_{i1})r_{i1} & -r_{i1}r_{i2} & \dots & -r_{i1}r_{iK} \\ -r_{i2}r_{i1} & (1 - r_{i2})r_{i2} & \dots & -r_{i2}r_{iK} \\ \vdots & & \ddots & \\ -r_{iK}r_{i1} & -r_{iK}r_{i2} & \dots & (1 - r_{iK})r_{iK} \end{pmatrix},$$

os processos  $W_i^\alpha$ ,  $\alpha \in \{a, d\}$ , são processos de Wiener padrão, e

$$(\sigma_i^\alpha)^2 \triangleq \lim \text{var}(\Delta_{i,l}^{\alpha,n}) / (\bar{\Delta}_i^{\alpha,n})^2.$$

Defina  $y_i^n(t) \triangleq (1/\sqrt{n}\bar{\Delta}_i^{d,n})I_i^n(nt)$ , e observe que pode-se escrever o seguinte em notação matricial:

$$x^n(t) = c^n t + W^n(t) + (I - R') y^n(t) + \varepsilon^n(t),$$

onde  $W^n(t)$  tem componentes dadas por  $W_i^n(t) \triangleq W_i^{a,n}(N_i^n(nt)/n) - W_i^{d,n}(S_i^n(nt)/n) + \sum_j (r_{ji} W_j^{d,n}(S_j^n(nt)/n) + W_{ji}^{r,n}(S_j^n(nt)/n))$  e  $\varepsilon^n$  é um erro pequeno que converge para o processo zero. Observe que  $y_i^n$  é não-decrescente, contínuo, e somente aumenta nos instantes  $t$  tais que  $x_i(t) = 0$ . Por isso, é possível mostrar que  $(x^n, z^n) = \Psi(\psi^n)$ , onde  $\Psi$  é o mapeamento de reflexão da Definição A.1.2, com  $z^n \triangleq (I - R') y^n$  e  $\psi^n(t) \triangleq c^n t + W^n(t) + \varepsilon^n(t)$ .

Como o mapeamento  $\Psi$  é contínuo sob o espaço das funções contínuas  $C(\mathbb{R}^K; 0, \infty)$ , tem-se o resultado com a aplicação do Teorema do Mapeamento Contínuo, chegando ao limite:

$$x(t) = ct + W(t) + (I - R') y(t),$$

onde  $W_i(t) \triangleq \sigma_i^a W_i^a(\lambda_i^a t) - \sigma_i^d W_i^d(\lambda_i^d t) + \sum_j (r_{ji} \sigma_j^d W_j^d(\lambda_j^d t) + W_{ji}^r(\lambda_j^d t))$ .

Em problemas práticos, usa-se a aproximação da seguinte forma: em primeiro lugar, um valor grande de  $n$  é escolhido de tal forma que  $c_i^n$  possua um “tamanho moderado”, em seguida, aproxima-se o número de clientes na fila em cada instante  $t$  por  $X(t) \approx \sqrt{n} x_{c_n}(t/n)$ , onde  $x_{c_n}(t) \triangleq c_n t + W(t) + (I - R') y(t)$ .

Para exemplificar o uso da aproximação, considere um sistema composto por

3 filas em rede, com matriz de roteamento dada por

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ .5 & 0 & 0 \end{pmatrix}.$$

Suponha que somente a primeira fila receba clientes de uma fonte exógena com tempo entre a chegada de clientes dada por uma distribuição de hiperexponencial com variação 2 e média 2. O tempo de serviço de cada fila é distribuído de acordo com uma distribuição hiperexponencial com médias tomando valores no conjunto  $\{0.80, 0.85, 0.90, 0.95\}$  para cada experimento, e o coeficiente de variação é fixo em 2. Estes valores são escolhidos de tal forma que  $\rho_1 = \rho_2 = \rho_3$ .

Tabela 2.1: Tabela mostrando o número médio de clientes na rede para cada fila em regime estacionário. Para a simulação, o intervalo de 95% t-confiância está em parênteses.

	Fila 1	Fila 2	Fila 3
$\rho = .80$			
QNET	5.84	6.06	6.13
SIM	6.40 (0.21)	6.70 (0.20)	6.81 (0.21)
$\rho = .85$			
QNET	8.79	9.13	9.23
SIM	9.52 (0.46)	9.69 (0.34)	10.12 (0.42)
$\rho = .90$			
QNET	14.79	15.35	15.52
SIM	15.63 (0.97)	15.83 (0.96)	15.51 (0.82)
$\rho = .95$			
QNET	32.96	34.20	34.58
SIM	29.36 (2.69)	29.41 (2.18)	30.19 (2.45)

Será comparado o número médio de clientes no sistema em regime estacionário obtido usando a aproximação por tráfego pesado e uma simulação computacional. Para calcular o valor esperado de um movimento Browniano refletido em regime estacionário (em dimensão  $K = 3$ ), será usado o algoritmo QNET (Dai, 1990) que é capaz de calcular a distribuição estacionária destes processos (o método BNAfm também pode ser usado para o mesmo propósito, veja (Shen et al.,

2002; Chen e Shen, 2003)). O resultado é dado pela Tabela 2.1.

Como neste exemplo somente o número médio de clientes do sistema em regime estacionário é computado, a escolha do parâmetro de escalonamento  $n$  não afeta o resultado. Por exemplo, no caso unidimensional o valor médio do movimento Browniano refletido em regime estacionário é dado por  $\sigma^2/2|c|$ , onde  $\sigma^2$  seria sua variância e  $c$  o valor do coeficiente de deriva, veja a Equação (2.17). Como o valor do coeficiente de deriva na aproximação, dado por  $c^n$ , é uma constante vezes  $\sqrt{n}$ , isto é,  $c^n = \sqrt{n}\beta$ , para  $\beta \in \mathbb{R}$ , tem-se que o número médio de clientes em regime estacionário, dado por  $\bar{X}$ , é igual à  $\bar{X} = \sqrt{n}\sigma^2/2|c^n| = \sigma^2/2|\beta|$ , em que o fator de escalonamento  $n$  desaparece.

## 2.8 Aproximação por Fluido

Um outro modelo aproximado que possui várias similaridades com a aproximação por tráfego pesado é a de fluido. A principal vantagem deste modelo é que não é necessário impor a condição do tráfego pesado. Portanto, mesmo em sistemas que não estejam operando neste regime a aproximação ainda pode ser usada. Por outro lado, a aproximação perde o caráter estocástico, pois o limite toma a forma de uma equação diferencial. Desta forma, pode-se argumentar que a aproximação por tráfego pesado é mais fiel à dinâmica do sistema.

Para deixar claro como esta aproximação toma forma, será mostrado brevemente como ela é obtida. Sejam  $\Delta_{i,l}^{a,n}$ ,  $\Delta_{i,l}^{d,n}$ ,  $X^n$ ,  $N_i^n$ ,  $S_i^n$ ,  $S_{ij}^n$ , e  $\mathbb{I}_{ij}^n$  definidos como na última seção. Mas redefina o processo escalonado  $x^n$  como o seguinte:

$$x^n(t) = X^n(nt)/n,$$

observe que desta vez  $X^n$  é dividido por  $1/n$  e não por  $1/\sqrt{n}$  como feito na seção anterior. Seja também  $X^n(0)$  a condição inicial do sistema.

**Condição 2.8.1.** *Suponha as seguintes condições:*

(a) Os processos  $\tilde{W}_i^{\alpha,n}$ ,  $\alpha \in \{a, d\}$ , dados por  $\tilde{W}_i^{\alpha,n}(t) \triangleq n^{-1} \sum_{l=1}^{nt} (\bar{\Delta}_i^{a,n} - \Delta_{i,l}^{\alpha,n})$

convergem para o processo zero. Além disso,  $\bar{\Delta}_i^{\alpha,n} \rightarrow \bar{\Delta}_i^\alpha$ .

(b) Existem constantes  $r_{ij}^n$  e  $r_{ij}$  tais que  $\mathbb{P}(\mathbb{I}_{ij,l}^n = 1) = r_{ij}^n \rightarrow r_{ij}$ , e o raio espectral da matriz  $R = (r_{ij})$  é menor do que um.

(c) Existem  $c_i$  tais que

$$\left[ \frac{1}{\bar{\Delta}_i^{a,n}} + \sum_{j=1}^K \frac{r_{ji}^n}{\bar{\Delta}_j^{d,n}} - \frac{1}{\bar{\Delta}_i^{d,n}} \right] \triangleq c_i^n \rightarrow c_i.$$

Suponha que  $x^n(0) \triangleq X^n(0)/n$  converge fracamente para  $x(0)$ . Visto o desenvolvimento da aproximação por tráfego pesado da seção anterior, e sabendo que neste caso  $\tilde{W}_i^{\alpha,n}$  converge para o processo zero, tem-se que  $x^n$  converge fracamente para o processo  $x$  dado por

$$x(t) = x(0) + ct + (I - R')y(t),$$

em que  $y$  é o processo de reflexão. O desenvolvimento acima foi retirado de (Kushner, 2001) Seção 6.5, na página 264.

# Capítulo 3

## Modelos de Fila com Sinais

Sistemas de filas com sinais, usualmente chamados de *G-networks* ou *G-Queue*, no caso de uma fila única, foram introduzidos por Gelenbe em (Gelenbe, 1991) como uma extensão das redes de Jackson. Nestes modelos, dois tipos de clientes circulam no sistema: os clientes usuais (também chamados de “clientes positivos”) e os sinais. O efeito que cada sinal provoca sobre o sistema que o recebe depende do seu tipo. Existem vários tipos de sinais, como por exemplo: clientes negativos, *triggers* (Gelenbe, 1993), desastres (Chao, 1995), *resets* (Gelenbe e Fourneau, 2002), entre outros. Os dois principais tipos de sinais são os clientes negativos, que serão chamados simplesmente de sinais neste trabalho, e os *triggers*. Um cliente negativo remove um cliente usual da fila que o recebe e deixa o sistema imediatamente após sua chegada. O *trigger* move um cliente da fila que o recebe para uma outra fila presente no sistema, e também deixa o sistema após sua chegada.

Cada sinal pode entrar em uma fila por uma fonte externa ou pode ser enviado por outras filas presentes no sistema. Neste último caso, sinais são clientes usuais que, após completar seu serviço, são roteados como sinais. Para deixar claro o funcionamento destes modelos, considere um sistema de filas com sinais dado por duas filas em sequência, como mostra a Figura 3.1. Suponha que a primeira fila receba sinais (do tipo “cliente negativo”) e que também envie sinais para a segunda fila. Quando um sinal externo chega à fila 1, ele remove um cliente usual da fila,

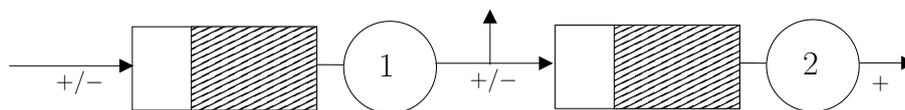


Figura 3.1: Duas filas em sequência com sinais. O símbolo + (resp., -) indica a possibilidade de chegada de clientes (resp., sinais).

caso não esteja vazia. O sinal deixa a fila 1 imediatamente após sua chegada. Suponha agora que um cliente (usual) termine seu serviço na primeira fila. Este cliente pode ser enviado para fora do sistema, como mostra a Figura 3.1, ou pode ser roteado para a segunda fila. Caso seja roteado, o cliente pode chegar na fila 2 como um sinal ou como um cliente regular, dependendo da escolha de roteamento. Se chegar como um sinal, ele removerá um cliente, caso algum estiver presente, e deixará o sistema logo em seguida. Desta forma, em uma rede mais complexa, onde todas as filas participam no envio e recebimento de sinais, cada fila possui um certo controle sobre o sistema.

Sistemas de filas com sinais vêm sendo extensivamente estudados nos últimos anos (alguns exemplos incluem: (Shin, 2007; Chakka e Do, 2007; Gómez-Corral e Martos, 2006; Li e Zhao, 2004; Harrison, 2004; Gómez-Corral, 2002; Fourneau et al., 1996; Harrison e Pitel, 1996; Fourneau e Verchère, 1995; Gelenbe e Schassberger, 1992)), que são motivados principalmente por uma série de aplicações. Uma das aplicações com maior sucesso, e que também foi a motivação inicial destes modelos, é a modelagem de redes neurais (Gelenbe, 1994; Gelenbe e Stafylopatis, 1991; Gelenbe, 1989). Neste contexto, cada fila representa um neurônio e clientes positivos e negativos representam sinais de excitação e inibição, respectivamente. Uma outra aplicação interessante é a de balanceamento de carga de trabalho em uma rede de computadores. O artigo (Artalejo, 2000) traz várias outras aplicações destas redes. Mais recentemente, *G-networks* vem sendo usadas em modelos de sistemas de regulação genético, veja por exemplo (Gelenbe, 2007; Arazi et al., 2004).

A formulação usual destes modelos considera tempos entre chegada de clientes e tempos de serviço exponencialmente distribuídos. A análise é geralmente

feita como na Seção 2.7, isto é, a distribuição estacionária da cadeia de Markov correspondente é encontrada. Esta distribuição descreve a probabilidade de encontrar um certo número de clientes no sistema. No caso de sistemas contendo somente sinais do tipo “clientes negativos”, tem-se que esta probabilidade é dada por

$$p(X_1 = n_1, \dots, X_K = n_K) = \prod_{i=1}^K \rho_i^{n_i} (1 - \rho_i),$$

para  $(n_1, \dots, n_k) \in \mathbb{N}^K$ , e

$$\begin{aligned} \rho_i &\triangleq \lambda_i^{a,+} / (\lambda_i^d + \lambda_i^{a,-}) \\ \lambda_i^{a,+} &\triangleq \sum_{j=1}^K \rho_j \lambda_i^d r_{ji}^+ + \lambda_i^a \\ \lambda_i^{a,-} &\triangleq \sum_{j=1}^K \rho_j \lambda_i^d r_{ji}^- + \lambda_i^s, \end{aligned} \tag{3.1}$$

onde  $\lambda_i^a$  é a taxa de chegada de clientes exógenos na fila  $i$ ;  $\lambda_i^s$  é a taxa de chegada de sinais exógenos na fila  $i$ ;  $\lambda_i^d$  é a taxa de serviço para a fila  $i$ ;  $r_{ij}^+$  é a probabilidade de um cliente saindo da fila  $j$  entrar na fila  $i$ ;  $r_{ij}^-$  é a probabilidade de um cliente saindo da fila  $j$  entrar na fila  $i$  como um sinal. É claro que a existência de tal distribuição depende de algumas condições, sendo uma delas a de que  $\rho_i < 1$ , que nada mais é do que o parâmetro de intensidade de tráfego destes sistemas.

Um dos motivos da popularidade das *G-networks* deve-se também ao fato de sua distribuição estacionária possuir esta forma simples de produto, que não é uma característica fácil de ser encontrada. Apesar desta propriedade importante, pouco se pode dizer sobre a evolução transiente destes sistemas, e pouco deste aspecto foi estudado sobre estes modelos. Esta análise é de extrema importância para ser possível pensar em problemas de controle, que é tão inerente nestes modelos.

Recentemente, alguns estudos nesta direção foram feitos usando aproximações por fluido (Guffens et al., 2006; Arazi et al., 2004, 2005). Em (Arazi et al., 2004, 2005), a seguinte aproximação para *G-networks* foi desenvolvida: sejam

$\{q_{j1}^+(X), \dots, q_{jR}^+(X)\}$  taxas dependentes do estado  $X$  de processos que aumentam o tamanho da fila  $j$  (i.e., chegadas de clientes exógenos e chegada de clientes vindos de outras filas); e sejam  $\{q_{j1}^-(X), \dots, q_{jQ}^-(X)\}$  taxas dependentes do estado representando processos que reduzem o tamanho da fila  $j$  (i.e., saída de clientes e chegada de clientes negativos). Usando a notação dos autores, a probabilidade de aumentar em 1 o número de clientes em uma fila  $j$  no intervalo  $(t, t + h]$  é dada por

$$P_j^+(X) = \sum_{i=1}^R q_{ji}^+(X)h + o(h),$$

e a probabilidade de reduzir o valor de  $X$  em 1 é dada por

$$P_j^-(X) = \sum_{i=1}^Q q_{ji}^-(X)h + o(h).$$

Como isto, aproxima-se o valor médio do sistema, dado por  $\hat{X}$ , fazendo o seguinte:

$$\begin{aligned} \hat{X}_j(t+h) &= \hat{X}_j(t) + P_j^+(\hat{X}(t)) - P_j^-(\hat{X}(t)) \\ &= \hat{X}_j(t) + h \left[ \sum_{i=1}^R q_{ji}^+(\hat{X}(t)) - \sum_{i=1}^Q q_{ji}^-(\hat{X}(t)) \right] + o(h), \end{aligned}$$

para cada fila  $j$ . Dividindo a equação acima por  $h$  e fazendo  $h \rightarrow 0$ , tem-se que

$$\frac{d\hat{X}_j}{dt} = \sum_{i=1}^R q_{ji}^+(\hat{X}(t)) - \sum_{i=1}^Q q_{ji}^-(\hat{X}(t)).$$

Neste ponto, os autores usam uma hipótese de que o sistema está cheio e, portanto, a entrada ou saída de um único cliente na fila tem pouco efeito no sistema. Desta forma, um fator de escalonamento  $\Delta$  é introduzido para representar o efeito da entrada ou saída de um único cliente no sistema, onde  $\Delta \ll 1$ . Esta abordagem leva à aproximação

$$\frac{d\hat{X}_j}{dt} = \Delta \left[ \sum_{i=1}^R q_{ji}^+(\hat{X}(t)) - \sum_{i=1}^Q q_{ji}^-(\hat{X}(t)) \right].$$

Uma crítica sobre este modelo é que o parâmetro  $\Delta$  é introduzido sem muita justificativa. Além disso, o requerimento de que  $\Delta$  seja “pequeno” força o modelo em um cenário de tráfego pesado, onde as aproximações por difusão são mais apropriadas. Uma outra observação é que nenhum termo de reflexão aparece, e, desta forma, o modelo pode tomar valores negativos se a escolha das funções  $q_{ji}^-(\cdot)$  não forem apropriadas.

Os autores de (Guffens et al., 2006) têm uma abordagem diferente que consiste em uma adaptação de uma aproximação por fluido, introduzida por (Tipper e Sundareshan, 1990) e estudada por Guffens em (Guffens, 2005), para modelos de filas com sinais. Para introduzir este modelo, considere primeiramente uma única fila  $M/M/1$ , com taxa de entrada dada por  $\lambda^a$  e de serviço dada por  $\lambda^d$ . A probabilidade de encontrar  $k$  clientes no sistema no instante  $t + h$  é dada por

$$\begin{aligned} p(k, t + h) &= p(k, t)(1 - \lambda^a h - \lambda^d h) \\ &\quad + p(k - 1, t)\lambda^a h \mathbb{I}\{k > 0\} + p(k + 1, t)\lambda^d h + o(h). \end{aligned}$$

Multiplicando por  $k$  a equação acima, somando em  $k$ , dividindo por  $h$ , e fazendo  $h \rightarrow 0$ , chega-se à:

$$\frac{d\bar{k}}{dt} = \lambda^a - (1 - p(0, t))\lambda^d,$$

em que  $\bar{k}(t) \triangleq \sum_{k=0}^{\infty} kp(k, t)$ , e  $1 - p(0, t)$  é a probabilidade do servidor estar ocupado no instante  $t$ , que não pode ser escrita somente em função de  $\bar{k}$ . A aproximação consiste em expressar a probabilidade do servidor estar ocupado em função do estado atual do sistema, de tal forma que, em equilíbrio, o valor exato de  $\bar{k}$  é obtido. Seja  $x$  a média aproximada dada por

$$\dot{x} = \lambda^a - \theta(x)\lambda^d,$$

em que  $\theta(\xi) \triangleq \xi/(1 + \xi)$ . Desta forma, quando  $x$  está em equilíbrio, dado por  $x^*$ ,

tem-se que

$$x^* = \lim_{t \rightarrow \infty} \bar{k}(t) = \rho/(1 - \rho), \quad (3.2)$$

com  $\rho = \lambda^a/\lambda^d$ , veja a Equação (2.3).

Esta mesma análise é estendida para *G-networks*. O modelo considerado pelos autores inclui os sinais do tipo “trigger” e “cliente negativo”. Para simplificar a notação, somente os clientes negativos são considerados abaixo. Analisando o balanço de fluxo de clientes, pode-se escrever:

$$\dot{x}_i = - \left( \lambda_i^d + \lambda_i^s + \sum_{j=1}^K \lambda_j^d \theta(x_j) r_{ji}^- \right) \theta(x_i) + \lambda_i^a + \sum_{j=1}^K \lambda_j^d \theta(x_j) r_{ji}^+,$$

onde a notação empregada aqui é a mesma usada nas Equações (3.1). Fazendo  $\lambda_i^{a,-}(x) \triangleq \lambda_i^s + \sum_{j=1}^K \lambda_j^d \theta(x_j) r_{ji}^-$  e  $\lambda_i^{a,+}(x) \triangleq \lambda_i^a + \sum_{j=1}^K \lambda_j^d \theta(x_j) r_{ji}^+$ , pode-se escrever

$$\dot{x}_i = - (\lambda_i^d + \lambda_i^{a,-}(x)) \theta(x_i) + \lambda_i^{a,+}(x).$$

O equivalente do que foi feito na Equação (3.2) pode ser verificado para este caso. Portanto, chega-se a aproximação:

$$\dot{x} = G(x)x + \lambda^a,$$

onde  $(G)_{ii}(x) \triangleq - (\lambda_i^d + \lambda_i^{a,-}(x)) \tilde{\theta}(x_i)$ ,  $(G)_{ij}(x) \triangleq \lambda^d r_{ji}^+ \tilde{\theta}(x_j)$ , para  $i \neq j$ , e  $\tilde{\theta}(\xi) \triangleq 1/(1 + \xi)$ , para  $\xi \in \mathbb{R}$ .

Este modelo é bastante interessante, contudo, não se sabe se o comportamento transiente do sistema será de fato bem representado. Isto é, a introdução da função  $\theta$  não garante o comportamento adequado do modelo quando o sistema ainda está longe de atingir o “regime estacionário”. A introdução de dependência do estado para as taxas dos processos de entrada e saída para este modelo não foi explorada pelos autores.

Nos capítulos que seguem serão mostrados aproximações por tráfego pesado

para estas filas com sinais que foram desenvolvidas durante este trabalho de tese.

# Capítulo 4

## Filas com Sinais Sob Tráfego Pesado

Como primeiro passo no roteiro proposto, será tratado a aproximação por tráfego pesado para sistemas compostos por uma única fila e um servidor. Nenhuma condição sobre a distribuição do tempo de serviço ou do tempo entre chegadas de clientes será assumida, além de algumas hipóteses gerais necessárias para a convergência fraca de processos que serão definidos. O conteúdo deste capítulo é referente aos trabalhos (Leite e Fragoso, 2007, 2008d).

Antes de iniciar a descrição do sistema para a obtenção da aproximação, algumas considerações relacionadas às filas com sinais são necessárias. Primeiramente, é preciso deixar claro se o sistema permitirá que clientes sendo servidos sejam removidos por sinais. Caso seja permitido, as discrepâncias geradas no tempo de serviço destes clientes precisam ser computadas e utilizadas no modelo. Para ilustrar essa afirmação, considere a seguinte linha de eventos dada pela Figura 4.1. No instante  $t_k$  o  $k$ -ésimo cliente entra no sistema que está vazio, e portanto, o servidor atende este cliente imediatamente. No instante  $\tau$ , um sinal é recebido pelo sistema e o cliente é removido imediatamente. Em seguida, em  $t_{k+1}$  um outro cliente entra no sistema e começa a ser servido. O modelo terá que ser ajustado para refletir o instante correto em que o cliente  $k + 1$  inicia o seu serviço no servidor. Ou seja, será necessário computar  $t_k + \Delta_k - t_{k+1}$  que é o “tempo residual” e usá-lo de forma adequada no modelo. Contudo, o cômputo desta quantidade e a sua adaptação no modelo sem hipóteses mais restritivas nas distribuições de tempo de serviço e

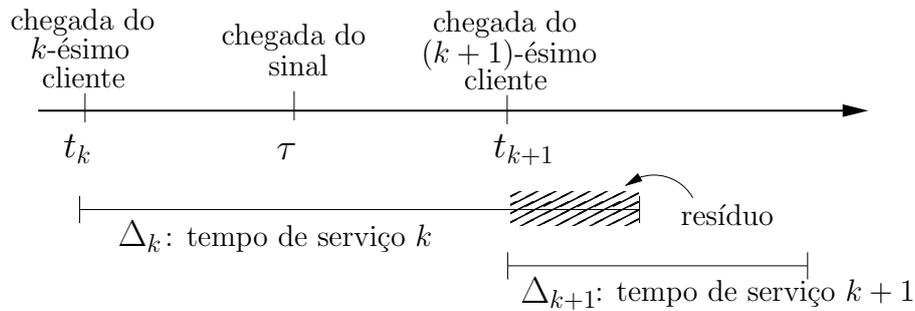


Figura 4.1: Uma linha de eventos ilustrando a discrepância gerada quando um cliente é removido em serviço.

entrada não são problemas fáceis. Desta forma, este trabalho se restringe à sistemas onde sinais não podem remover clientes que estão em serviço. Este tipo de rotina para remoção de clientes é chamada de *immune servicing* (e.g., (Chakka e Harrison, 2001)).

Uma outra consideração importante a ser feita, e a que de fato tem maior importância devido à dificuldade que este problema gera para o tratamento do caso de redes de filas com sinais, é referente a representação do processo de contagem a partir dos tempos entre chegadas consecutivas de sinais no sistema. A primeira vista, uma fila com sinais pode parecer igual a uma fila com dois servidores, já que existem dois processos que “removem” clientes do sistema. Em um sistema com dois servidores, o processo que conta o número de clientes servidos é descrito a partir do tempo em que cada cliente permanece no servidor. Por outro lado, em um sistema com sinais, o processo que conta o número de clientes removidos terá que ser descrito a partir do tempo entre chegada de sinais. Contudo, é importante observar que nem sempre a distribuição do tempo entre clientes removidos por um sinal é igual a distribuição do tempo entre chegadas de sinais. Além disso, a relação entre estas duas quantidades não é clara para o caso de distribuições gerais como é o caso tratado aqui.

Para ilustração, considere a seguinte linha de eventos dada pela Figura 4.2. Suponha que  $t_k$ ,  $t_{k+1}$ , e  $t_{k+2}$  são os instantes de chegada de sinais em um sistema,  $\Delta_{k+1}$  e  $\Delta_{k+2}$  são os intervalos entre chegada destes sinais. Considere o caso onde o

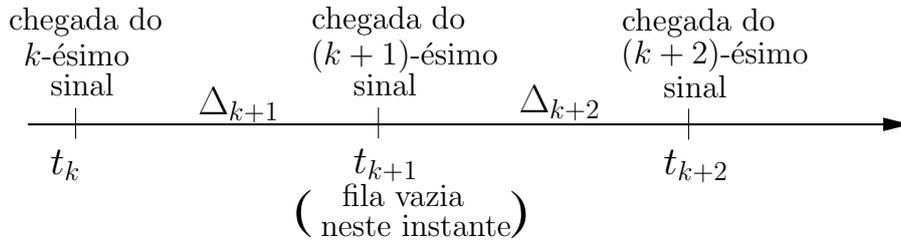


Figura 4.2: Uma linha de eventos ilustrando a diferença entre o tempo entre clientes removidos e chegada de sinais.

sinal  $k$  e  $k+2$  removem clientes do sistema, mas o sinal  $k+1$  não remove devido à falta de clientes naquele instante. Portanto, neste caso, o tempo entre os clientes removidos é  $\Delta_{k+1} + \Delta_{k+2}$ . Em um outro caso onde todos os sinais removem clientes, o tempo entre clientes removidos coincide com o tempo entre chegada de sinais.

Nesta seção serão tratados dois modelos diferentes de filas com sinais. Primeiramente, será considerado o caso onde a chegada de sinais é independente do processo de chegada de clientes. A aproximação de tráfego pesado será desenvolvida para o processo do número de clientes na fila e para o processo de carga de trabalho do servidor. Em seguida, será tratado um modelo onde a chegada de sinais e de clientes é modulada por uma mesma cadeia de Markov. Para este modelo, somente o processo do número de clientes será aproximado por uma difusão.

#### 4.1 Modelo para o Número de Clientes no Servidor

Como é comum em aproximações de tráfego pesado, assume-se a existência de uma sequência de filas indexadas por um parâmetro  $n$ . À medida que este parâmetro aumenta, a diferença entre a taxa de clientes entrando e deixando o sistema fica menor, tendendo para o que é chamado de “situação de tráfego pesado”. Para cada  $n > 1$ , seja  $\Delta_l^{a,n}$  uma variável aleatória representando o  $l$ -ésimo intervalo de tempo entre chegada de clientes. Seja também  $\Delta_l^{r,n}$  uma variável aleatória representando o intervalo de tempo entre chegada de sinais. O tempo de serviço para o  $l$ -ésimo cliente a ser servido é dado por  $\Delta_l^{d,n}$  para cada  $n > 0$ . Suponha que estas variáveis aleatórias satisfazem a seguinte condição.

**Condição 4.1.1.** *As variáveis aleatórias  $\Delta_l^{\alpha,n}$  são mutuamente independentes para cada  $l$  e  $n$  e são identicamente distribuídas com média dada por  $\bar{\Delta}^{\alpha,n}$  satisfazendo  $0 < \bar{\Delta}^{\alpha,n} < \infty$ , onde  $\alpha \in \{a, r, d\}$ . Seja  $\sigma^{\alpha,n}$  o coeficiente de variação de  $\Delta_l^{\alpha,n}$ . Existem constantes  $\sigma^\alpha$  e  $\bar{\Delta}^\alpha$  tais que  $\sigma^{\alpha,n} \rightarrow \sigma^\alpha$  e  $\bar{\Delta}^{\alpha,n} \rightarrow \bar{\Delta}^\alpha$ , onde  $\alpha \in \{a, r, d\}$ . Além disso,  $\{\Delta_l^{a,n}\}$ ,  $\{\Delta_l^{r,n}\}$ , e  $\{\Delta_l^{d,n}\}$  são independentes e  $\{|\Delta^{a,n}|^2, |\Delta^{d,n}|^2, |\Delta^{r,n}|^2; n\}$  é uniformemente integrável.*

A condição sobre a integrabilidade uniforme pode ser substituída pela condição de que  $\Delta_l^{\alpha,n}$  converge fracamente para uma variável aleatória com média  $\bar{\Delta}^\alpha$  e coeficiente de variação  $\sigma^\alpha$  (veja Teorema 5.4 em (Billingsley, 1968)).

**Condição 4.1.2.** *Existe uma constante  $b \in \mathbb{R}$  tal que*

$$\sqrt{n} \left( \frac{1}{\bar{\Delta}^{a,n}} - \frac{1}{\bar{\Delta}^{r,n}} - \frac{1}{\bar{\Delta}^{d,n}} \right) \triangleq b_n \rightarrow b.$$

A condição acima é chamada de “condição do tráfego pesado”. Esta condição exige que a taxa de clientes entrando no sistema seja cada vez mais próxima da taxa em que clientes deixam o sistema. Defina  $\lambda^{\alpha,n} \triangleq 1/\bar{\Delta}^{\alpha,n}$ , para  $\alpha \in \{a, r, d\}$ .

Seja  $S^{a,n}(t)$  (resp.,  $S^{r,n}(t)$ )  $1/n$  vezes o número de clientes (resp., sinais) que chegaram no sistema até o instante  $nt$ . Seja também  $S^{d,n}(t)$  definido como  $1/n$  vezes o número de clientes servidos até o tempo  $nt$ . Estes processos podem ser definidos como:

$$\begin{aligned} S^{a,n}(t) &\triangleq \frac{1}{n} \max \left\{ m \in \mathbb{N}_0 : \sum_{l=1}^m \Delta_l^{a,n} \leq nt \right\}, \\ S^{r,n}(t) &\triangleq \frac{1}{n} \max \left\{ m \in \mathbb{N}_0 : \sum_{l=1}^m \Delta_l^{r,n} \leq nt \right\}, \\ S^{d,n}(t) &\triangleq \frac{1}{n} \max \left\{ m \in \mathbb{N}_0 : \sum_{l=1}^m \Delta_l^{d,n} \leq nt - T^n(t) \right\}, \end{aligned} \quad (4.1)$$

onde  $T^n(t)$  é definido como o tempo ocioso total do servidor até o instante  $nt$ .

Observe que o processo  $x^n(t)$  que representa  $1/\sqrt{n}$  vezes o número de clientes

presentes no sistema no instante  $nt$  pode ser escrito como:

$$x^n(t) = x^n(0) + \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} 1 - \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{r,n}(t)} 1 - \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{d,n}(t)} 1 + z_0^n(t), \quad (4.2)$$

onde  $x^n(0)$  é  $1/\sqrt{n}$  vezes o número de clientes no sistema no instante inicial e  $z_0^n(t)$  representa  $1/\sqrt{n}$  vezes o número total de sinais que chegaram no sistema quando a fila estava vazia até o tempo  $nt$ . Note que  $z_0^n$  somente aumenta nos instantes  $t$  tais que  $x^n(t) \leq 1/\sqrt{n}$ .

**Teorema 4.1.3.** *Suponha que  $x^n(0)$  converge fracamente para  $x(0)$  e que  $x^n(0)$  é independente das variáveis aleatórias representando o tempo entre chegada de sinais e clientes e do tempo de serviço. Além disso suponha as Condições 4.1.1 e 4.1.2. Então o processo  $x^n$  converge fracamente para o processo  $x$  que toma valores*

$$x(t) = x(0) + w^a(\lambda^a t) - w^r(\lambda^r t) - w^d(\lambda^d t) + bt + z(t), \quad (4.3)$$

onde  $w^a(\cdot)$ ,  $w^r(\cdot)$  e  $w^d(\cdot)$  são processos de Wiener independentes com variâncias dada por  $(\sigma^a)^2$ ,  $(\sigma^r)^2$ , e  $(\sigma^d)^2$ , respectivamente. A constante  $b$  é a definida na Condição 4.1.2, e  $z(\cdot)$  é o processo de reflexão, isto é,  $z(0) = 0$ ,  $z(\cdot)$  é não-decrescente e somente aumenta nos instantes  $t$  tais que  $x^n(t) = 0$ .

*Demonstração.* A prova deste teorema segue as linhas do Teorema 5.1.1 em (Kushner, 2001). Observe que

$$\frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{\alpha,n}(t)} 1 = \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{\alpha,n}(t)} \left(1 - \frac{\Delta_l^{\alpha,n}}{\bar{\Delta}^{\alpha,n}}\right) + \frac{1}{\bar{\Delta}^{\alpha,n}\sqrt{n}} \sum_{l=1}^{nS^{\alpha,n}(t)} \Delta_l^{\alpha,n}, \quad (4.4)$$

para  $\alpha = a, r, d$ . Defina  $w^{\alpha,n}$  como

$$w^{\alpha,n}(t) \triangleq \frac{1}{\sqrt{n}} \sum_{l=1}^{nt} \left(1 - \frac{\Delta_l^{\alpha,n}}{\bar{\Delta}^{\alpha,n}}\right). \quad (4.5)$$

Com a aplicação do Teorema B.1.9, que é uma extensão do Teorema de Donsker, é possível verificar que  $w^{\alpha,n}$  converge fracamente para um processo de Wiener com

variância dada por

$$\lim_n \mathbb{E} \left[ \left( 1 - \frac{\Delta_l^{\alpha,n}}{\bar{\Delta}^{\alpha,n}} \right)^2 \right] = (\sigma^\alpha)^2.$$

Usando o Teorema B.2.6 e a Equação (4.1), observa-se que  $S^{a,n}$  converge fracamente para um processo  $S^a$  que toma valores  $S^a(t) = \lambda^a t$ . O mesmo é válido para  $S^{r,n}$ . De forma análoga, usando a Equação (4.1) e a primeira parte do Teorema B.2.6,  $S^{d,n}(t)$  é limitado em probabilidade e qualquer subsequência convergente tem como limite um processo que possui (quase certamente) funções amostras contínuas. Portanto, já que o processo de Wiener tem funções amostras quase certamente contínuas, o processo  $w^{\alpha,n}(S^{\alpha,n}(\cdot))$  converge fracamente para  $w^\alpha(\lambda^\alpha \cdot)$ , onde  $\alpha \in \{a, r\}$ , e  $w^{d,n}(S^{d,n}(\cdot))$  é assintoticamente contínuo.

Observe que para  $\alpha \in \{a, r\}$ ,

$$\frac{1}{\bar{\Delta}^{\alpha,n} \sqrt{n}} \sum_{l=1}^{nS^{\alpha,n}(t)} \Delta_l^{\alpha,n} = \frac{nt}{\bar{\Delta}^{\alpha,n} \sqrt{n}} + \epsilon_\alpha^n(t),$$

onde  $\epsilon_\alpha^n(t)$  denota  $1/(\bar{\Delta}^{\alpha,n} \sqrt{n})$  vezes o tempo desde a chegada do último cliente, que será negligenciável a medida que  $n \rightarrow \infty$ . Para  $\alpha = d$ , é preciso levar em consideração o tempo ocioso do servidor, portanto,

$$\frac{1}{\bar{\Delta}^{d,n} \sqrt{n}} \sum_{l=1}^{nS^{d,n}(t)} \Delta_l^{d,n} = \frac{nt - T^n(t)}{\bar{\Delta}^{d,n} \sqrt{n}} + \epsilon_d^n(t)$$

onde  $\epsilon_d^n(t)$  é um erro negligenciável.

Usando a expressão (4.2), é possível escrever o seguinte:

$$\begin{aligned} x^n(t) &= x^n(0) + w^{a,n}(S^{a,n}(t)) - w^{r,n}(S^{r,n}(t)) - w^{d,n}(S^{d,n}(t)) \\ &\quad + \sqrt{n} \left( \frac{1}{\bar{\Delta}^{a,n}} - \frac{1}{\bar{\Delta}^{r,n}} - \frac{1}{\bar{\Delta}^{d,n}} \right) t + z^n(t) + \epsilon^n(t), \end{aligned}$$

onde  $z^n(t) = z_0^n(t) + T^n(t)/(\sqrt{n} \bar{\Delta}^{d,n})$ , e  $\epsilon^n(t)$  é a soma dos termos de erros negligenciáveis.

O fato que  $\{z^n\}$  é limitada em probabilidade e que converge fracamente para o termo de reflexão segue da aplicação do Teorema A.2.3, já que  $z^n(0) = 0$ ,  $z_0^n(t)$  é não-decrescente, possui saltos de tamanhos  $1/\sqrt{n}$ , e  $z^n(t)$  somente aumenta nos instantes  $t$  tais que  $x^n(t) \leq 1/\sqrt{n}$ . Isso implica que  $\{T^n/\sqrt{n}\}$  é limitada em probabilidade, e portanto  $S^{d,n}$  converge fracamente para o processo que toma valores  $\lambda^d t$ , pelo Teorema B.2.6.  $\square$

É importante observar que o limite dado pela Equação (4.3) é igual ao limite obtido através da análise via tráfego pesado uma fila com dois servidores independentes (veja Teorema 5.4.2 (Kushner, 2001)) com tempo de permanência no servidor dado por  $\Delta_l^{d,n}$  e  $\Delta_l^{r,n}$ , apesar dos sistemas serem diferentes como foi discutido no início do capítulo. Para esse modelo de uma fila simples, essa diferença não fica aparente e é “escondida” no termo de reflexão.

## 4.2 Modelo para o Processo de Carga de Trabalho

A carga de trabalho em um dado instante  $t$  é definida como o tempo total que o servidor precisa trabalhar para atender todos os clientes presentes na fila naquele instante. Seguindo a forma usual de escalonamento, seja  $wl^n(t)$  definido como  $1/\sqrt{n}$  vezes a carga de trabalho do sistema no instante  $nt$ .

**Teorema 4.2.1.** *Considere o modelo usado na seção anterior e assuma as condições do Teorema 4.1.3. Defina  $wl^n(0) \triangleq x^n(0)\bar{\Delta}^{d,n}$ . Então, a diferença  $\bar{\Delta}^{d,n}x^n - wl^n$  converge fracamente para o processo zero.*

*Demonstração.* Esta prova segue as idéias apresentadas no Teorema 5.3.3 de (Kushner, 2001). Seja  $I^n(t)$  definido como o número de clientes que deixaram o sistema até o tempo  $nt$ . Observe que  $I^n(t)$  é limitado por  $n(S^{d,n}(t) + S^{r,n}(t))$ . Além disso, os processos  $S^{d,n}$  e  $S^{r,n}$  são limitados com alta probabilidade em qualquer intervalo  $[0, T]$ , pelo Teorema B.2.6. Reorganize os índices  $l$  de  $\Delta_l^{d,n}$  de forma que sejam ordenados na ordem em que os clientes deixam o sistema. Desta forma, para cada

$t > 0$ , tem-se que

$$\frac{1}{\sqrt{n}} \sum_{l=I^n(t)+2}^{I^n(t)+\sqrt{n}x^n(t)} \Delta_l^{d,n} \leq \tilde{w}l^n(t) \leq \frac{1}{\sqrt{n}} \sum_{l=I^n(t)+1}^{I^n(t)+\sqrt{n}x^n(t)} \Delta_l^{d,n}. \quad (4.6)$$

Observe que

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{l=I^n(t)+1}^{I^n(t)+\sqrt{n}x^n(t)} \Delta_l^{d,n} &= \frac{1}{\sqrt{n}} \left[ \sum_{l=I^n(t)+1}^{I^n(t)+\sqrt{n}x^n(t)} (\Delta_l^{d,n} - \bar{\Delta}^{d,n}) \right] + \bar{\Delta}^{d,n} x^n(t). \\ &= \left[ -\bar{\Delta}^{d,n} w^{d,n}(I^n(t)/n + x^n(t)/\sqrt{n}) + \Delta^{d,n} w^{d,n}(I^n(t)/n) \right] \\ &\quad + \bar{\Delta}^{d,n} x^n(t). \end{aligned} \quad (4.7)$$

Como  $\{x^n\}$  é limitada em probabilidade,  $\{x^n/\sqrt{n}\}$  converge para o processo zero. Já que  $S^{d,n} + S^{r,n}$  são limitados com alta probabilidade em qualquer intervalo  $[0, T]$ , sabe-se que  $I^n/n$  também é limitado com alta probabilidade em  $[0, T]$ . Desta forma, como o termo em chaves do lado direito da Equação (4.7) converge fracamente para o processo zero, já que  $\{w^{d,n}\}$  é assintoticamente contínua. O mesmo pode ser feito para o lado esquerdo da Inequação (4.6).  $\square$

Considere agora um modelo onde o efeito de um sinal sob o sistema é diferente do que foi feito anteriormente. Ao invés de remover um cliente da fila, um sinal remove uma quantidade fixa de trabalho denotada por  $\bar{W}$ . Seja  $\tilde{w}l^n(t)$  definido como  $1/\sqrt{n}$  vezes a carga de trabalho no instante  $nt$  para esse novo sistema. De forma análoga ao que foi feito na Seção 4.1, é possível escrever o seguinte:

$$\tilde{w}l^n(t) = \tilde{w}l^n(0) + \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} \Delta_l^{d,n} - \sqrt{nt} - \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{r,n}(t)} \bar{W} + z^n(t), \quad (4.8)$$

onde  $\tilde{w}l^n(0)$  é a quantidade escalonada de carga de trabalho inicial,  $z^n(t) \triangleq z_0^n(t) + T^n(t)/\sqrt{n}$ , e o processo  $T^n(t)$  é definido como o total de tempo ocioso do servidor até o instante  $nt$ . O processo  $z_0^n(\cdot)$  é adicionado para manter  $\tilde{w}l^n(\cdot)$  positivo. Esse processo irá aumentar nos instantes em que um sinal chegar no sistema e encontrar uma carga de trabalho inferior à  $\bar{W}/\sqrt{n}$ .

**Condição 4.2.2.** *Existe uma constante  $b \in \mathbb{R}$  tal que:*

$$\sqrt{n}\bar{\Delta}^{d,n} \left( \frac{1}{\bar{\Delta}^{a,n}} - \frac{1}{\bar{\Delta}^{d,n}} - \frac{\bar{W}}{\bar{\Delta}^{r,n}\bar{\Delta}^{d,n}} \right) \triangleq b_n \rightarrow b \in \mathbb{R} \quad (4.9)$$

**Teorema 4.2.3.** *Suponha que  $\tilde{w}^n(0)$  converge fracamente para  $\tilde{w}l(0)$  e é independente dos processos de chegada e serviço. Assuma as Condições 4.1.1 e 4.2.2. Então o processo  $\tilde{w}l^n(\cdot)$  converge fracamente para o processo  $wl(\cdot)$ , definido como*

$$\tilde{w}l(t) = \tilde{w}l(0) + \bar{\Delta}^d w^a(\lambda^a t) - \bar{W} w^r(\lambda^r t) - \bar{\Delta}^d w^d(\lambda^d t) + bt + z(t) \quad (4.10)$$

onde  $w^a(\cdot)$ ,  $w^r(\cdot)$  e  $w^d(\cdot)$  são processos de Wiener independentes com variâncias dadas por  $(\sigma^a)^2$ ,  $(\sigma^r)^2$ , e  $(\sigma^d)^2$ , respectivamente. A constante  $b$  é a definida na Condição 4.2.2, e  $z(\cdot)$  é o processo de reflexão.

*Demonstração.* Observe que é possível expandir a primeira soma da Equação (4.8) como:

$$\frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} \Delta_l^{d,n} = \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}} \left( \Delta_l^{d,n} - \bar{\Delta}^{d,n} \right) + \frac{\bar{\Delta}^{d,n}}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} 1. \quad (4.11)$$

Note que o primeiro termo da equação acima é  $-\bar{\Delta}^{d,n} w^{d,n}(S^{a,n}(t))$ , onde  $w^{\alpha,n}(\cdot)$  é definido como na Equação (4.5), para  $\alpha \in \{a, r, d\}$ . A segunda soma da Equação (4.8) e o último termo da Equação (4.11) podem ser expandido como feito no Teorema 4.1.3. Portanto,

$$\begin{aligned} \tilde{w}l^n(t) &= \tilde{w}l^n(0) + \bar{\Delta}^{d,n} [w^{a,n}(S^{a,n}(t)) - w^{d,n}(S^{a,n}(t))] - \bar{W} w^{r,n}(S^{r,n}(t)) \\ &\quad + b_n t + z^n(t) + \epsilon^n(t). \end{aligned}$$

O fato que  $w^{\alpha,n}(\cdot)$ , para  $\alpha \in \{a, r\}$ , converge fracamente para um processo de Wiener segue as linhas do que foi feito no Teorema 4.1.3. O mesmo pode ser dito sobre a convergência fraca de  $S^{\alpha,n}(\cdot)$ ,  $\alpha \in \{a, r\}$ , para um processo que toma valores  $\lambda^{\alpha} t$ . A prova de que  $\{z^n\}$  é limitada em probabilidade e que converge para o

processo de reflexão segue com a aplicação do Teorema A.2.3 e o fato que o processo é não-decrescente, aumenta somente nos instantes em que  $\tilde{w}l^n(t) < \bar{W}/\sqrt{n}$ , e  $z_0^n(t)$  tem saltos com tamanho limitado por  $\bar{W}/\sqrt{n}$ .  $\square$

Observe que se  $\bar{W} = \bar{\Delta}^d$  e supondo que  $\tilde{w}l^n(0) = wl^n(0)$ , para todo  $n > 1$ , a equação limite do primeiro modelo (dado pelo Teorema 4.2.1) e a do modelo do Teorema 4.2.3 somente diferem no termo de escalonamento do tempo do processo de Wiener  $w^d(\cdot)$ .

### 4.3 Modelo com Chegadas Moduladas por um Processo de Markov

Nesta seção será considerado um modelo de filas com sinais, onde a entrada de clientes e sinais é modulada por uma cadeia de Markov. O modelo considerado aqui pode ser visto como uma extensão do modelo apresentado em (Gelenbe et al., 1991). Filas com sinais moduladas por uma cadeia de Markov também foram estudadas mais recentemente em (Chakka e Do, 2007; Shin e Choi, 2003; Chakka e Harrison, 2001) através de métodos analíticos matriciais (veja a Seção 2.4). Estes modelos são motivados pelo fato de que o fluxo de tráfego pela Internet é geralmente correlacionado e apresenta um comportamento que varia no tempo, algumas vezes a intensidade de chegada de clientes é alta e outras vezes baixa. Este tipo de comportamento do processo de entrada é geralmente modelado através de uma cadeia de Markov que modula a entrada. Ou seja, a taxa de entrada de clientes varia de acordo com o estado da cadeia de Markov.

Para o modelo que será tratado aqui, suponha que existe apenas um fluxo de entrada que alimenta o sistema com clientes e sinais. A decisão se uma entrada será um sinal ou um cliente depende do estado de uma cadeia de Markov. Esta cadeia também interfere na distribuição de  $\Delta_l^{a,n}$ . Da mesma forma que foi feito na Seção 4.1, um sinal não poderá remover um cliente que esteja em serviço.

Seja  $\mathbb{I}_l^{+,n}$  (resp.,  $\mathbb{I}_l^{-,n}$ ) uma função indicadora do evento que a  $l$ -ésima entrada é um cliente (resp., um sinal). Seja  $\{M_l^n; l\}$  um processo de Markov irredutível, estacionário, homogêneo no tempo, e tomando valores no espaço de estados  $\mathcal{S}$ , para

cada  $n > 1$ . Suponha que a probabilidade condicional de  $p$ -passos desta cadeia de Markov converge geometricamente para sua distribuição estacionária que é dada por  $(\pi_k^n, k \in \mathcal{S})$ . Ou seja, assume-se que existe uma constante  $C_n$  e um  $\epsilon_n \in (0, 1)$  tal que

$$\sum_k \left| \sum_z \mathbb{P}(M_p^n = k | M_0^n = z) \mu(z) - \pi_k^n \right| \leq C_n (1 - \epsilon_n)^p,$$

para qualquer distribuição inicial  $\mu$ . Além disso, assume-se que  $\pi_k^n \rightarrow \pi_k$  para cada  $k \in \mathcal{S}$  quando  $n \rightarrow \infty$ . Uma cadeia de Markov tem probabilidades de transição que convergem geometricamente quando, por exemplo, ela satisfaz a condição de Doeblin (e.g., Capítulo 3 de (Stroock, 2005)). Para garantir tal convergência, basta assumir que  $\mathcal{S}$  é finito e a cadeia é aperiódica.

Assuma que os processos  $\{M_l^n; l\}$ ,  $\{\Delta_l^{a,n}; l\}$ , e  $\{\mathbb{I}_l^{\alpha,n}; l, \alpha = +, -\}$  são independentes de  $\{\Delta_l^{d,n}; l\}$ . Além disso, as variáveis aleatórias  $\{\Delta_l^{a,n}, \mathbb{I}_l^{\alpha,n}; l, \alpha = +, -\}$  são mutuamente independentes quando condicionadas ao estado da cadeia.

Seja  $w^{d,n}$  definido como na Equação (4.5). Suponha que  $w^{d,n}$  converge fracamente para um processo de Wiener com variâncias  $\sigma_d^2$ , onde  $\bar{\Delta}^{d,n}$  é uma constante tal que  $\bar{\Delta}^{d,n} \rightarrow \bar{\Delta}^d \in \mathbb{R}_{>0}$ . Esta condição é atendida se, por exemplo, as variáveis aleatórias  $\Delta_l^{d,n}$  satisfazem a Condição 4.1.1.

Defina  $\mathcal{F}_l^{a,n}$  como a mínima  $\sigma$ -álgebra que mensura todos os processos definidos acima até o instante da chegada da  $l$ -ésima entrada, não incluindo a variável aleatória  $\Delta_l^{a,n}$  e  $\mathbb{I}_l^{\alpha,n}$ , para  $\alpha \in \{+, -\}$ . Suponha que a distribuição de  $\Delta_l^{a,n}$  dado  $\mathcal{F}_l^{a,n}$  e o evento  $\{M_l^n = k\}$  não dependa de  $l$  ou de  $\mathcal{F}_l^{a,n}$  e convirja fracamente quando  $n \rightarrow \infty$ . Além disso, seja

$$\begin{aligned} \bar{\Delta}^{a,n}(k) &\triangleq \mathbb{E}[\Delta_l^{a,n} | \mathcal{F}_l^{a,n}, M_l^n = k] \\ &= \mathbb{E}[\Delta_l^{a,n} | M_l^n = k] \rightarrow \bar{\Delta}^a(k) \in \mathbb{R}_{>0}, \\ v^{a,n}(k) &\triangleq \mathbb{E}[(\Delta_l^{a,n})^2 | \mathcal{F}_l^{a,n}, M_l^n = k] \\ &= \mathbb{E}[(\Delta_l^{a,n})^2 | M_l^n = k] \rightarrow v^a(k) \in \mathbb{R}_{>0}, \end{aligned}$$

quando  $n \rightarrow \infty$  para cada valor de  $k$ , e

$$\sup_n \sup_{k \in \mathcal{S}} |v^{a,n}(k)| < \infty \quad \sup_n \sup_{k \in \mathcal{S}} |\bar{\Delta}^{a,n}(k)| < \infty.$$

Suponha ainda que  $\{|\Delta_l^{a,n}|^2; n, l\}$  seja uniformemente integrável.

Assuma que para cada valor de  $k \in \mathcal{S}$ , existam constantes  $q^\alpha(k) \in \mathbb{R}_{>0}$  tais que

$$\mathbb{E} [\mathbb{I}_l^{\alpha,n} | \mathcal{F}_l^{a,n}, M_l^n = k] = \mathbb{E} [\mathbb{I}_l^{\alpha,n} | M_l^n = k] = q^{\alpha,n}(k) \rightarrow q^\alpha(k) \quad (4.12)$$

para  $\alpha \in \{+, -\}$ , quando  $n \rightarrow \infty$ . Note que  $(q^{+,n}(k) + q^{-,n}(k))$  e  $(q^+(k) + q^-(k))$  podem tomar valores menores do que 1. Isto é feito para permitir problemas onde existe a possibilidade de uma entrada não chegar no sistema quando a cadeia de Markov está no estado  $k$ .

Será usado também a seguinte condição de tráfego pesado:

$$\sqrt{n} \left( \frac{\bar{q}^{+,n} - \bar{q}^{-,n}}{\bar{\Delta}^{a,n}} - \frac{1}{\bar{\Delta}^{d,n}} \right) \triangleq b_n \rightarrow b \in \mathbb{R}. \quad (4.13)$$

De forma similar ao que é feito nas Equações (4.2) e (4.8), é possível escrever o processo  $x^n(t)$  como

$$x^n(t) = x^n(0) + \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} (\mathbb{I}_l^{+,n} - \mathbb{I}_l^{-,n}) - \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{d,n}(t)} 1 + z_0^n(t), \quad (4.14)$$

onde  $z_0^n(t)$  representa  $1/\sqrt{n}$  vezes o número total de sinais que chegaram quando a fila estava vazia até o instante  $nt$ . Defina as seguintes constantes

$$\begin{aligned} \bar{q}^{\alpha,n} &\triangleq \sum_k q^{\alpha,n}(k) \pi_k^n, & \bar{\Delta}^{a,n} &\triangleq \sum_k \bar{\Delta}^{a,n}(k) \pi_k^n, \\ h_a^{\alpha,n} &\triangleq \sum_k \bar{\Delta}^{a,n}(k) q^{\alpha,n}(k) \pi_k^n, & \bar{v}^{a,n} &\triangleq \sum_k v^{a,n}(k) \pi_k^n, \end{aligned}$$

e denote por  $\bar{q}^\alpha$ ,  $\bar{\Delta}^a$ ,  $h_a^\alpha$ , e  $\bar{v}^a$  seus respectivos limites quando  $n \rightarrow \infty$ . Além disso

defina as matrizes

$$\begin{aligned}\Sigma_0 &\triangleq \begin{pmatrix} \bar{q}^+ - (\bar{q}^+)^2 & -\bar{q}^+\bar{q}^- & h_a^+ - \bar{q}^+\bar{\Delta}^a \\ -\bar{q}^+\bar{q}^- & \bar{q}^- - (\bar{q}^-)^2 & h_a^- - \bar{q}^-\bar{\Delta}^a \\ h_a^+ - \bar{q}^+\bar{\Delta}^a & h_a^- - \bar{q}^-\bar{\Delta}^a & \bar{v}^a - (\bar{\Delta}^a)^2 \end{pmatrix}, \\ \Sigma_1 &\triangleq \lim_n \sum_{u=1}^{\infty} \sum_{k,m} \zeta^n(k) \zeta^n(m)' D^{n,k,m}(u),\end{aligned}\quad (4.15)$$

onde  $D^{n,k,m}(u) \triangleq [\mathbb{P}(M_u^n = m | M_0^n = k) - \pi_m^n] \pi_k^n$ , e  $\zeta^n(k) \triangleq (q^{+,n}(k), q^{-,n}(k), \bar{\Delta}^{a,n}(k))$ .

Observe que todas as hipóteses acima têm um caráter bem geral e permitem vários cenários diferentes. Além disso, elas são padrão quando se trata de análise de tráfego pesado, veja por exemplo o Capítulo 5 de (Kushner, 2001).

**Teorema 4.3.1.** *Assuma que  $x^n(0)$  converge fracamente para  $x(0)$  e que é independente do processo de chegada e de serviço. Com as hipóteses listadas acima, o processo  $x^n(\cdot)$  converge fracamente para o processo  $x(\cdot)$  que toma valores*

$$x(t) = x(0) - \lambda^a(\bar{q}^+ - \bar{q}^-)w^a(\lambda^a t) + w^+(\lambda^a t) - w^-(\lambda^a t) - w^d(\lambda^d t) + bt + z(t)\quad (4.16)$$

onde  $w^d$  é um processo de Wiener com variância  $(\sigma^d)^2$ ,  $\tilde{w}^a \triangleq (w^+, w^-, w^a)$  é um processo de Wiener com matriz de covariância  $\Sigma \triangleq \Sigma_0 + 2\Sigma_1$ ,  $b$  é a constante definida na Equação (4.13), e  $z$  é o processo de reflexão.

*Demonstração.* Considere o seguinte

$$\begin{aligned}\frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} (\mathbb{I}_l^{+,n} - \mathbb{I}_l^{-,n}) &= \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} (\mathbb{I}_l^{+,n} - \bar{q}^{+,n}) - \frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} (\mathbb{I}_l^{-,n} - \bar{q}^{-,n}) \\ &\quad + \frac{(\bar{q}^{+,n} - \bar{q}^{-,n})}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} 1.\end{aligned}\quad (4.17)$$

Defina, para  $\alpha \in \{+, -\}$ ,

$$w^{\alpha,n}(t) \triangleq \frac{1}{\sqrt{n}} \sum_{l=1}^{nt} (\mathbb{I}_l^{\alpha,n} - \bar{q}^{\alpha,n}), \quad \text{e} \quad w^{a,n}(t) \triangleq \frac{1}{\sqrt{n}} \sum_{l=1}^{nt} (\Delta_l^{a,n} - \bar{\Delta}^{a,n}).$$

Observe que a última soma do lado direito da Equação (4.17) pode ser expandida como foi feito na Equação (4.4), e também a última soma da Equação (4.14). Portanto, tem-se que

$$x^n(t) = x^n(0) - \frac{(\bar{q}^{+,n} - \bar{q}^{-,n})}{\bar{\Delta}^{a,n}} w^{a,n}(S^{a,n}(t)) + w^{+,n}(S^{a,n}(t)) - w^{-,n}(S^{a,n}(t)) \\ - w^{d,n}(S^{d,n}(t)) + \sqrt{n} \left( \frac{\bar{q}^{+,n} - \bar{q}^{-,n}}{\bar{\Delta}^{a,n}} - \frac{1}{\bar{\Delta}^{d,n}} \right) + z^n(t) + \epsilon^n(t),$$

onde  $z^n(t) = z_0^n(t) + T^n(t)/(\sqrt{n}\bar{\Delta}^{d,n})$ , e  $\epsilon^n(t)$  é um erro negligenciável. A maior dificuldade nesta prova é mostrar que  $\tilde{w}^{a,n}(\cdot) \triangleq (w^{+,n}(\cdot), w^{-,n}(\cdot), w^{a,n}(\cdot))$  converge para um processo de Wiener com matriz de covariância  $\Sigma$ .

Para mostrar esta convergência, será aplicado o Teorema B.1.10, da mesma forma como foi feito no Teorema 5.5.1 de (Kushner, 2001). A primeira condição do Teorema B.1.10 é atendida pela integrabilidade uniforme de  $\Delta_l^{a,n}$ . Para mostrar a segunda condição, seja

$$\xi_l^n \triangleq \begin{pmatrix} \mathbb{I}_l^{+,n} - \bar{q}^{+,n} \\ \mathbb{I}_l^{-,n} - \bar{q}^{-,n} \\ \Delta_l^{a,n} - \bar{\Delta}^{a,n} \end{pmatrix}.$$

Observe que

$$\mathbb{E}[\xi_l^n | \mathcal{F}_i^{a,n}] = \begin{pmatrix} \mathbb{E}[\mathbb{I}_l^{+,n} | M_i^n] - \bar{q}^{+,n} \\ \mathbb{E}[\mathbb{I}_l^{-,n} | M_i^n] - \bar{q}^{-,n} \\ \mathbb{E}[\Delta_l^{a,n} | M_i^n] - \bar{\Delta}^{a,n} \end{pmatrix} = \begin{pmatrix} \sum_k q^{+,n}(k) \mathbb{P}(M_l^n = k | M_i^n) - \bar{q}^{+,n} \\ \sum_k q^{-,n}(k) \mathbb{P}(M_l^n = k | M_i^n) - \bar{q}^{-,n} \\ \sum_k \bar{\Delta}^a(k) \mathbb{P}(M_l^n = k | M_i^n) - \bar{\Delta}^{a,n} \end{pmatrix}$$

As probabilidades condicionais  $\mathbb{P}(M_l^n = k | M_i^n)$  convergem geometricamente para  $\pi_k^n$  quando  $l - i \rightarrow \infty$ . Portanto a soma

$$\sum_{l=i+1}^{nT} \mathbb{E}[\xi_l^n | \mathcal{F}_i^{a,n}]$$

é limitada por uma constante uniformemente em  $T, i, n$ , e a segunda condição é

satisfeita. Para calcular  $\mathbb{E} [\xi_l^n(\xi_l^n)' | \mathcal{F}_i^{a,n}]$ , considere o seguinte:

$$\begin{aligned} \mathbb{E} [(\mathbb{I}_l^{\alpha,n} - \bar{q}^{\alpha,n})^2 | \mathcal{F}_i^{a,n}] &= \mathbb{E} [\mathbb{I}_l^{\alpha,n} | M_i^n] - 2\bar{q}^{\alpha,n}\mathbb{E} [\mathbb{I}_l^{\alpha,n} | M_i^n] + (\bar{q}^{\alpha,n})^2 \\ &= \sum_k q^{\alpha,n}(k)\mathbb{P}(M_l^n = k | M_i^n) \\ &\quad - 2\bar{q}^{\alpha,n} \sum_k q^{\alpha,n}(k)\mathbb{P}(M_l^n = k | M_i^n) + (\bar{q}^{\alpha,n})^2, \end{aligned}$$

$$\begin{aligned} \mathbb{E} [(\Delta_l^{a,n} - \bar{\Delta}^{a,n})^2 | \mathcal{F}_i^{a,n}] &= \mathbb{E} [(\Delta_l^{a,n})^2 | M_i^n] - 2\bar{\Delta}^{a,n}\mathbb{E} [\Delta_l^{a,n} | M_i^n] + (\bar{\Delta}^{a,n})^2 \\ &= \sum_k v^{a,n}(k)\mathbb{P}(M_l^n = k | M_i^n) \\ &\quad - 2\bar{\Delta}^{a,n} \sum_k \bar{\Delta}^{a,n}(k)\mathbb{P}(M_l^n = k | M_i^n) + (\bar{\Delta}^{a,n})^2, \end{aligned}$$

$$\begin{aligned} \mathbb{E} [(\mathbb{I}_l^{+,n} - \bar{q}^{+,n})(\mathbb{I}_l^{-,n} - \bar{q}^{-,n}) | \mathcal{F}_i^{a,n}] &= -\bar{q}^{+,n} \sum_k q^{-,n}(k)\mathbb{P}(M_l^n = k | M_i^n) \\ &\quad - \bar{q}^{-,n}(k) \sum_k q^{+,n}(k) \times \mathbb{P}(M_l^n = k | M_i^n) \\ &\quad + \bar{q}^{+,n}\bar{q}^{-,n}, \end{aligned}$$

$$\begin{aligned} \mathbb{E} [(\Delta_l^{a,n} - \bar{\Delta}^{a,n})(\mathbb{I}_l^{\alpha,n} - \bar{q}^{\alpha,n}) | \mathcal{F}_i^{a,n}] &= \sum_k \bar{\Delta}^{a,n}(k)q^{\alpha,n}(k)\mathbb{P}(M_l^n = k | M_i^n) \\ &\quad - \bar{\Delta}^{a,n} \sum_k q^{\alpha,n}(k)\mathbb{P}(M_l^n = k | M_i^n) \\ &\quad - \bar{q}^{\alpha,n} \sum_k \bar{\Delta}^{a,n}(k)\mathbb{P}(M_l^n = k | M_i^n) \\ &\quad + \bar{q}^{\alpha,n}\bar{\Delta}^{a,n}. \end{aligned}$$

Como as probabilidades de transição convergem geometricamente, tem-se que para

qualquer  $N_n$  que atende as condições do Teorema B.1.10:

$$\frac{1}{N_n} \sum_{l=i+1}^{i+N_n} \mathbb{E} [\xi_l^n (\xi_l^n)' | \mathcal{F}_i^{\alpha,n}] \rightarrow \Sigma_0, \quad (\text{as } n, i \rightarrow \infty)$$

onde  $\Sigma_0$  é definido como na Equação (4.15). Agora será mostrado que  $\sum_{u=l+1}^p \mathbb{E} [\xi_l^n (\xi_u^n)']$  converge quando  $p-l \rightarrow \infty$  e  $n, l \rightarrow \infty$ . Portanto, considere o seguinte

$$\mathbb{E} [(\mathbb{I}_l^{\alpha,n} - \bar{q}^{\alpha,n})(\mathbb{I}_u^{\alpha,n} - \bar{q}^{\alpha,n})] = \mathbb{E} [\mathbb{I}_l^{\alpha,n} \mathbb{I}_u^{\alpha,n}] - \bar{q}^{\alpha,n} (\mathbb{E} [\mathbb{I}_l^{\alpha,n}] + \mathbb{E} [\mathbb{I}_u^{\alpha,n}]) + (\bar{q}^{\alpha,n})^2. \quad (4.18)$$

Pode-se escrever o primeiro termo à direita da Equação (4.18) como

$$\sum_{k,m} \mathbb{E} [\mathbb{I}_l^{\alpha,n} | M_l^n = k] \mathbb{E} [\mathbb{I}_u^{\alpha,n} | M_u^n = m] \mathbb{P} (M_l^n = k, M_u^n = m).$$

Esta passagem é possível pela independência condicional dos termos dado o estado da cadeia de Markov. Desta forma a Equação (4.18) pode ser escrita como

$$\begin{aligned} & \mathbb{E} [(\mathbb{I}_l^{\alpha,n} - \bar{q}^{\alpha,n})(\mathbb{I}_l^{\alpha,n} - \bar{q}^{\alpha,n})] \\ &= \sum_{k,m} q^{\alpha,n}(k) q^{\alpha,n}(m) [\mathbb{P} (M_l^n = k, M_u^n = m) - \pi_k^n \mathbb{P} (M_u^n = m) - \pi_m^n \mathbb{P} (M_l^n = k) + \pi_m^n \pi_k^n] \\ &= \sum_{k,m} q^{\alpha,n}(k) q^{\alpha,n}(m) D^{n,k,m}(u-l), \end{aligned}$$

onde  $D^{n,k,m}(u) \triangleq [\mathbb{P} (M_u^n = m | M_0^n = k) - \pi_m^n] \pi_k^n$ , e a estacionariedade da cadeia de Markov foi usada. Os mesmos argumentos podem ser usados para mostrar que

$$\begin{aligned} \mathbb{E} [(\mathbb{I}_l^{+,n} - \bar{q}^{+,n})(\mathbb{I}_u^{-,n} - \bar{q}^{-,n})] &= \sum_{k,m} q^{+,n}(k) q^{-,n}(m) D^{n,k,m}(u-l), \\ \mathbb{E} [(\Delta_l^{a,n} - \bar{\Delta}^{a,n})(\mathbb{I}_u^{\alpha,n} - \bar{q}^{\alpha,n})] &= \sum_{k,m} \bar{\Delta}^{a,n}(k) q^{\alpha,n}(m) D^{n,k,m}(u-l), \\ \mathbb{E} [(\Delta_l^{a,n} - \bar{\Delta}^{a,n})(\Delta_u^{a,n} - \bar{\Delta}^{a,n})] &= \sum_{k,m} \bar{\Delta}^{a,n}(k) \bar{\Delta}^{a,n}(m) D^{n,k,m}(u-l). \end{aligned}$$

Já que a probabilidade de transição converge geometricamente, a soma a seguir é

finita para cada  $n$

$$\sum_{u=1}^{\infty} \sum_{k,m} \zeta^n(k) \zeta^n(m)' D^{n,k,m}(u) \triangleq \Sigma_1^n,$$

onde, como definido previamente,  $\zeta^n(k) \triangleq (q^{+,n}(k), q^{-,n}(k), \bar{\Delta}^{a,n}(k))$ . A última condição pode ser verificada usando os mesmo argumentos acima. O fato que  $\{S^{a,n}\}$ ,  $\{S^{d,n}\}$  e  $\{z^n\}$  são limitados em probabilidade e o resto da prova segue com os mesmos argumentos usados no Teorema 4.1.3.  $\square$

A condição do tráfego pesado, dada pela Equação (4.13), será estudada em relação ao resultado apresentado em (Gelenbe et al., 1991), onde foi estabelecido condições de estabilidade para filas com sinais moduladas por uma cadeia de Markov. Colocado na notação utilizada aqui, o cenário tratado em (Gelenbe et al., 1991) foi o seguinte:  $\{M_l^n; l\}$  é uma cadeia de Markov definida sob o espaço de estados  $\mathcal{S} = \{0, 1\}$ . Esta cadeia é idêntica para cada  $n$ , portando não será carregado este índice na notação. A matriz de transição desta cadeia é dada por

$$\begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

onde  $p \in (0, 1)$  e  $q = 1 - p$ . A decisão de que uma entrada será um cliente ou um sinal é baseada no estado da cadeia de Markov. Se o estado for 1 então a entrada será um cliente, caso contrário será um sinal, ou seja,

$$q^+(0) = 0, \quad q^+(1) = 1, \quad q^-(0) = 1, \quad q^-(1) = 0,$$

onde  $q^+(k)$  e  $q^-(k)$  são definidos na Equação (4.12). A distribuição estacionária desta cadeia pode ser facilmente computada e é dada por  $\pi = (q/(1+q), 1/(1+q))$ . Portanto, tem-se que  $\bar{q}^+ = 1/(1+q)$  e  $\bar{q}^- = q/(1+q)$ . O tempo entre chegadas consecutivas são independentes da cadeia de Markov, portanto  $\bar{\Delta}^{a,n}(k)$  é somente o tempo médio entre chegadas para qualquer  $k$ .

O resultado em (Gelenbe et al., 1991) diz que  $\{X^n\}$ , onde  $X^n(t)$  é o número de clientes no sistema no instante  $t$ , é um processo regenerativo recorrente positivo se e somente se

$$\frac{1}{\bar{\Delta}^{a,n}} < \frac{1}{\bar{\Delta}^{d,n}} \left( \frac{1+q}{1-q} \right).$$

A condição dada pela Equação (4.13) diz que  $\bar{\Delta}^{d,n}(\bar{q}^+ - \bar{q}^-)/\bar{\Delta}^{a,n} \uparrow 1$ , ou para esse caso específico

$$\frac{\bar{\Delta}^{d,n}}{\bar{\Delta}^{a,n}} \left( \frac{1-q}{1+q} \right) \uparrow 1.$$

Portanto, quando  $n \rightarrow \infty$  o sistema aproxima o limite em estabilidade, o que caracteriza um cenário de tráfego pesado.

Agora considere o caso onde nenhuma entrada é perdida no momento de chegada. Isto é, existem somente duas possibilidades quando uma entrada chega no sistema: a entrada será um cliente ou um sinal, não há a possibilidade de perda desta entrada. Neste caso,  $\mathbb{I}_l^{+,n} + \mathbb{I}_l^{-,n} = 1$  para cada  $l$ , o que simplifica o modelo consideravelmente já que pode-se reescrever a Equação (4.17) como

$$\frac{1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} (\mathbb{I}_l^{+,n} - \mathbb{I}_l^{-,n}) = \frac{2}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} (\mathbb{I}_l^{+,n} - \bar{q}^{+,n}) + \frac{2\bar{q}^{+,n} - 1}{\sqrt{n}} \sum_{l=1}^{nS^{a,n}(t)} 1.$$

Usando esta expansão, a mesma análise gera o seguinte limite

$$x(t) = x(0) - \lambda^a(2\bar{q}^+ - 1)w^a(\lambda^a t) + 2w^+(\lambda^a t) - w^d(\lambda^d t) + bt + z(t),$$

onde o processo de Wiener  $(w^+(\cdot), w^a(\cdot))$  tem matriz de covariância dada por  $\Sigma$ , que é definida como antes, mas com

$$\Sigma_0 \triangleq \begin{pmatrix} \bar{q}^+ - (\bar{q}^+)^2 & h_a^+ - \bar{q}^+ \bar{\Delta}^a \\ h_a^+ - \bar{q}^+ \bar{\Delta}^a & \bar{v}^a - (\bar{\Delta}^a)^2 \end{pmatrix},$$

e  $\zeta^n(k) \triangleq (q^{+,n}(k), \bar{\Delta}^{a,n}(k))$ .

#### 4.4 Experimentos Numéricos

Em aplicações práticas, as aproximações de tráfego pesado são usadas da seguinte forma. Primeiro, escolhe-se um  $n \in \mathbb{N}$  relativamente “grande” e aproxima-se o número de clientes (ou a carga de trabalho) no instante  $t$ , denotado por  $X(t)$ , usando  $X(nt) \sim \sqrt{n}x(t; b_n)$ , onde  $x(t; b_n)$  é a expressão do limite para o processo de interesse (isto é, Equações (4.3), (4.10), ou (4.16)) com o termo de deriva  $b$  substituído por  $b_n$ .

Com o efeito meramente ilustrativo, será calculado o número médio de clientes e de carga de trabalho sobre o regime estacionário em dois cenários diferentes usando a aproximação por tráfego pesado e uma simulação computacional para diferentes intensidades de tráfego  $\rho$ . Observe que o valor esperado do movimento Browniano refletido,  $x(t) = \sigma w(t) + bt + z(t)$ , em regime estacionário é dado por  $\sigma^2/2|b|$  (veja a Equação (2.17)).

Considere primeiro os modelos que não são modulados por uma cadeia de Markov. Como  $\rho \triangleq \lambda^a/(\lambda^r + \lambda^d)$ , seja  $\lambda^a = \rho$  e  $(\lambda^r + \lambda^d) = 1$  para os exemplos considerados aqui. As distribuições do tempo entre chegadas consecutivas de clientes, sinais e de tempo de serviço são hipereexponenciais. Para o primeiro problema, o quadrado dos coeficientes de variação são fixados em  $((\sigma^a)^2, (\sigma^r)^2, (\sigma^d)^2) = (1, 1.5, 2)$ . Para o segundo problema, os seguintes números são usados  $((\sigma^a)^2, (\sigma^r)^2, (\sigma^d)^2) = (2, 1.5, 1)$ . Para os dois problemas, os valores  $1/4$  e  $3/4$  são dados para as taxas  $\lambda^r$  e  $\lambda^d$ , respectivamente.

Somente será usado o segundo modelo de carga de trabalho, com limite dado pelo Teorema 4.2.3, e com a constante  $\bar{W}$  fixa em  $3/4$ . Portanto, tem-se que  $\sigma^2 = \lambda^a(\sigma^a)^2 + \lambda^d(\sigma^d)^2 + \lambda^r(\sigma^r)^2$  e  $b = \lambda^a - \lambda^d - \lambda^r = \rho - 1$ , para o número de clientes na fila, e  $\sigma^2 = (\lambda^a(\sigma^a)^2 + \lambda^a(\sigma^d)^2 + \lambda^r(\sigma^r)^2)/(\lambda^d)^2$  e  $b = (\lambda^a - \lambda^d - \lambda^r)/\lambda^d$ , para o processo de carga de trabalho. Os resultados estão na Figura 4.3 e na Figura 4.4.

Para o modelo com entrada modulada por uma cadeia de Markov, será considerado o cenário de (Gelenbe et al., 1991), mas será adicionado dependência do fluxo de entrada na cadeia de Markov. Como discutido na seção anterior, a intensidade de tráfego é definida como  $\rho \triangleq \lambda^a(1 - q)/\lambda^d(1 + q)$ . Para simplificar, seja  $\lambda^d = \lambda^a(1 - q)/\rho(1 + q)$ . Serão considerados dois exemplos diferentes onde  $(\bar{\Delta}^a(1), \bar{\Delta}^a(0), (\sigma^a)^2, (\sigma^d)^2, q) = (2, 1, 2, 1, .1)$  para o primeiro exemplo e  $(\bar{\Delta}^a(1), \bar{\Delta}^a(0), (\sigma^a)^2, (\sigma^d)^2, q) = (1, 2, 1, 2, .9)$  para o segundo exemplo. Para estes dois problemas, a matriz  $\Sigma_1$  é dada por:

$$\Sigma_1 = \frac{-q^2}{(1 + q)^3} \sum_{k,m} (-1)^k (-1)^m \zeta(k) \zeta(m)'$$

Para calcular o número médio de clientes em regime estacionário, seja  $\tilde{\Sigma}$  definido como  $\Sigma = \Sigma_0 + 2\Sigma_1$  mas com os elementos da última linha e coluna multiplicados por  $(\lambda^a(\bar{q}^+ - \bar{q}^-))^2$ . Seja  $A$  a matriz tal que  $\tilde{\Sigma} = AA'$ , e defina  $B = \nu A$ , onde  $\nu$  é o vetor coluna  $(1, -1, -1)$ . Desta forma, tem-se que  $\sigma^2 = BB'\lambda^a + (\sigma^d)^2\lambda^d$  e  $b = \lambda^a(\bar{q}^+ - \bar{q}^-) - \lambda^d$ . Os resultados estão na Figura 4.5.

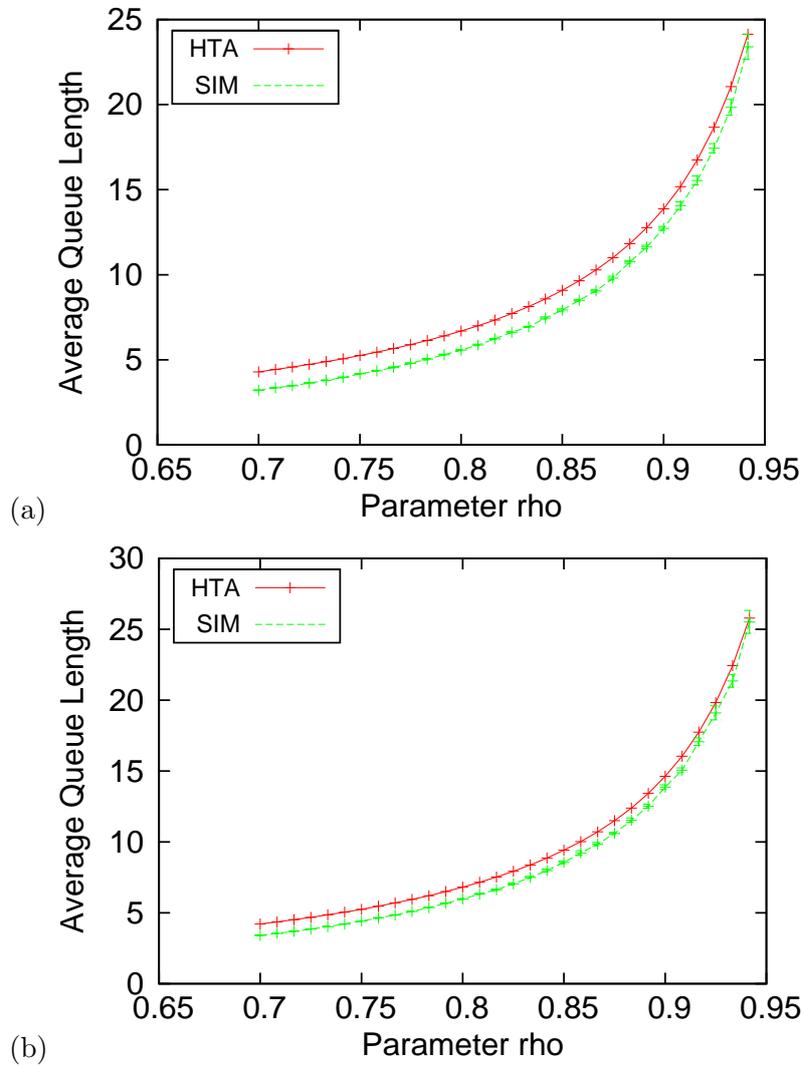


Figura 4.3: Gráfico do número médio de clientes no sistema para diferentes valores de intensidade de tráfego  $\rho$ . Os resultados foram obtidos usando uma simulação computacional (SIM) e através da aproximação por tráfego pesado (HTA). Para a simulação, também é mostrado o intervalo de 95% t-confiança (veja (Gross e Harris, 1998), página 392). As Figuras (a) e (b) mostram o resultado para o primeiro e segundo exemplo respectivamente

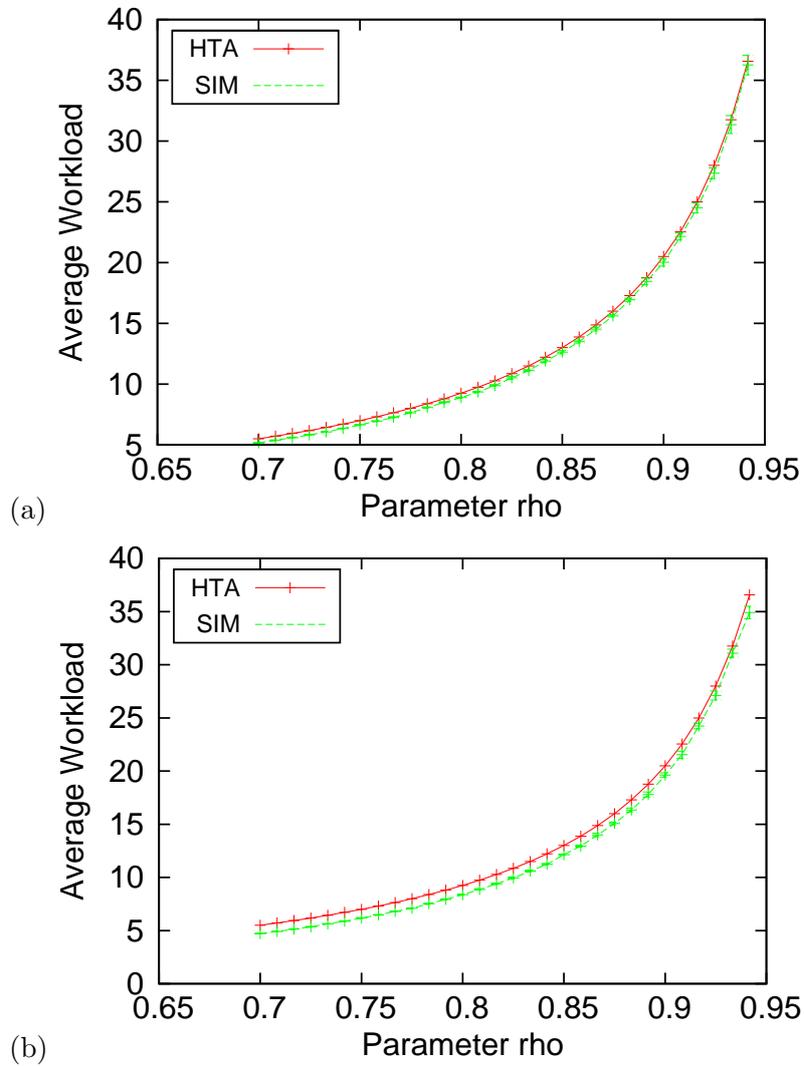


Figura 4.4: Gráfico do número médio de carga de trabalho para diferentes valores de intensidade de tráfego  $\rho$ . Os resultados foram obtidos usando uma simulação computacional (SIM) e através da aproximação por tráfego pesado (HTA). Para a simulação, também é mostrado o intervalo de 95% t-confiança (veja (Gross e Harris, 1998), página 392). As Figuras (a) e (b) mostram o resultado para o primeiro e segundo exemplo respectivamente.

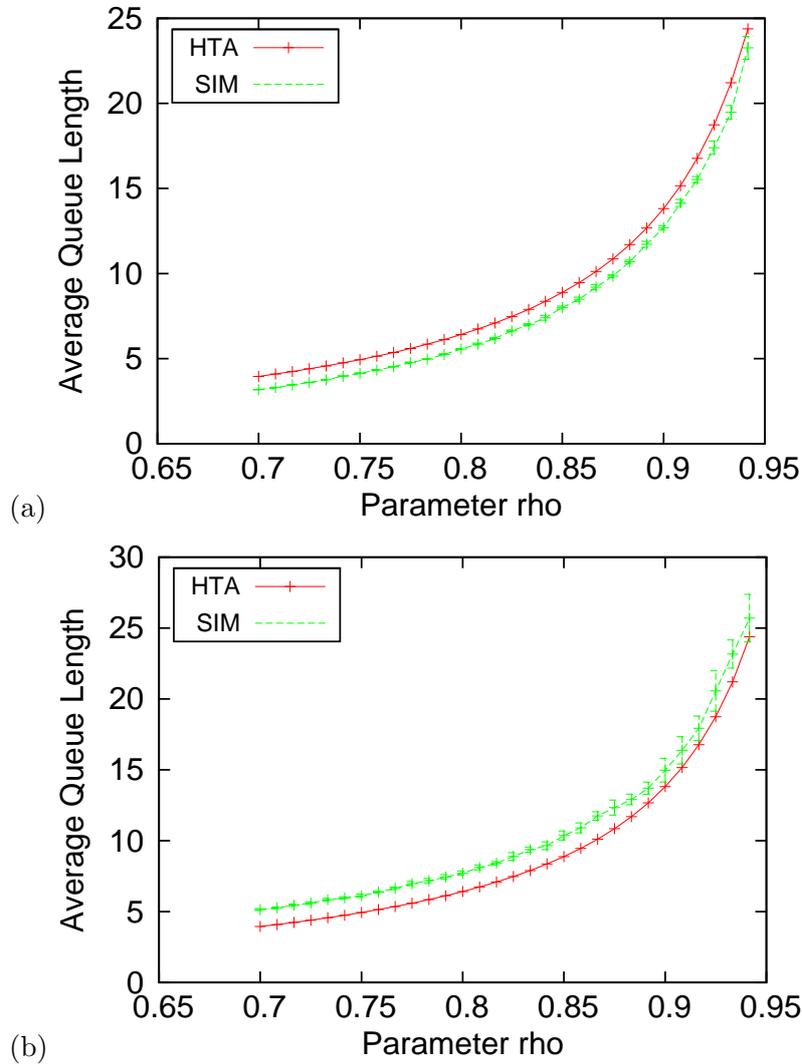


Figura 4.5: Gráfico do número médio de clientes na fila para o modelo com processo de entrada modulado por uma cadeia de Markov para diferentes valores de intensidade de tráfego  $\rho$ . Os resultados foram obtidos usando uma simulação computacional (SIM) e através da aproximação por tráfego pesado (HTA). Para a simulação, também é mostrado o intervalo de 95% t-confiança (veja (Gross e Harris, 1998), página 392). Figuras (a) e (b) mostram o resultado para o primeiro e segundo exemplo respectivamente.

# Capítulo 5

## Redes de Filas com Sinais Sob Tráfego

### Pesado

Neste capítulo será estudado aproximações de tráfego pesado para redes de filas com sinais. O tipo de sinal tratado aqui será aquele que remove um cliente ao chegar no sistema, ou seja o “cliente negativo”. Os sinais são introduzidos de duas formas diferentes: eles podem entrar por uma fonte externa, ou podem vir de outras filas presentes no sistema. No segundo caso, os sinais são clientes que quando terminam o serviço em uma fila são roteados para outras filas como sinais.

No capítulo anterior, foram apontadas duas dificuldades trazidas quando se introduz sinais em sistemas de filas. A primeira delas era em relação ao tempo de serviço residual deixado no modelo quando um cliente era removido durante o serviço. A segunda era em relação à representação do processo de contagem do número de clientes removidos a partir do tempo entre chegada de sinais.

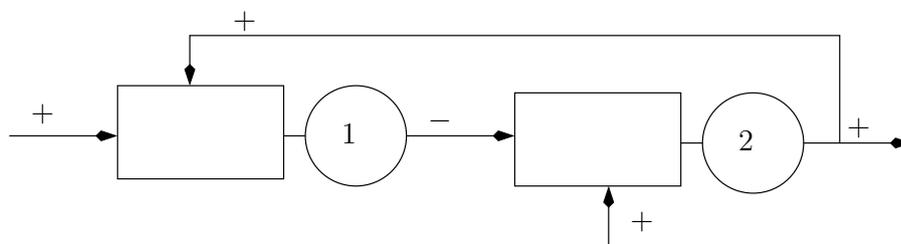


Figura 5.1: Duas filas em sequência com sinais. O símbolo “+” indica a entrada de clientes, já o símbolo “-” indica a entrada de sinais.

No caso de uma fila única, o primeiro problema foi resolvido permitindo que um sinal somente poderia remover clientes que não estavam em serviço. No entanto, este problema não é tão facilmente evitado no caso de redes de filas com sinais. A razão disto vem da dificuldade de obtenção da direção de reflexão no limite. Este mesmo problema foi observado e discutido na Seção 7.1.3 de (Kushner, 2001) para sistemas de manufatura. Para deixar a origem deste problema clara, um exemplo semelhante ao que foi usado na discussão presente em (Kushner, 2001) será colocado aqui no cenário de filas com sinais. Considere o sistema dado pela Figura 5.1. Nesta figura, tem-se dois sistemas de filas em sequência. A primeira fila envia sinais à segunda fila toda vez em que um cliente termina o serviço. Suponha que estes sinais somente removam clientes que não estão sendo servidos.

Agora considere o gráfico dado pela Figura 5.2. Suponha que a fila 2 recebe um sinal quando apenas um cliente em serviço está presente. Neste caso, adotando o escalonamento usual, o processo  $x^n$  representando o estado do sistema escalonado moveria de um ponto (a) (veja a Figura 5.2), que está à distância de  $1/\sqrt{n}$  da fronteira  $\partial G_2 \triangleq \{\xi \in \mathbb{R}^2 : \xi_2 = 0\}$ , para o ponto (b). Como o sinal não pode remover o cliente em serviço, o processo seria empurrado de volta do ponto (b) para o ponto (a).

Caso o sistema estiver no ponto (b) e um sinal é recebido, o processo teria que ir para o ponto (c), que é inviável e portanto o processo seria imediatamente devolvido ao ponto (b). Existem duas possibilidades caso o servidor da fila 2 tente iniciar serviço com o sistema vazio. Se o cliente fosse roteado à fila 1 quando o serviço terminasse, o processo seria levado ao ponto (d). Como este ponto é inviável, ele seria imediatamente devolvido ao ponto (b). Se o cliente deixasse o sistema após concluir o serviço, o processo sairia do ponto (b) para o ponto (c), que é inviável, e portanto o processo retornaria ao ponto (b).

A direção de reflexão que irá aparecer sobre a face da fronteira  $\partial G_2$  seria uma “média” das direções  $d_1$  e  $d_2$  de acordo com a frequência com que elas são usadas. O problema se torna calcular a frequência em que  $d_1$  é usada quando  $x_2^n(t) = 1/\sqrt{n}$

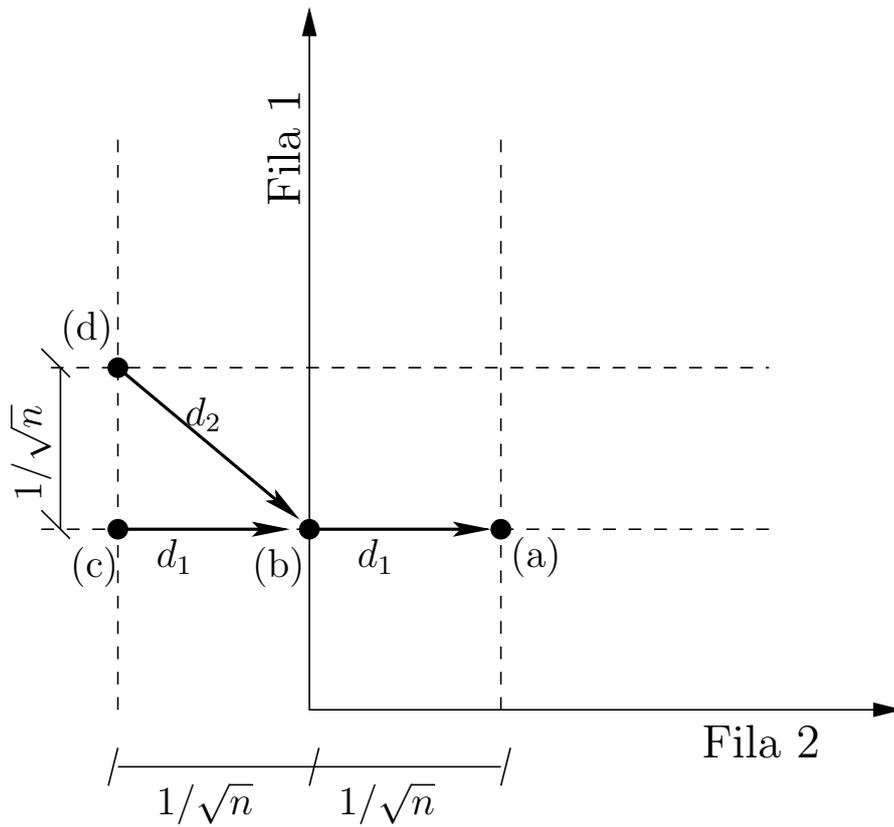


Figura 5.2: Gráfico representando um problema na representação da direção de reflexão na fronteira  $\partial G_2 = \{\xi \in \mathbb{R}^2 : \xi_2 = 0\}$ . Cada eixo representa o número de clientes no sistema

relativo à quando  $x_2^n(t) = 0$ , que não é uma quantidade fácil de ser computada. É importante ressaltar que a dificuldade não está em mostrar que o processo de reflexão é limitado em probabilidade, mas sim em obter uma representação do limite de qualquer subsequência fracamente convergente. Ou seja, no cenário usado aqui, não é possível calcular a direção de reflexão que aparecerá sobre a face de  $\partial G_2$  no limite de forma imediata.

O segundo problema, mencionado no início do capítulo anterior, não permitiu que o processo que conta o número de clientes removidos do sistema por um sinal fosse escrito da forma como  $S^{d,n}$  foi escrito na Equação (4.1). Ao invés disso, o processo  $S^{r,n}$  que conta o número de sinais recebidos pelo sistema foi utilizado e em seguida este processo foi corrigido por  $z_0^n$  para representar o número de clientes removidos do sistema. Apesar desta dificuldade na representação do processo, nenhuma dificuldade na obtenção do limite e na representação do processo de reflexão foi encontrada para o caso de uma fila única. No entanto, no caso tratado aqui, em que o sistema é composto por várias filas em rede, este problema se torna mais complicado. O problema é novamente na representação do processo de reflexão no limite. Caso fosse utilizado a mesma estratégia usada para um sistema composto de somente uma fila, seria necessário estabelecer uma relação entre o tempo em que a fila está vazia com o número de sinais que não removeram clientes. Esta relação não é fácil de se obter, especialmente porque não se conhece a distribuição dos processos envolvidos.

Uma forma de contornar os dois problemas acima e conseguir estabelecer uma aproximação para filas com sinais sob tráfego pesado é impondo a condição que os processos de contagem envolvidos sejam de Poisson. Ou seja, o tempo de serviço e o tempo entre chegada de sinais sejam exponencialmente distribuídos.

Caso o tempo de serviço seja exponencialmente distribuído, o cálculo do tempo de serviço residual gerado por remover um cliente em serviço não seria um problema. Tomando como exemplo novamente o cenário descrito pela Figura 4.1,

tem-se que para qualquer  $t > 0$ ,

$$\begin{aligned}\mathbb{P}(\Delta_{k+1} > t) &= \mathbb{P}(\Delta_k > t) = \mathbb{P}(\Delta_k > t + (t_{k+1} - t_k) | \Delta_k > (t_{k+1} - t_k)) \\ &= \mathbb{P}(t_k + \Delta_k - t_{k+1} > t | \Delta_k > (t_{k+1} - t_k)),\end{aligned}$$

onde foi usado o fato de que os tempos de serviços para cada cliente são identicamente distribuídos, e também foi usada a propriedade “sem memória” de distribuições exponenciais. Portanto, dado que  $\Delta_k > (t_{k+1} - t_k)$ , a distribuição do resíduo  $t_k + \Delta_k - t_{k+1}$  é igual à do tempo de serviço para o próximo cliente. Desta forma, o tempo residual pode ser usado como o tempo de serviço do próximo cliente caso o cliente em serviço seja removido por um sinal.

O problema de representação do processo de contagem do número de clientes removidos da fila até um certo tempo  $t > 0$  também é simplificado. Suponha que  $S_i^r(t)$  represente o número de sinais que chegaram na fila  $i$  até o tempo  $t$ . Então, sendo  $S_i^r$  um processo de Poisson, e  $X_i$  o processo representando o número de clientes na fila  $i$ , pode-se fazer

$$\int_0^t \mathbb{I}\{X_i(s) > 0\} dS_i^r(s)$$

para representar o número de clientes removidos da fila  $i$  até o tempo  $t > 0$ . Em seguida, um teorema de representação (que será discutido com mais detalhes na próxima seção) é usado para separar o “ruído” da “regularidade estatística”, resolvendo assim o problema de representação.

No modelo apresentado aqui, será considerado processos de Poisson com taxas de saltos dependentes do estado do sistema. Além disso, as probabilidades de roteamento também são dependente do estado. A forma de representação do modelo usada aqui se baseia no modelo apresentado em (Mandelbaum e Pats, 1998). O ganho de introduzir esta dependência está na possibilidade de introduzir controles de realimentação. O modelo contido neste capítulo é referente ao trabalho (Leite e Fragoso, 2008a).

## 5.1 Modelo de Filas

Para um sistema de  $K$  filas em rede, o processo que conta o número de clientes presente na  $i$ -ésima fila, para  $i \leq K$ , em um certo instante de tempo  $t \geq 0$  é dado por:

$$X_i(t) = X_i(0) + A_i(t) - D_i(t) - S_i(t) + \sum_{j \leq K} [D_{ji}^+(t) - D_{ji}^-(t)], \quad (5.1)$$

onde  $X_i(0)$  é uma variável aleatória e cada processo de contagem  $A_i$ ,  $D_i$ ,  $S_i$ ,  $D_{ij}^+$ , e  $D_{ij}^-$ , para  $(i, j) \in \{1, \dots, K\}^2$ , será definido abaixo. Sejam os processos de Poisson dependentes do estado dados por

$$\mathcal{N}_i^\alpha(t) \triangleq N_i^\alpha \left( \int_0^t \Lambda_i^\alpha(X(s)) ds \right),$$

onde  $X = (X_1, \dots, X_K)'$ ,  $\Lambda_i^\alpha : \mathbb{R}_{\geq 0}^K \rightarrow \mathbb{R}_{>0}$  são funções mensuráveis, e  $N_i^\alpha$  é um processo de Poisson padrão (i.e., com taxa igual à um), para  $i \in \{1, \dots, K\}$  e  $\alpha \in \{a, r, d\}$ . Além disso, defina o processo  $\mathbb{I}_{ij}^+$  como

$$\mathbb{I}_{ij}^+(t) \triangleq \begin{cases} 1 & \text{- Caso um cliente complete seu serviço na fila } i \text{ no instante } t \\ & \text{e seja roteado para a fila } j \text{ como um cliente;} \\ 0 & \text{- Caso contrário,} \end{cases}$$

e o processo  $\mathbb{I}_{ij}^-$  como

$$\mathbb{I}_{ij}^-(t) \triangleq \begin{cases} 1 & \text{- Caso um cliente complete seu serviço na fila } i \text{ no instante } t \\ & \text{e seja roteado para a fila } j \text{ como um sinal;} \\ 0 & \text{- Caso contrário.} \end{cases}$$

É claro que se assume  $\sum_{j=1}^K (\mathbb{I}_{ij}^+(t) + \mathbb{I}_{ij}^-(t)) \leq 1$ , para cada  $i \in \{1, \dots, K\}$  e  $t \geq 0$ .

Com estas definições, é possível fazer:

$$\begin{aligned} A_i(t) &\triangleq \mathcal{N}_i^a(t) \\ D_i(t) &\triangleq \int_0^t \mathbb{I}\{X_i(s-) > 0\} d\mathcal{N}_i^d(s) \\ S_i(t) &\triangleq \int_0^t \mathbb{I}\{X_i(s-) > 0\} d\mathcal{N}_i^r(s) \\ D_{ij}^+(t) &\triangleq \int_0^t \mathbb{I}_{ij}^+(s) \mathbb{I}\{X_i(s-) > 0\} d\mathcal{N}_i^d(s) \\ D_{ij}^-(t) &\triangleq \int_0^t \mathbb{I}_{ij}^-(s) \mathbb{I}\{X_i(s-) > 0, X_j(s-) > 0\} d\mathcal{N}_i^d(s). \end{aligned}$$

Todos os processos definidos acima estão definidos no mesmo espaço de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$ . Seja  $\mathcal{F}_t$  a mínima  $\sigma$ -álgebra que mensura todos os elementos aleatórios definidos acima até o instante  $t$ . Seja também a filtragem  $\mathcal{F}_t^r$  definida como  $\mathcal{F}_t$  exceto que esta não contém nenhuma decisão de roteamento efetuada no instante  $t$ .

As interpretações dos processos de contagem acima são as seguintes:  $A_i(t)$  denota o número de clientes que chegaram na fila  $i$  de uma fonte externa até o tempo  $t$ ;  $D_i(t)$  denota o número de clientes servidos na fila  $i$  até o tempo  $t$ ;  $S_i(t)$  denota o número de clientes removidos da fila  $i$  através da chegada de um sinal externo até o tempo  $t$ ;  $D_{ij}^+$  conta o número de clientes roteados da fila  $i$  para a fila  $j$  até o instante  $t$ ; e  $D_{ij}^-$  denota o número de clientes que deixaram a fila  $i$  e entraram na fila  $j$  como um sinal e removeram um cliente até o instante  $t$ .

Caso seja necessário considerar filas com tamanho limitado, pode-se fazer o seguinte:

$$\begin{aligned} A_i(t) &\triangleq \int_0^t \mathbb{I}\{X_i(s-) < \tilde{B}_i\} d\mathcal{N}_i^a(s), \\ D_{ij}^+(t) &\triangleq \int_0^t \mathbb{I}_{ij}^+(s) \mathbb{I}\{X_i(s-) > 0, X_j(s-) < \tilde{B}_j\} d\mathcal{N}_i^d(s) \end{aligned}$$

onde  $\tilde{B}_i \in \mathbb{R}_{>0}$  é uma constante que denota a capacidade máxima da fila  $i$ . Para

tratar ambos os casos simultaneamente, suponha que  $\tilde{B}_i \in \mathbb{R}_{>0} \cup \{\infty\}$ . Caso a fila seja ilimitada, faça  $\tilde{B}_i = \infty$ , e portanto

$$\mathbb{I}\{X_i(s) < \infty\} \equiv 1, \quad \mathbb{P} - q.c.,$$

para qualquer  $s < \infty$ , já que todos os processos de salto considerados aqui serão não-explosivos.

Apesar da definição recursiva de  $X$  em função de  $A_i$ ,  $D_i$ ,  $S_i$ ,  $D_{ij}^+$ , e  $D_{ij}^-$  e dos mesmos em função de  $X$ , não é difícil mostrar a existência e unicidade dos processos que satisfazem as definições acima. Um argumento de construção de funções amostras pode ser usado para este propósito da mesma forma como é feito no Teorema 4.1 na página 327 de (Ethier e Kurtz, 1986). O mesmo argumento também aparece em (Mandelbaum e Pats, 1998).

De fato, seja  $X_i^{(0)} \triangleq X_i(0)$  para cada  $i \in \{1, \dots, K\}$ . Defina

$$X_i^{(k)}(t) \triangleq X_i(0) + A_i^{(k-1)}(t) - D_i^{(k-1)}(t) - S_i^{(k-1)}(t) + \sum_{j \leq K} \left[ D_{ji}^{+, (k-1)}(t) - D_{ji}^{-, (k-1)}(t) \right],$$

onde  $A_i^{(k-1)}$ ,  $D_i^{(k-1)}$ ,  $S_i^{(k-1)}$ ,  $D_{ji}^{+, (k-1)}$ , e  $D_{ji}^{-, (k-1)}$  são definidos como  $A_i$ ,  $D_i$ ,  $S_i$ ,  $D_{ij}^+$ , e  $D_{ij}^-$  mas com  $X^{(k-1)}$  no lugar de  $X$  nas suas respectivas definições. Defina o tempo de parada  $\tau_k$  como o instante do  $k$ -ésimo salto de qualquer componente de  $X^{(k)}$ . Ou seja,  $\tau_k$  é definido da seguinte forma: seja  $\tau_1^{a,i}$ ,  $\tau_1^{d,i}$ , e  $\tau_1^{r,i}$  dados como abaixo:

$$\begin{aligned} \tau_1^{a,i} &= \inf\{s > 0 : |A_i^{(1)}(s) - A_i^{(1)}(s-)| > 0\} \\ \tau_1^{d,i} &= \inf\{s > 0 : |D_i^{(1)}(s) - D_i^{(1)}(s-)| > 0\} \\ \tau_1^{r,i} &= \inf\{s > 0 : |S_i^{(1)}(s) - S_i^{(1)}(s-)| > 0\}, \end{aligned}$$

então  $\tau_1 \triangleq \min\{\tau_1^{\alpha,i}; \alpha \in \{a, r, d\}, i \in \{1, \dots, K\}\}$ . Para  $k > 1$ , o seguinte é feito

de forma análoga:

$$\begin{aligned}\tau_k^{a,i} &= \inf\{s > \tau_k : |A_i^{(k)}(s) - A_i^{(k)}(s-)| > 0\} \\ \tau_k^{d,i} &= \inf\{s > \tau_k : |D_i^{(k)}(s) - D_i^{(k)}(s-)| > 0\} \\ \tau_k^{r,i} &= \inf\{s > \tau_k : |S_i^{(k)}(s) - S_i^{(k)}(s-)| > 0\},\end{aligned}$$

e então  $\tau_k \triangleq \min\{\tau_k^{\alpha,i}; \alpha \in \{a, r, d\}, i \in \{1, \dots, K\}\}$ . Observe que para  $t < \tau_k$ :

$$\begin{aligned}A_i^{(k)}(t) &= A_i^{(k-1)}(t), & D_i^{(k)}(t) &= D_i^{(k-1)}(t), \\ S_i^{(k)}(t) &= S_i^{(k-1)}(t), & D_{ij}^{+, (k)}(t) &= D_{ij}^{+, (k-1)}(t), \\ D_{ij}^{-, (k)}(t) &= D_{ij}^{-, (k-1)}(t), & X^{(k)}(t) &= X^{(k-1)}(t).\end{aligned}$$

Logo  $X(t) = \lim_{k \rightarrow \infty} X^{(k)}(t)$  existe para todo  $t < \lim_{k \rightarrow \infty} \tau_k$ , e este  $X$  satisfaz a Equação (5.1) (é claro que o mesmo pode ser dito para  $A_i^{(k)}$ ,  $D_i^{(k)}$ ,  $S_i^{(k)}$ ,  $D_{ji}^{+, (k)}$ , e  $D_{ji}^{-, (k)}$  e seus respectivos limites). Como as funções  $\Lambda_i^\alpha$  sempre tomam valores positivos, tem-se que  $\lim_{k \rightarrow \infty} \tau_k = \infty$ . A prova da unicidade segue facilmente por um argumento de contradição.

A condição abaixo garante, entre outras coisas, que os processos de contagem definidos acima são não-explosivos e possuem uma representação de martingale, que será apresentada abaixo. A condição colocada nas funções  $\Lambda_i^\alpha$  que as fazem limitadas podem ser relaxadas e isto será feito mais tarde na próxima seção.

**Condição 5.1.1.** *Considere as seguintes condições para cada  $\alpha \in \{a, r, d\}$ ,  $\beta \in \{+, -\}$ , e  $(i, j) \in \{1, \dots, K\}^2$ :*

- (a) *Os elementos aleatórios  $\{X_i(0), N_i^\alpha; i \in \{1, \dots, K\}, \alpha \in \{a, r, d\}\}$  são mutuamente independentes.*
- (b) *As funções  $\Lambda_i^\alpha : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{> 0}$  são mensuráveis e limitadas.*
- (c) *Os processos  $\mathbb{I}_{ij}^+$  e  $\mathbb{I}_{ij}^-$  satisfazem*

$$\mathbb{E} \left[ \mathbb{I}_{ij}^\beta(t) \middle| \mathcal{F}_t^r \right] = Q_{ij}^\beta(X(t-)),$$

onde  $Q_{ij}^\beta : \mathbb{R}_{\geq 0}^K \rightarrow [0, 1]$  é uma função mensurável.

Observe que o Lema B.2.3 em combinação com o Teorema B.2.2 possibilita a decomposição dos processos  $A_i$ ,  $D_i$ , e  $S_i$  em

$$\begin{aligned} A_i(t) &= M_i^a(t) + \int_0^t \Lambda_i^a(X(s)) \mathbb{I}\{X_i(s) < \tilde{B}_i\} ds \\ D_i(t) &= M_i^d(t) + \int_0^t \Lambda_i^d(X(s)) \mathbb{I}\{X_i(s) > 0\} ds \\ S_i(t) &= M_i^r(t) + \int_0^t \Lambda_i^r(X(s)) \mathbb{I}\{X_i(s) > 0\} ds, \end{aligned}$$

onde  $M_i^a$ ,  $M_i^d$ , e  $M_i^r$  são  $\mathcal{F}_t$ -martingales. Para obter uma decomposição para os termos  $D_{ij}^+$  e  $D_{ij}^-$ , define-se os processos

$$\begin{aligned} M_{ij}^+(t) &\triangleq \int_0^t (\mathbb{I}_{ij}^+(s) - Q_{ij}^+(X(s-))) \mathbb{I}\{X_i(s-) > 0, X_j(s-) < \tilde{B}_j\} d\mathcal{N}_i^d(s) \\ M_{ij}^-(t) &\triangleq \int_0^t (\mathbb{I}_{ij}^-(s) - Q_{ij}^-(X(s-))) \mathbb{I}\{X_i(s-) > 0, X_j(s-) > 0\} d\mathcal{N}_i^d(s). \end{aligned}$$

Observe que  $M_{ij}^+$  e  $M_{ij}^-$  são  $\mathcal{F}_t$ -martingales. De fato, seja  $0 \leq s \leq t < \infty$  e defina o tempo de parada  $\tau_l$  como o instante do  $l$ -ésimo salto de  $\mathcal{N}_i^d$  após ou no instante  $s$ , então pode-se escrever:

$$\begin{aligned} &\mathbb{E} [M_{ij}^+(t) - M_{ij}^+(s) | \mathcal{F}_s] \\ &= \mathbb{E} \left[ \sum_{l=1}^{\infty} \mathbb{I}\{\tau_l \leq t\} \mathbb{I}\{X_i(\tau_l-) > 0, X_j(\tau_l-) < \tilde{B}_j\} (\mathbb{I}_{ij}^+(\tau_l) - Q_{ij}^+(X(\tau_l-))) \middle| \mathcal{F}_s \right] \\ &= \sum_{l=1}^{\infty} \mathbb{E} \left[ \mathbb{I}\{\tau_l \leq t\} \mathbb{I}\{X_i(\tau_l-) > 0, X_j(\tau_l-) < \tilde{B}_j\} \mathbb{E} [\mathbb{I}_{ij}^+(\tau_l) - Q_{ij}^+(X(\tau_l-)) | \mathcal{F}_{\tau_l}^r] \middle| \mathcal{F}_s \right] \\ &= 0, \end{aligned}$$

para todo  $(i, j) \in \{1, \dots, K\}^2$ , onde foi usada a “regra da torre” (veja Equação (1.7.1) na página 19 de (Arnold, 1974)) e a Condição 5.1.1 (c). O mesmo é válido

para  $M_{ij}^-$ . Desta forma, tem-se

$$\begin{aligned}
D_{ij}^+(t) &= \int_0^t (\mathbb{I}_{ij}^+(s) - Q_{ij}^+(X(s-))) \mathbb{I}\{X_i(s-) > 0, X_j(s-) < \tilde{B}_j\} d\mathcal{N}_i^d(s) \\
&\quad + \int_0^t Q_{ij}^+(X(s-)) \mathbb{I}\{X_i(s-) > 0, X_j(s-) < \tilde{B}_j\} d\mathcal{N}_i^d(s) \\
&= M_{ij}^+(t) + \int_0^t Q_{ij}^+(X(s-)) \mathbb{I}\{X_j(s) < \tilde{B}_j\} dM_i^d(s) \\
&\quad + \int_0^t Q_{ij}^+(X(s-)) \mathbb{I}\{X_i(s) > 0, X_j(s) < \tilde{B}_j\} \Lambda_i^d(X(s)) ds,
\end{aligned}$$

onde foi usada a decomposição em martingale do Teorema B.2.2 novamente. O mesmo desenvolvimento usado acima pode ser feito para  $M_{ij}^-$ , e portanto,

$$\begin{aligned}
D_{ij}^-(t) &= M_{ij}^-(t) + \int_0^t Q_{ij}^-(X(s-)) \mathbb{I}\{X_j(s-) > 0\} dM_i^d(s) \\
&\quad + \int_0^t Q_{ij}^+(X(s-)) \mathbb{I}\{X_i(s) > 0, X_j(s) > 0\} \Lambda_i^d(X(s)) ds.
\end{aligned}$$

Desta forma, o processo  $X = (X_1, \dots, X_K)'$  aceita a seguinte representação

$$X(t) = X(0) + P(t) + M(t)$$

onde  $P = (P_1, \dots, P_k)'$  é um processo  $\mathcal{F}_t$ -previsível e  $M = (M_1, \dots, M_K)'$  é um  $\mathcal{F}_t$ -martingale, que são dados por:

$$\begin{aligned}
P_i(t) &\triangleq \int_0^t \mathbb{I}\{X_i(s) < \tilde{B}_i\} \Lambda_i^a(X(s)) - (\Lambda_i^d(X(s)) + \Lambda_i^r(X(s))) \mathbb{I}\{X_i(s) > 0\} \\
&\quad + \sum_{j \leq K} \left( Q_{ji}^+(X(s)) \mathbb{I}\{X_j(s) > 0, X_i(s) < \tilde{B}_i\} \right. \\
&\quad \left. - Q_{ji}^-(X(s)) \mathbb{I}\{X_j(s) > 0, X_i(s) > 0\} \right) \Lambda_j^d(X(s)) ds,
\end{aligned}$$

e

$$\begin{aligned}
M_i(t) &\triangleq M_i^a(t) - M_i^d(t) - M_i^r(t) \\
&\quad + \sum_{j \leq K} \left( M_{ji}^+(t) - M_{ji}^-(t) + \int_0^t Q_{ji}^+(X(s-)) \mathbb{I}\{X_i(s) < \tilde{B}_i\} dM_j^d(s) \right. \\
&\quad \left. - \int_0^t Q_{ji}^-(X(s-)) \mathbb{I}\{X_i(s-) > 0\} dM_j^d(s) \right),
\end{aligned}$$

para cada  $i \in \{1, \dots, K\}$ .

## 5.2 Limite em Regime de Tráfego Pesado

Como é usual em aproximações de filas em regime de tráfego pesado, assume-se a existência de uma sequência de filas  $\{X^n\}$  indexadas pelo parâmetro  $n$ . Isto é, para cada  $n \in \mathbb{N}$ ,  $X^n$  satisfaz a Equação (5.1). A medida que  $n$  aumenta, o sistema aproxima um cenário de tráfego pesado, no sentido que a taxa em que clientes entram em cada fila do sistema fica mais próxima da taxa com que os clientes saem da fila.

Seja  $x^n$  o processo escalonado dado por

$$x^n(t) \triangleq X^n(nt)/\sqrt{n},$$

para cada  $t \geq 0$  e  $n \in \mathbb{N}$ . Observe que sob este escalonamento pode-se escrever a decomposição do processo  $A_i^n$  em martingale como:

$$\begin{aligned}
\frac{1}{\sqrt{n}} A_i^n(nt) &= \frac{1}{\sqrt{n}} M_i^{a,n}(nt) + \frac{1}{\sqrt{n}} \int_0^{nt} \Lambda_i^{a,n}(X^n(s)) \mathbb{I}\{X_i^n(s) < \tilde{B}_i^n\} ds \\
&= m_i^{a,n}(t) + \sqrt{n} \int_0^t \Lambda_i^{a,n}(\sqrt{n}x^n(s)) \mathbb{I}\{\sqrt{n}x^n(s) < \tilde{B}_i^n\} ds \\
&= m_i^{a,n}(t) + \sqrt{n} \int_0^t \lambda_i^{a,n}(x^n(s)) \mathbb{I}\{x^n(s) < B_i\} ds,
\end{aligned}$$

onde  $m_i^{a,n}(t) \triangleq M_i^{a,n}(nt)/\sqrt{n}$ , e  $\lambda_i^{a,n}(\xi) \triangleq \Lambda_i^{a,n}(\sqrt{n}\xi)$ , para todo  $t \geq 0$ ,  $n \in \mathbb{N}$ , e  $\xi \in \mathbb{R}_{\geq 0}^K$ . Além disso foi assumido que a capacidade de cada fila  $i$  cresce com o parâmetro  $n$  da seguinte forma  $B_i \triangleq \tilde{B}_i^n/\sqrt{n}$ , onde  $B_i \in \mathbb{R}_{>0}$  é uma constante fixa

independente de  $n$ .

O mesmo pode ser feito de forma análoga para os outros processos de salto, definindo  $m_i^{\alpha,n} \triangleq M_i^{\alpha,n}(nt)/\sqrt{n}$ ,  $m_{ij}^{\beta,n} \triangleq M_{ij}^{\beta,n}(nt)/\sqrt{n}$ ,  $\lambda_i^{\alpha,n}(\xi) \triangleq \Lambda_i^{\alpha,n}(\sqrt{n}\xi)$ , e  $q_{ij}^{\beta,n}(\xi) \triangleq Q_{ij}^{\beta,n}(\sqrt{n}\xi)$ , para cada  $t \geq 0$ ,  $n \in \mathbb{N}$ ,  $(i, j) \in \{1, \dots, K\}^2$ ,  $\alpha \in \{d, r\}$ , e  $\beta \in \{+, -\}$ . Desta forma, pode-se escrever  $x^n$  como:

$$x^n(t) = x^n(0) + p^n(t) + m^n(t)$$

onde  $x_i^n(0) \triangleq X_i^n(0)/\sqrt{n}$ , para cada  $i \in \{1, \dots, K\}$ ,

$$\begin{aligned} p_i^n(t) &\triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) < B_i\} \lambda_i^{a,n}(x^n(s)) - \left( \lambda_i^{d,n}(x^n(s)) + \lambda_i^{r,n}(x^n(s)) \right) \mathbb{I}\{x_i^n(s) > 0\} \\ &+ \sum_{j \leq K} \left( q_{ji}^{+,n}(x^n(s)) \mathbb{I}\{x_j^n(s) > 0, x_i^n(s) < B_i\} - q_{ji}^{-,n}(x^n(s)) \mathbb{I}\{x_j^n(s) > 0, x_i^n(s) > 0\} \right) \\ &\times \lambda_j^{d,n}(x^n(s)) ds, \end{aligned} \tag{5.2}$$

e

$$\begin{aligned} m_i^n(t) &\triangleq m_i^{a,n}(t) - m_i^{d,n}(t) - m_i^{r,n}(t) + \sum_{j \leq K} [m_{ji}^{+,n}(t) - m_{ji}^{-,n}(t) \\ &+ \int_0^t q_{ji}^{+,n}(x^n(s-)) \mathbb{I}\{x_i^n(s) < B_i\} dm_j^{d,n}(s) \\ &- \int_0^t q_{ji}^{-,n}(x^n(s-)) \mathbb{I}\{x_i^n(s) > 0\} dm_j^{d,n}(s)] \end{aligned}$$

A seguinte condição precisa ser introduzida. Ela define formalmente a hipótese do tráfego pesado e estabelece como a dependência de estado afeta o sistema. Apesar da dependência do estado parecer pequena quando  $n$  é grande, ela tem um efeito significativo no limite. Como será visto em seguida, as funções  $f_i^\alpha$  e  $f_{ij}^\beta$  (definidas na condição abaixo) aparecem no termo de deriva da equação limite. A mesma condição imposta sob as taxas  $\lambda^{\alpha,n}$  foram usadas no modelo da Seção 8.2 de (Kushner, 2001).

**Condição 5.2.1.** *Para  $o(\cdot)$  uniforme em  $\xi \in \mathbb{R}^K$ , assume-se que:*

(a) Existem constantes não-negativas  $r_i^\alpha$  e  $r_{ij}^\beta$  e funções limitadas e contínuas  $f_i^\alpha$  e  $f_{ij}^\beta$ , com  $(i, j) \in \{1, \dots, K\}^2$ ,  $\alpha \in \{a, d, r\}$ , e  $\beta \in \{+, -\}$ , tais que

$$\begin{aligned}\lambda_i^{\alpha,n}(\xi) &= r_i^\alpha + f_i^\alpha(\xi)/\sqrt{n} + o(1/\sqrt{n}) \\ q_{ij}^{\beta,n}(\xi) &= r_{ij}^\beta + f_{ij}^\beta(\xi)/\sqrt{n} + o(1/\sqrt{n})\end{aligned}$$

para cada  $\xi \in \mathbb{R}_{\geq 0}^K$ .

(b) Para qualquer  $i \in \{1, \dots, K\}$

$$r_i^a + \sum_{j \leq K} r_j^d r_{ji}^+ = r_i^d + r_i^r + \sum_{j \leq K} r_j^d r_{ji}^-$$

que é a condição comumente chamada de “condição do tráfego pesado”.

Observe que a Condição 5.2.1 acima nada mais é que condições para a convergência do termo de deriva. Note que para  $\xi \in \mathbb{R}_{\geq 0}^K$ , tem-se que

$$\sqrt{n} \left( \lambda_i^{a,n}(\xi) - \lambda_i^{d,n}(\xi) - \lambda_i^{r,n}(\xi) + \sum_{j \leq K} (q_{ji}^{+,n}(\xi) - q_{ji}^{-,n}(\xi)) \lambda_j^{d,n}(\xi) \right) \triangleq b_i^n(\xi) \rightarrow b_i(\xi),$$

onde  $b_i$  é definido no Teorema 5.2.3 abaixo.

A próxima condição será sobre as direções de reflexão. Antes de apresentar a condição, sejam as seguintes matrizes em  $\mathbb{R}^{K \times K}$ :

$$\begin{aligned}I_r &\triangleq \text{diag}(r_i^d + r_i^r)_{i=1, \dots, K} \\ \Omega^S &\triangleq \text{diag} \left( \sum_{j \in Z \setminus S \cup \{i\}} r_{ji}^- r_j^d \right)_{i=1, \dots, K} \\ \Theta_{ij} &\triangleq r_i^d (r_{ij}^+ - r_{ij}^-) \\ R^S &\triangleq I_r - \Theta' + \Omega^S \\ R &\triangleq R^\emptyset\end{aligned}$$

onde  $S \subseteq Z \triangleq \{1, \dots, K\}$ , e  $\text{diag}(a_i)_{i=1, \dots, m} \in \mathbb{R}^{m \times m}$  representa uma matriz

diagonal com entradas  $a_i$ .

**Condição 5.2.2.**

(a) A matriz  $R$  satisfaz a condição “completamente- $\mathcal{S}$ ” dada pela Condição A.1.13.

(b) Para cada  $S \subseteq \{1, \dots, K\}$ , com  $|S| \geq 2$ , existe um  $\alpha = (\alpha_1, \dots, \alpha_{|S|})' \neq 0$ , onde  $|S|$  denota o número de elementos do conjunto  $S$ , tal que  $\alpha_i \geq 0$  e

$$\tilde{R}\alpha = \tilde{R}^S e,$$

onde  $e \triangleq (1, \dots, 1)'$ , e  $\tilde{R}$  (resp.,  $\tilde{R}^S$ ) é a matriz construída com as colunas da matriz  $R$  (resp.,  $R^S$ ) que possuem índices no conjunto  $S$ .

A interpretação usual que se dá para a matriz  $R$  é que suas colunas, dadas por vetores  $d_i$ , são direções de reflexão. Isto é, caso o processo  $x^n$  tente ultrapassar qualquer fronteira do espaço de estado, como por exemplo  $\partial G_i \triangleq \{\xi \in \mathbb{R}^K \mid \xi_i = 0\}$ , ele será “empurrado” de volta na direção  $d_i$ . Quando  $x^n(t) \in \partial G_i \cap \partial G_j$ , com  $i \neq j$ , a hipótese usual é que a direção de reflexão neste instante seja uma combinação linear positiva das direções  $d_i$  e  $d_j$ . O mesmo vale quando  $x$  atinge a intercessão de mais de duas fronteiras. Isso possibilita escrever o termo de reflexão na forma usual como  $z(t) = \sum_{i=1}^K d_i y_i(t) = Ry(t)$ , onde  $y = (y_1, \dots, y_K)'$ . Esta condição que as direções de reflexão nas “quinas” do espaço de estado sejam combinações lineares positivas das direções nas faces adjacentes é atendida naturalmente na grande maioria dos sistemas de filas, como é visto em (Kushner, 2001). Contudo, no caso de filas com sinais, esta condição não é sempre atendida, e portanto ela é imposta no item (b) da Condição 5.2.2. Na seção seguinte são apresentadas duas topologias de rede que satisfazem esta condição.

Durante o desenvolvimento desta tese, foi descoberta uma outra forma de tratar este problema que não impõe esta condição forçosamente. Contudo, isto só será tratado mais tarde no Capítulo 7 sob um contexto mais geral. Este capítulo segue o desenvolvimento apresentado em (Leite e Fragoso, 2008a).

**Teorema 5.2.3.** *Suponha que  $x^n(0) \Rightarrow x(0)$ , e considere as Condições 5.1.1, 5.2.1, e 5.2.2. Então  $\{x^n\}$  é limitado em probabilidade e qualquer subsequência fracamente convergente satisfaz:*

$$x(t) = x(0) + \int_0^t b(x(s))ds + w(t) + z(t) \quad (5.3)$$

onde  $x_i(t) \in [0, B_i]$ , caso  $B_i < \infty$ , e  $x_i(t) \in [0, \infty)$ , caso contrário, para  $i \in \{1, \dots, K\}$  e  $t \geq 0$ . Além disso, o processo de reflexão é dado por

$$z(t) \triangleq Ry(t) - u(t),$$

onde  $y_i(0) = 0$ ,  $y_i$  possui funções amostras contínuas, não-decrescente quase certamente, e que somente aumentam nos instantes  $t$  tais que  $x_i(t) = 0$ . De forma análoga, caso  $B_i < \infty$ ,  $u_i(0) = 0$ ,  $u_i$  possui funções amostras contínuas, não-decrescentes, e que somente aumentam nos instantes  $t$  tais que  $x_i(t) = B_i$ , quase certamente. Caso  $B_i = \infty$ ,  $u_i \equiv 0$ .

O termo de deriva  $b$  é dado por

$$b_i(\xi) = f_i^a(\xi) - f_i^d(\xi) - f_i^r(\xi) + \sum_{j=1}^K [f_j^d(\xi)(r_{ji}^+ - r_{ji}^-) + r_j^d(f_{ji}^+(\xi) - f_{ji}^-(\xi))],$$

para cada  $\xi \in \mathbb{R}_{\geq 0}^K$ . O processo  $w$  é dado por

$$w_i(t) = w_i^a(t) - w_i^d(t) - w_i^r(t) + \sum_{j=1}^K [(r_{ji}^+ - r_{ji}^-)w_j^d(t) + w_{ji}^+(t) - w_{ji}^-(t)].$$

Seja  $w_i^s \triangleq (w_{i1}^+, \dots, w_{iK}^+, w_{i1}^-, \dots, w_{iK}^-)'$ . Os processos  $w_i^\alpha$ , para  $\alpha \in \{a, d, r, s\}$  e  $i \in \{1, \dots, K\}$ , são processos de Wiener mutuamente independentes, onde  $w_i^a$ ,  $w_i^d$ , e  $w_i^r$  possuem variâncias dadas por  $r_i^a$ ,  $r_i^d$ , e  $r_i^r$ , respectivamente. O processo  $w_i^s$

possui matriz de covariância dada por  $\Sigma_i$ , onde

$$\Sigma_i \triangleq r_i^d \begin{pmatrix} (\Sigma_i^+) & (\Sigma_i^{+-}) \\ (\Sigma_i^{+-})' & (\Sigma_i^-) \end{pmatrix}$$

com

$$(\Sigma_i^\beta)_{jk} \triangleq \begin{cases} (1 - r_{ij}^\beta) r_{ij}^\beta & \text{se } j = k \\ -r_{ij}^\beta r_{ik}^\beta & \text{caso contrário} \end{cases}$$

e  $(\Sigma^{+-})_{jk} = -r_{ij}^+ r_{ik}^-$ , para  $\beta \in \{+, -\}$ , e onde  $\Sigma_i^{+-}, \Sigma_i^\beta \in \mathbb{R}^{K \times K}$ .

*Demonstração.* Usando as igualdades  $\mathbb{I}\{x_i^n(s) < B_i\} + \mathbb{I}\{x_i^n(s) = B_i\} = 1$ , para  $i \in \{1, \dots, K\}$  tal que  $B_i < \infty$ , e  $\mathbb{I}\{x_i^n(s) > 0\} + \mathbb{I}\{x_i^n(s) = 0\} = 1$ , para todo  $i \in \{1, \dots, K\}$ , e definindo os processos  $y_i^n(t) \triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = 0\} ds$ ,  $y_{ij}^n(t) \triangleq y_{ji}^n(t) \triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = 0, x_j^n(s) = 0\} ds$ ,  $\tilde{u}_i^n(t) \triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = B_i\} ds$ , e  $\tilde{u}_{ij}^n(t) \triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_j^n(s) > 0, x_i^n(s) = B_i\} ds$  tem-se que o processo  $p_i^n$ , dado pela Equação (5.2), pode ser escrito como:

$$\begin{aligned} p_i^n(t) = & \sqrt{n} \int_0^t b_i^n(x^n(s)) ds + \sqrt{n} \int_0^t -\lambda_i^{a,n}(x^n(s)) \mathbb{I}\{x_i^n(s) = B_i\} \\ & + \lambda_i^{d,n} \mathbb{I}\{x_i^n(s) = 0\} + \lambda_i^{r,n}(x^n(s)) \mathbb{I}\{x_i^n(s) = 0\} \\ & + \sum_{j \leq K} \left\{ q_{ji}^{n,+}(x^n(s)) \lambda_j^{d,n}(x^n(s)) [\mathbb{I}\{x_j^n(s) = 0, x_i^n(s) = B_i\} - \mathbb{I}\{x_j^n(s) = 0\}] \right. \\ & - \mathbb{I}\{x_i^n(s) = B_i\} - q_{ji}^{n,-}(x^n(s)) \lambda_j^{d,n}(x^n(s)) [\mathbb{I}\{x_i^n(s) = 0, x_j^n(s) = 0\} \\ & \left. - \mathbb{I}\{x_i^n(s) = 0\} - \mathbb{I}\{x_j^n(s) = 0\}] \right\} ds \end{aligned}$$

$$\begin{aligned}
&= \sqrt{n} \int_0^t b_i^n(x^n(s)) ds + \sqrt{n} \int_0^t -\lambda_i^{a,n}(x^n(s)) \mathbb{I}\{x_i^n(s) = B_i\} \\
&\quad - \sum_{j \leq K} (q_{ji}^+(x^n(s)) \lambda_j^n(x^n(s)) \mathbb{I}\{x_j^n(s) > 0, x_i^n(s) = B_i\}) \\
&\quad + \left\{ \lambda_i^{d,n}(x^n(s)) + \lambda_i^{r,n}(x^n(s)) - [q_{ii}^{n,+}(x^n(s)) - q_{ii}^{n,-}(x^n(s))] \lambda_i^{d,n}(x^n(s)) \right. \\
&\quad \left. + \sum_{j \leq K (j \neq i)} q_{ji}^{n,-}(x^n(s)) \lambda_j^{d,n}(x^n(s)) \right\} \mathbb{I}\{x_i^n(s) = 0\} \\
&\quad - \sum_{j \leq K (j \neq i)} (q_{ji}^{n,+}(x^n(s)) - q_{ji}^{n,-}(x^n(s))) \lambda_j^{d,n}(x^n(s)) \mathbb{I}\{x_j^n(s) = 0\} \\
&\quad - \sum_{j \leq K (j \neq i)} q_{ji}^{n,-}(x^n(s)) \lambda_j^{d,n}(x^n(s)) \mathbb{I}\{x_i^n(s) = 0, x_j^n(s) = 0\} ds \\
&= \int_0^t b_i(x^n(s)) ds - \left( r_i^a \tilde{u}_i^n(t) + \sum_{j \leq K} r_{ji}^+ r_j^d \tilde{u}_{ij}^n(t) \right) \\
&\quad + \left( r_i^d + r_i^r - (r_{ii}^+ - r_{ii}^-) r_i^d + \sum_{j \leq K (j \neq i)} r_{ji}^- r_j^d \right) y_i^n(t) \\
&\quad - \sum_{j \leq K (j \neq i)} [r_{ji}^+ - r_{ji}^-] r_j^d y_j^n(t) - \sum_{j \leq K (j \neq i)} r_{ji}^- r_j^d y_j^n(t) \\
&\quad + O(\tilde{u}_i^n(t)/\sqrt{n}) + \sum_{j \leq K} O(y_j^n(t)/\sqrt{n}) + \sqrt{n} o(1/\sqrt{n}),
\end{aligned}$$

onde foi usado a Condição 5.2.1 na última passagem. Os termos  $O(y_j^n(t)/\sqrt{n})$  (resp.,  $O(\tilde{u}_j^n(t)/\sqrt{n})$ ) denotam todos os processos que podem ser majorados por  $y_j^n(t)/\sqrt{n}$  (resp.,  $\tilde{u}_j^n(t)/\sqrt{n}$ ) vezes uma constante, como por exemplo,  $|\int_0^t f_i^d(x^n(s)) \mathbb{I}\{x_i^n(s) = 0\} ds| \leq C y_i(t)/\sqrt{n}$ , para uma constante  $C > 0$ .

Defina o processo  $u_i^n(t) \triangleq \left( r_i^a \tilde{u}_i^n(t) + \sum_{j \leq K} r_{ji}^+ r_j^d \tilde{u}_{ij}^n(t) \right)$  e o processo  $z_i^n$  como:

$$\begin{aligned}
z_i^n(t) &\triangleq \left( r_i^d + r_i^r - (r_{ii}^+ - r_{ii}^-) r_i^d + \sum_{j \leq K (j \neq i)} r_{ji}^- r_j^d \right) y_i^n(t) \\
&\quad - \sum_{j \leq K (j \neq i)} [r_{ji}^+ - r_{ji}^-] r_j^d y_j^n(t) - \sum_{j \leq K (j \neq i)} r_{ji}^- r_j^d y_j^n(t).
\end{aligned}$$

Sejam  $z^n = (z_1^n, \dots, z_K^n)'$  e  $y^n = (y_1^n, \dots, y_K^n)'$ , observe que é possível escrever em

notação matricial o seguinte:

$$z^n(t) = \int_0^t R^{\mathcal{E}(x^n(s))} dy^n(s),$$

onde o conjunto  $\mathcal{E}(\xi) \triangleq \{i \in \{1, \dots, K\} | \xi_i = 0\}$  para qualquer  $\xi \in \mathbb{R}^K$ . Defina o processo  $\mathcal{I}^n(\cdot; \alpha) \triangleq (\mathcal{I}_1^n(\cdot; \alpha), \dots, \mathcal{I}_K^n(\cdot; \alpha))'$ , para  $\alpha \in \mathbb{R}^K$  e

$$\mathcal{I}_i^n(\cdot; \alpha) \triangleq \alpha_i \sqrt{n} \mathbb{1}\{x_i^n(\cdot) = 0\}.$$

Usando a hipótese dada pela Condição 5.2.2(b), tem-se que para cada  $t \in [0, \infty)$ , existe um  $\alpha(t) \in \mathbb{R}^K$  tal que

$$R^{\mathcal{E}(x^n(t))} \mathcal{I}^n(t; e) = R \mathcal{I}^n(t; \alpha(t)),$$

onde  $e \triangleq (1, \dots, 1)' \in \mathbb{R}^K$ . Desta forma,

$$\int_0^t R^{\mathcal{E}(x^n(s))} dy^n(s) = \int_0^t R^{\mathcal{E}(x^n(s))} \mathcal{I}^n(s; e) ds = \int_0^t R \mathcal{I}^n(s; \alpha(s)) ds = R \bar{y}^n(t),$$

onde  $\bar{y}^n(t) \triangleq \int_0^t \mathcal{I}^n(s; \alpha(s)) ds$ .

Seja  $\psi_i^n(t) = m_i^n(t) + \int_0^t b_i(x^n(s)) ds + O(\tilde{u}_i^n(t)/\sqrt{n}) + \sum_{j \leq K} O(y_j^n(t)/\sqrt{n}) + \sqrt{n} o(1/\sqrt{n})$ . O Teorema A.2.3 será usado para mostrar que  $\{z^n\}$  é limitado em probabilidade. Para isso, é necessário mostrar que  $\psi_i^n$  é de fato assintoticamente contínuo. Através do critério dado pela Proposição B.1.5, pode-se facilmente verificar que os termos  $O(\tilde{u}_i^n(t)/\sqrt{n})$  e  $O(y_i^n(t)/\sqrt{n})$  são assintoticamente contínuos. Usando o mesmo critério, pode-se verificar que, para qualquer sequência  $\{\phi^n\}$  de processos  $\phi^n$  com funções amostras em  $D(\mathbb{R}_{\geq 0}^K; 0, \infty)$ , o processo  $\int_0^\cdot b_i(\phi^n(s)) ds$  é assintoticamente contínuo, já que  $b_i$  é limitado. Fica faltando mostrar que  $m_i^n$  é assintoticamente contínuo.

Do Lema B.2.4, sabe-se que os processos de Doob-Meyer associados à  $m_i^{a,n}$ ,

$m_i^{d,n}$ , e  $m_i^{r,n}$  são dados por

$$\begin{aligned}\langle m_i^{a,n} \rangle(t) &= \int_0^t \mathbb{I}\{x_i^n(s) < B_i\} \lambda_i^{a,n}(x^n(s)) ds \\ \langle m_i^{d,n} \rangle(t) &= \int_0^t \mathbb{I}\{x_i^n(s) > 0\} \lambda_i^{d,n}(x^n(s)) ds \\ \langle m_i^{r,n} \rangle(t) &= \int_0^t \mathbb{I}\{x_i^n(s) > 0\} \lambda_i^{r,n}(x^n(s)) ds.\end{aligned}$$

Além disso,  $\langle m_i^{\alpha,n}, m_j^{\alpha,n} \rangle \equiv 0$  e  $\langle m_i^{\alpha,n}, m_k^{\beta,n} \rangle \equiv 0$ , para  $(\alpha, \beta) \in \{a, r, d\}^2$ ,  $\alpha \neq \beta$ ,  $(i, j, k) \in \{1, \dots, K\}^3$ ,  $i \neq k$ , já que, pela hipótese da independência dos processos de Poisson  $N_i^{a,n}$ ,  $N_i^{d,n}$ , e  $N_i^{r,n}$ , nenhum salto em  $m_i^{a,n}$ ,  $m_i^{d,n}$ , e  $m_i^{r,n}$  coincide com probabilidade um. Seja  $m_i^{s,n} \triangleq (m_{i1}^{+,n}, \dots, m_{iK}^{+,n}, m_{i1}^{-,n}, \dots, m_{iK}^{-,n})'$ , com a aplicação do Lema B.2.5 tem-se que

$$\begin{aligned}\langle m_{ij}^{+,n} \rangle(t) &= \int_0^t (1 - q_{ij}^{+,n}(x^n(s)) q_{ij}^{+,n}(x^n(s)) \mathbb{I}\{x_i^n(s) > 0, x_j^n(s) < B_j\} \\ &\quad \times \lambda_i^{d,n}(x^n(s)) ds \\ \langle m_{ij}^{+,n}, m_{ik}^{+,n} \rangle(t) &= - \int_0^t q_{ij}^{+,n}(x^n(s)) q_{ik}^{+,n}(x^n(s)) \mathbb{I}\{x_i^n(s) > 0\} \\ &\quad \times \mathbb{I}\{x_j^n(s) < B_j, x_k^n(s) < B_k\} \lambda_i^{d,n}(x^n(s)) ds \\ \langle m_{ij}^{-,n} \rangle(t) &= \int_0^t (1 - q_{ij}^{-,n}(x^n(s)) q_{ij}^{-,n}(x^n(s)) \\ &\quad \times \mathbb{I}\{x_i^n(s) > 0, x_j^n(s) > 0\} \lambda_i^{d,n}(x^n(s)) ds \\ \langle m_{ij}^{-,n}, m_{ik}^{-,n} \rangle(t) &= - \int_0^t q_{ij}^{-,n}(x^n(s)) q_{ik}^{-,n}(x^n(s)) \\ &\quad \times \mathbb{I}\{x_i^n(s) > 0, x_j^n(s) > 0, x_k^n(s) > 0\} \lambda_i^{d,n}(x^n(s)) ds \\ \langle m_{ij}^{+,n}, m_{ik}^{-,n} \rangle(t) &= - \int_0^t q_{ij}^{+,n}(x^n(s)) q_{ik}^{-,n}(x^n(s)) \\ &\quad \times \mathbb{I}\{x_i^n(s) > 0, x_k^n(s) > 0, x_j^n(s) < B_j\} \lambda_i^{d,n}(x^n(s)) ds.\end{aligned}$$

O mesmo argumento usado no Lema B.2.5 pode ser usado em conjunção com a Condição 5.1.1(c) para mostrar que  $\langle m_i^{d,n}, m_{ij}^{\beta,n} \rangle \equiv 0$  onde  $(i, j) \in \{1, \dots, K\}^2$  e  $\beta \in \{+, -\}$ . De fato, observe que pode-se escrever o seguinte, para  $t, \tau > 0$ ,

$$M_i^{d,n}(t + \tau) - M_i^{d,n}(t) = \sum_{l=1}^{\infty} \mathbb{I}\{s_l \leq t + \tau\} + \int_t^{t+\tau} g_i^n(s) ds,$$

onde  $s_l$  é o  $l$ -ésimo salto de  $D_i^n$  após o instante  $t$ , e  $g_i^n$  é um processo  $\mathcal{F}_t$ -mensurável.

Portanto,

$$\begin{aligned} M_i^{d,n}(t + \tau) - M_i^{d,n}(t) &= \sum_{l=1}^{\infty} \left( 1 + \int_{s_{l-1}}^{s_l} g_i^n(s) ds \right) \mathbb{I}\{s_l \leq t + \tau\} + \int_{\tilde{s}}^{t+\tau} g_i^n(s) ds \\ &= \sum_{l=1}^{\infty} (M_i^{d,n}(s_l) - M_i^{d,n}(s_{l-1})) \mathbb{I}\{s_l \leq t + \tau\} \\ &\quad + (M_i^{d,n}(t + \tau) - M_i^{d,n}(\tilde{s})) \end{aligned}$$

onde  $s_0 \triangleq t$  e  $\tilde{s}$  é o instante do último salto de  $D_i^n$  antes de  $t + \tau$ . Suponha, sem perda de generalidade, que  $\int_{\tilde{s}}^{t+\tau} g_i^n(s) ds \equiv 0$ . Seja  $\Delta f \triangleq f(t + \tau) - f(t)$ , para qualquer  $f$ . Desta forma,

$$\begin{aligned} \mathbb{E} \left[ \left( \Delta(M_i^{d,n} + M_{ij}^{+,n}) \right)^2 \middle| \mathcal{F}_t^n \right] &= \mathbb{E} \left[ \Delta \langle M_i^{d,n} \rangle \middle| \mathcal{F}_t^n \right] + \mathbb{E} \left[ \Delta \langle M_{ij}^{+,n} \rangle \middle| \mathcal{F}_t^n \right] \\ + 2\mathbb{E} \left[ \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \left( 1 + \int_{s_{l-1}}^{s_l} g_i^n(s) ds \right) \left( \mathbb{I}_{ij}^{+,n}(s_m) - Q_{ij}^{+,n}(s_m-) \right) \right. \\ &\quad \left. \times \mathbb{I}\{X_i^n(s-) > 0, X_j^n(s-) < \tilde{B}_j\} \mathbb{I}\{s_l \leq t + \tau\} \mathbb{I}\{s_m \leq t + \tau\} \middle| \mathcal{F}_t^n \right], \end{aligned}$$

usando a “regra da torre” (i.e.,  $\mathbb{E} [\mathbb{E} [\dots | \mathcal{F}_{s_h}^n] | \mathcal{F}_t^n] = \mathbb{E} [\dots | \mathcal{F}_t^n]$ , onde  $h = \max\{l, m\}$ ) e a Condição 5.1.1(c) tem-se que

$$\mathbb{E} \left[ \left( \Delta(M_i^{d,n} + M_{ij}^{+,n}) \right)^2 \middle| \mathcal{F}_t^n \right] = \mathbb{E} \left[ \Delta \langle M_i^{d,n} \rangle \middle| \mathcal{F}_t^n \right] + \mathbb{E} \left[ \Delta \langle M_{ij}^{+,n} \rangle \middle| \mathcal{F}_t^n \right]$$

e o resultado segue com o uso da identidade de polarização (veja (Protter, 1995) página 98).

Além disso, como  $N_i^{a,n}$ ,  $N_i^{d,n}$ , e  $N_i^{r,n}$  são independentes, tem-se que  $\langle m_i^{\alpha,n}, m_{kj}^{\beta,n} \rangle \equiv 0$ , para  $\alpha \in \{a, r\}$ ,  $\beta \in \{+, -\}$ , e  $(i, j, k) \in \{1, \dots, K\}^3$ , já que os saltos não coincidem com probabilidade um.

Visto isso, e com a aplicação do Teorema B.1.12 sabe-se que  $\{m_i^{\alpha,n}\}$ ,  $\alpha \in$

$\{r, d, a, s\}$ , são assintoticamente contínuos. Observe também que

$$\begin{aligned} & \int_0^t q_{ji}^{+,n}(x^n(s-))\mathbb{I}\{x_i^n(s) < B_i\}dm_j^{d,n}(s) \\ &= \int_0^t q_{ji}^{+,n}(x^n(s-))dm_j^{d,n}(s) + \int_0^t q_{ji}^{+,n}(x^n(s-))\mathbb{I}\{x_i^n(s) = B_i\}dm_j^{d,n}(s). \end{aligned} \quad (5.4)$$

Além disso, tem-se que

$$\int_0^t q_{ji}^{+,n}(x^n(s-))dm_j^{d,n}(s) = r_{ji}^+m_j^{d,n}(t) + \frac{1}{\sqrt{n}} \int_0^t f_{ji}^+(x^n(s-))dm_j^{d,n}(s) + o(1/\sqrt{n}). \quad (5.5)$$

Logo, o fato de  $\{m_i^{d,n}\}$  ser assintoticamente contínua implica que

$$1/\sqrt{n} \int_0^t f_{ji}^+(x^n(s-))dm_j^{d,n}(s) + o(1/\sqrt{n}) \Rightarrow 0,$$

onde 0 é o processo zero. Portanto, a integral em relação ao martingale do lado esquerdo da Equação (5.5) também é assintoticamente contínua. O último termo do lado direito da Equação (5.4) converge fracamente para o processo zero pelo Lema B.1.13. Pode-se repetir o argumento acima para o processo  $\int_0^t \mathbb{I}\{x_i^n(s-) > 0\}q_{ji}^{-,n}(x^n(s-))dm_j^{d,n}(s)$ .

Desta forma, o Teorema A.2.3 pode ser aplicado e implica que  $\{z^n, y^n, u^n\}$  é assintoticamente contínua, onde  $u^n \triangleq (u_1^n, \dots, u_k^n)'$ . Do Teorema B.1.12, sabe-se que  $m_i^{\alpha,n} \Rightarrow w_i^\alpha$  para  $\alpha \in \{a, r, d, s\}$ . Além disso, como  $\{y^n/\sqrt{n}\}$  e  $\{u^n/\sqrt{n}\}$  convergem fracamente para o processo zero, tem-se que  $\langle m_i^{\alpha,n} \rangle \Rightarrow \int_0^\cdot r_i^d ds$ , para  $\alpha \in \{a, r, d\}$ , e  $\langle m_i^{s,n} \rangle \Rightarrow \int_0^\cdot \Sigma_i ds$ .

Portanto, foi estabelecido que  $\{x^n, m^{\alpha,n}, z^n, u^n; \alpha \in \{a, d, r, s\}\}$  é assintoticamente contínua, e o limite de qualquer subsequência convergente satisfaz a Equação (5.3).  $\square$

### 5.2.1 Enfraquecendo as Condições Sob as Funções $f_i^\alpha$ e $f_{ij}^\beta$

É comum a necessidade de ter as funções  $f_i^\alpha$  e  $f_{ij}^\beta$  descontínuas. De fato, em se tratando de problemas de controle, como será visto na Seção 5.4 e nos capítulos futuros, o controle ótimo é geralmente uma função descontínua, tipo “curva de chaveamento”, isto é, o controle é aplicado na intensidade máxima em uma região do espaço de estados e não é aplicado nas outras regiões. O teorema abaixo estende o resultado do Teorema 5.2.3, incluindo o caso onde as funções  $f_i^\alpha$  e  $f_{ij}^\beta$  podem ser descontínuas.

**Teorema 5.2.4.** *Suponha as condições do Teorema 5.2.3, mas troque a hipótese sobre a continuidade das funções  $f_i^\alpha$ ,  $f_{ij}^\beta$ , onde  $\alpha \in \{a, r, d\}$ ,  $\beta \in \{+, -\}$ , e  $(i, j) \in \{1, \dots, K\}^2$ , por mensurabilidade. Suponha ainda que os mapeamentos de  $D(\mathbb{R}^K; 0, \infty)$  a  $D(\mathbb{R}; 0, \infty)$  dados por*

$$\phi \mapsto \int_0^\cdot f_i^\alpha(\phi(s))ds \text{ e } \phi \mapsto \int_0^\cdot f_{ij}^\beta(\phi(s))ds,$$

*sejam contínuos com probabilidade um relativo à medida de probabilidade induzida por qualquer limite no sentido fraco de  $x^n$ . Então as conclusões do Teorema 5.2.3 ainda são válidas.*

*Demonstração.* Esta prova segue as idéias do Teorema 8.4.1 na página 327 de (Kushner, 2001). Note que no Teorema 5.2.3 a continuidade das funções  $f_i^\alpha$  e  $f_{ij}^\beta$  somente foram usadas para caracterizar o termo de deriva no limite. Isto é, após mostrar que  $\{x^n\}$  é limitado em probabilidade, qualquer subsequência convergente, como por exemplo  $\{x^{n_i}\}$  com limite  $x$ , satisfaz o seguinte:

$$b_i(x^{n_i}) \Rightarrow b_i(x),$$

pela continuidade de  $b_i$ . Logo,  $\int_0^\cdot b_i(x^{n_i}(s))ds \Rightarrow \int_0^\cdot b_i(x(s))ds$ .

Mas a convergência acima ainda é garantida com hipóteses mais fracas (sob a hipótese deste teorema) através do “Teorema da função contínua” (veja por exemplo

Teorema 2.7 de (Billingsley, 1999) na página 21).  $\square$

Uma condição mais simples de ser verificada e que implica na condição do teorema acima é a seguinte: Seja  $G_d$  o conjunto de pontos de descontinuidade das funções  $f_i^\alpha$  e  $f_{ij}^\beta$  (isto é, de  $b_i$ ). Então, para cada  $\delta > 0$  e  $t > 0$

$$\lim_{\epsilon \rightarrow 0} \mathbb{P}(m\{s \in [0, t] : x_i(s) \in N_\epsilon(G_d)\} \geq \delta) = 0,$$

onde  $N_\epsilon(G_d)$  é uma vizinhança de  $G_d$  e  $m$  é a medida de Lebesgue. Esta mesma condição pode ser encontrada na página 327 de (Kushner, 2001).

Um outro caso de possível interesse é tratar modelos onde as funções  $f_i^\alpha$  não são limitadas mas possuem no máximo um crescimento linear. Para tratar este caso é necessário introduzir uma condição mais restritiva sob as direções de reflexão, a chamada de “condição do conjunto  $B$ ”, dada pela Condição A.1.7. Isto é necessário pois sob esta hipótese é possível obter a Inequação (A.4), que será usada na prova do teorema abaixo.

**Teorema 5.2.5.** *Suponha as condições do Teorema 5.2.3. Suponha ainda que as funções  $f_i^\alpha$ ,  $\alpha \in \{a, r, d\}$ , podem ser ilimitadas mas possuem no máximo um crescimento linear, isto é, existe uma constante  $L > 0$  tal que*

$$\max_{\alpha} \max_i |f_i^\alpha(\xi)| \leq L(1 + |\xi|),$$

para cada  $\xi \in \mathbb{R}_{\geq 0}^K$ . Assuma a Condição A.1.7. Então as conclusões do Teorema 5.2.3 ainda são válidas.

*Demonstração.* Seja  $q_C : \mathbb{R}_{\geq 0}^K \rightarrow [0, 1]$ , para  $C \in \mathbb{N}$ , uma função suave tal que  $0 \leq q_C(\xi) \leq 1$ , se  $C < |\xi| < C + 1$ ,  $q_C(\xi) \equiv 1$ , se  $|\xi| \leq C$ , e  $q_C(\xi) \equiv 0$ , se  $|\xi| \geq C + 1$ . Sejam as funções truncadas dadas por

$$f_i^{\alpha, C}(\xi) \triangleq f_i^\alpha(\xi)q_C(\xi),$$

para  $\alpha \in \{a, r, d\}$ . Além disso, denote por  $\lambda_i^{\alpha, n, C}(\xi) \triangleq r_i^\alpha + f_i^{\alpha, C}(\xi)/\sqrt{n} + o(1/\sqrt{n})$ ,

$\alpha \in \{r, d, a\}$ , e seja  $x_i^{n,C}$  o processo do número de clientes na fila  $i$  escalonado para um sistema de filas que usa as taxas truncadas  $\lambda_i^{\alpha,n,C}$ .

Portanto, pela prova do Teorema 5.2.3 sabe-se que  $\{x^{n,C}\}$  é assintoticamente contínua e possui limite que satisfaz

$$x^C(t) = x(0) + \int_0^t b^C(x^C(s))ds + w(t) + z(t)$$

onde  $b_i^C$  é definida como  $b_i$  mas onde  $f_i^{\alpha,C}$  são usadas no lugar de  $f_i^\alpha$ . Observe também que  $x^{n,C}(t) = x^n(t)$  para todo  $t \leq \tau_C^n \triangleq \inf\{s \geq 0 : |x^n(s)| \geq C\}$ . Portanto, se for estabelecido que qualquer solução da equação

$$x(t) = x(0) + \int_0^t b(x(s))ds + w(t) + z(t)$$

satisfaz

$$\lim_{C \rightarrow \infty} \mathbb{P} \left( \sup_{s \leq t} |x(s)| \geq C \right) = 0, \quad (5.6)$$

o resultado desejado segue.

Desta forma, fica faltando apenas demonstrar a validade da Equação (5.6).

Observe que utilizando o Teorema A.1.9 tem-se que existe um  $C_1 > 0$  tal que

$$\begin{aligned} \sup_{s \leq t} |x(s)| &\leq C_1 \sup_{s \leq t} |x(0) + \int_0^s b(x(u))du + M(s)| \\ &\leq C_1 \sup_{s \leq t} \left[ |x(0)| + \int_0^s |b(x(u))|du + |M(s)| \right]. \end{aligned}$$

Usando o fato que  $\max_k |b_k(\xi)| \leq L_1(1 + |\xi|)$ , para um  $L_1 > 0$ , tem-se que

$$\sup_{s \leq t} |x(s)| \leq C_2 \sup_{s \leq t} \left[ |x(0)| + \int_0^s 1 + |x(u)|du + |M(s)| \right],$$

onde  $C_2 > 0$ . Tomando a esperança de ambos os lados, tem-se

$$\mathbb{E} \left[ \sup_{s \leq t} |x(s)| \right] \leq C_2 \left[ \mathbb{E} [|x(0)|] + \mathbb{E} [|t|] + \mathbb{E} \left[ \sup_{s \leq t} \int_0^s |x(u)|du \right] + \mathbb{E} \left[ \sup_{s \leq t} |M(s)| \right] \right].$$

Observe que, usando o fato de que para qualquer número  $a \in \mathbb{R}$ ,  $a \leq 1 + a^2$ , tem-se que

$$\mathbb{E} \left[ \sup_{s \leq t} |M(s)| \right] \leq 1 + \left( \mathbb{E} \left[ \sup_{s \leq t} |M(s)| \right] \right)^2 \leq 1 + \mathbb{E} \left[ \sup_{s \leq t} |M(s)| \right]^2 \leq 1 + 4\mathbb{E} [M(t)^2],$$

onde na segunda passagem foi usada a inequação de Jensen (veja página 254 de (Ash e Doléans-Dade, 2000)) com  $g(x) = x^2$ , e em seguida o Teorema 20 de (Protter, 1995) na página 12. Então:

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \leq t} |x(s)| \right] &\leq C_3 \left[ 1 + t + \mathbb{E} \left[ \sup_{s \leq t} \int_0^s |x(u)| du \right] \right] \\ &\leq C_3 \left[ 1 + t + \mathbb{E} \left[ \int_0^t \sup_{\nu \leq u} |x(\nu)| du \right] \right]. \end{aligned}$$

Usando o Teorema de Fubini (veja página 108 de (Ash e Doléans-Dade, 2000)), tem-se

$$\mathbb{E} \left[ \sup_{s \leq t} |x(s)| \right] \leq C_3 \left[ 1 + t + \int_0^t \mathbb{E} \left[ \sup_{\nu \leq u} |x(\nu)| \right] du \right].$$

Defina  $y(t) = \mathbb{E} \left[ \sup_{s \leq t} |x(s)| \right]$ , então

$$y(t) \leq C_3 \left( 1 + t + \int_0^t y(u) du \right).$$

Aplicando o Lema de Bellman-Gronwall (veja página 107 de (Arnold, 1974)), chega-se à:

$$y(t) \leq C_4 \left( 1 + t + \int_0^t e^{C_3(t-s)} (1 + s) ds \right) \triangleq H(t),$$

onde  $C_4 > 0$ . Aplicando a inequação de Chebychev (veja página 192 de (Ash e Doléans-Dade, 2000)), tem-se que para cada  $t$ :

$$\mathbb{P} \left( \sup_{s \leq t} |x(s)| \geq C \right) \leq \mathbb{E} \left[ \sup_{s \leq t} |x(s)| \right] / C \leq H(t) / C.$$

Logo tem-se o resultado quando  $C \rightarrow \infty$ . □

### 5.3 Topologias de Rede Satisfazendo a Condição 5.2.2(b)

Nesta seção são colocados dois exemplos de topologia de redes de filas que satisfazem a Condição 5.2.2(b).

#### 5.3.1 Duas Filas em Sequência

Considere duas filas em sequência onde cada fila pode enviar clientes e sinais para a outra, como é descrito pela Figura 5.3. As duas filas também podem receber sinais e clientes de uma fonte externa, e retroalimentação não é permitido, isto é, clientes que saem da fila  $i$  não podem ser roteados imediatamente de volta para fila  $i$  após completarem serviço.

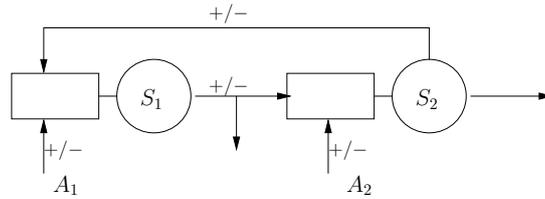


Figura 5.3: Duas filas em sequência com sinais. O símbolo + (resp., -) indica a possibilidade de chegada de clientes (resp., sinais).

Neste caso, as matrizes  $I_r$ ,  $\Theta$ ,  $\Omega^S$  e  $R$  são definidas da seguinte forma:

$$I_r = \begin{pmatrix} r_1^d + r_1^r & 0 \\ 0 & r_2^d + r_2^r \end{pmatrix} \quad \Omega^\emptyset = \begin{pmatrix} r_{21}^- r_2^d & 0 \\ 0 & r_{12}^- r_1^d \end{pmatrix} \quad \Omega^{\{1,2\}} = 0$$

$$\Theta = \begin{pmatrix} 0 & r_1^d(r_{12}^+ - r_{12}^-) \\ r_2^d(r_{21}^+ - r_{21}^-) & 0 \end{pmatrix}$$

$$R = \begin{pmatrix} r_1^d + r_1^r + r_{21}^- r_2^d & -r_2^d(r_{21}^+ - r_{21}^-) \\ -r_1^d(r_{12}^+ - r_{12}^-) & r_2^d + r_2^r + r_{12}^- r_1^d \end{pmatrix}$$

$$R^{\{1,2\}} = \begin{pmatrix} r_1^d + r_1^r & -r_2^d(r_{21}^+ - r_{21}^-) \\ -r_1^d(r_{12}^+ - r_{12}^-) & r_2^d + r_2^r \end{pmatrix}.$$

Portanto, a condição é verificada se existir um  $\alpha = (\alpha_1, \alpha_2)'$  com componentes positivas tais que  $R\alpha = R^{\{1,2\}}e$ . Esta propriedade é facilmente verificada, desde que  $r_1^d, r_2^d > 0$ .

### 5.3.2 Redes com Duas Camadas

Considere agora uma rede de duas camadas onde as filas na primeira camada podem enviar clientes e sinais para as filas da segunda camada, mas o contrário não é possível, veja a Figura 5.4 para referência. Cada fila pode também receber sinais e clientes de uma fonte externa. Suponha que existam  $K_1$  filas na primeira camada e  $K_2$  filas na segunda camada. Defina  $K = K_1 + K_2$ . Ordene os índices das filas de forma que se  $K_1 < i \leq K$ , então a fila  $i$  está na segunda camada.

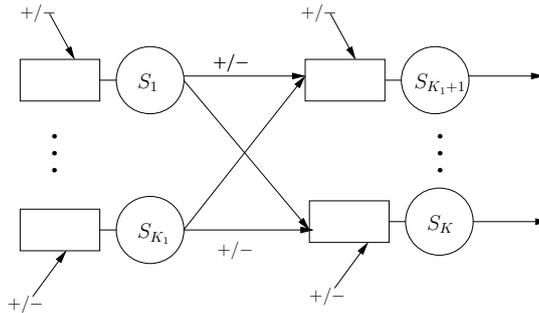


Figura 5.4: Uma rede de filas com duas camadas.

Defina  $Z_1 = \{1, \dots, K_1\}$ . Neste cenário, a matriz  $R^S$  é dada por

$$R^S = \begin{pmatrix} \text{diag}(r_i^d + r_i^r)_{i=1, \dots, K_1} & 0 \\ -\tilde{\Theta}' & \text{diag}\left(r_i^d + r_i^r + \sum_{j \in Z_1 \setminus S} r_{ji}^- r_j^d\right)_{i=K_1+1, \dots, K} \end{pmatrix}$$

onde  $\tilde{\Theta} \in \mathbb{R}^{K_1 \times K_2}$  é definida como

$$\tilde{\Theta} = \begin{pmatrix} r_1(r_{1(K_1+1)}^+ - r_{1(K_1+1)}^-) & \cdots & r_1(r_{1K}^+ - r_{1K}^-) \\ \vdots & & \vdots \\ r_{K_1}(r_{K_1(K_1+1)}^+ - r_{K_1(K_1+1)}^-) & \cdots & r_{K_1}(r_{K_1K}^+ - r_{K_1K}^-) \end{pmatrix}.$$

Para verificar a Condição 5.2.2(b), seja  $S \subseteq \{1, \dots, K\}$  um conjunto ordenado, onde  $|S| \geq 2$ . Defina  $S_1 \subseteq \{1, \dots, K_1\}$  e  $S_2 \subseteq \{K_1 + 1, \dots, K\}$  como conjuntos ordenados tais que  $S_1 \cup S_2 = S$ . Então escolha  $\alpha = (\alpha_1, \dots, \alpha_{|S_1|}, \beta_1, \dots, \beta_{|S_2|})'$  tal que  $\alpha_i = 1$ , para  $i = 1, \dots, |S_1|$ , e

$$\beta_j = \frac{r_k^d + r_k^r + \sum_{l \in Z_1 \setminus S_1} r_{lk}^- r_l^d}{r_k^d + r_k^r + \sum_{l \in Z_1} r_{lk}^- r_l^d}, \quad \text{para } j = 1, \dots, |S_2|,$$

onde  $k$  é o  $j$ -ésimo elemento de  $S_2$ . Agora é possível verificar que  $\tilde{R}\alpha = \tilde{R}^S e$ . Já que isso é válido para qualquer escolha de  $S$ , a Condição 5.2.2(b) é satisfeita.

#### 5.4 Experimentos Numéricos

Para ilustrar o uso da aproximação obtida neste capítulo, considere o sistema da Seção 5.3.1, dado pela Figura 5.3. Será assumido que todo cliente saindo da fila 1 entra na fila 2 como um cliente (e não como um sinal), e a fila 2 não recebe clientes externos. Suponha também que a capacidade máxima de cada fila é finita. Para evitar que a fila 2 fique cheia e perca clientes por causa disto, ela pode enviar sinais para a fila 1. Neste exemplo, será mostrado como o modelo desenvolvido aqui pode ser usado para obter uma estratégia ótima de roteamento de sinais para um sistema operando sob um regime de tráfego pesado.

Como já foi mencionado, um valor de  $n$  grande é escolhido de tal forma que as taxas dos processos de salto do problema que será tratado satisfazem

$$\begin{aligned} \Lambda_i^\alpha(\sqrt{n}\xi) &\triangleq \lambda_i^{\alpha,n}(\xi) \approx r_i^\alpha + f_i^\alpha(\xi)/\sqrt{n}, & \alpha \in \{a, d, r\} \\ Q_{ij}^\beta(\sqrt{n}\xi) &\triangleq q_{ij}^{\beta,n}(\xi) \approx r_{ij}^\beta + f_{ij}^\beta(\xi)/\sqrt{n}, & \beta \in \{+, -\} \end{aligned}$$

e onde a hipótese do tráfego pesado é válida (isto é, Condição 5.2.1(b)). Desta forma, pode-se aproximar o número de clientes em cada fila no instante  $t$  por  $\sqrt{n}x(t/n)$ , onde  $x$  é o processo limite dado pela Equação (5.3).

Para o exemplo tratado aqui, suponha que, para  $\xi \in \mathbb{R}_{\geq 0}^K$ ,

$$\lambda_1^{a,n}(\xi) = \lambda \quad \lambda_2^{a,n}(\xi) = 0 \quad \lambda_1^{r,n}(\xi) = 0 \quad \lambda_2^{r,n}(\xi) = 0$$

e que  $\lambda_1^{d,n}(\xi) = \mu_1$  e  $\lambda_2^{d,n}(\xi) = \mu_2$ , onde  $\mu_1$ ,  $\mu_2$  e  $\lambda$  são constantes positivas. Pela hipótese de tráfego pesado, existem constantes (“pequenas”)  $b_1$  e  $b_2$  tais que  $b_1 = \sqrt{n}(\mu_1 - \lambda)$  e  $b_2 = \sqrt{n}(\mu_2 - \mu_1)$ . Isto é, a taxa de clientes entrando cada fila é bem próxima da taxa de clientes deixando a fila. Portanto,  $\lambda_1^{d,n}(\xi) = \lambda + b_1/\sqrt{n}$  e  $\lambda_2^{d,n}(\xi) = \lambda + (b_1 + b_2)/\sqrt{n}$ . Além disso, supõe-se que

$$\begin{aligned} q_{12}^{+,n}(\xi) &= 1 & q_{21}^{+,n}(x) &= 0 \\ q_{12}^{-,n}(\xi) &= 0 & q_{21}^{-,n}(x) &= g(\xi)/\sqrt{n}, \end{aligned}$$

onde  $g : \mathbb{R}^2 \rightarrow [0, 1]$ . A capacidade de cada fila é dada por  $B_1^n = \sqrt{n}B_1$  para a primeira fila e  $B_2^n = \sqrt{n}B_2$  para a segunda. Desta forma, a aproximação é dada por

$$dx(t) = \begin{pmatrix} -b_1 - \lambda g(x(t)) \\ -b_2 \end{pmatrix} dt + \begin{pmatrix} 2\lambda & -\lambda \\ -\lambda & 2\lambda \end{pmatrix}^{1/2} dw(t) + \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} dy(t) - du(t).$$

É importante ressaltar que a função  $g$  acima somente age na primeira componente de  $x$ , ou seja, na fila 1, apesar da intenção ser em controlar a entrada de clientes na fila 2. Contudo, a segunda componente de  $x$  será afetada indiretamente através do termo de reflexão.

Agora deseja-se escolher um  $g$  que irá reduzir o número de perda de clientes geradas quando a fila 2 está na capacidade máxima. Suponha que  $\lambda = 1$ ,  $b_1 =$

$b_2 = 0.1$ ,  $B_1 = B_2 = 25.6$ , e defina a seguinte função degrau

$$s(\xi) = \begin{cases} 1 & \text{se } x_2 > 15 \\ 0 & \text{caso contrário.} \end{cases}$$

A Figura 5.5 compara a mesma realização de  $x$  com  $g$  igual à  $s$  e  $g$  igual à função zero, para a condição inicial  $x(0) = (B_1, B_2)'$ . Observe que a função amostra sem controle (isto é,  $g \equiv 0$ ) atinge a capacidade máxima da fila 2 mais frequentemente. As funções amostras foram construídas com o método de Euler (veja (Kloeden et al., 1994) página 110). O parâmetro de discretização do tempo foi fixado à  $h = 0.01$ . O termo de reflexão foi implementado empurrando o processo de volta ao espaço de estado (na direção do vetor de reflexão) toda vez que o processo cruzava uma fronteira.

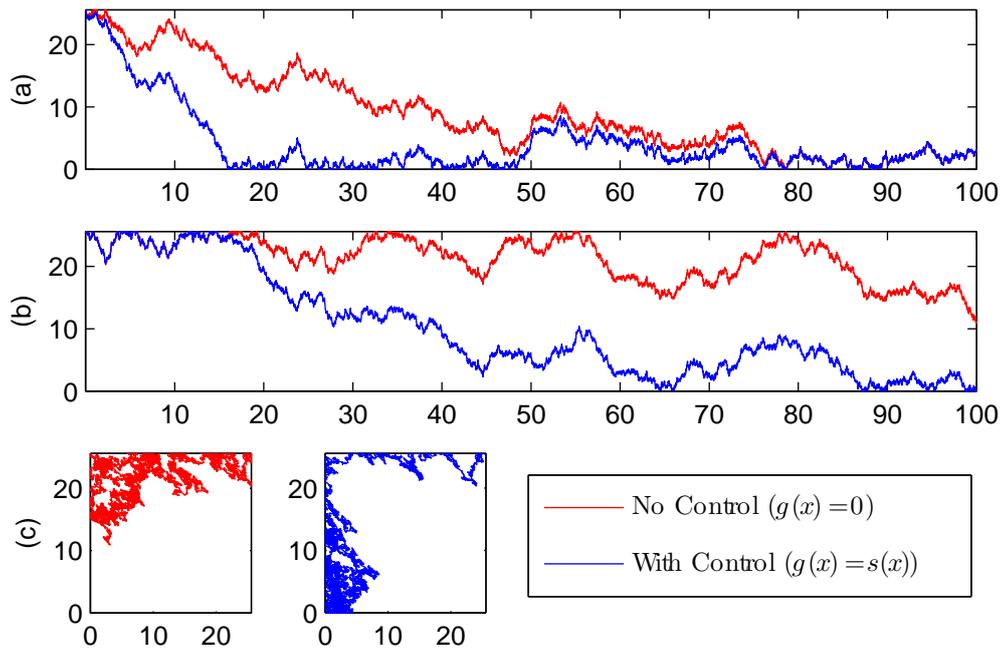


Figura 5.5: Função amostra de  $x$  com  $g \equiv s$  (azul) e  $g \equiv 0$  (vermelho). (a) Gráfico de  $x_1$  no tempo. (b) Gráfico de  $x_2$  no tempo. (c) Gráfico de  $x_1$  versus  $x_2$ .

É possível também calcular a escolha ótima de  $g$  com relação à uma função

de custo. Para este exemplo será usado a seguinte função de custo descontado:

$$W(x_0, g) = \mathbb{E}_{x_0}^g \left[ \int_0^\infty e^{-\beta t} [cg(x(t))dt + vdu_2(t)] \right],$$

onde  $x_0$  é a condição inicial, e  $c$  e  $v$  são constantes associadas ao custo de rotar um sinal (ou de remover um cliente na primeira fila) e o custo de perder um cliente na fila 2, respectivamente.

O método da aproximação por cadeia de Markov (veja, (Kushner e Dupuis, 1992; Jarvis e Kushner, 1996)) foi usado para encontrar o controle ótimo (ou aproximadamente ótimo) numericamente. Os parâmetros usados foram os seguintes:  $\beta = 0.01$ , e  $h = 0.1$ , onde  $h$  é o parâmetro de discretização. A Figura 5.6 mostra o controle obtido com escolhas diferentes de  $c$  e  $v$ . Observe que o controle ótimo é uma “curva de chaveamento”, isto é, ele é descontínuo e aplica o controle na força máxima em uma região do espaço de estado.

É interessante ver a forma das curvas dadas pela Figura 5.6. Observe que as curvas movem para cima no lado direito do gráfico. Isto pode ser explicado pelo atraso na ação do controle na fila 2. Quando a fila 1 e a fila 2 estão quase cheias, provavelmente haverá perda de clientes na fila 2 mesmo se sinais forem enviados para a fila 1. Portanto, neste caso, não há necessidade de enviar sinais.

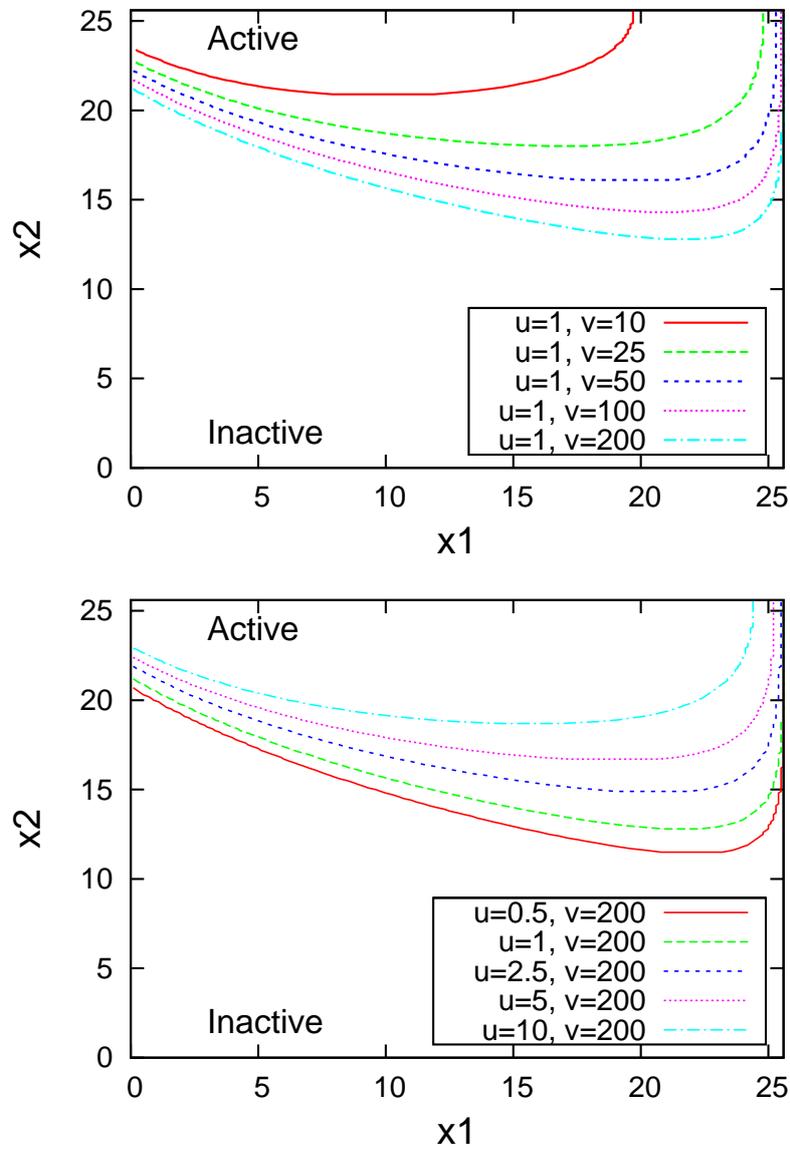


Figura 5.6: Curvas de chaveamento dadas pelo controle ótimo para diferentes valores de  $c$  e  $v$ . O controle é aplicado em força máxima na região denominada de *Active* e não é aplicado na região denominada de *Inactive*.

# Capítulo 6

## Estudo de Caso: Aplicação para Sistemas de Processamento Paralelo

O foco principal deste capítulo é a utilização prática da aproximação por tráfego pesado de um sistema de filas com sinais. A aplicação considerada aqui é a de um servidor de busca na Internet. Contudo, como ficará claro ao decorrer do capítulo, o mesmo modelo tem aplicações em vários outros sistemas de processamento paralelo. A idéia é usar sinais para balancear a rede e usar a aproximação por tráfego pesado para encontrar a estratégia ótima para este balanceamento. Para isso, um tipo de sinal que ainda não foi considerado nos capítulos anteriores será usado. Este sinal moverá um cliente de uma fila para outra. Na literatura, este sinal é usualmente chamado de “*trigger*” (Gelenbe, 1993; Artalejo, 2000). O conteúdo deste capítulo é referente aos trabalhos (Leite e Fragoso, 2008b,c).

O sistema de filas considerado aqui será do tipo *fork-join* que é usualmente encontrado em problemas de análise de performance em sistemas computacionais de processamento paralelo. Usualmente, este tipo de sistema é constituído por  $K$  filas em paralelo. Quando um cliente chega no sistema, ele é dividido em  $K$  frações (ou tarefas) e cada uma destas frações é enviada para uma fila (ou processador) diferente. O serviço de cada cliente só termina quando todas as tarefas deixam as suas respectivas filas (para maiores detalhes sobre sistemas de filas do tipo *fork-join* veja, por exemplo, (Boxma et al., 1994)).

Em várias aplicações, o tempo de resposta de um sistema de processamento

paralelo é de crucial importância. Como cada cliente somente deixa o sistema quando cada uma de suas  $K$  tarefas completam o serviço, o tempo de resposta será maior quando o sistema estiver *desbalanceado*. Devido à natureza estocástica destes sistemas, o desbalanceamento pode aparecer tanto em sistemas heterogêneos (redes compostas por filas com velocidade de processamento diferente) como nos homogêneos. Para reduzir este desbalanceamento, uma estratégia natural é distribuir melhor os trabalhos dados a cada fila. Existem várias formas de se fazer isso. O controle pode ser aplicado no momento da chegada do cliente atribuindo mais de uma tarefa a uma fila se uma outra estiver sobrecarregada. Esta estratégia é usualmente chamada de controle centralizado. De forma alternativa, o controle pode ser aplicado por cada processador individualmente. Neste caso, uma fila que não está sobrecarregada pode “pegar” trabalho de uma fila vizinha sobrecarregada, este é usualmente chamado de controle distribuído (Eager et al., 1985).

Controle distribuído parece a estratégia mais apropriada para reduzir o tempo de resposta em sistemas do tipo *fork-join*. Isso é devido a como a prioridade de serviço pode ser implementada em cada esquema. Por exemplo, em uma rede homogênea desbalanceada, sabe-se que existem tarefas atrasadas já que cada fila recebe serviço no mesmo instante. Desta forma, uma fila subutilizada deveria dar prioridade para tarefas de clientes que chegaram primeiro. Em um esquema de controle centralizado, as estações subutilizadas recebem tarefas de clientes que chegaram por último, ao contrário do esquema de controle distribuído, onde prioridade pode ser dada para tarefas mais antigas.

Para ilustrar a afirmação do parágrafo acima, suponha que existam duas estações em paralelo (chamadas de A e B). Cada fila recebe uma fração do trabalho dos clientes que chegam no sistema. Considere o cenário onde três clientes chegam no sistema e a fila A termina sua fração do serviço antes que B termine a primeira tarefa, veja a Figura 6.1(a) para referência. Em seguida, um quarto cliente chega no sistema e três casos diferentes são considerados. No primeiro caso, dado pela Figura 6.1(b), um sistema sem estratégia de balanceamento é considerado. No

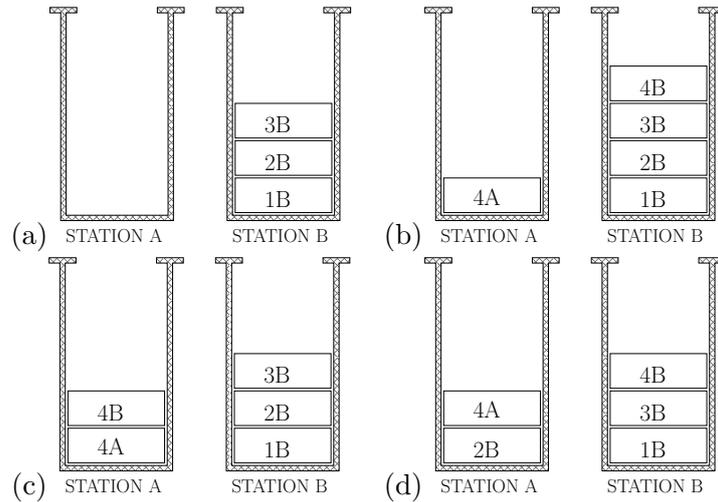


Figura 6.1: Diagrama de duas estações em paralelo. (a) Filas antes da chegada do quarto cliente. (b) Um sistema sem estratégia de balanceamento. (c) Um sistema com controle centralizado. (d) Um sistema com controle distribuído.

segundo caso, Figura 6.1(c), tem-se um sistema que utiliza um controle centralizado que atribui à fila A as duas frações do trabalho do quarto cliente. No último caso, Figura 6.1(d), é considerado um sistema com controle distribuído, onde cada tarefa é atribuída às filas correspondentes mas em seguida a fila A move um cliente da fila B para tentar manter o sistema balanceado. Apesar dos dois sistemas com controle balancear as filas, somente o controle distribuído pode dar prioridade para trabalhos mais antigos.

Este controle distribuído corresponde aos sinais do tipo *trigger*, já que eles movem um cliente de uma fila para a outra. Contudo, como mover tarefas entre estações de processamento pode ser custoso, por causa do atraso na comunicação entre as filas, ou no atraso no processamento (isto é, cada estação pode ser especializada na sua fração do trabalho, mas pode efetuar o trabalho de uma fila vizinha sob um custo computacional maior), a questão de quando é o momento ideal para mover uma tarefa entre as estações não é simples. E através da aproximação por difusão é possível tratar este problema como um de controle ótimo.

Um problema similar ao que é tratado aqui já foi estudado por alguns autores na literatura sobre o nome de “*resource pooling*” (veja por exemplo, (Harrison, 1998; Harrison e López, 1999; Williams, 1999; Bell e Williams, 2001; Kushner,

2001)). Este problema consiste em atribuir tempo de processadores à diferentes filas em paralelo. Geralmente nestes trabalhos, o processo de carga de trabalho é considerado como um invariante no sistema, e um controle ótimo assintótico é buscado. A diferença principal entre os modelos é a introdução dos sinais *triggers* que move cliente entre as filas ao invés de tratar o problema de atribuir tempo de processador à filas diferentes.

## 6.1 Modelo de Filas

Como foi feito no capítulo anterior, o modelo de filas considerado aqui será do tipo Poisson com taxas dependentes do estado. Os clientes chegam no sistema de uma mesma fonte externa e seus trabalhos são divididos entre as filas imediatamente. Cada fração do trabalho de cada cliente é assumido possuir tempo de serviço independente. Quando uma tarefa é terminada, a estação de onde ela saiu pode enviar um sinal para uma outra estação. Será usada a interpretação de que a tarefa foi “roteada” como um sinal para a outra estação de processamento. Como discutido anteriormente, o sinal tem o efeito de mover uma tarefa da fila que recebe o sinal para a fila que envia o sinal.

O modelo se restringe ao caso de somente duas filas em paralelo. Utilizando este modelo simplificado, é possível evitar complicações referentes às direções de reflexão presentes nas quinas do espaço de estados não serem combinações lineares positivas das direções nas faces adjacentes, como foi discutido no capítulo anterior (o qual deu a origem para a Condição 5.2.2(b)). Além disso, é importante mencionar que os métodos numéricos para encontrar controles ótimos para sistemas contendo várias componentes (isto é, centenas ou milhares, como é o caso de sistemas de processamento paralelos modernos) é impraticável no ponto de vista de custo computacional. Apesar de impor esta restrição no modelo teórico, na prática é possível usar a estratégia ótima obtida para um modelo com duas filas em um sistema com várias estações. A razão disto é que a decisão de mover tarefas entre filas é sempre feita entre duas filas. Na Seção 6.3.1 será discutido com mais

detalhes como isto pode ser feito.

Como feito anteriormente, é assumido que cada fila pode ter uma capacidade máxima finita. Neste caso, uma tarefa só será movida através de um sinal se a fila que receberá a tarefa possuir espaço suficiente disponível. A capacidade de cada fila é dada pela constante  $\tilde{B}_i \in \mathbb{R}_{>0} \cup \{\infty\}$ . Caso a fila contenha capacidade ilimitada,  $\tilde{B}_i \equiv \infty$ .

O processo que conta o número de cliente em cada fila é dado por

$$X_i(t) \triangleq X_i(0) + E(t) - D_i(t) + C_i(t) - C_j(t),$$

para  $i \in \{1, 2\}$ ,  $i \neq j$ ,

$$\begin{aligned} A(t) &\triangleq N^a \left( \int_0^t \Lambda^a(X(s)) ds \right), \\ S_i(t) &\triangleq N_i^d \left( \int_0^t \Lambda_i^d(X(s)) ds \right), \\ D_i(t) &\triangleq \int_0^t \mathbb{I}\{X_i(s-) > 0\} dS_i(s), \\ E(t) &\triangleq \int_0^t \mathbb{I}\{X_i(s-) < \tilde{B}_i, X_j(s-) < \tilde{B}_j\} dA(s), \\ C_i(t) &\triangleq \int_0^t \mathbb{I}_i^c(s) \mathbb{I}\{X_j(s-) > 0, X_i(s-) < \tilde{B}_i\} dD_i(s), \end{aligned}$$

onde  $N_i^d$  e  $N^a$  são processos de Poisson padrões, e  $\Lambda^a : \mathbb{R}_{>0}^2 \rightarrow \mathbb{R}_{>0}$  são funções mensuráveis. O processo estocástico tomando valores  $\mathbb{I}_i^c(t)$  denota a função indicadora do evento que uma tarefa deixando a fila  $i$  no instante  $t$  é roteado para a outra fila como um *trigger*.

Os processos  $E$ ,  $D_i$  e  $C_i$  são interpretados como o seguinte:  $E(t)$  denota o número total de clientes que chegaram no sistema da fonte externa até o tempo  $t$ , e  $D_i(t)$  denota o número de tarefas completadas pela estação  $i$  até o tempo  $t$ . O processo  $C_1(t)$  (resp.,  $C_2(t)$ ) denota o número total de tarefas que foram movidas da fila 1 (resp., fila 2) para entrar na fila 2 (resp., fila 1) por causa de um sinal até o tempo  $t$ .

Seja a filtração  $\mathcal{F}_t$  a mínima  $\sigma$ -álgebra que mensura todos os processos descritos acima até o instante  $t$ . Seja também  $\mathcal{F}_t^r$  a  $\sigma$ -álgebra definida como  $\mathcal{F}_t$  mas que não inclui qualquer decisão de roteamento no instante  $t$ .

**Condição 6.1.1.** *Para  $i \in \{1, 2\}$ , suponha que:*

- (a) *Os elementos aleatórios  $\{X_i(0), N^a, N_i^d; i \in \{1, 2\}\}$  são mutuamente independentes;*
- (b) *As funções  $\Lambda^a$  e  $\Lambda_i^d$  são mensuráveis e limitadas;*
- (c)  *$\mathbb{E}[\mathbb{I}_i^c(t) | \mathcal{F}_t^r] = Q_i^c(X(t-))$ , onde  $Q_i^c : \mathbb{R}_+^2 \rightarrow [0, 1]$  é uma função mensurável.*

Como feito anteriormente, os processos de salto definidos acima tem uma decomposição de martingale da seguinte forma:  $M^a = E - \hat{E}$  e  $M_i^d = D_i - \hat{D}_i$ , onde  $M^a$  e  $M^d$  são  $\mathcal{F}_t$ -martingales, e  $\hat{E}$ ,  $\hat{D}_i$  são processos  $\mathcal{F}_t$ -previsíveis dados por

$$\begin{aligned}\hat{E}(t) &= \int_0^t \mathbb{I}\{X_i(s) < \tilde{B}_i, X_j(s) < \tilde{B}_j\} \Lambda^a(X(s)) ds, \\ \hat{D}_i(t) &= \int_0^t \mathbb{I}\{X_i(s) > 0\} \Lambda_i^d(X(s)) ds.\end{aligned}$$

Defina  $M_i^c(t) = \int_0^t (\mathbb{I}_i^c(s) - Q_i^c(X(s-))) \mathbb{I}\{X_j(s-) > 0, X_i(s-) < \tilde{B}_i\} dD_i(s)$ .

Usando o mesmo argumento usado na Seção 5.1 tem-se que  $M_i^c$  é um  $\mathcal{F}_t$ -martingale.

Além disso, pode-se escrever o seguinte:

$$\begin{aligned}C_i(t) &= M_i^c(t) + \int_0^t Q_i^c(X(s-)) \mathbb{I}\{X_j(s-) > 0, X_i(s-) < \tilde{B}_i\} dD_i(s) \\ &= M_i^c(t) + \int_0^t Q_i^c(X(s-)) \mathbb{I}\{X_j(s-) > 0, X_i(s-) < \tilde{B}_i\} dM_i^d(s) \\ &\quad + \int_0^t Q_i^c(X(s)) \mathbb{I}\{X_i(s) > 0, X_j(s) > 0, X_i(s) < \tilde{B}_i\} \Lambda_i^d(X(s)) ds.\end{aligned}$$

Portanto,  $X$  possui a seguinte decomposição em semi-martingale:  $X_i(t) = X_i(0) +$

$B_i(t) + M_i(t)$ , onde

$$\begin{aligned}
B_i(t) &= \int_0^t \mathbb{I}\{X_i(s) < \tilde{B}_i, X_j(s) < \tilde{B}_j\} \Lambda^a(X(s)) - \mathbb{I}\{X_i(s) > 0\} \Lambda_i^d(X(s)) \\
&\quad + Q_i^c(X(s)) \mathbb{I}\{X_i(s) > 0, X_j(s) > 0, X_i(s) < \tilde{B}_i\} \Lambda_i^d(X(s)) \\
&\quad - Q_j^c(X(s)) \mathbb{I}\{X_j(s) > 0, X_i(s) > 0, X_j(s) < \tilde{B}_j\} \Lambda_j^d(X(s)) ds \\
M_i(t) &= M^a - M_i^d + M_i^c(t) + \int_0^t Q_i^c(X(s-)) \mathbb{I}\{X_j(s-) > 0, X_i(s-) < \tilde{B}_i\} dM_i^d(s) \\
&\quad - M_j^c(t) - \int_0^t Q_j^c(X(s-)) \mathbb{I}\{X_i(s-) > 0, X_j(s-) < \tilde{B}_j\} dM_j^d(s).
\end{aligned}$$

## 6.2 Limite em Regime de Tráfego Pesado

Seja  $\{X^n\}$  uma sequência de sistemas de filas como definido na seção anterior para cada  $n \in \mathbb{N}$ . Defina o processo escalonado  $x^n$  como  $x^n(t) \triangleq X^n(t)/\sqrt{n}$ . Além disso, sejam  $m^{a,n}$ ,  $m_i^{d,n}$ , e  $m_i^{c,n}$  os martingales  $M^{a,n}$ ,  $M_i^{d,n}$ , e  $M_i^{c,n}$  com o mesmo escalonamento usando para  $x^n$ . Assuma ainda que o tamanho da capacidade máxima de cada fila é dado por  $\tilde{B}_i^n \triangleq \sqrt{n}B_i$ , para  $i \in \{1, 2\}$ , onde  $B_i$  é uma constante positiva.

A seguinte condição precisa ser introduzida. Ela é equivalente à Condição 5.2.1 usada no capítulo anterior.

**Condição 6.2.1.** *Para  $o(\cdot)$  uniforme em  $\xi \in \mathbb{R}^K$ , assuma que:*

- (a) *Existem constantes  $r^a, r_i^d \in (0, \infty)$  e  $r_i^c \in [0, 1)$ , e funções limitadas e contínuas  $f^a(\cdot)$ ,  $f_i^d(\cdot)$  e  $f_i^c(\cdot)$ ,  $i = 1, 2$ , tais que para qualquer  $\xi \in \mathbb{R}_+^2$*

$$\begin{aligned}
\lambda^{a,n}(\xi) &= r^a + f^a(\xi)/\sqrt{n} + o(1/\sqrt{n}), \\
\lambda_i^{d,n}(\xi) &= r_i^d + f_i^d(\xi)/\sqrt{n} + o(1/\sqrt{n}), \\
q_i^{c,n}(\xi) &= r_i^c + f_i^c(\xi)/\sqrt{n} + o(1/\sqrt{n}).
\end{aligned}$$

- (b)  *$r^a + r_i^d r_i^c = r_i^d + r_j^d r_j^c$ , para  $i, j \in \{1, 2\}$ , onde  $i \neq j$ , que é a condição de tráfego pesado.*

Neste modelo não é necessário introduzir condições sobre as direções de reflexão. Isso acontece pois o modelo contém somente duas filas e, portanto, satisfaz automaticamente as condições necessárias quando as constantes  $r^a$  e  $r_i^d$  são estritamente positivas e  $r_i^c$  é estritamente menor do que um, como é o caso aqui.

**Teorema 6.2.2.** *Suponha que  $x^n(0) \Rightarrow x(0)$ . Além disso, assuma as Condições 6.1.1 e 6.2.1. Então  $\{x^n\}$  é limitado em probabilidade e qualquer subsequência convergente de  $\{x^n\}$  tem como limite um processo que satisfaz:*

$$x(t) = x(0) + \int_0^t b(x(s))ds + w(t) + z(t), \quad (6.1)$$

$0 \leq x_i(t) \leq B_i$ , se  $B_i < \infty$ , e  $0 \leq x_i(t) < \infty$ , se  $B_i \equiv \infty$ ,  $i \in \{1, 2\}$ , onde

$$\begin{aligned} b_i(\xi) &= f^a(\xi) - f_i^d(\xi)(1 - r_i^c) + r_i^d f_i^c(\xi) - r_j^d f_j^c(\xi) - r_j^c f_j^d(\xi), \\ w_i(t) &= w^a(t) - w_i^d(t)(1 - r_i^c) - w_i^c(t) - w_j^c(t) - r_j^c w_j^d(t), \end{aligned}$$

para  $i, j \in \{1, 2\}$ ,  $i \neq j$ . Os processos  $w^a(\cdot)$ ,  $w_i^d(\cdot)$ , e  $w_i^c(\cdot)$  são processos de Wiener independentes com variâncias  $r^a$ ,  $r_i^d$ , e  $r_i^c(1 - r_i^c)$ , respectivamente. O processo  $z(t) \triangleq R\zeta(t)$  é o processo de reflexão, onde

$$R \triangleq \begin{pmatrix} 1 & 1 - r_1^d/r^a & -1 - r_1^c r_1^d/r^a & -1 + r_2^c r_2^d/r^a \\ 1 - r_2^d/r^a & 1 & -1 + r_1^c r_1^d/r^a & -1 - r_2^c r_2^d/r^a \end{pmatrix},$$

$\zeta(\cdot) = (y_1(\cdot), y_2(\cdot), u_1(\cdot), u_2(\cdot))'$ , e  $y_i(\cdot)$  (resp.,  $u_i(\cdot)$ ) é não-decrescente, contínuo, e aumenta somente nos instante  $t$  tais que  $x_i(t) = 0$  (resp.,  $x_i(t) = B_i$ ), além disso,  $y_i(0) = 0$  e  $u_i(0) = 0$ .

*Demonstração.* Sem perda de generalidade, suponha que  $B_i < \infty$ ,  $i \in \{1, 2\}$ . Como feito na prova do Teorema 5.2.3, o caso geral segue de forma análoga fazendo  $\sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = B_i\}ds \equiv 0$ , caso  $B_i = \infty$ . Usando a decomposição do processo

$X^n$  pode-se escrever o seguinte:

$$x_i^n(t) = x_i^n(0) + \sqrt{n} \int_0^t b_i^n(x^n(s)) ds + m_i^n(t),$$

onde

$$\begin{aligned} \sqrt{n} \int_0^t b_i^n(x^n(s)) ds &= \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) < B_i, x_j^n(s) < B_j\} \lambda^{a,n}(x^n(s)) \\ &\quad - \mathbb{I}\{x_i^n(s) > 0\} \lambda_i^{d,n}(x^n(s)) + \mathbb{I}\{x_i^n(x) > 0, x_j^n(s) > 0\} \\ &\quad \times \left[ \mathbb{I}\{x_i^n(s) < B_i\} q_i^{c,n}(x^n(s)) \lambda_i^{d,n}(x^n(s)) - \mathbb{I}\{x_j^n(s) < B_j\} q_j^{c,n}(x^n(s)) \lambda_j^{d,n}(x^n(s)) \right] ds \end{aligned} \quad (6.2)$$

para  $i, j \in \{1, 2\}$ ,  $i \neq j$ , e

$$\begin{aligned} m_i^n(t) &= m^{a,n} - m_i^{d,n} \\ &\quad + m_i^{c,n}(t) + \int_0^t q_i^{c,n}(x^n(s-)) \mathbb{I}\{x_j^n(s-) > 0, x_i^n(s-) < B_i\} dm_i^{d,n}(s) \\ &\quad - m_j^{c,n}(t) - \int_0^t q_j^{c,n}(x^n(s-)) \mathbb{I}\{x_i^n(s-) > 0, x_j^n(s-) < B_j\} dm_j^{d,n}(s) \end{aligned}$$

é um martingale. Defina os seguintes processos

$$\begin{aligned} y_i^n(t) &\triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = 0\} ds, \\ u_i^n(t) &\triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = B_i\} ds, \\ y_{ij}^n(t) &\triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = 0, x_j^n(s) = 0\} ds, \\ u_{ij}^n(t) &\triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = B_i, x_j^n(s) = B_j\} ds, \\ w_{ij}^n(t) &\triangleq \sqrt{n} \int_0^t \mathbb{I}\{x_i^n(s) = B_i, x_j^n(s) = 0\} ds. \end{aligned}$$

Observe que  $y_i^n(t)$  denota o tempo total (até o instante  $t$ ) que o processo  $x^n$  passa na fronteira do espaço de estado  $\partial G_i \triangleq \{\xi \in \mathbb{R}^2 : \xi_i = 0\}$ . De forma análoga,  $u_i^n(t)$  denota o tempo total (até o instante  $t$ ) que o processo  $x^n(\cdot)$  passa na fronteira

$\partial G_i^b \triangleq \{\xi \in \mathbb{R}^2 : \xi_i = B_i\}$ . Os processos  $u_{ij}^n(t) = u_{ji}^n(t)$ ,  $y_{ij}^n(t) = y_{ji}^n(t)$ , e  $w_{ij}^n(t)$ , para  $i \neq j$ , denotam o tempo total (até o tempo  $t$ ) que o processo  $x^n$  está em cada quina do espaço de estado.

De forma similar ao que foi feito na prova do Teorema 5.2.3, reescreve-se a Equação (6.2) usando a Condição 6.2.1 para obter o seguinte:

$$\begin{aligned}
& \int_0^t f^a(x^n(s)) - f_i^d(x^n(s)) + r_i^d f_i^c(x^n(s)) + r_i^c f_i^d(x^n(s)) - r_j^c f_j^d(x^n(s)) - r_j^d f_j^c(x^n(s)) ds \\
& + (r_i^d - r_i^c r_i^d + r_j^c r_j^d) y_i^n(t) + (-r_i^c r_i^d + r_j^c r_j^d) y_j^n(t) + (r_i^c r_i^d - r_j^c r_j^d) y_{ij}^n(t) \\
& - (r^a + r_i^c r_i^d) u_i^n(t) - (r^a - r_j^c r_j^d) u_j^n(t) + r^a u_{ij}^n(t) + r_i^c r^d w_{ij}^n(t) - r_j^c r_j^d w_{ji}^n(t) \\
& + O(y_i^n(t)/\sqrt{n}) + O(y_j^n(t)/\sqrt{n}) + O(u_i^n(t)/\sqrt{n}) + O(u_j^n(t)/\sqrt{n}) + \sqrt{no}(1/\sqrt{n}).
\end{aligned} \tag{6.3}$$

Para escrever a equação acima de forma mais compacta, seja

$$\zeta^n(\cdot) \triangleq (y_1^n(\cdot), y_2^n(\cdot), u_1^n(\cdot), u_2^n(\cdot))',$$

e  $M_k \in \mathbb{R}^{2 \times 4}$ ,  $k \in \{0, 1, 2, 3, 4\}$ , definidos como  $M_0 \triangleq (d_1^0, d_2^0, d_3^0, d_4^0)$ ,  $M_1 \triangleq (d_1^1, d_2^1, d_3^0, d_4^0)$ ,  $M_2 \triangleq (d_1^0, d_2^2, d_3^2, d_4^0)$ ,  $M_3 \triangleq (d_1^0, d_2^0, d_3^3, d_4^3)$ , e  $M_4 \triangleq (d_1^4, d_2^0, d_3^0, d_4^4)$ , onde os vetores coluna  $d_q^p$ ,  $p, q \in \{1, 2, 3, 4\}$  são dados por:

$$\begin{aligned}
d_1^0 &= (r_1^d(1 - r_1^c) + r_2^c r_2^d, -r_2^c r_2^d + r_1^c r_1^d)' & d_2^0 &= (-r_1^c r_1^d + r_2^c r_2^d, r_2^d(1 - r_2^c) + r_1^c r_1^d)' \\
d_3^0 &= (-r^a - r_1^c r_1^d, -r^a + r_1^c r_1^d)' & d_4^0 &= (-r^a + r_2^c r_2^d, -r^a - r_2^c r_2^d)' \\
d_1^1 &= (r_1^d, -r_2^c r_2^d + r_1^c r_1^d)' & d_2^1 &= (-r_1^c r_1^d + r_2^c r_2^d, r_2^d)' \\
d_2^2 &= (-r_1^c r_1^d + r_2^c r_2^d, r_2^d(1 - r_2^c))' & d_3^2 &= (-r^a, -r^a + r_1^c r_1^d)' \\
d_3^3 &= (-r_1^c r_1^d, -r^a + r_1^c r_1^d)' & d_4^3 &= (-r^a + r_2^c r_2^d, -r_2^c r_2^d)' \\
d_1^4 &= (r_1^d(1 - r_1^c), -r_2^c r_2^d + r_1^c r_1^d)' & d_2^4 &= (-r^a + r_2^c r_2^d, -r^a)'.
\end{aligned}$$

Os termos presentes na Equação (6.3) relacionados com  $y_i$ ,  $y_{ij}$ ,  $u_i$ ,  $u_{ij}$ , e  $w_{ij}$ , podem ser escrito com notação matricial como  $\int_0^t M(x^n(s)) d\zeta^n(s)$  onde a função que toma

valores matriciais  $M : \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 4}$  é definida como

$$M(\xi) \triangleq \begin{cases} M_1 & \text{se } \xi_1 = 0, \xi_2 = 0 \\ M_2 & \text{se } \xi_1 = B_1, \xi_2 = 0 \\ M_3 & \text{se } \xi_1 = B_1, \xi_2 = B_2 \\ M_4 & \text{se } \xi_1 = 0, \xi_2 = B_2 \\ M_0 & \text{caso contrário} \end{cases}$$

Ignorando os termos  $O(\cdot)$  e  $o(\cdot)$  que convergem para o processo zero (como visto na prova do Teorema 5.2.3), tem-se que

$$\begin{aligned} x^n(t) &= x^n(0) + \int_0^t b(x^n(s))ds + m^n(t) + z^n(t) \\ z^n(t) &= \int_0^t M(x^n(s))\zeta^n(s). \end{aligned}$$

Como as constantes  $r_i^d$  e  $r^a$  são estritamente positivas, e já que  $0 \leq r_i^c < 1$ , pode-se verificar que existem  $\alpha^p \in \mathbb{R}^2 \setminus \{0\}$ , tais que  $\alpha_i^p \geq 0$ ,  $p \in \{1, 2, 3, 4\}$ , e

$$\begin{aligned} \alpha_1^1 d_1^0 + \alpha_2^1 d_2^0 &= d_1^1 + d_2^1, & \alpha_1^2 d_2^0 + \alpha_2^2 d_3^0 &= d_2^2 + d_3^2 \\ \alpha_1^3 d_3^0 + \alpha_2^3 d_4^0 &= d_3^3 + d_4^3, & \alpha_1^4 d_1^0 + \alpha_2^4 d_4^0 &= d_1^4 + d_4^4. \end{aligned}$$

Portanto, usando um argumento similar ao usado no Teorema 5.2.3, pode-se definir  $\bar{\zeta}^n(\cdot)$  de tal forma que

$$z^n(t) = \int_0^t M(x^n(s))\zeta^n(s) = M_0 \bar{\zeta}^n(t).$$

Além disso, usando as igualdades requeridas na condição do tráfego pesado, pode-se reescrever  $z^n$  na forma mais simples e conveniente, dada por  $z^n(t) = R\tilde{\zeta}^n(t)$ .

Para mostrar que  $m_i^n$  é assintoticamente contínuo, segue-se o roteiro usado no Teorema 5.2.3 e computa-se os processos de Doob-Meyer associados à  $m^{a,n}$ ,

$m_i^{d,n}$  e  $m_i^{c,n}$ , que são dados respectivamente por:

$$\begin{aligned}
\langle m^{a,n} \rangle(t) &= \int_0^t \mathbb{I}\{x_i^n(s) < B_i, x_j^n(s) < B_j\} \lambda^{a,n}(x^n(s)) ds, \\
\langle m_i^{d,n} \rangle(t) &= \int_0^t \mathbb{I}\{x_i^n(s) > 0\} \lambda_i^{d,n}(x^n(s)) ds, \\
\langle m_i^{c,n} \rangle(t) &= \int_0^t (1 - q_i^{c,n}(x^n(s))) q_i^{c,n}(x^n(s)) \mathbb{I}\{x_i^n(s) > 0, x_j^n(s) > 0, x_i^n(s) < B_i\} \\
&\quad \times \lambda_i^{d,n}(x^n(s)) ds. \tag{6.4}
\end{aligned}$$

Além disso,  $\langle m^a, m_i^d \rangle = \langle m^a, m_i^c \rangle = \langle m_i^d, m_i^c \rangle = 0$ , para  $i, j \in \{1, 2\}$ , pelos mesmos argumentos usados no Teorema 5.2.3. Pode-se então usar o Teorema B.1.12 juntamente com o fato de  $\lambda^a(\cdot)$ ,  $\lambda_i^d(\cdot)$  e  $q_i^c(\cdot)$  serem limitados, e que os termos de martingale terem saltos do tamanho  $1/\sqrt{n}$  para mostrar que eles são assintoticamente contínuos.

O Teorema A.2.3 pode ser usado agora para mostrar que  $\{z^n\}$  é limitado em probabilidade, onde a condição A.1.14 é satisfeita sob a hipótese estabelecida na Condição 6.2.1(a) para as constantes  $r^a$ ,  $r_i^d$ , e  $r_i^c$ . Observe que qualquer subsequência convergente de  $\{z^n\}$  tem como limite fraco o processo  $z(\cdot) \triangleq R\zeta(\cdot)$ , onde  $\zeta$  é definido no enunciado. Além disso, observe que  $\{y_i^n/\sqrt{n}\}$  e  $\{u_i^n/\sqrt{n}\}$  convergem fracamente para o processo zero e, portanto, as funções indicadoras em (6.4) podem ser descartadas sem alteração no limite, e usando a Condição 6.2.1(a), tem-se que

$$\int_0^t q_i^{c,n}(x^n(s-)) \mathbb{I}\{x_j^n(s-) > 0, x_i^n(s-) < B_i\} dm_i^{d,n}(s) = r_i^c m_i^d(s)$$

modulo um erro negligenciável que converge para o processo zero a medida que  $n \rightarrow \infty$ . Para mostrar que qualquer limite no sentido fraco de  $m_i^n$  é dado por  $w_i$ , o Teorema B.1.12 pode ser usado novamente.  $\square$

**Observação 6.2.3.** *As condições sobre as funções  $f^a$ ,  $f_i^c$ , e  $f_i^d$  podem ser relaxadas de forma análoga ao que foi feito nos Teoremas 5.2.4 e 5.2.5,*

### 6.3 Aplicação: Sistemas de Busca na Internet

Sistemas de busca na Internet são aqueles que indexam o conteúdo de páginas encontradas na *World Wide Web* e que podem ser usados para buscar sítios de interesse. Alguns exemplos comerciais destes sistemas são: Google, Yahoo, MSN Live Search, entre outros. De acordo com (Barroso et al., 2003), estes sistemas são tipicamente implementadas em *clusters* de computadores em paralelo para conseguir processar todas as buscas que o sistema recebe. O processamento em um destes sistemas é geralmente dividido em duas partes principais. A primeira parte, como descrito por (Gonçalves et al., 2007b) e (Barroso et al., 2003), consiste em recolher os documentos relacionados à busca dos bancos de dados. A segunda parte, consiste em ordenar estes documentos de forma relevante para exibí-los ao usuário. A primeira fase deste processamento é muito custosa computacionalmente e por isso este processo é feito em paralelo. Cada estação do *cluster* de computadores recebe uma fração do banco de dados. Quando uma busca chega no sistema, cada estação de processamento é responsável por fazer uma busca local na sua fração do banco de dados. A resposta ao usuário é dada somente quando cada estação de processamento termina sua busca local. Como foi observado por (Gonçalves et al., 2007b), mesmo quando o *cluster* de computadores é homogêneo, existe desbalanceamento entre as estações, contribuindo para aumentar o atraso na resposta ao usuário.

Considere a seguinte estratégia para reduzir o desbalanceamento: suponha que cada estação em paralelo tenha mais de uma fração da coleção de documentos. Por exemplo, uma estação pode ter acesso à sua fração do banco de dados e à fração de uma estação vizinha. Então, se esta estação estiver subutilizada e a vizinha estiver sobrecarregada, a estação subutilizada pode pegar uma tarefa da estação sobrecarregada para adiantar a resposta para o usuário. No caso considerado aqui, supõe-se que todas as estações tenham acesso à todas as frações da coleção de documentos.

Como estes sistemas são do tipo *fork-join* e operam em processamento para-

lelo, o modelo de filas descrito nas seções anteriores pode ser usado neste problema. Além disso, estes sistemas de busca geralmente operam sobre uma forte demanda, justificando a hipótese da aproximação, que requer que o sistema opere sobre tráfego pesado. A aproximação será usada para encontrar o melhor momento para mover tarefas entre as estações.

### 6.3.1 Simulação Computacional

Para testar e validar os controles obtidos e estabelecer comparações, foi necessário construir uma simulação de um sistema de busca onde as estratégias de balanceamento pudessem ser facilmente implementadas. Esta simulação foi construída a partir das observações e dados coletados por (Gonçalves et al., 2007a) em um sistema de busca na Internet.

Para construir esta simulação, foi assumido que o tempo entre a chegada de clientes é distribuído exponencialmente, como foi observado por (Gonçalves et al., 2007a). Seguindo ainda o modelo apresentado em (Gonçalves et al., 2007a), o tempo de serviço é dividido em dois casos: no primeiro caso, o servidor (ou a estação de processamento) tem que acessar o disco, e no segundo caso, o servidor consegue obter as informações necessárias da memória. Nos dois casos, o tempo de serviço é exponencialmente distribuído. Desta forma, com probabilidade  $h$ , o serviço será exponencialmente distribuído com média  $m_d$ , quando o servidor precisar acessar o disco, e exponencialmente distribuído com média  $m_c$  quando o processador encontrar a informação na memória.

Dos dados coletados por (Gonçalves et al., 2007a) para um sistema (pequeno) de 8 servidores homogêneos em paralelo, tem-se que  $h = 0.17$ ,  $m_d = 38.12ms$  e  $m_c = 9.20ms$ .

Como o modelo construído aqui se limita a um sistema com apenas duas filas, a política de movimentação de tarefas entre processadores é implementada da seguinte forma em um sistema com mais de dois processadores: quando uma tarefa é terminada em um processador, a maior fila no sistema é encontrada (isto é,

a fila com o maior número de tarefas à serem trabalhadas). Em seguida, o controle é usado para decidir se uma tarefa deve ser movida da maior fila para a fila que acabou de terminar a sua tarefa. Caso a decisão seja de mover a tarefa, então a tarefa mais antiga no sistema, que não esteja sendo processada, é movida.

Para fazer a simulação mais realista, cada tarefa movida terá sempre a distribuição de serviço do primeiro tipo, isto é, o servidor terá que acessar o disco e a distribuição será exponencial com média  $m_d$ . Como o banco de dados da tarefa que foi movida é diferente da que o processador utiliza, é muito improvável que exista informação disponível em *cache*.

A simulação computa o tempo de resposta do sistema depois de entrar em regime estacionário. Para isto, a simulação é executada até  $10^4$  clientes serem atendidos e após disso as médias começam a ser computadas. O programa termina depois de  $5 \times 10^4$  clientes serem atendidos. O intervalo de 95% *t*-confiança para as médias também é calculado durante a simulação (veja (Gross e Harris, 1998)).

### 6.3.2 Aplicação do Modelo Teórico

Suponha que a taxa de chegada de clientes seja constante. Isto é,

$$\Lambda^a(\sqrt{n}\xi) \triangleq \lambda^a(\xi) = \lambda$$

onde  $\lambda$  é uma constante positiva. A taxa de serviço também é constante e dada por  $\Lambda^d(\sqrt{n}\xi) \triangleq \lambda^d(\xi) = \mu_i$ . A aproximação é usada escolhendo um  $n$  grande, onde  $b_i^n \triangleq \sqrt{n}(\mu_i - \lambda)$  seja relativamente “pequeno”. Observe que pode-se escrever o seguinte:

$$\lambda^d(\xi) = \mu_i = \lambda + b_i/\sqrt{n}.$$

Além disso, a política de roteamento será dada por:

$$Q_i^c(\sqrt{n}\xi) \triangleq q_i^c(\xi) = f_i^c(\xi)/\sqrt{n},$$

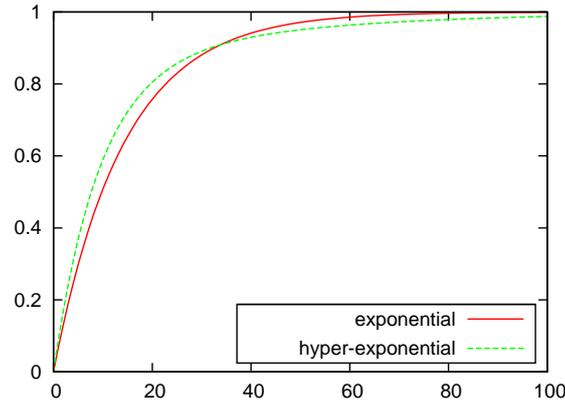


Figura 6.2: Gráfico da função de distribuição cumulativa exponencial e hiper-exponencial

onde a função  $f_i^c : \mathbb{R}_{\geq 0}^2 \rightarrow [0, 1]$  será escolhida mais tarde.

Portanto, o processo  $x$  é dado por

$$\begin{aligned}
 dx(t) = & \begin{pmatrix} -b_1^n + \lambda [f_1^c(x(t)) - f_2^c(x(t))] \\ -b_2^n + \lambda [f_2^c(x(t)) - f_1^c(x(t))] \end{pmatrix} dt + \begin{pmatrix} 2\lambda & \lambda \\ \lambda & 2\lambda \end{pmatrix}^{1/2} dw(t) \\
 & + \begin{pmatrix} 1 & 0 & -1 & -1 \\ 0 & 1 & -1 & -1 \end{pmatrix} d\zeta(t), \tag{6.5}
 \end{aligned}$$

É importante mencionar uma pequena discrepância entre modelo e aplicação. O modelo desenvolvido assume tempo de serviço exponencial. Contudo, observe que na aplicação, o tempo de serviço é na verdade hiper-exponencial, já que o serviço é dividido em duas partes (a que acessa o disco, e a que obtém as informações da memória). Apesar disso, se escolhermos a taxa  $\mu_i$  para ser definida como  $\mu_i \triangleq 1/(h \cdot m_c + (1-h) \cdot m_d)$ , onde as constantes  $h$ ,  $m_c$ , e  $m_d$  foram definidas no capítulo anterior, esta discrepância se torna mais aceitável pois a diferença entre as distribuições é pequena, como mostra a Figura 6.2.

Nos experimentos feitos aqui, foi utilizado  $n = 100$ .

### 6.3.3 Formulação do Problema de Controle Ótimo

A função custo que será utilizada para encontrar a melhor política de balanceamento será o seguinte:

$$W(x_0, f^c) = \mathbb{E}_{x_0}^{f^c} \left[ \int_0^\infty e^{-\beta t} (|x_1(t) - x_2(t)| + \sum_{i=1,2} \eta_i f_i^c(x(t))) dt \right],$$

onde  $\eta_1$  e  $\eta_2$  são constantes dependentes do sistema,  $x = (x_1, x_2)'$  é a equação que descreve o sistema dada pela Equação (6.5), e  $x(0) = x_0$  é a condição inicial. Portanto, procura-se uma solução do seguinte problema:

$$\begin{cases} \text{encontrar } f^* \in \mathcal{U} \text{ que minimiza o custo} \\ V(x_0, f^*) = \inf_{f \in \mathcal{U}} W(x_0, f), \end{cases}$$

onde  $\mathcal{U}$  é o conjunto de controles de realimentação admissíveis do tipo:  $f = (f_1, f_2)'$  e  $f_i : \mathbb{R}_+^2 \rightarrow [0, 1]$ ,  $i = 1, 2$ .

Para calcular o controle ótimo numericamente, foi utilizado o método da cadeia de Markov aproximada (Kushner e Dupuis, 1992) (utilizando a implementação feita por Jarvis, veja (Jarvis e Kushner, 1996)). O parâmetro de discretização usado foi de  $1/\sqrt{n}$ , que é sugerido em (Kushner et al., 1995). Além disso, foi escolhido  $\beta = 0.001$ ,  $\eta_1 = \eta_2 \triangleq \eta$ , onde a constante  $\eta$  será escolhida mais tarde.

Para utilizar o método da cadeia de Markov, foi necessário estipular uma capacidade máxima para cada fila. Foram escolhidos os seguintes valores  $B_i = 5$ , para  $i \in \{1, 2\}$ . Estes valores não interferem no resultado final e poderiam ser maiores, ou menores (desde que ainda não interferissem na escolha do controle ótimo).

### 6.3.4 Resultados Numéricos

Em primeiro lugar, é importante validar a simulação implementada contra dados obtidos de um sistema de busca real para poder servir de fato como uma plataforma para testes. Neste sentido, o cenário dado em (Gonçalves et al., 2007a)

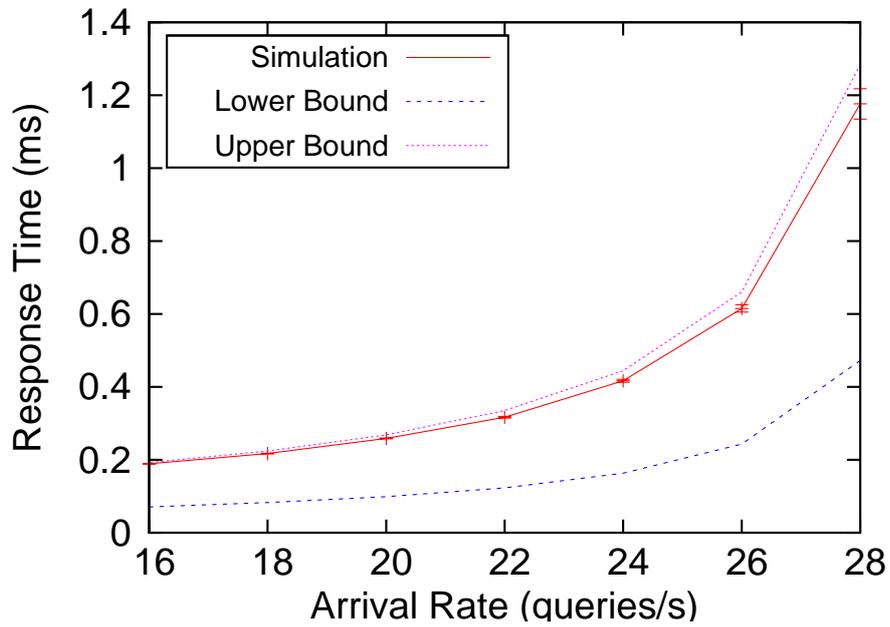
foi considerado, isto é, um sistema de busca com 8 servidores idênticos com tempo de serviço dado pelas constantes da Seção 6.3.1. O tempo de resposta da simulação foi calculado para taxas de entradas diferentes, e o resultado é dado pela Figura 6.3(a). Também foi calculado o limite superior e inferior usados em (Gonçalves et al., 2007a), dados abaixo, respectivamente,

$$U = \frac{H_8 \cdot m}{1 - \rho} \quad L = \frac{m}{1 - \rho},$$

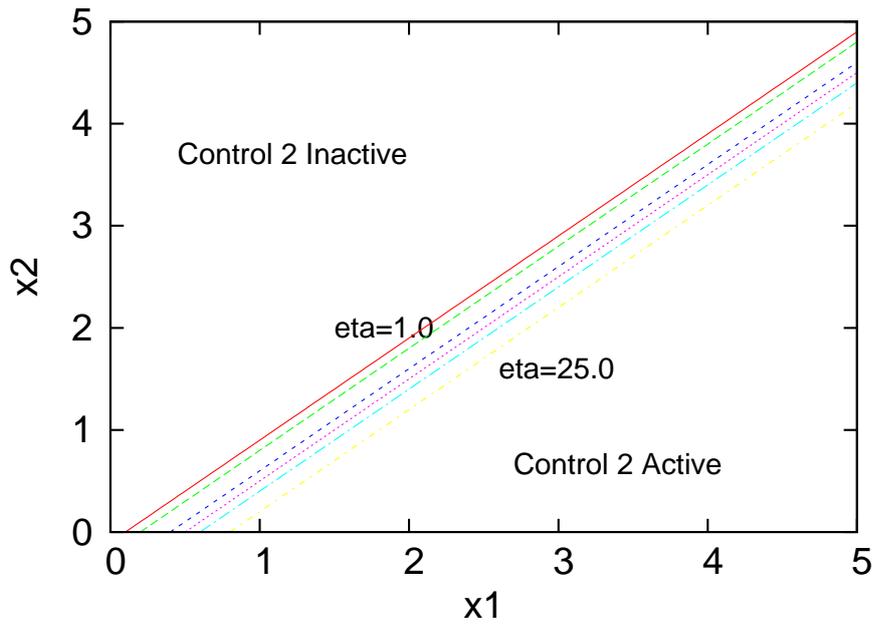
onde  $m \triangleq h \cdot m_c + (1 - h) \cdot m_d$ ,  $\rho$  é a intensidade de tráfego, e  $H_p$  é o número  $\sum_{i=1}^p 1/i$ . Todos estes valores são dados pela Figura 6.3(a). Estes mesmos cálculos foram feitos para o sistema de busca de (Gonçalves et al., 2007a). A simulação usada aqui reflete bem o mesmo comportamento observado no sistema real.

Neste ponto, a simulação será usada para ajudar a escolher o parâmetro  $\eta$ , que é a constante relacionada ao custo de mover uma tarefa de uma fila à outra. Para isso, foram escolhidos vários valores de  $\eta$  e o controle ótimo foi computado para cada escolha. Neste processo, foi usado a uma taxa de entrada de 28 buscas (ou perguntas) por segundo. Em seguida, o controle foi aplicado na simulação e o tempo de resposta médio do sistema computado. A Tabela 6.1 contém os resultados obtidos. Como a escolha de  $\eta = 15$  gerou o menor tempo de resposta médio, este valor será usado no resto do capítulo.

A Tabela 6.1 também contém a curva de chaveamento do controle ótimo. Se o processo  $X$  estiver acima desta curva, o controle é aplicado, caso contrário ele não é aplicado. Por exemplo, para  $\eta = 15$ , se o número de clientes na fila 2 (dado por  $X_2$ ) satisfazer  $X_1 - 5 > X_2$ , quando comparado ao número de tarefas na fila 1, uma tarefa poderá ser movida da fila 1 para a fila 2 com probabilidade  $1/\sqrt{n} = 0.1$ . Desta forma, quando uma tarefa deixar a fila 2, ela moverá uma tarefa da fila 1 somente se  $X_1 - 5 > X_2$ , com probabilidade 0.1. As curvas de chaveamento estão na Figura 6.3(b) para as diferentes escolhas de  $\eta$ . É importante mencionar que estas curvas são regressões lineares aproximadas das curvas obtidas como solução do problema de controle ótimo. Como exemplo, as regiões do espaço



(a)



(b)

Figura 6.3: (a) Gráfico do tempo de resposta do sistema simulado para taxas de entradas variadas (o intervalo de 95%  $t$ -confiança para simulação também está presente no gráfico). (b) Curvas de chaveamento para o controle ótimo  $f_2^{c,*}$  para valores variados de  $\eta \in \{1, 5, 10, 15, 20, 25\}$  (o espaço de estado mostrado na figura está escalonado por  $1/\sqrt{100}$ ).

de estado onde cada controle é aplicado é dado pela Figura 6.4(a), para o caso de  $\eta = 15$ .

Tabela 6.1: Tabela mostrando o tempo de resposta médio obtido da simulação para um sistema com controle ótimo calculado usando diferentes valores para o parâmetro  $\eta$  na função de custo.

$\eta$	Tempo Médio de Resposta	95% $t$ -confiança	Curva de <i>Switching</i>
1	0.629651	0.012978	$X_1 - 1 > X_2$
5	0.601555	0.009725	$X_1 - 2 > X_2$
10	0.599692	0.011384	$X_1 - 4 > X_2$
15	0.593476	0.007581	$X_1 - 5 > X_2$
20	0.600476	0.008314	$X_1 - 6 > X_2$
25	0.614856	0.007030	$X_1 - 8 > X_2$

O tempo de resposta do sistema sem controle comparado com o controle ótimo para diferentes taxas de chegada  $\lambda$  é dado pela Figura 6.4(b). Observe que, a medida que as taxas de chegada aumentam (isto é, a medida que o sistema aproxima tráfego pesado), o tempo de resposta para o sistema com controle aproxima do limite inferior  $L$  (dado pela Figura 6.3(a)) e, por outro lado, o tempo de resposta para o sistema sem controle aproxima do limite superior  $U$ .

Para estabelecer uma base de comparação, suponha a seguinte política de controle: mover uma tarefa da fila  $i$  para a fila  $j$  com probabilidade 1 se  $X_i - 1.0 > X_j$ . Este controle reflete a idéia geral que deve-se mover uma tarefa se ela está em uma fila maior (ignorando o cliente em serviço). Esta política será chamada de SIB (*Switch If Bigger*). Para comparar o ganho entre tempo de resposta de sistemas diferentes, será calculado o ganho relativo de um sistema em relação ao outro com o seguinte cálculo:

$$R(a,b) = \frac{a - b}{a} \cdot 100\%,$$

onde  $R(a, b)$  é o ganho da quantidade  $b$  em relação à quantidade  $a$ . A Figura 6.5(a) tem o ganho relativo do sistema usando o controle ótimo em relação ao sistema usando o controle SIB. Observe que para valores de taxas de chegada menores, o sistema com controle SIB é melhor. Contudo, a medida que o sistema aproxima do

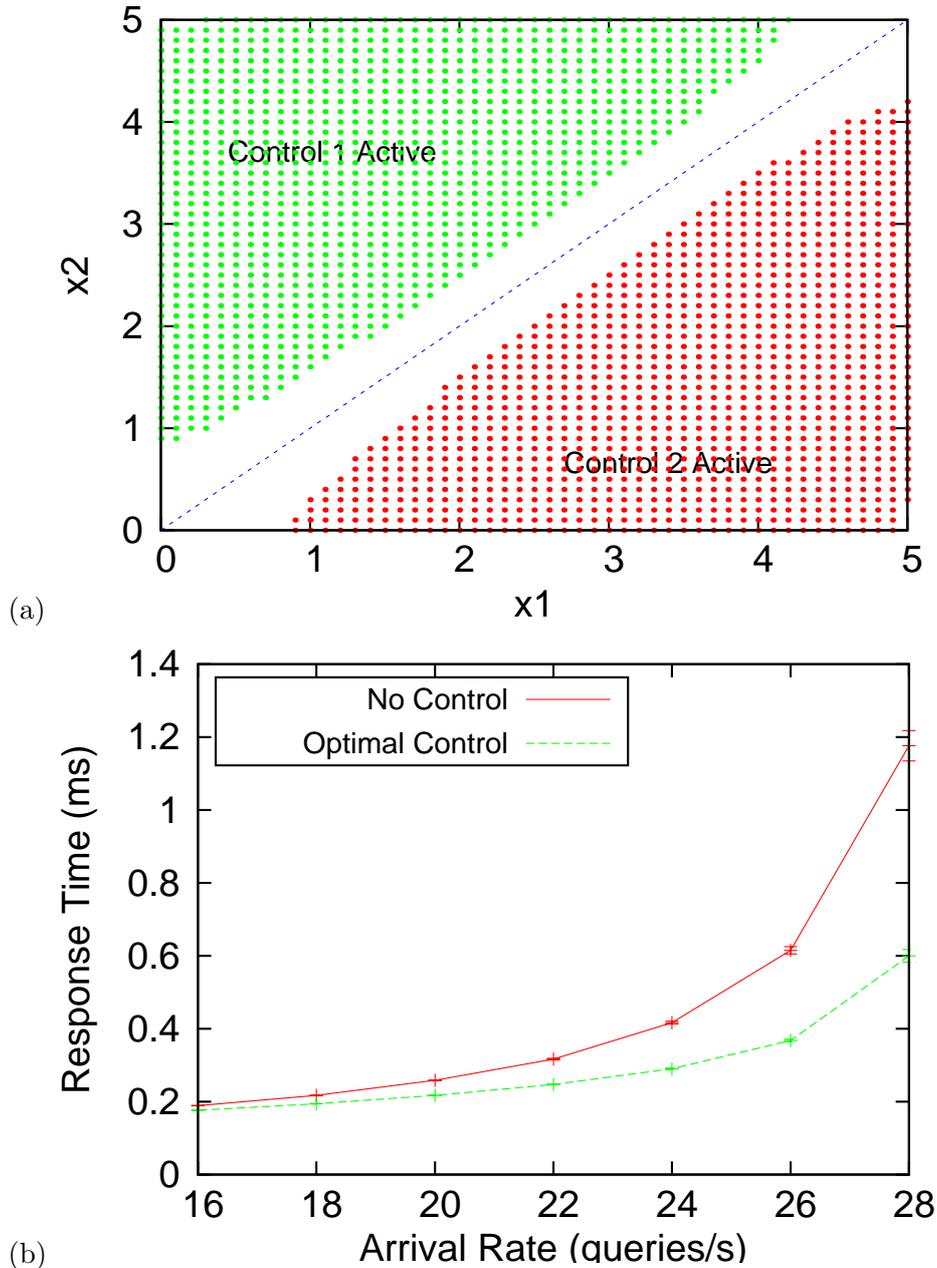


Figura 6.4: (a) Região do espaço de estado (escalonado) onde cada controle age. Controle 1 e 2 são referentes à  $f_1^c$  e  $f_2^c$ , respectivamente. As regiões são dadas aproximadamente por  $R_1 = \{x \in \mathbb{R}_+^2 : x_1 + 0.5 > x_2\}$  e  $R_2 = \{x \in \mathbb{R}_+^2 : x_1 - 0.5 < x_2\}$ . A linha azul atravessando o meio do espaço de estado é a linha  $x_1 = x_2$ . (b) O gráfico do tempo de resposta do sistema calculado usando a simulação com valores variados de taxa de chegada para o sistema com controle e o sistema sem controle (o intervalo de 95%  $t$ -confiança para simulação também está presente no gráfico)

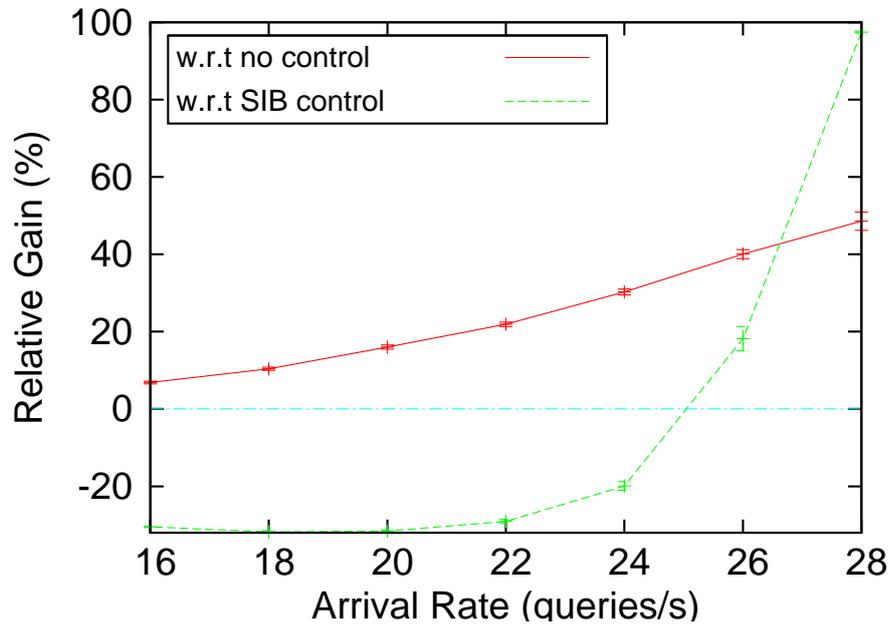
regime de tráfego pesado, o sistema usando o controle ótimo é melhor, com ganho relativo acima de 80% para o caso de taxa de chegada de 28 perguntas/segundo. A Figura 6.5(a) também mostra o ganho relativo do sistema usando o controle ótimo em relação ao sistema sem controle. Observe que neste caso o ganho é próximo do 50% quando a taxa de entrada está em 28 perguntas/segundo.

A Figura 6.5(b) contém o ganho relativo do sistema usando o controle ótimo em relação ao sistema sem controle para variados números de servidores em paralelo e taxa de entrada fixa em 28 perguntas/segundo. Observe que não há perda de ganho a medida que o número de servidores aumenta, sugerindo que esta estratégia pode ser aplicada em sistemas com um grande número de servidores em paralelo.

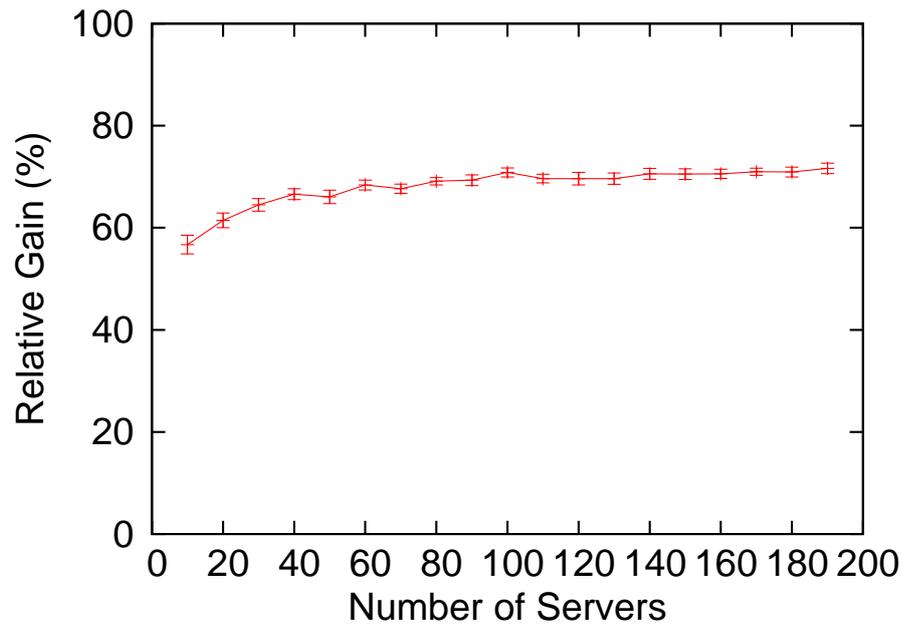
Suponha agora que 4 novos servidores mais rápidos sejam adicionados ao sistema composto de 8 servidores. Para estes novos servidores, seja  $h = 0.17$ ,  $m_d = 30.0ms$ , e  $m_c = 7.0ms$ . A política ótima de balanceamento para este novo sistema heterogêneo será computada. Quando a decisão de balanceamento é feita entre dois servidores com a mesma velocidade de processamento, o controle ótimo para o sistema homogêneo é usado. Quando a decisão de balanceamento é feita entre dois servidores com velocidade de processamento diferente, será usado o controle para o sistema heterogêneo.

A política ótima para este sistema heterogêneo com taxa de chegada dada por 28 perguntas/segundo, é a seguinte: se a fila 2 é mais rápida, move-se uma tarefa da fila 1 com probabilidade 0.1 se  $1.34 \cdot X_1 - 2 > X_2$ ; caso contrário, se a fila 2 é mais lenta, move-se uma tarefa da fila 1 para a fila 2 com probabilidade 0.1 se  $0.71 \cdot X_1 - 10.9 > X_2$ . A Figura 6.6(a) tem as regiões do espaço de estado onde cada controle age.

O tempo de resposta do sistema usando o controle heterogêneo e o sistema sem controle é dado pela Figura 6.6(b). Observe que a medida que o sistema aproxima tráfego pesado, a diferença entre o tempo de resposta entre o sistema com controle e o sistema sem controle se torna muito grande. Isto é explicado pelo fato que os servidores mais lentos se tornam um gargalo para o sistema sem



(a)



(b)

Figura 6.5: (a) Ganho relativo do sistema com controle em relação ao (*w.r.t*) sistema sem controle e o sistema usando o controle SIB. (b) Ganho relativo do sistema com controle em relação ao sistema sem controle variando o número de estações em paralelo com taxa de entrada fixa em 28 perguntas/segundo (o intervalo de 95% *t*-confiança para simulação também está presente no gráfico).

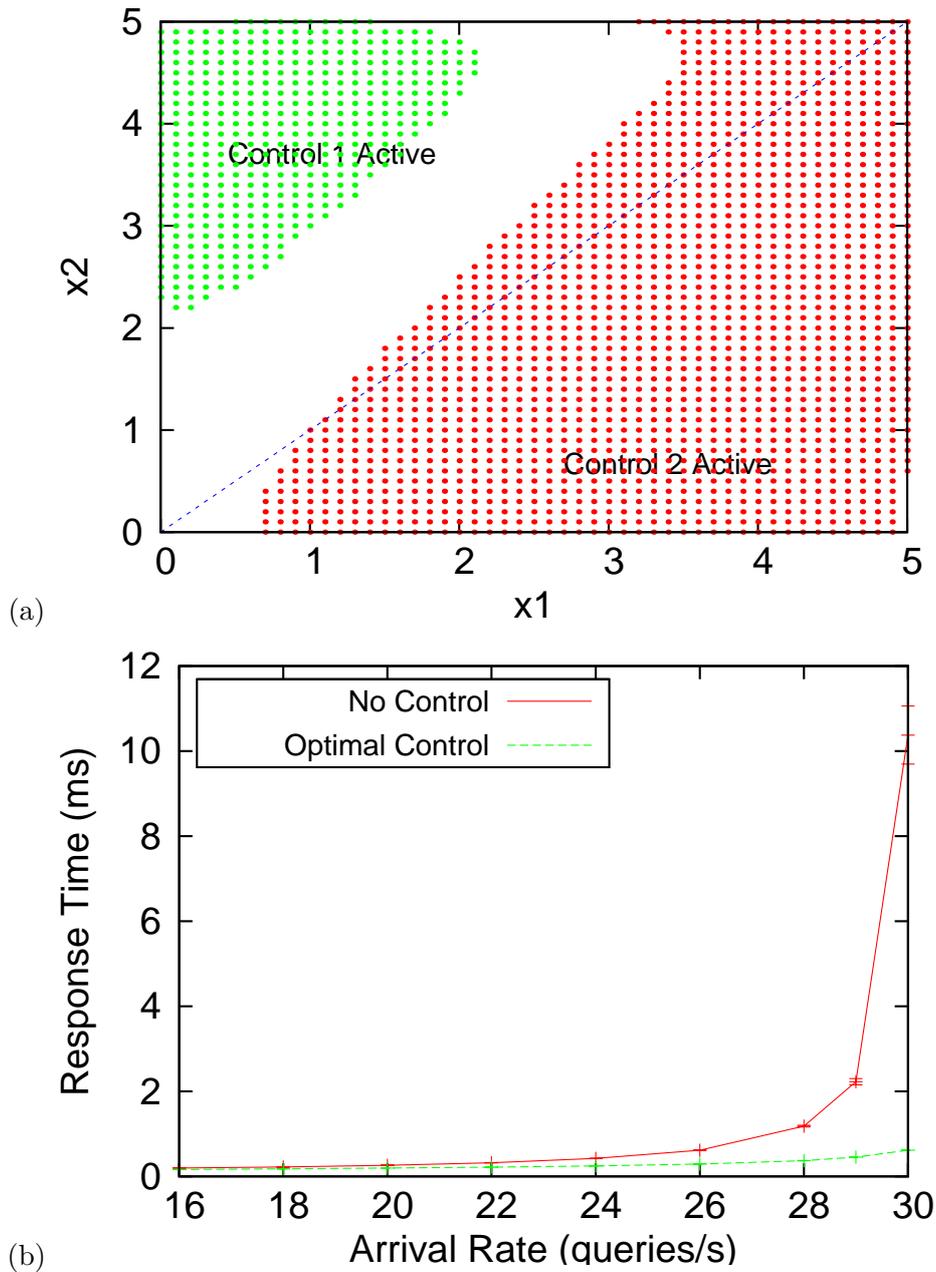
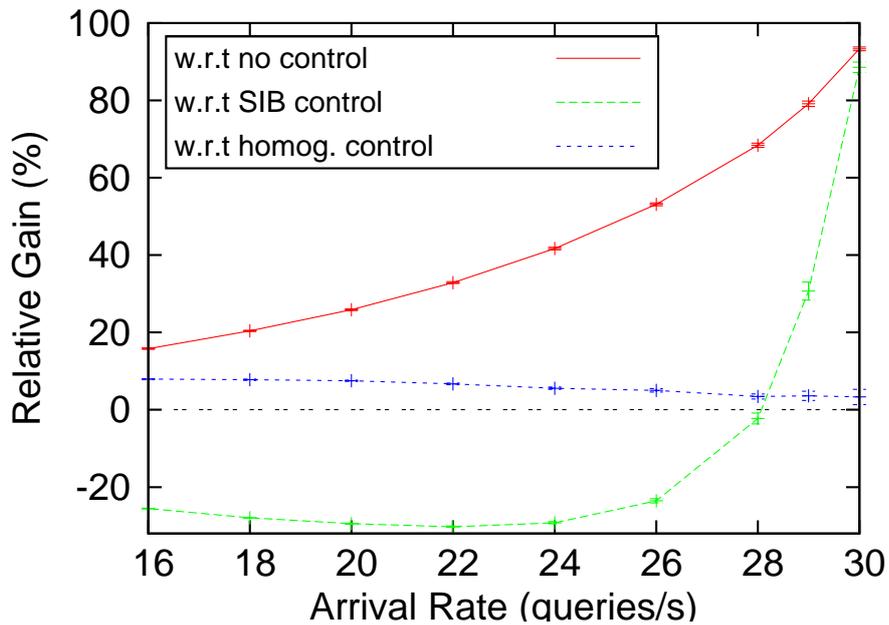


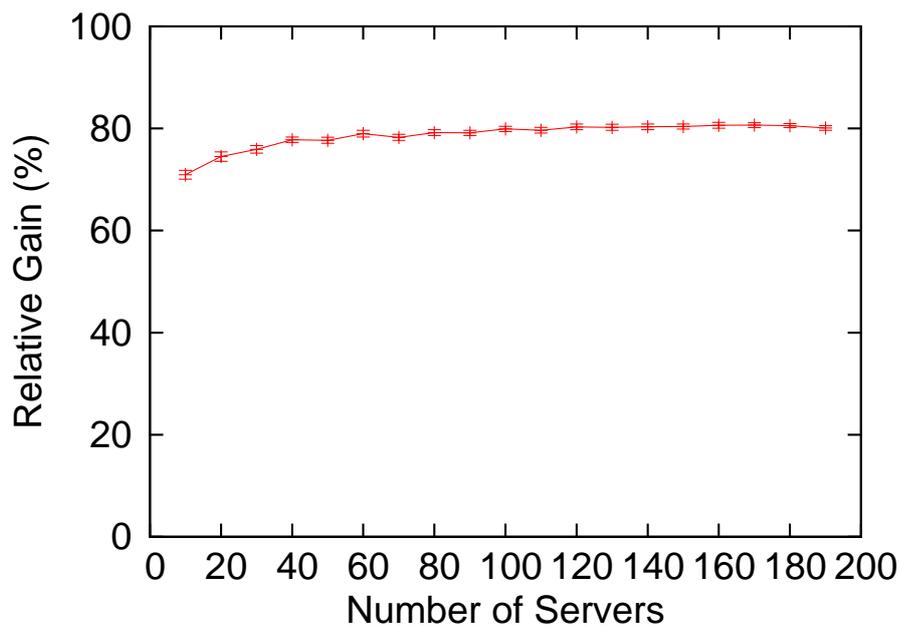
Figura 6.6: (a) Região do estado de espaço (escalonado por  $1/\sqrt{n}$ ) onde cada controle ótimo age. Os controles 1 e 2 são referentes às funções  $f_1^c$  e  $f_2^c$ , respectivamente. As regiões são aproximadamente dadas por  $R_1 = \{x \in \mathbb{R}_+^2 : 1.34x_1 + 0.2 > x_2\}$  e  $R_2 = \{x \in \mathbb{R}_+^2 : 1.34x_1 - 1.6 < x_2\}$ . A linha azul que cruza o espaço de estados é a linha  $x_1 = x_2$ . Neste gráfico, a fila 2 é mais rápida que a fila 1. (b) Gráfico do tempo de resposta do sistema com diferentes valores para a taxa de chegada para o sistema com controle e sem controle. (o intervalo de 95%  $t$ -confiança para a simulação também está presente no gráfico).

controle para taxas de chegada acima de 28 perguntas/segundo. O mesmo não acontece com o sistema com controle já que os servidores mais rápidos ajudam reduzir o congestionamento movendo clientes dos servidores mais lentos.

A Figura 6.7(a) tem o ganho do tempo de resposta do sistema com controle em relação ao sistema sem controle. Além disso, o ganho do sistema usando o controle ótimo em relação ao sistema usando o controle SIB e o sistema usando somente o controle homogêneo também é mostrado na Figura 6.7(a). Novamente, quando o sistema está próximo de uma situação de tráfego pesado, o controle ótimo é melhor do que o controle SIB. Observe também que o sistema usando o controle ótimo tem ganho sobre o sistema usando somente o controle para sistemas homogêneos. A Figura 6.7(b) mostra o ganho em tempo de resposta do sistema com controle sobre o sistema sem controle para taxa de entrada fixa em 28 perguntas/segundo e com número de servidores aumentando. Neste caso, para facilitar o experimento, metade dos servidores são do tipo “rápido” e metade do tipo “lento” (ou original). Como pode-se ver, não há redução do ganho, sugerindo que esta estratégia pode ser usada em sistemas com um número grande de servidores em paralelo.



(a)



(b)

Figura 6.7: (a) Ganho relativo do sistema com controle em relação ao sistema sem controle, com o controle SIB, e usando somente o controle para sistemas heterogêneos. (b) Ganho relativo do sistema com controle em relação ao sistema sem controle variando o número de servidores no sistema com taxa de entrada fixa em 28 perguntas/segundo (o intervalo de 95%  $t$ -confiança também está presente no gráfico).

# Capítulo 7

## Uma Estrutura Unificante para Redes com Sinais

Neste capítulo, uma forma mais geral de descrever sistemas de filas com sinais será introduzida. A partir desta generalização, será possível ampliar o modelo para atender uma classe maior de problemas. O foco desta abordagem será de caracterizar o sistema de interesse como um sistema composto por certas quantidades que sofrem o efeito de trocas (ou transformações) estocásticas entre elas. Para atingir tal objetivo, uma representação do sistema através de redes de Petri estocásticas será adotado. O conteúdo deste capítulo é referente ao trabalho (Leite e Fragoso, 2009).

Redes de Petri (Petri, 1962) são modelos baseados em grafos de sistemas de comunicação tais como os que possuem características de serem concorrentes, assíncronos, e distribuídos (Murata, 1989). O objetivo principal das redes de Petri é estudar propriedades qualitativas destes sistemas que podem ser detectadas através deste formalismo, que incluem: *deadlocks*, ciclos, estados alcançáveis, entre outras (Reisig, 1985; Petri, 1996; Murata, 1989). Além do ganho em entendimento teórico, redes de Petri também trazem consigo uma representação gráfica que facilita a aplicação da teoria na prática. Dentre as primeiras aplicações destes modelos se destacam as seguintes: desenvolvimento de protocolos de comunicação e avaliação de performance de sistemas computacionais (Murata, 1989).

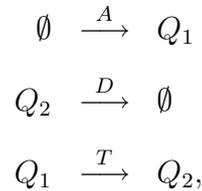
A idéia de introduzir o tempo na descrição destes modelos veio mais tarde

com Ramchandani em (Ramchandani, 1973), e com ela veio também a idéia de estudar a dinâmica de tais sistemas. Nestes modelos, cada transição que troca o estado do sistema acontece em intervalos fixos de tempo. As redes de Petri estocásticas são extensões naturais destes modelos, onde o tempo entre as transições são dadas por variáveis aleatórias. As redes de Petri estocásticas foram introduzidas por Symons em (Symons, 1980), Natkin em (Natkin, 1980), e Molloy em (Molloy, 1981, 1982). Em todas suas formas diferentes, as redes de Petri tem sido largamente utilizadas em várias áreas tais como: análise de performance de sistemas computacionais, sistemas distribuídos, e mais recentemente, modelagem de redes biológicas (moleculares) (alguns exemplos da literatura incluem (Hardy e Robillard, 2008; Chen et al., 2007; Chaouiya, 2007; Wilkinson, 2006; Hardy e Robillard, 2004; Pinney et al., 2003)).

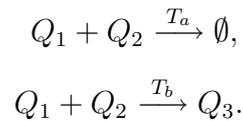
Redes de Petri têm várias similaridades com sistemas de filas. De fato, se conceitos relacionados diretamente a problemas de filas não forem importantes, como por exemplo, ordem de processamento, disciplina da fila, e tempo de permanência de clientes, pode-se interpretar estes sistemas simplesmente como sistemas compostos por “caixas” contendo uma certa quantidade de alguma coisa (clientes no caso de filas) e transições que fazem trocas destas quantidades nestas caixas. Por exemplo, pode-se descrever uma fila única com o formalismo das redes de Petri, e até mesmo duas filas em sequência. Contudo, um aspecto importante de redes de filas que não pode ser modelado com redes de Petri é o roteamento, que estabelece um certo grau de correlação entre as transições com efeitos diferentes da rede. Para tratar este problema de representação, uma nova classe de redes de Petri estocástica é introduzida aqui. Nesta classe, considera-se redes de Petri estocásticas onde existe um número contável de possíveis efeitos de cada transição, que é escolhida de forma aleatória. Com esta generalização, pode-se descrever todas as redes de filas do tipo de Jackson, e até descrever as redes de Gelenbe (*G-networks*), com sinais do tipo “cliente negativo” e “triggers”.

Para manter o modelo em um caráter abstrato, será usado uma notação

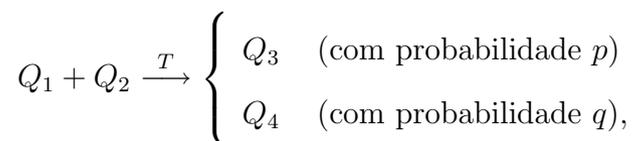
similar a usada em reações químicas. Por exemplo, uma rede de filas composta por duas filas em sequência será descrita pelas quantidades  $Q_1$  e  $Q_2$  e as transições estocásticas  $A$ ,  $D$ , e  $T$ . Estas transições têm o seguinte efeito no sistema:



que são interpretadas como a chegada de um cliente de uma fonte externa (onde  $\emptyset$  denota essa fonte externa) na fila 1 ( $Q_1$ ), a saída de clientes da fila 2 ( $Q_2$ ) por serviço, e a saída de clientes da fila 1 por serviço que são enviados para a fila 2, respectivamente. Esta descrição também permite a representação de filas com sinais. Por exemplo, pode-se pensar nas seguintes transições:



A transição  $T_a$  consome uma quantidade do tipo  $Q_1$  e uma do tipo  $Q_2$ . Esta transição pode ser interpretada como um “cliente negativo” que sai da fila 1 ( $Q_1$ ) e remove um cliente da fila 2 ( $Q_2$ ). De forma similar, a transição  $T_b$  pode ser interpretada como o efeito de um “trigger”, onde um cliente na fila 1 ( $Q_1$ ) move um cliente da fila 2 ( $Q_2$ ) para a fila 3 ( $Q_3$ ). Além disso, um sistema com roteamento pode ser representado da seguinte forma,



onde  $p + q = 1$ . Neste caso o efeito da transição  $T$  no sistema pode produzir uma quantidade do tipo  $Q_3$  com probabilidade  $p$  e uma quantidade do tipo  $Q_4$  com probabilidade  $q$ .

O objetivo principal desta abordagem é obter um modelo geral o bastante para que sua aproximação por difusão possa ser aplicada diretamente em vários problemas. Para construir tal aproximação, será usada a mesma descrição que vem sendo usada nos capítulos anteriores. Isto é, será considerado um modelo onde o tempo de transições é exponencialmente distribuído e possui taxas dependentes do estado. Além disso a probabilidade de cada efeito de uma transição ocorrer também será dependente do estado.

O capítulo é organizado da seguinte forma: na Seção 7.1, será feita uma breve revisão sobre redes de Petri. Em seguida, na Seção 7.2, será desenvolvida a aproximação por tráfego pesado das redes de Petri estocásticas. Na Seção 7.3, a aproximação para redes de Petri estocástica com transições que possuem efeito aleatório no sistema será introduzida. E na última seção, alguns exemplos práticos serão estudados para ilustrar o uso do modelo.

## 7.1 Redes de Petri

Redes de Petri são descritas por um conjunto de *nós*, *transições*, e *arcos direcionados*. Cada nó possui uma *marca* que indica a quantidade de *itens* presentes. Por exemplo, marcas podem representar o número de clientes em um sistema de filas, ou o número de moléculas em um certo tipo de rede bioquímica. Cada arco dirigido liga uma transição a um nó, ou um nó a uma transição. Quando uma transição é *disparada*, itens pertencentes aos nós que estão ligados a esta transição podem ser movidos. Um arco dirigido que liga um nó a uma transição indica que os itens contidos neste nó são necessários para que a transição ocorra. Arcos ligando transições aos nós indicam que estes nós recebem itens como o efeito das transições. O número de itens que são movidos são dados pelos *pesos* associados a cada arco. Uma transição somente pode ser disparada quando cada nó necessário para a transição possuir uma marca maior ou igual ao peso do arco.

No exemplo da Figura 7.1, tem-se três nós  $p_1$ ,  $p_2$ , e  $p_3$  que possuem marcas 3, 4, e 2, respectivamente, antes da execução da transição  $t_1$ . A transição  $t_1$  precisa

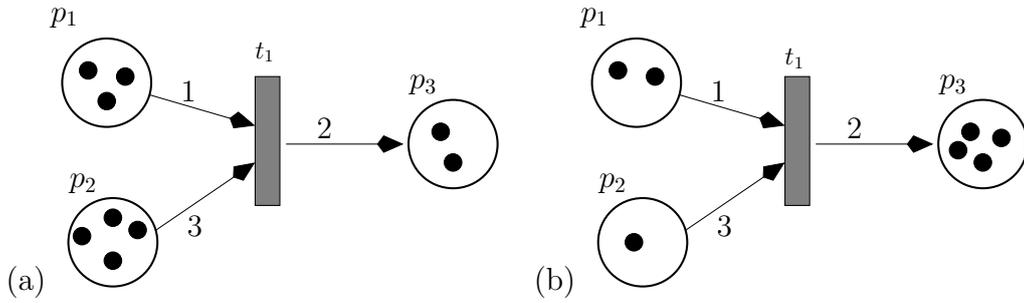


Figura 7.1: Representação gráfica de uma rede de Petri. (a) Os círculos em preto dentro dos nós representam item antes da execução da transição  $t_1$ . (b) Os círculos em preto indicam a marca da rede após a execução da transição  $t_1$ . No estado presente da rede, a transição  $t_1$  não pode ser disparada novamente já que não existem itens suficientes no nó  $p_2$ .

de um item de  $p_1$  e três de  $p_2$  para produzir dois itens de  $p_3$ . De forma alternativa, o sistema pode ser representado na forma simplificada

$$p_1 + 3p_2 \xrightarrow{t_1} 2p_3,$$

como discutido no início do capítulo.

Chama-se um nó  $p$  de *nó de entrada* da transição  $t$  se existir um arco direcionado saindo de  $p$  e chegando à  $t$ . De forma similar, chama-se  $p$  de um *nó de saída* de  $t$  se existir um arco direcionado saindo de  $t$  e indo ao  $p$ . Por exemplo, na Figura 7.1, os nós  $p_1$  e  $p_2$  são nós de entrada para a transição  $t_1$ , e  $p_3$  é um nó de saída.

Formalmente, uma rede de Petri é definida como um grafo bipartite, direcionado, e com pesos. Isto é,

**Definição 7.1.1.** (Murata, 1989) *Uma rede de Petri é a quintupla  $(P, T, F, W, M_0)$  onde:*

$$P = \{p_1, \dots, p_u\}, u \in \mathbb{N}, \text{ é um conjunto finito de nós,}$$

$$T = \{t_1, \dots, t_v\}, v \in \mathbb{N}, \text{ é um conjunto finito de transições,}$$

$$F \subseteq (P \times T) \cup (T \times P) \text{ é um conjunto de arcos direcionados,}$$

$W : F \rightarrow \mathbb{N}$  é uma função de peso,

$M_0 : P \rightarrow \mathbb{N}_0$  é a marca do sistema.

onde  $P \cap T = \emptyset$ .

Seja  $X_0 \in \mathbb{N}^u$  o vetor dado por  $(X_0)_i \triangleq M_0(p_i)$ , para  $i \in \{1, \dots, u\}$ . Seja também  $Pre \in \mathbb{N}^{v \times u}$  a matriz construída com os pesos dos arcos ligando os nós às transições, isto é,  $(Pre)_{ij} \triangleq W(p_j, t_i)$ , para  $(i, j) \in \{1, \dots, v\} \times \{1, \dots, u\}$ . De forma similar, seja  $Post \in \mathbb{N}^{v \times u}$  a matriz construída com os pesos dos arcos ligando as transições aos nós, isto é  $(Post)_{ij} \triangleq W(t_i, p_j)$ , para  $(i, j) \in \{1, \dots, v\} \times \{1, \dots, u\}$ . A proposição abaixo é um resultado elementar e bem conhecido das redes de Petri e que será usado neste capítulo.

**Proposição 7.1.2.** *Seja  $r \in \{0, 1\}^v$  um vetor indicando o conjunto de transições que serão disparadas (isto é, se  $r_i = 1$  então a transição  $t_i$  será disparada, caso contrário  $r_i = 0$ ). Seja  $X_0$  o vetor contendo a marca inicial do sistema. Suponha que existam itens suficientes nos nós de entrada para as transições serem disparadas. Então a nova marca, dada por  $X$ , do sistema após a execução das transições dadas por  $r$  satisfaz:*

$$X = X_0 + A'r,$$

onde  $A = Post - Pre$ .

Para maiores detalhes sobre rede de Petri, veja por exemplo o livro (Reisig, 1985).

## 7.2 Redes de Petri Estocásticas

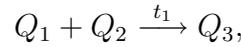
Uma rede de Petri estocástica é definida como acima exceto que cada transição possui um período aleatório de execução (Marsan, 1990; Marsan et al., 1984). Suponha que  $T(t)$  é um vetor aleatório tomando valores em  $\mathbb{N}_0^v$  onde cada componente  $T_i(t)$ ,  $i \in \{1, \dots, v\}$ , conta o número de execuções da transição  $t_i$  até o

instante  $t$ . Desta forma, pode-se escrever a marca do sistema em qualquer instante  $t \in \mathbb{R}_{\geq 0}$  como:

$$X(t) = X(0) + A'T(t),$$

onde  $X(0)$  é a marca inicial do sistema e a matriz  $A$  é definida como na Proposição 7.1.2. A capacidade máxima de itens que um nó pode guardar será dada por  $\tilde{B}_i \in \mathbb{R}_{>0} \cup \{\infty\}$ ,  $i \in \{1, \dots, u\}$ . Observe que é permitido que um nó  $i$  possua um número infinito de itens, neste caso  $\tilde{B}_i \equiv \infty$ .

Geralmente, ao trabalhar com redes de Petri, uma transição somente pode ser disparada se existir itens suficientes nos nós de entrada e espaço suficiente nos nós de saída para os novos itens criados. Contudo, este tipo de comportamento não representa bem todo tipo de cenário desejado. Por exemplo, considere uma transição dada por



que representa o efeito de um “*trigger*” em uma rede de Gelenbe. Isto é, a fila 1 ( $Q_1$ ) envia um sinal para a fila 2 ( $Q_2$ ) para mover um cliente para a fila 3 ( $Q_3$ ). Neste caso, se não existirem itens em  $Q_2$ , a transição será disparada, removendo um cliente de  $Q_1$  mas não criando nenhum item novo em  $Q_3$ , já que não haverá itens em  $Q_2$  para serem movidos. Para considerar todos os casos possíveis, os conjuntos  $E_{ki}, F_{ki} \subseteq \{1, \dots, u\}$  são definidos para cada  $k \in \{1, \dots, v\}$  e  $i \in \{1, \dots, u\}$  como:  $E_{ki} \triangleq \{\text{índices dos nós que, quando vazios, impedem o efeito da transição } k \text{ de mudar o estado do nó } i\}$  e  $F_{ki} \triangleq \{\text{índices dos nós que, quando cheios, impedem o efeito da transição } k \text{ de mudar o estado do nó } i\}$ . No exemplo acima, tem-se que

$$E_{11} = \{1\}, \quad F_{11} = \emptyset,$$

$$E_{12} = \{1, 2\}, \quad F_{12} = \emptyset,$$

$$E_{13} = \{1, 2\}, \text{ e } F_{13} = \{3\}.$$

Para garantir que o número de itens de um nó será positivo, assume-se que  $i \in E_{ki}$  se  $i$  é um nó de entrada para a transição  $t_k$ , isto é,  $(Pre)_{ki} > 0$ . Além disso, se  $i$  é um nó de saída para a transição  $t_k$ , isto é,  $(Post)_{ki} > 0$ , assume-se que  $i \in F_{ki}$ . Desta forma, o número de itens de um nó não excederá sua capacidade máxima.

Desta forma, escreve-se a marca do sistema em qualquer instante de tempo  $t \geq 0$  como:

$$X(t) = X(0) + \int_0^t \mathbb{A}'(X(s-)) dT(s), \quad (7.1)$$

onde

$$\mathbb{A}_{ki}(\xi) \triangleq A_{ki} \mathbb{I}\{\xi_j \geq (Pre)_{kj}, \text{ para } j \in E_{ki}, \text{ e } \xi_j \leq \tilde{B}_j - (Post)_{kj}, \text{ para } j \in F_{ki}\}, \quad (7.2)$$

para cada  $\xi \in \mathbb{R}^u$ .

A idéia de *backlogging* será adotada aqui para evitar uma dificuldade técnica na representação das direções de reflexão no limite. Esta dificuldade é a mesma discutida no início do Capítulo 5 e ilustrado pela Figura 5.2. *Backlogging* consiste em permitir que uma transição seja disparada mesmo quando não existam itens suficientes nos nós de entrada ou não exista espaço nos nós de saída. Ou seja, uma transição somente deixa de ser disparada por falta de itens se um dos nós de entrada desta transição possuir um número de itens igual ou menor do que zero. De forma análoga, uma transição será impedida de disparar por falta de espaço nos nós de saída se ao menos um deles estiver com uma marca maior ou igual que sua capacidade máxima. Portanto, estes nós poderão conter “temporariamente” mais itens do que o máximo permitido. Em termos práticos, isto significa que  $\mathbb{A}$  será dado por

$$\mathbb{A}_{ki}(\xi) \triangleq A_{ki} \mathbb{I}\{\xi_j > 0, \text{ para } j \in E_{ki}, \text{ e } \xi_j < \tilde{B}_j, \text{ para } j \in F_{ki}\},$$

para  $\xi \in \mathbb{R}^u$ , ao invés de usar a definição dada pela Equação 7.2. *Backlogging* já foi usado em aproximações de tráfego pesado para sistemas de manufatura, veja a Seção 7.1 de (Kushner, 2001) e as referências contidas lá. Esta simplificação do problema não é ideal, mas como será adotado o escalonamento usual em  $X$ , o número de marcas temporariamente negativas ou que excedem a capacidade máxima dos nós será pequeno para  $n$  grande. Portanto, a discrepância entre o modelo e o problema físico não serão grandes. Além disso, é importante ressaltar que *backlogging* somente é necessário nos casos onde uma transição pode movimentar mais de um item do mesmo nó, caso contrário, o modelo não é alterado.

O processo  $T$  será dado por

$$T_k(t) \triangleq N_k \left( \int_0^t \Lambda_k(X(s)) ds \right),$$

onde  $N_k$  é um processo de Poisson padrão (i.e., com taxa igual a 1), para cada  $k \in \{1, \dots, v\}$ . Além disso, as seguintes condições precisam ser introduzidas:

**Condição 7.2.1.**

- (a) Os elementos aleatórios  $\{X_i(0), N_k; i \in \{1, \dots, u\}, k \in \{1, \dots, v\}\}$  são mutuamente independentes.
- (b) As funções  $\Lambda_k : \mathbb{R}^u \rightarrow \mathbb{R}_{\geq 0}$  são mensuráveis e limitada.

De forma análoga ao que foi feito nos capítulos anteriores, usa-se o Lema B.2.3 e o Teorema B.2.2 para decompor  $T_k$  em um martingale  $\mathcal{F}_t$ -mensurável e um processo  $\mathcal{F}_t$ -previsível:

$$T_k(t) = M_k(t) + \int_0^t \Lambda_k(X(s)) ds,$$

onde  $M_k$  é o martingale. Portanto, pode-se escrever  $X$  em notação vetorial como:

$$X(t) = X(0) + \int_0^t \mathbb{A}'(X(s-)) dM(s) + \int_0^t \mathbb{A}'(X(s)) \Lambda(X(s)) ds,$$

onde  $M \triangleq (M_1, \dots, M_v)'$  e  $\Lambda(\cdot) \triangleq (\Lambda_1(\cdot), \dots, \Lambda_v(\cdot))'$ .

Defina, para cada conjunto  $S \subseteq \{1, \dots, 2u\}$ , o seguinte:

$$\mathcal{I}_{ik}(S) \triangleq \begin{cases} 1 & \text{se } S \cap (E_{ki} \cup \{u + F_{ki}\}) \neq \emptyset \\ 0 & \text{caso contrário,} \end{cases}$$

para cada  $k \in \{1, \dots, v\}$  e  $i \in \{1, \dots, u\}$ . Além disso, seja o conjunto

$$\tilde{\mathcal{E}}(\xi) \triangleq \mathbb{I}\{i \in \{1, \dots, 2u\} | \xi_i \leq 0, \text{ para } i \leq u, \text{ ou } \xi_{i-u} \geq \tilde{B}_{i-u}, \text{ para } i > u\},$$

para cada  $\xi \in \mathbb{R}^u$ . Então, observe que

$$\mathbb{A}'_{ik}(\xi) = A'_{ik} - A'_{ik} \mathcal{I}_{ik}(\tilde{\mathcal{E}}(\xi)),$$

e, portanto, pode-se escrever

$$\begin{aligned} X(t) = X(0) + \int_0^t \mathbb{A}'(X(s-)) dM(s) + A' \int_0^t \Lambda(X(s)) ds \\ - \int_0^t \mathcal{R}(\tilde{\mathcal{E}}(X(s))) \Lambda(X(s)) ds, \end{aligned} \quad (7.3)$$

onde  $\mathcal{R}_{ik}(S) \triangleq A'_{ik} \mathcal{I}_{ik}(S)$ . Desta forma,  $X$  foi decomposto em um termo de martingale, que dará origem ao processo de Wiener no limite, um termo que dará origem ao termo de deriva, e o último termo que está associado ao processo de reflexão, que mantém  $X$  preso no espaço de estados.

### 7.2.1 Limite em Regime de Tráfego Pesado

Seja a sequência de filas indexadas pelo parâmetro  $n$ , dada por  $\{X^n\}$ , onde para cada  $n$ ,  $X^n$  é definido como na Equação (7.1) e a matriz  $A$  é independente do parâmetro  $n$  (ou seja, a topologia da rede permanece constante). Suponha também que a capacidade máxima de um nó  $i$  presente na rede é dado por  $\tilde{B}_i^n \triangleq \sqrt{n} B_i$ , onde  $B_i \in \mathbb{R}_{>0} \cup \{\infty\}$  é independente do parâmetro  $n$ .

Agora pode-se definir o processo escalonado da forma usual:

$$x^n(t) \triangleq X^n(nt)/\sqrt{n}, \quad (7.4)$$

e usando a decomposição dada pela Equação (7.3), escreve-se o seguinte:

$$\begin{aligned} x^n(t) = x^n(0) + \int_0^t \mathbb{A}'(x^n(s-)) dm^n(s) + \sqrt{n} A' \int_0^t \lambda(x^n(s)) ds \\ - \sqrt{n} \int_0^t \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) \lambda^n(x^n(s)) ds, \end{aligned} \quad (7.5)$$

onde, como feito nos capítulos anteriores,  $m^n(t) \triangleq M^n(nt)/\sqrt{n}$ ,  $\lambda_k^n(\xi) \triangleq \Lambda_k^n(\sqrt{n}\xi)$ , e  $x^n(0) \triangleq X^n(0)/\sqrt{n}$ , para  $\xi \in \mathbb{R}_{\geq 0}^u$  e  $t \in \mathbb{R}_{\geq 0}$ . Além disso,  $\tilde{\mathcal{E}}$  é redefinido como  $\tilde{\mathcal{E}}(\xi) \triangleq \{i \in \{1, \dots, 2u\} | \xi_i \leq 0, \text{ se } i \leq u, \text{ ou } \xi_{i-u} \geq B_{i-u}, \text{ se } i > u\}$  e  $\mathbb{A}_{ki}(\xi) \triangleq A_{ki} \mathbb{I}\{\xi_j > 0, \text{ para } j \in E_{ki}, \text{ e } \xi_j < B_j, \text{ para } j \in F_{ki}\}$ , onde a única mudança está na troca de  $\tilde{B}_i$  por  $B_i$ .

Será necessário definir o espaço de estados  $G$  onde o processo limite tomará valores. Para isso, seja para cada  $i \in \{1, \dots, u\}$ , o seguinte:

$$G_i \triangleq \{\xi \in \mathbb{R}^u | \xi_i \geq 0\},$$

e para cada  $i \in \{u+1, \dots, 2u\}$  tal que  $B_{i-u} < \infty$ , defina o seguinte:

$$G_i \triangleq \{\xi \in \mathbb{R}^u | \xi_{i-u} \leq B_{i-u}\}.$$

Defina o conjunto de índices  $\mathfrak{F} \triangleq \{\{1\}, \dots, \{u\}\} \cup \{\{i+u\} | i \in \{1, \dots, u\} \text{ e } B_i < \infty\}$ . Então  $G \triangleq \bigcap_{\{i\} \in \mathfrak{F}} G_i$ . Defina também  $\mathfrak{E} \triangleq \{K \subset \{1, \dots, 2u\} | \bigcap_{k \in K} \partial G_k \neq \emptyset \text{ e } |K| \geq 2\}$ , que contém os conjuntos de índices correspondentes as “quinas” e “arestas” do espaço de estado. Além disso, será necessário definir o seguinte:  $\tilde{\partial} G_i \triangleq \{\xi \in \mathbb{R}^u | \xi_i \leq 0\}$  para  $i \in \{1, \dots, u\}$  e  $\tilde{\partial} G_i \triangleq \{\xi \in \mathbb{R}^u | \xi_{i-u} \geq B_{i-u}\}$  para  $i \in \{u+1, \dots, 2u\}$  tal que  $B_{i-u} < \infty$ .

Na condição que será dada abaixo, seja  $\gamma(S) \triangleq -\mathcal{R}(S)r$ , para cada conjunto

$S \subseteq \{1, \dots, 2u\}$ , onde  $r \in \mathbb{R}_{\geq 0}^v$  será definido abaixo. Além disso, seja

$$D(\xi) \triangleq \{\gamma(S) \mid S \neq \emptyset \text{ e } S \text{ é um subconjunto de } \mathcal{E}(\xi)\}, \quad (7.6)$$

para cada  $\xi \in \mathbb{R}^n$ , onde  $\mathcal{E}(\xi) \triangleq \{i \in \{1, \dots, 2u\} \mid \xi_i = 0, \text{ se } i \leq u, \text{ ou } \xi_{i-u} = B_{i-u}, \text{ se } i > u\}$ . Será visto na prova do teorema abaixo que para cada ponto na fronteira  $\xi \in \partial G$  o conjunto  $D(\xi)$  representa as direções de reflexão que aparecem no ponto  $\xi$ .

**Condição 7.2.2.** *Para  $o(\cdot)$  uniforme em  $\xi \in \mathbb{R}^u$ , assume-se que:*

- (a)  $\lambda^n(\xi) = r + f(\xi)/\sqrt{n} + o(1/\sqrt{n})$ , onde  $r \in \mathbb{R}_{\geq 0}^v$ ,  $f_k : \mathbb{R}^u \rightarrow \mathbb{R}$  são funções contínuas e limitadas quando restritas ao  $G$ , para  $k \in \{1, \dots, v\}$ .
- (b)  $A'r = 0$ , que é a condição do tráfego pesado.
- (c) Para cada  $\xi \in \partial G$ , existe um  $\nu \in \text{Cone}\{n_i, i \in \mathcal{E}(\xi)\}$  tal que  $\nu'd > 0$  para todo  $d \in D(\xi)$ , onde  $n_i$  é o vetor normal aparecendo na face de  $\partial G_i$ .

O item (a) da condição acima é a mesma condição que já foi adotada nos capítulos anteriores. O item (b) é a condição de tráfego pesado, que diz que a taxa com que itens são introduzidos nos nós é próxima da taxa com que itens são retirados destes nós. A última condição é uma generalização da condição “completamente- $\mathcal{S}$ ” que foi usada no Capítulo 5, que é exatamente a mesma condição dada pela Condição A.1.12.

**Teorema 7.2.3.** *Assuma que  $x^n(0)$  converge fracamente para  $x(0)$  e as Condições 7.2.1 e 7.2.2. Então  $\{\Psi^n\}$  é limitado em probabilidade, onde*

$$\Psi^n(\cdot) \triangleq (x^n(\cdot), m^n(\cdot), z^n(\cdot)),$$

e  $z^n(\cdot) \triangleq \sqrt{n} \int_0^\cdot \gamma(\tilde{\mathcal{E}}(x^n(s))) ds$ . Tome qualquer subsequência convergente de  $\{\Psi^n\}$  e denote seu limite no sentido fraco por  $\Psi(\cdot) \triangleq (x(\cdot), m(\cdot), z(\cdot))$ . Além disso, seja  $\mathcal{G}_t$  a mínima  $\sigma$ -álgebra que mensura  $\{x(0), m(s), z(s); s \leq t\}$ . Então, para  $t \geq 0$ ,

$\Psi(\cdot)$  satisfaz:

$$x(t) = x(0) + A' \int_0^t f(x(s))ds + A'm(t) + z(t) \quad (7.7)$$

onde  $x(t) \in G$ , para todo  $t \geq 0$ ,  $m$  é um processo de Wiener tomando valores em  $\mathbb{R}^v$  e  $\mathcal{G}_t$ -adaptado com matriz de covariância dada por  $\mathbb{E}[m(1)m(1)'] = \text{diag}(r)$ . O processo  $z$  é o processo de reflexão, onde

$$z(t) = \sum_{e \in \mathfrak{S} \cup \mathfrak{E}} \gamma(e) \zeta_e(t),$$

$\zeta_e$  é contínuo, não-decrescente,  $\zeta_e(0) = 0$ , e somente aumenta nos instantes  $t$  tais que  $x(t) \in \cap_{i \in e} \partial G_i$ .

*Demonstração.* Suponha primeiramente que  $f$  é limitada em todo o espaço  $\mathbb{R}^u$ . Usando a decomposição dada pela Equação (7.5) com as Condições 7.2.2(a) e 7.2.2(b), pode-se escrever  $x^n$  como:

$$\begin{aligned} x^n(t) &= x^n(0) + \int_0^t \mathbb{A}'(x^n(s-)) dm^n(s) + \sqrt{n} A' r t + \int_0^t \left[ A' - \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) \right] f(x^n(s)) ds \\ &\quad - \sqrt{n} \int_0^t \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) r ds + \sqrt{n} o(1/\sqrt{n}) \\ &= x^n(0) + \int_0^t \mathbb{A}'(x^n(s-)) dm^n(s) + \int_0^t \left[ A' - \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) \right] f(x^n(s)) ds \\ &\quad + \sqrt{n} \int_0^t \gamma(\tilde{\mathcal{E}}(x^n(s))) ds + \sqrt{n} o(1/\sqrt{n}). \end{aligned}$$

O Teorema A.2.3 será usado para mostrar que  $\{z^n\}$  é limitada em probabilidade.

Para isso, precisa-se mostrar que

$$\psi^n(\cdot) \triangleq \int_0^\cdot \mathbb{A}'(x^n(s-)) dm^n(s) + \int_0^\cdot \left[ A' - \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) \right] f(x^n(s)) ds$$

é assintoticamente contínuo. Primeiramente, usa-se o critério dado pela Proposição B.1.5 para mostrar que o processo tomando valores  $\int_0^t \left[ A' - \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) \right] f(x^n(s)) ds$  é limitado em probabilidade. De fato, como  $f$  é limitada, para qualquer sequência

$\{\phi^n\}$  de processos com funções amostras càdlàg, existe um  $K > 0$  tal que

$$\begin{aligned} \lim_{\delta \rightarrow 0} \limsup_n \mathbb{P} \left( \sup_{t \leq T} \sup_{s \leq \delta} \left| \int_t^{t+s} \left[ A' - \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) \right] f(x^n(s)) ds \right| \geq \nu \right) \\ \leq \lim_{\delta \rightarrow 0} \mathbb{P}(K\delta \geq \nu) = 0, \end{aligned}$$

para qualquer constante  $\nu > 0$ , e  $T > 0$ .

Observe que os martingales  $m_k^n$  possuem o seguinte processo de Doob-Meyer associado:

$$\langle m_k^n \rangle(t) = \int_0^t \lambda_k^n(x^n(s)) ds, \quad (7.8)$$

que pode ser verificado com o uso do resultado dado pelo Teorema B.2.4. Além disso, tem-se que  $\langle m_k^n, m_j^n \rangle = 0$ , para  $k \neq j$ , já que os instantes de salto de  $m_k^n$  e  $m_j^n$  são distintos com probabilidade um. Como  $\lambda^n$  é limitado, tem-se que  $\{m^n\}$  é assintoticamente contínua pelo Teorema B.1.12. Usando o critério do Teorema 2.5.6 de (Kushner, 2001) na página 70, e a inequação

$$\mathbb{E} \left[ \left( \int_{\tau}^{\tau+\delta} \mathbb{A}'_{ik}(x^n(s-)) dm^n(s) \right)^2 \right] \leq (A'_{ik})^2 \mathbb{E} \left[ \int_{\tau}^{\tau+\delta} \lambda_k^n(x^n(s)) ds \right] + \epsilon_n,$$

onde  $\tau$  é um tempo de parada limitado,  $\delta > 0$ , e  $\epsilon_n \rightarrow 0$  a medida que  $n \rightarrow \infty$ , pode-se verificar que o processo que toma valores  $\int_0^t \mathbb{A}'(x^n(s-)) dm^n(s)$  é limitado em probabilidade. Já que o tamanho dos saltos de  $m^n$  são da ordem de  $1/\sqrt{n}$ , o processo também é assintoticamente contínuo.

Neste ponto, o termo  $z^n(\cdot)$  será reescrito. Observe que pode-se escrever o

seguinte:

$$\begin{aligned}
z^n(t) &\triangleq \sqrt{n} \int_0^t \gamma(\tilde{\mathcal{E}}(x^n(s))) ds \\
&= \sum_{\{l\} \in \mathfrak{F}} \sqrt{n} \int_0^t \gamma(\tilde{\mathcal{E}}(x^n(s))) \mathbb{I}\{x^n(s) \in \tilde{\partial}_l, \text{ e } x^n(s) \notin \tilde{\partial}_j \text{ para } j \neq l\} ds \\
&+ \sum_{e \in \mathfrak{E}} \sqrt{n} \int_0^t \gamma(\tilde{\mathcal{E}}(x^n(s))) \mathbb{I}\{x^n(s) \in \cap_{i \in e} \tilde{\partial}_i, \text{ e } x^n(s) \notin \tilde{\partial}_j \text{ para } j \notin e\} ds.
\end{aligned}$$

Como  $\gamma(\tilde{\mathcal{E}}(x^n(s)))$  é constante dentro destas integrais, pode-se escrever

$$\sqrt{n} \int_0^t \gamma(\tilde{\mathcal{E}}(x^n(s))) ds = \sum_{\{l\} \in \mathfrak{F}} \gamma(\{l\}) \zeta_{\{l\}}^n(t) + \sum_{e \in \mathfrak{E}} \gamma(e) \zeta_e^n(t),$$

onde  $\zeta_S^n(t) \triangleq \sqrt{n} \int_0^t \mathbb{I}\{x^n(s) \in \cap_{i \in S} \tilde{\partial}_i, \text{ e } x^n(s) \notin \tilde{\partial}_j \text{ para } j \notin S\} ds$ , para qualquer conjunto  $S \in \mathfrak{F} \cup \mathfrak{E}$ .

Para qualquer  $e \in \mathfrak{E}$ , seja  $G_e \triangleq \{\xi \in \mathbb{R}^u : n'_e \xi \geq -\sum_{i \in e, i > u} B_{i-u}\}$ , onde o vetor normal  $n_e$  é dado por

$$(n_e)_i \triangleq \begin{cases} 1 & \text{se } i \in e \\ -1 & \text{se } i + u \in e \\ 0 & \text{caso contrário,} \end{cases}$$

para  $i \in \{1, \dots, u\}$ . Observe que o espaço de estados  $G$  pode ser escrito como

$$G = \cap_{\{i\} \in \mathfrak{F}} G_i \cap_{e \in \mathfrak{E}} G_e,$$

onde os vetores  $\gamma(e)$ ,  $e \in \mathfrak{E}$ , podem ser interpretados como sendo direções de reflexão na face de  $\{\xi \in \mathbb{R}^u : n'_e \xi = -\sum_{i \in e, i > u} B_{i-u}\}$ . Desta forma, o Teorema A.2.3 pode ser usado para mostrar que  $\{z^n\}$  é limitado em probabilidade. Além disso, pode-se verificar que qualquer subsequência fracamente convergente de  $\{z^n\}$

converge para  $z$  que satisfaz:

$$z(t) = \sum_{\{l\} \in \mathfrak{F}} \gamma(\{l\}) \zeta_{\{l\}}(t) + \sum_{e \in \mathfrak{E}} \gamma(e) \zeta_e(t) = \sum_{e \in \mathfrak{F} \cup \mathfrak{E}} \gamma(e) \zeta_e(t),$$

onde  $\zeta_e$  são contínuos, não-decrescentes,  $\zeta_e(0) = 0$ , e aumentam somente nos instantes  $t$  tais que  $x(t) \in \cap_{i \in e} \partial_i$ .

Como  $\{x^n(0)\}$ ,  $\{\psi^n\}$  e  $\{z^n\}$  são limitados em probabilidade, a sequência  $\{x^n\}$  também é, e pela continuidade quase certamente das funções amostras, tem-se que  $\{x^n(0), m^n, z^n, x^n\}$  é limitado em probabilidade. Agora resta caracterizar o limite no sentido fraco.

Como  $f$  é limitado, verifica-se que

$$\left| \int_0^t \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) f(x^n(s)) ds \right| \leq K \sum_{e \in \mathfrak{F} \cup \mathfrak{E}} \zeta_e(t) / \sqrt{n},$$

onde  $K$  é uma constante positiva. Como  $\{\zeta_e^n\}$  é limitado em probabilidade, o processo tomando valores  $\int_0^t \mathcal{R}(\tilde{\mathcal{E}}(x^n(s))) f(x^n(s)) ds$  converge fracamente para o processo “zero”. Além disso, pela continuidade de  $f$ , tem-se que se  $\phi^n \Rightarrow \phi$  então  $\int_0^t A' f(\phi^n(s)) ds \Rightarrow \int_0^t A' f(\phi(s)) ds$ .

Observe também que para qualquer  $t \geq 0$ :

$$\mathbb{E} \left[ \left( \int_0^t A'_{ik} \mathcal{I}_{ik}(\tilde{\mathcal{E}}(x^n(s))) dm^n(s) \right)^2 \right] \leq (A'_{ik})^2 \mathbb{E} \left[ \int_0^t \mathcal{I}_{ik}(\tilde{\mathcal{E}}(x^n(s))) \lambda^n(x^n(s)) ds \right] + \epsilon_n,$$

onde o termo do lado direito converge para zero a medida que  $n \rightarrow \infty$ , já que  $\{\zeta_e\}$  é limitado em probabilidade. Agora o Teorema B.1.12 pode ser aplicado junto com o fato de  $m^n = (m_1^n, \dots, m_v^n)'$  possuir descontinuidades na ordem de  $1/\sqrt{n}$  e com a Condição 7.2.2(a) para obter o resultado desejado.

Agora suponha que  $f_k$ ,  $k \in \{1, \dots, v\}$ , é limitado quando restrito ao  $G$ , mas pode ser ilimitado próximo das fronteiras. Pela definição de  $x^n$ , e tendo em mente a discussão sobre *backlogging* na seção anterior, se  $x^n(t) \notin G$  então  $\text{dist}(x^n(t), G) \leq C/\sqrt{n}$ , onde  $C = \max_{ji} |A_{ij}| - 1$ . Então, pode-se usar a técnica

de truncagem da mesma forma como foi usada na prova do Teorema 5.2.5 para chegar ao resultado.  $\square$

O teorema abaixo simplifica consideravelmente a representação do processo de reflexão  $z$  para os casos onde a difusão não é degenerada (i.e., o processo de Wiener associado possui matriz de covariância definida positiva).

**Teorema 7.2.4.** *Considere as hipóteses usadas no Teorema 7.2.3 e suponha ainda que a matriz  $\Sigma \triangleq A' \text{diag}(r) A$  é positiva definida. Então o limite no sentido fraco de  $\Psi^n$  (definido como no Teorema 7.2.3) satisfaz:*

$$x(t) = x(0) + A'm(t) + A' \int_0^t f(x(s))ds + \sum_{i=1}^{2u} \gamma(\{i\})\zeta_i(t),$$

onde  $m$  é definido como no Teorema 7.2.3,  $\zeta_i$  são contínuos, não-decrescentes,  $\zeta_i(0) = 0$ , e aumentam somente nos instantes  $t$  tais que  $x(t) \in \partial G_i$ .

*Demonstração.* Este resultado segue com a aplicação do Teorema A.3.5 que afirma que para cada  $t > 0$ :

$$\zeta_S(t) = \int_0^t \mathbb{I}\{x(s) \in \cap_{i \in S} \partial G_i\} d\zeta_S(s) = 0 \text{ q.c.},$$

para  $S \in \mathfrak{E}$ . Então pode-se simplificar o termo de reflexão da seguinte forma:

$$z(t) = \sum_{i=1}^{2u} \gamma(\{i\})\zeta_i(t).$$

$\square$

**Observação 7.2.5.** *As condições sobre as funções  $f_k$ ,  $k \in \{1, \dots, v\}$ , podem ser relaxadas de forma análoga ao que foi feito nos Teoremas 5.2.4 e 5.2.5.*

### 7.3 Redes de Petri Estocásticas com Transições Aleatórias

Para considerar uma classe mais geral de redes de Petri que podem representar completamente o roteamento presente em sistemas de filas, introduz-se uma

rede de Petri onde o efeito de cada transição é aleatório. Por exemplo, suponha as seguintes transições:

$$\begin{aligned} p_1 + 3p_2 &\xrightarrow{t_1} 2p_3 \quad (\text{com probabilidade } q_1) \\ 2p_1 + 2p_2 &\xrightarrow{t_1} 2p_3 \quad (\text{com probabilidade } q_2) \\ 2p_1 + 3p_2 &\xrightarrow{t_1} 2p_3 + p_4 \quad (\text{com probabilidade } q_3), \end{aligned}$$

onde  $q_1 + q_2 + q_3 \leq 1$ ,  $p_1, p_2, p_3$  e  $p_4$  são nós de uma rede de Petri e  $t_1$  é uma transição. Cada uma destas possibilidades é chamada de possível *efeito* da transição  $t_k$ .

Suponha que para cada  $k \in \{1, \dots, v\}$  existam  $c_k \in \mathbb{N}$  possíveis efeitos da transição  $t_k$ . Seja  $Pre_k \in \mathbb{N}^{c_k \times u}$  a matriz construída com os pesos dos arcos ligando nós com a transição  $t_k$ , para todo efeito possível. Isto é, para  $(i, j) \in \{1, \dots, c_k\} \times \{1, \dots, u\}$ ,  $(Pre_k)_{ij}$  é o número de itens removidos do nó  $p_j$  quando a transição  $t_k$  é disparada sob o efeito  $i$ . De forma análoga, defina  $Post_k \in \mathbb{N}^{c_k \times u}$  como a matriz construída com os pesos de arcos ligando a transição  $t_k$  com os nós para todo o efeito possível. Novamente, para  $(i, j) \in \{1, \dots, c_k\} \times \{1, \dots, u\}$ ,  $(Post_k)_{ij}$  é o número de itens recebidos pelo nó  $p_j$  quando a transição  $t_k$  é disparada sob o efeito  $i$ . Então defina  $A_k \triangleq Post_k - Pre_k$ .

De forma análoga ao caso anterior, define-se os conjuntos  $E_{kij} \triangleq \{ \text{índices dos nós que, quando vazios, impedem o efeito } i \text{ da transição } k \text{ de mudar o estado do nó } j \}$  e  $F_{kij} \triangleq \{ \text{índices dos nós que, quando cheios, impedem o efeito } i \text{ da transição } k \text{ de mudar o estado do nó } j \}$ .

Desta vez, define-se o processo  $T_k(t)$  como sendo um vetor aleatório tomando valores em  $\mathbb{N}^{c_k}$  onde cada componente  $T_{ki}(t)$ ,  $i \in \{1, \dots, c_k\}$ , conta o número de vezes que a transição  $t_k$  foi disparada sob o efeito  $i$  até o instante  $t \in \mathbb{R}_{\geq 0}$ . Formalmente, define-se

$$T_k(t) \triangleq \int_0^t I_k(s) d\mathcal{N}_k(s), \quad \text{e} \quad \mathcal{N}_k(t) \triangleq N_k \left( \int_0^t \Lambda(X(s)) ds \right),$$

onde  $N_k$  é um processo de Poisson, e  $I_{ki}$ , para  $i \in \{1, \dots, c_k\}$ , são funções indicado-

ras  $\mathcal{F}_t$ -adaptadas de eventos disjuntos tais que  $\sum_{i=1}^{c_k} I_{ki}(t) \leq 1$ , para todo  $t \in \mathbb{R}_{\geq 0}$  e  $k \in \{1, \dots, v\}$ .

Seja a matriz  $\mathbb{A}_k(\xi)$  definida como segue, para cada  $k \in \{1, \dots, v\}$  e  $\xi \in \mathbb{R}^K$ :

$$(\mathbb{A}_k)_{ij}(\xi) \triangleq (A_k)_{ij} \mathbb{I}\{\xi_l > 0, \text{ para } l \in E_{kij}, \text{ e } \xi_l < \tilde{B}, \text{ para } l \in F_{kij}\}.$$

Portanto, o sistema é dado por

$$X(t) = X(0) + \sum_{k=1}^v \int_0^t \mathbb{A}'_k(X(s-)) dT_k(s).$$

Considere as seguintes condições:

**Condição 7.3.1.**

- (a) As funções  $\Lambda_k : \mathbb{R}^u \rightarrow \mathbb{R}_{\geq 0}$  são mensuráveis e limitadas superiormente.
- (b) Os elementos aleatórios  $\{X(0), N_k, I_k \triangleq (I_{k1}, \dots, I_{kc_k})'; k \in \{1, \dots, v\}\}$  são mutuamente independentes.
- (c)  $\mathbb{E}[I_{ki}(t) | \mathcal{F}_t^r] = Q_{ki}(X(t-))$ , onde  $Q_{ki} : \mathbb{R}^u \rightarrow [0, 1]$  é uma função mensurável e  $\mathcal{F}_t^r$  é a mínima  $\sigma$ -álgebra que mensura todos os processos definidos acima até o instante  $t$  exceto  $I_k(t)$ .

Defina

$$M_{ki}^Q(t) \triangleq \int_0^t (I_{ki}(s) - Q_{ki}(X(s-))) d\mathcal{N}_k(s). \quad (7.9)$$

Pode-se verificar que  $M_{ki}^Q$  é um  $\mathcal{F}_t$ -martingale usando os mesmos argumentos usados no Capítulo 5, ou seja, observando que se  $s_l$  denota o instante do  $l$ -ésimo salto de  $\mathcal{N}_k$  após o tempo  $s$ , então  $\mathbb{E}[I_{ki}(s_l) - Q_{ki}(X(s_l-)) | \mathcal{F}_{s_l}^r] = 0$ . Portanto, pode-se escrever o seguinte:

$$\begin{aligned} T_{ki}(t) &= M_{ki}^Q(t) + \int_0^t Q_{ki}(X(s-)) d\mathcal{N}_k(s) \\ &= M_{ki}^Q(t) + \int_0^t Q_{ki}(X(s-)) dM_k(s) + \int_0^t Q_{ki}(X(s)) \Lambda_k(X(s)) ds, \end{aligned}$$

onde foi usado o Teorema de decomposição B.2.2 em conjunto como o Lema B.2.3.

Portanto, pode-se escrever a equação do sistema como:

$$\begin{aligned} X(t) = & X(0) + \sum_{k=1}^v \int_0^t \mathbb{A}'_k(X(s-)) dM_k^Q + \int_0^t \mathbb{A}'_k(X(s-)) Q_k(X(s-)) dM_k(s) \\ & + \int_0^t \mathbb{A}'_k(X(s)) Q_k(X(s)) \Lambda_k(X(s)) ds. \end{aligned} \quad (7.10)$$

De forma análoga ao que foi feito anteriormente, define-se o seguinte para cada conjunto  $S \subseteq \{1, \dots, 2u\}$ :

$$(\mathcal{I}_k)_{ji}(S) = \begin{cases} 1 & \text{se } S \cap (E_{kij} \cup \{u + F_{kij}\}) \neq \emptyset \\ 0 & \text{caso contrário,} \end{cases}$$

para cada  $k \in \{1, \dots, v\}$ ,  $i \in \{1, \dots, c_k\}$ , e  $j \in \{1, \dots, u\}$ . Portanto, é possível escrever:

$$\begin{aligned} X(t) = & X(0) + \sum_{k=1}^v \int_0^t \mathbb{A}'_k(X(s-)) dM_k^Q(s) + \int_0^t \mathbb{A}'_k(X(s-)) Q_k(X(s-)) dM_k(s) \\ & + A'_k \int_0^t Q_k(X(s)) \Lambda_k(X(s)) ds - \int_0^t \mathcal{R}_k(\tilde{\mathcal{E}}(X(s))) Q_k(X(s)) \Lambda_k(X(s)) ds, \end{aligned}$$

onde  $(\mathcal{R}_k)_{ji}(\cdot) \triangleq (A'_k)_{ji}(\mathcal{I}_k)_{ji}(\cdot)$ .

### 7.3.1 Limite em Regime de Tráfego Pesado

Suponha a sequência  $\{X^n\}$ , onde para cada  $n \in \mathbb{N}$  as matrizes  $A_k$  e as constantes  $c_k$  são fixas, para cada  $k \in \{1, \dots, v\}$ . Seja o escalonamento  $x^n(t) \triangleq X^n(nt)/\sqrt{n}$ , então pode-se escrever a expansão dada pela Equação (7.10) como:

$$\begin{aligned} x^n(t) = & x^n(0) + \sum_{k=1}^v \left[ \int_0^t \mathbb{A}'_k(x^n(s-)) dm_k^{Q,n}(s) + \int_0^t \mathbb{A}'_k(x^n(s-)) q_k^n(x^n(s-)) dm_k^n(s) \right. \\ & \left. + \sqrt{n} A'_k \int_0^t q_k^n(x^n(s)) \lambda_k^n(x^n(s)) ds - \sqrt{n} \int_0^t \mathcal{R}_k(\tilde{\mathcal{E}}(x^n(s))) q_k^n(x^n(s)) \lambda_k^n(x^n(s)) ds \right], \end{aligned}$$

onde  $m_{ki}^{Q,n}(t) \triangleq M_{ki}^{Q,n}(nt)/\sqrt{n}$ ,  $m_{ki}^n(t) \triangleq M_{ki}^n(nt)/\sqrt{n}$ ,  $x^n(0) \triangleq X^n(0)/\sqrt{n}$ ,  $\lambda_k^n(\xi) \triangleq \Lambda_k^n(\sqrt{n}\xi)$  e  $q_{ki}^n(\xi) \triangleq Q_{ki}^n(\sqrt{n}\xi)$ , para  $\xi \in \mathbb{R}^u$ ,  $t \in \mathbb{R}_{\geq 0}$ , e  $i \in \{1, \dots, c_k\}$ . Da mesma forma como foi feito na Seção 7.2.1, redefini-se  $\mathbb{A}_k$  e  $\tilde{\mathcal{E}}$  para que possuam  $B_i$  em suas respectivas definições ao invés de  $\bar{B}_i$ .

Defina o seguinte vetor  $\gamma(S) \triangleq -\sum_{k=1}^v \mathcal{R}_k(S) p_k r_k$ , onde  $p_k$  e  $r_k$  serão definidos abaixo, e considere as seguintes condições:

**Condição 7.3.2.** *Para  $o(\cdot)$  uniforme em  $\xi \in \mathbb{R}^u$ , assume-se que*

- (a)  $\lambda^n(\xi) = r + f(\xi)/\sqrt{n} + o(1/\sqrt{n})$ , onde  $r \in \mathbb{R}_{\geq 0}^v$ ,  $f_k : \mathbb{R}^u \rightarrow \mathbb{R}$  são contínuas e limitadas quando restritas ao  $G$ , para  $k \in \{1, \dots, v\}$ .
- (b)  $q_k^n(\xi) = p_k + g_k(\xi)/\sqrt{n} + o(1/\sqrt{n})$ , onde  $p_k \in \mathbb{R}_{\geq 0}^{c_k}$  e  $g_{ki} : \mathbb{R}^u \rightarrow \mathbb{R}$ , para  $k \in \{1, \dots, v\}$  e  $i \in \{1, \dots, c_k\}$ , são funções contínuas.
- (c)  $\sum_{k=1}^v A'_k p_k r_k = 0$ , que é a condição do tráfego pesado.
- (d) Para cada  $\xi \in \partial G$ , existe um  $\nu \in \text{Cone}\{n_i, i \in \mathcal{E}(\xi)\}$  tal que  $\nu'd > 0$  para todo  $d \in D(\xi)$ , onde  $D(\xi)$  é definido como na Equação (7.6).

As condições acima têm a mesma interpretação das que foram introduzidas na Seção 7.2.1, com somente a adição da Condição 7.3.2(b) que diz como a dependência de estado é introduzida para a probabilidade com que cada efeito de uma transição ocorre.

**Teorema 7.3.3.** *Assuma que  $x^n(0)$  converge fracamente para  $x(0)$  e as Condições 7.3.1 e 7.3.2. Então  $\{\Psi^n\}$  é limitada em probabilidade onde*

$$\Psi^n(\cdot) \triangleq (x^n(\cdot), m^{Q,n}(\cdot), m^n(\cdot), z^n(\cdot)),$$

$m^{Q,n} \triangleq (m_1^{Q,n}, \dots, m_v^{Q,n})'$ ,  $m^n \triangleq (m_1^n, \dots, m_v^n)'$ , e  $z^n(\cdot) \triangleq \sqrt{n} \int_0^\cdot \gamma(\tilde{\mathcal{E}}(x^n(\cdot))) ds$ . Tome uma sequência fracamente convergente e denote seu limite fraco por  $\Psi(\cdot) \triangleq (x(\cdot), m^Q(\cdot), m(\cdot), z(\cdot))$ . Além disso, seja  $\mathcal{G}_t$  a mínima  $\sigma$ -álgebra que mensura

$\{x(0), m^Q(s), m(s), z(s); s \leq t\}$ . Então, para  $t \geq 0$ ,  $\Psi(\cdot)$  satisfaz:

$$x(t) = x(0) + \sum_{k=1}^v A'_k \left[ m_k^Q(t) + p_k m_k(t) + \int_0^t g_k(x(s)) r_k + p_k f_k(x(s)) ds \right] + z(t) \quad (7.11)$$

onde  $x(t) \in G$ ,  $m_k^Q$  é um processo de Wiener  $\mathcal{G}_t$ -mensurável, tomando valores em  $\mathbb{R}^{c_k}$ , e com matriz de covariância dada por  $\mathbb{E} \left[ m_k^Q(1) m_k^Q(1)' \right] = \Sigma_k r_k$ , onde

$$(\Sigma_k)_{ij} \triangleq \begin{cases} (1 - p_{ki}) p_{ki} & \text{if } i = j \\ -p_{ki} p_{kj} & \text{if } i \neq j, \end{cases}$$

para cada  $k \in \{1, \dots, v\}$ , e  $m$  é um processo de Wiener  $\mathcal{G}_t$ -mensurável, tomando valores em  $\mathbb{R}^v$ , independente de  $m_k^Q$ , para  $k \in \{1, \dots, v\}$ , e com matriz de covariância dada por  $\mathbb{E} [m(1) m(1)'] = \text{diag}(r)$ . O processo  $z$  é o processo de reflexão, onde

$$z(t) = \sum_{e \in \mathfrak{F} \cup \mathfrak{E}} \gamma(e) \zeta_e(t)$$

$\zeta_e$  são contínuos, não-decrescentes,  $\zeta_e(0) = 0$ , e aumentam somente nos instantes  $t$  tais que  $x(t) \in \cap_{i \in e} \partial G_i$ .

*Demonstração.* Suponha que  $f_k$  é limitado em todo o  $\mathbb{R}^u$  para cada  $k \in \{1, \dots, v\}$ , caso contrário pode-se usar o argumento de truncagem como foi discutido na prova do Teorema 7.2.3. Primeiramente será mostrado que  $\{m_k^{Q,n}\}$  é limitado em probabilidade. Novamente, isto é feito computando o processo de Doob-Meyer associado, que é dado por

$$\langle m_k^{Q,n} \rangle(t) = \int_0^t \tilde{\Sigma}_k^n(x^n(s)) \lambda_k(x^n(s)) ds,$$

usando o Lema B.2.5, onde  $(\tilde{\Sigma}_k^n)_{ij}(\xi) = \delta_{ij} q_{ki}^n(\xi) - q_{ki}^n(\xi) q_{kj}^n(\xi)$ , para  $i, j \in \{1, \dots, v\}$ ,  $\xi \in \mathbb{R}^u$ , e  $\delta_{ij} = 1$  se  $i = j$  e 0 caso contrário. Como  $\lambda^n$  e  $q^n$  são limitados e usando

o Teorema B.1.12, tem-se que  $\{m_k^{Q,n}\}$  é limitada em probabilidade. Além disso, já que  $m_k^{Q,n}$  possui descontinuidades da ordem de  $1/\sqrt{n}$ , o processo é assintoticamente contínuo. Usando os mesmos argumentos dados na prova do Lema B.2.5, em conjunto com a Condição 7.3.1(c) (como foi feito na prova do Teorema 5.2.3), verifica-se que  $\langle m_{ki}^{Q,n}, m_k^n \rangle = 0$ . Observe também que  $\langle m_{ki}^{Q,n}, m_{lj}^{Q,n} \rangle = 0$  para  $l \neq k$  já que os instantes dos saltos de  $m_{ki}^{Q,n}$  e  $m_{lj}^{Q,n}$  são distintos com probabilidade um.

Usando os mesmos argumentos do Teorema 7.2.3, tem-se que

$$\langle m_k^n \rangle(t) = \int_0^t \lambda_k^n(x^n(s)) ds$$

e  $\langle m_k^n, m_l^n \rangle = 0$  para  $k \neq l$ . Além disso,  $\{m_k^n\}$  é assintoticamente contínua.

Observe que pode-se escrever para cada  $k \in \{1, \dots, v\}$

$$\begin{aligned} & \sum_{k=1}^v \sqrt{n} \int_0^t \left( A'_k - \mathcal{R}_k(\tilde{\mathcal{E}}(x^n(s))) \right) q_k^n(x^n(s)) \lambda_k^n(x^n(s)) ds \\ &= \sum_{k=1}^v \left[ \sqrt{n} A'_k p_k r_k t - \sqrt{n} \int_0^t \mathcal{R}_k(\tilde{\mathcal{E}}(x^n(s))) p_k r_k ds \right. \\ & \quad \left. + \int_0^t \left( A'_k - \mathcal{R}_k(\tilde{\mathcal{E}}(x^n(s))) \right) (g_k(x^n(s)) r_k + p_k f_k(x^n(s))) ds + \sqrt{n} o(1/\sqrt{n}) \right] \\ &= \sqrt{n} \int_0^t \gamma(\tilde{\mathcal{E}}(x^n(s))) ds + \sum_{k=1}^v \int_0^t \left( A'_k - \mathcal{R}_k(\tilde{\mathcal{E}}(x^n(s))) \right) \\ & \quad \times (g_k(x^n(s)) r_k + p_k f_k(x^n(s))) ds + \sqrt{n} o(1/\sqrt{n}). \end{aligned} \tag{7.12}$$

Portanto, usando o mesmo argumento usado na prova do Teorema 7.2.3, mostra-se que  $\{z^n\}$  é limitado em probabilidade, onde qualquer limite no sentido fraco de  $z^n$  é dado por  $z$ . Além disso, é possível mostrar que

$$\sum_{k=1}^v \int_0^t \left( A'_k - \mathcal{R}_k(\tilde{\mathcal{E}}(x^n(s))) \right) (g_k(x^n(s)) r_k + p_k f_k(x^n(s))) ds + \sqrt{n} o(1/\sqrt{n})$$

converge fracamente para o processo tomando valores

$$\sum_{k=1}^v \int_0^t A'_k (g_k(x^n(s)) r_k + p_k f_k(x^n(s))) ds.$$

Pode-se obter a caracterização do limite no sentido fraco de  $m_k^n$  usando os argumentos do Teorema 7.2.3. Além disso, a caracterização do limite de  $m_k^{Q,n}$  como  $m_k^Q$  é obtido usando as Condições 7.3.2(a) e (b) e o Teorema B.1.12.

□

O teorema abaixo é equivalente ao Teorema 7.2.4 para o modelo considerado nesta seção. O resultado simplifica a caracterização do processo  $z$ , removendo as direções de reflexão que aparecem nas “quinas” e “arestas” do espaço de estado.

**Teorema 7.3.4.** *Considere as hipóteses usadas no Teorema 7.3.3 e suponha ainda que a matriz  $\Sigma \triangleq \sum_{k=1}^v r_k A'_k \text{diag}(p_k) A_k$  é positiva definida. Então o limite de  $\{\Psi^n\}$  (definido no Teorema 7.3.3) satisfaz*

$$x(t) = x(0) + w(t) + \sum_{k=1}^v \int_0^t g_k(x(s)) r_k + p_k f_k(x(s)) ds + \sum_{i=1}^{2u} \gamma(\{i\}) \zeta_i(t),$$

onde  $w$  é um processo de Wiener  $\mathbb{R}^u$ -valorado e  $\mathcal{G}_t$ -mensurável com matriz de covariância dada por  $\Sigma$  (onde  $\mathcal{G}_t$  é a mínima  $\sigma$ -álgebra que mensura  $\{x(0), w(s), z(s); s \leq t\}$ ) e  $\zeta_i$  são contínuos, não-decrescentes,  $\zeta_i(0) = 0$ , e aumentam somente nos instantes  $t$  tais que  $x(t) \in \partial G_i$ .

*Demonstração.* A simplificação da representação do processo de reflexão segue os mesmos argumentos usados no Teorema 7.2.4. Além disso, pode-se trocar o processo de Wiener dado pela Equação (7.11) pelo que está dado na equação limite acima. Apenas observe que a matriz de covariância  $\Sigma = \sum_{k=1}^v r_k A'_k (\Sigma_k + p_k p'_k) A_k$  é simplificada para  $\sum_{k=1}^v r_k A'_k \text{diag}(p_k) A_k$  pois  $\Sigma_k + p_k p'_k = \text{diag}(p_k)$ . □

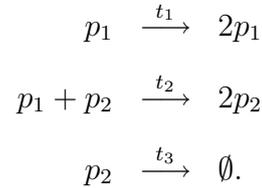
**Observação 7.3.5.** *O resultado dos Teoremas 5.2.4 e 5.2.5 podem ser aplicado no modelo desta seção. Seguindo as idéias do Teorema 5.2.4, permite-se descontinuidades em  $g_k$ , para  $k \in \{1, \dots, v\}$ , se for estabelecido que as funções  $\phi(\cdot) \mapsto \int_0^\cdot g_k(\phi(s)) ds$  são contínuas de  $D(\mathbb{R}^u; 0, \infty)$  a  $D(\mathbb{R}; 0, \infty)$  com probabilidade um em relação a medida induzida por qualquer limite no sentido fraco de  $x^n$ .*

## 7.4 Experimentos Numéricos

Nesta seção serão apresentados alguns problemas práticos, além daqueles já apresentados nos Capítulos 5 e 6, que ilustram a aplicação da aproximação por tráfego pesado desenvolvida aqui.

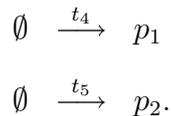
### 7.4.1 O Modelo de Predador-Presa de Lotka-Volterra

O modelo de presa-predador usado em (Wilkinson, 2006) será considerado nesta seção. O modelo pode ser descrito por dois nós  $p_1$  e  $p_2$  e três transições  $t_1$ ,  $t_2$ , e  $t_3$  que são dadas por:



Seja  $X_i(t)$  o número de itens de  $p_i$  no instante  $t \in \mathbb{R}_{\geq 0}$ , para  $i \in \{1, 2\}$ . Para usar a aproximação por tráfego pesado, procede-se da seguinte forma: um valor de  $n \in \mathbb{N}$  “grande” é escolhido de tal forma que as Condições 7.2.2(a) e (b) sejam atendidas. Desta forma, pode-se aproximar  $X(t)$  por  $x(tn^{-1})n^{1/2}$  onde  $x$  é o limite dado pela Equação (7.7).

Neste exemplo, a Condição 7.2.2(c) não é atendida. Isso acontece pois se o número de itens nos nós  $p_1$  ou  $p_2$  acabarem, o sistema não pode mais continuar. Portanto, as duas novas transições abaixo são inseridas no modelo:



Com estas transições, sempre existirá um fluxo constante de entrada de itens de  $p_1$  e de  $p_2$ . Neste sistema, a hipótese do tráfego pesado (i.e., Condição 7.2.2(b)) diz que  $r_1 + r_4 = r_2$  e  $r_2 + r_5 = r_3$ , onde as constantes  $r_i$ ,  $i \in \{1, \dots, 5\}$ , são aquelas

dada pela Condição 7.2.2. Além disso, considera-se a capacidade máxima de cada nó dada por  $\sqrt{n}B_1$  e  $\sqrt{n}B_2$ . Desta forma, tem-se as seguintes direções de reflexão em cada fronteira do espaço de estado:

$$\begin{aligned}
\text{em } \partial G_1 &\triangleq \{\xi \in \mathbb{R}^2 : \xi_1 = 0\}: \gamma(\{1\}) = (r_2 - r_1, -r_2)' \\
\text{em } \partial G_2 &\triangleq \{\xi \in \mathbb{R}^2 : \xi_2 = 0\}: \gamma(\{2\}) = (r_2, r_3 - r_2)' \\
\text{em } \partial G_3 &\triangleq \{\xi \in \mathbb{R}^2 : \xi_1 = B_1\}: \gamma(\{3\}) = (-r_1 - r_4, 0)' \\
\text{em } \partial G_4 &\triangleq \{\xi \in \mathbb{R}^2 : \xi_2 = B_2\}: \gamma(\{4\}) = (r_2, -r_2 - r_5)' \\
\text{em } \partial G_1 \cap \partial G_2: &\gamma(\{1, 2\}) = (r_2 - r_1, r_3 - r_2)' \\
\text{em } \partial G_1 \cap \partial G_4: &\gamma(\{1, 4\}) = (-r_1 + r_2, -r_2 - r_5)' \\
\text{em } \partial G_2 \cap \partial G_3: &\gamma(\{2, 3\}) = (-r_1 + r_2 - r_4, -r_2 + r_3)' \\
\text{em } \partial G_3 \cap \partial G_4: &\gamma(\{3, 4\}) = (-r_1 + r_2 - r_4, -r_2 + r_5)',
\end{aligned}$$

que satisfaz a Condição 7.2.2(c) se  $r_i > 0$  para todo  $i \in \{1, \dots, 5\}$ . A matriz de covariância do processo de Wiener,  $A'm(t)$ , (presente na Equação (7.7)) é dada por

$$\Sigma = \begin{pmatrix} 2r_2 & -r_2 \\ -r_2 & 2r_3 \end{pmatrix},$$

e supondo que é não-degenerada, pode-se descartar as direções de reflexão que aparecem nas quinas e lados do espaço de estado pelo Teorema 7.2.4. Portanto, o processo de reflexão toma a seguinte forma:

$$\begin{aligned}
z(t) &= \begin{pmatrix} r_2 - r_1 & r_2 & -r_1 - r_4 & r_2 \\ -r_2 & r_3 - r_2 & 0 & -r_2 - r_5 \end{pmatrix} \zeta(t) \\
&= \begin{pmatrix} r_4 & r_2 & -r_2 & r_2 \\ -r_2 & r_5 & 0 & -r_3 \end{pmatrix} \zeta(t),
\end{aligned}$$

onde foi usado as identidades requeridas pela condição de tráfego pesado na última passagem para simplificar a matriz acima.

Seja as funções  $f_1$ ,  $f_2$ , e  $f_3$ , dadas da seguinte forma para  $\xi \in \mathbb{R}^2$ :

$$\begin{aligned} f_1(\xi) &= k_1 \xi_1 \\ f_2(\xi) &= k_2 \xi_1 \xi_2 \\ f_3(\xi) &= k_3 \xi_2, \end{aligned}$$

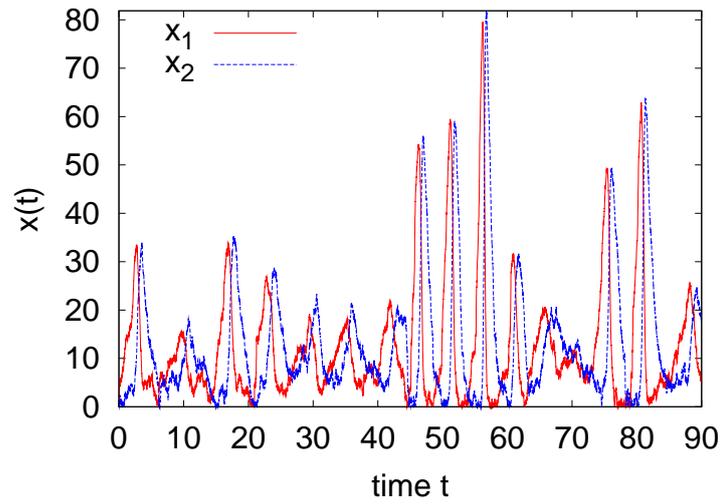
onde estas funções são aquelas requeridas pela Condição 7.2.2, e  $k_i$ ,  $i \in \{1, 2, 3\}$ , são constantes. Além disso, seja  $f_4(\cdot) = f_5(\cdot) = 0$ . Portanto,  $x$  é dado por:

$$\begin{aligned} x(t) = x(0) + \int_0^t \begin{pmatrix} k_1 x_1(s) - k_2 x_1(s)x_2(s) \\ k_2 x_1(s)x_2(s) - k_3 x_2(s) \end{pmatrix} ds + \begin{pmatrix} 2r_2 & -r_2 \\ -r_2 & 2r_3 \end{pmatrix}^{1/2} w(t) \\ + \begin{pmatrix} r_4 & r_2 & -r_2 & r_2 \\ -r_2 & r_5 & 0 & -r_3 \end{pmatrix} \zeta(t). \end{aligned}$$

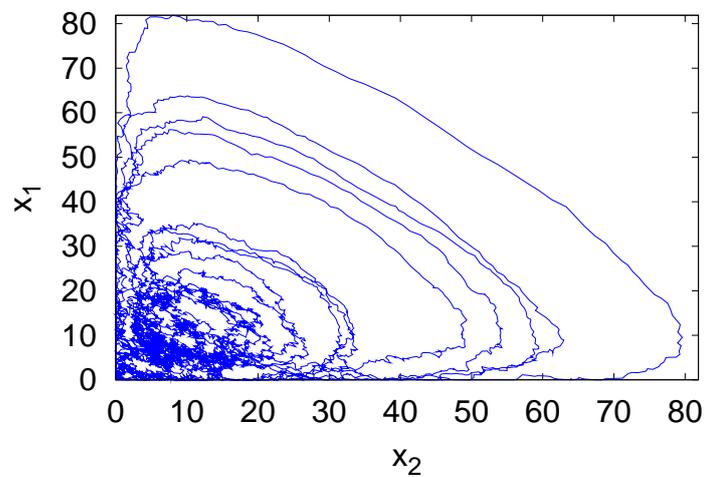
A Figura 7.2 tem funções amostras de  $x$  construída com o método de Euler (veja por exemplo (Kloeden et al., 1994)), onde o processo de reflexão é implementado empurrando o processo de volta ao espaço de estado (na direção do vetor de reflexão) caso atravesse uma das fronteiras. Neste exemplo, foi usado  $r_2 = 4$ ,  $r_3 = 5$ ,  $k_1 = 1$ ,  $k_2 = 0.1$ ,  $k_3 = 1$ ,  $x(0) = (0, 0)'$ , e as capacidades  $B_1$  e  $B_2$  foram escolhidas grandes o suficiente de forma a não atrapalhar o processo no intervalo de tempo considerado.

#### 7.4.2 Controle de Congestionamento de Rede

Nesta seção será considerado o sistema dado pela Figura 7.3. Neste sistema, existem duas filas em sequência e uma outra fila que não está diretamente conectada as outras. Suponha que cada fila possua uma capacidade máxima de clientes finita. Nos instantes em que as filas estiverem em suas capacidades máximas, os novos clientes que chegarem serão perdidos. Considere o cenário onde a fila 2 ( $Q_2$ ) tenta evitar esta perda de clientes enviando sinais para mover clientes da fila 1 ( $Q_1$ ) para a fila 3 ( $Q_3$ ). Estes sinais são enviados nos instantes em que clientes completam



(a)



(b)

Figura 7.2: (a) Função  $x$  versus o tempo para o modelo predador-presa. (b) Gráfico da componente  $x_1$  versus a componente  $x_2$  para a mesma função amostra.

seu serviço na fila 2. Ou seja, o sistema é dado pelas seguintes transições:

$$\begin{array}{l}
 \emptyset \xrightarrow{t_1} Q_1 \\
 Q_1 \xrightarrow{t_2} Q_2 \\
 \left. \begin{array}{l} Q_1 + Q_2 \\ Q_2 \end{array} \right\} \xrightarrow{t_3} \left\{ \begin{array}{l} Q_3 \\ \emptyset \end{array} \right. \\
 \emptyset \xrightarrow{t_4} Q_3 \\
 Q_3 \xrightarrow{t_5} \emptyset,
 \end{array}$$

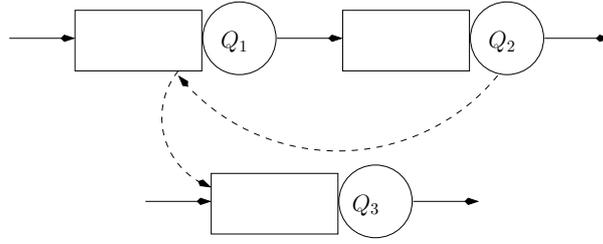


Figura 7.3: Diagrama de um sistema de filas com controle de congestionamento.

onde a transição  $t_3$  possui dois possíveis efeitos sobre o sistema. O primeiro efeito corresponde a saída de clientes da fila 2, e o segundo corresponde a saída de cliente da fila 2 junto com o deslocamento de um cliente da fila 1 para a fila 3.

Suponha que  $r_i > 0$ , para  $i \in \{1, 2, 3, 4, 5\}$ ,  $f_1(\cdot) = f_4(\cdot) = 0$ ,  $f_2(\cdot) = b_2$ ,  $f_3(\cdot) = b_3$ , e  $f_5(\cdot) = b_5$ , onde  $b_4$ ,  $b_3$ , e  $b_5$  são constantes. Além disso, seja  $p_{31} = 0$ ,  $p_{32} = 1$ ,  $g_{31}(\xi) = g(\xi)$ , e  $g_{32}(\xi) = -g(\xi)$ , para  $\xi \in \mathbb{R}^3$ , onde  $p_{31}$ ,  $p_{32}$ ,  $g_{31}$ , e  $g_{32}$  são como definidos na Condição 7.3.2, e  $g : \mathbb{R}_{\geq 0}^3 \rightarrow [0, 1]$  é uma função que será escolhida mais tarde. A condição do tráfego pesado exige que  $r_1 = r_2 = r_3$  e  $r_4 = r_5$ . Desta forma, a equação limite toma a seguinte forma:

$$x(t) = x(0) + \int_0^t \begin{pmatrix} -b_2 - g(x(s))r_3 \\ b_2 - b_3 \\ g(x(s))r_3 - b_3 \end{pmatrix} ds + \begin{pmatrix} r_1 + r_2 & -r_2 & 0 \\ -r_2 & r_2 + r_3 & 0 \\ 0 & 0 & r_4 + r_5 \end{pmatrix}^{1/2} w(t) + \begin{pmatrix} r_2 & 0 & 0 & -r_1 & 0 & 0 \\ -r_2 & r_3 & 0 & 0 & -r_2 & 0 \\ 0 & 0 & r_5 & 0 & 0 & -r_4 \end{pmatrix} \zeta(t),$$

onde foi assumido que a matriz de covariância do processo de Wiener é definida positiva e, desta forma, as direções de reflexão nas fronteiras podem ser ignoradas.

Desta vez, deseja-se escolher a função  $g$  de forma que minimize a perda de cliente na fila 2. Para isto, será usada a seguinte função custo:

$$W(x_0, g) = \mathbb{E}_{x_0}^g \left[ \int_0^\infty e^{-\beta t} (cg(x(t)))dt + v_1 d\zeta_5(t) + v_2 d\zeta_6(t) \right],$$

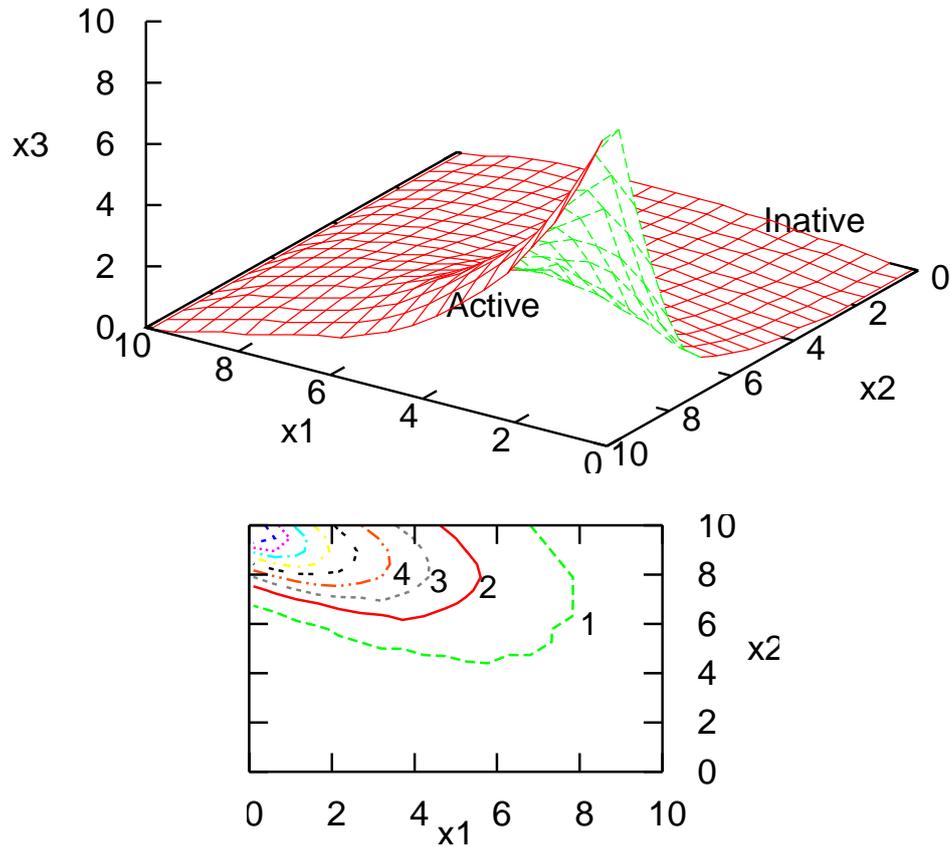


Figura 7.4: Superfície de chaveamento para o controle ótimo no espaço de estados escalonado. O controle é aplicado em força máxima na região debaixo da curva. Abaixo, tem-se o mapa de contorno desta superfície.

onde  $c$  é uma constante positiva associada ao custo de mover clientes,  $v_1$  é uma constante positiva associada ao custo de perder um cliente quando a capacidade máxima é atingida na fila 2,  $v_2$  é uma constante positiva associada ao custo da perda de clientes na fila 3, e  $x(0) = x_0$  é a condição inicial do sistema.

O método numérico da “Cadeia de Markov Aproximada” (veja (Kushner e Dupuis, 1992)) será usado para encontrar a solução do problema de controle dado acima. O parâmetro de discretização foi escolhido para ser  $h = 0.1$ , e as constantes  $\beta = 0.001$ ,  $c = 1$ ,  $v_1 = v_2 = 200$ ,  $r_1 = r_4 = 1$ ,  $b_1 = b_3 = 0.1$ , e  $b_2 = 0.2$ . Além disso, foram usados os seguintes valores para a capacidade máxima :  $B_1 = B_2 = B_3 = 10$ . A Figura 7.4 traz a curva de chaveamento encontrada como solução do problema de controle ótimo. Esta curva separa o espaço de estados em uma região onde o controle é aplicado em força máxima e um outra região onde o controle não é

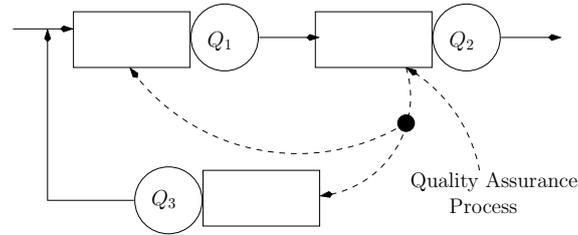


Figura 7.5: Diagrama de um sistema de linha de produção com certificador de qualidade. A linha pontilhada indica produtos que são enviados de volta para reprocessamento ou são enviados para uma terceira estação para serem restaurados.

aplicado.

### 7.4.3 Certificador de Qualidade em Linhas de Produção

Nesta seção, um sistema de linha de produção composto por duas filas em sequência será considerado, veja a Figura 7.5 para referência. Produtos que estão na segunda estação passam por um processo de certificação de qualidade. Os produtos que não atendem o requerimento mínimo são enviados de volta para a fila 1 para reprocessamento ou são enviados para uma terceira estação para restauração, antes de serem novamente enviados à linha de produção.

Este problema é representado pelas seguintes transições:

$$\begin{aligned}
 \emptyset & \xrightarrow{t_1} Q_1 \\
 Q_1 & \xrightarrow{t_2} Q_2 \\
 Q_2 & \xrightarrow{t_3} \emptyset \\
 Q_2 & \xrightarrow{t_4} \left\{ \begin{array}{l} Q_1 \\ Q_3 \end{array} \right. \\
 Q_3 & \xrightarrow{t_5} Q_1,
 \end{aligned}$$

onde a transição  $t_4$  possui dois efeitos possíveis. Suponha que as constantes  $r_i$ ,  $i \in \{1, 2, 3, 4, 5\}$ , são positivas, e que  $p_{41} + p_{42} = 1$  (onde estas constantes são aquelas definidas na Condição 7.3.2). Além disso, seja  $f_1(\cdot) = 0$ ,  $f_i(\cdot) = b_i$ ,  $i \in \{2, 3, 4, 5\}$ , onde  $b_i$  são constantes, e  $g_{41}(\cdot) = g_{42}(\cdot) = 0$ . Assuma também que a capacidade

máxima de cada fila seja finita. Para este exemplo, a condição do tráfego pesado exige que  $r_1 + p_1 r_4 + r_5 = r_2 = r_3 + r_4$ , e  $p_2 r_4 = r_5$ . Desta forma, o equação limite é escrita da seguinte forma:

$$\begin{aligned}
x(t) = & x(0) + \begin{pmatrix} -b_2 + b_5 + p_1 b_4 \\ b_2 - b_3 - b_4 \\ -b_5 + p_2 b_4 \end{pmatrix} t \\
& + \begin{pmatrix} r_1 + p_1 r_4 + r_2 + r_5 & -r_2 - p_1 r_4 & -r_5 \\ -r_2 - p_1 r_4 & r_2 + r_4 + r_3 & -p_2 r_4 \\ -r_5 & -p_2 r_4 & p_2 r_4 + r_5 \end{pmatrix}^{1/2} w(t) \\
& + \begin{pmatrix} r_2 & -p_1 r_4 & -r_5 & -r_1 - p_1 r_4 - r_5 & r_2 & 0 \\ -r_2 & r_3 + r_4 & 0 & p_1 r_4 & -r_2 & p_2 r_4 \\ 0 & -p_4 r_4 & r_5 & r_5 & 0 & -p_2 r_4 \end{pmatrix} \zeta(t),
\end{aligned}$$

onde, novamente, a matriz de covariância é assumida ser definida positiva e, desta forma, as direções de reflexão que aparecem nas quinas e lados do espaço de estado podem ser descartadas.

# Capítulo 8

## Conclusões e Trabalhos Futuros

O principal objetivo desse trabalho, foi abordar o problema de estabelecer aproximações para *rede de filas com sinais em regime de tráfego pesado*. O estudo apresentado aqui se inicia no modelo mais simples de uma fila única e termina com o estabelecimento de uma forma mais abrangente de descrever o modelo através de *redes de Petri com transições probabilistas*. Este estudo *estende significativamente* o conhecimento na área de redes com sinais, pois ainda não havia sido estudado aproximações por difusão nestes problemas. Esta aproximação possibilita o estudo da evolução transiente que é especialmente útil para trabalhar com problemas de controle, que são tão inerentes nestes sistemas de filas com sinais. O modelo é feito da forma mais geral que foi possível para possibilitar a aplicação em vários problemas práticos. Algumas aplicações foram ilustradas no decorrer desta tese, incluindo o *problema de balanceamento* em um sistema de processamento paralelo.

Embora parte secundária desse trabalho, um dos objetivos desta tese foi coletar um conjunto mínimo de resultados principais na área de aproximações por difusão para que pudessem guiar com maior rapidez futuros interessados no tema. O autor espera que este objetivo também tenha sido alcançado.

Como trabalhos futuros, destacam-se três temas principais:

- aplicações;
- extensão do modelo;

- tratar problemas que ficaram em aberto.

No que diz respeito as *aplicações*, um tema central de pesquisa é a aplicação do modelo em problemas computacionais, tais como o problema de *agendamento de tarefas em sistemas distribuídos ou redução de congestionamento*. Mas é preciso destacar potenciais aplicações na área da *Biologia*, como foi apontado por (Arazi et al., 2004), e que vem demandando crescente atenção da comunidade científica. Destaca-se também a possibilidade de aplicação em modelos de *redes neurais* (onde as filas de Gelenbe já foram aplicadas com bastante sucesso). O trabalho (Ratcliff e McKoon, 2008) reforça essa possibilidade. Nesse artigo, os autores descrevem um modelo de difusão para o comportamento do tempo de resposta em tarefas de discriminação de duas escolhas, e relacionam o efeito de “ruído” presente na resposta com a atividade de neurônios no cérebro. Obviamente, tudo isso demandará uma *maior e mais cuidadosa* investigação destes temas.

Em relação a possíveis *extensões do modelo*, algumas direções de trabalho são apontadas aqui. Primeiramente, uma possibilidade é a exploração das chamadas redes “fechadas”, onde os mesmos clientes circulam no sistema indefinidamente (veja (Kleinrock, 1975) para o tratamento clássico e (Kushner, 2001) para a análise via tráfego pesado destes sistemas). No caso de redes com sinais, onde clientes podem ser removidos do sistema, seria permitido a entrada de clientes externos, mas somente para “compensar” estas perdas. Este tipo de problema geraria dificuldades técnicas interessantes.

Outra possibilidade é tentar relaxar a exigência de como a dependência de estado é introduzida no modelo. Da forma como foi trabalhada aqui, esta dependência não interfere na matriz de difusão no limite ou nas direções de reflexão. Se a dependência no estado for mais “forte”, há possibilidade desta interferência existir, e novos problemas surgirão, principalmente, em relação a dependência das direções de reflexão no estado.

Finalmente, a última possível extensão do modelo que será apontada aqui diz respeito ao problema de tratar redes sinalizantes onde a *topologia da rede pode mu-*

*dar bruscamente.* Tomando como exemplo o modelo da Seção 7.2, isto implicaria que a matriz  $A$ , que carrega a informação de como os nós da rede estão interligados, seria dependente de um processo  $\theta(t)$  que saltaria entre possíveis conformações da rede, i.e., a matriz  $A$  seria dada por  $A_{\theta(t)}$ . Isso provavelmente levaria a complicações interessantes no estabelecimento de convergência fraca e principalmente na caracterização do limite do processo de reflexão.

Em relação ao tema de tratar *problemas que ficaram em aberto*, destaca-se a questão referente a extensão do resultado apresentado por (Piera et al., 2006), e descrito com maiores detalhes na Seção A.3 do Apêndice A.

## Referências Bibliográficas

- E. Altman e H.J. Kushner. Admission control for combined guaranteed performance and best effort communications systems under heavy traffic. **SIAM Journal on Control and Optimization**, 37(6):1780–1807, 1999.
- E. Altman e H.J. Kushner. Control of polling in presence of vacations in heavy traffic with applications to satellite and mobile radio systems. **SIAM J. Control Optim.**, 41(1):217–252, 2002.
- A. Arazi, E. Ben-Jacob, e U. Yechiali. Bridging genetic networks and queueing theory. **Physica A: Statistical Mechanics and its Applications**, 332:585–616, 2004.
- A. Arazi, E. Ben-Jacob, e U. Yechiali. Controlling an oscillating jackson-type network having state-dependent service rates. **Mathematical Methods of Operations Research**, 62(3):453–466, 2005.
- L. Arnold. **Stochastic Differential Equations: Theory and Applications**. John Wiley & Sons, New York, 1974.
- J.R Artalejo. G-networks: A versatile approach for work removal in queueing networks. **European Journal of Operational Research**, 126:233–249, 2000.
- R.B. Ash e C.A. Doléans-Dade. **Probability and Measure Theory (2nd Edition)**. Harcourt Academic Press, London, 2000.
- S. Asmussen. Matrix-analytic models and their analysis. **Scandinavian Journal of Statistics**, 27:193–226, 2000.

- L.A. Barroso, J. Dean, e U. Holzle. Web search for a planet: The google cluster architecture. **IEEE Micro**, 23(2):22–28, 2003.
- S.L. Bell e R.J. Williams. Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: Asymptotic optimality of a threshold policy. **The annals of Applied Probability**, 11(3):608–649, 2001.
- P. Billingsley. **Convergence of Probability Measures**. John Wiley & Sons, New York, 1968.
- P. Billingsley. **Convergence of Probability Measures (2nd Edition)**. John Wiley & Sons, New York, 1999.
- A. Borovkov. Some limit theorems in the theory of mass service i. **Theor. Probability Appl.**, 9:550–565, 1964.
- A. Borovkov. Some limit theorems in the theory of mass service ii. **Theor. Probability Appl.**, 10:375–400, 1965.
- O.J. Boxma, G. Koole, e Z. Liu. Queueing-theoretic solution methods for models of parallel and distributed systems. In: **Performance Evaluation of Parallel and Distributed Systems-Solution Methods**, eds. O.J. Boxma and G.M. Koole, páginas 1–24. CWI Tract, Amsterdam, 1994.
- L. Breiman. **Probability**. SIAM, Philadelphia, 1993.
- P. Brémaud. **Point Processes and Queues, Martingale Dynamics**. Springer-Verlag, New York, 1981.
- R. Buche e H.J. Kushner. Control of mobile communications with time-varying channels in heavy traffic. **IEEE Transactions on Automatic Control**, 47(6):992–1003, 2002.
- P.J. Burke. The output of a queueing system. **Operations Research**, 4:699–704, 1966.

- R. Chakka e T.V. Do. The  $MM \sum_{k=1}^K CPP_k/GE/c/L$  G-queue with heterogeneous servers: Steady state solution and an application to performance evaluation. **Performance Evaluation**, 64(3):191–209, 2007.
- R. Chakka e P.G. Harrison. A markov modulated multi-server queue with negative customers—the MM CPP/GE/c/L G-queue. **Acta Informatica**, 37(11-12): 881–919, 2001.
- C.S. Chang. On the input-output map of a  $G/G/1$  queue. **J. Appl. Prob.**, 31: 1128–1133, 1994.
- X. Chao. A queueing network model with catastrophes and product form solution. **Operations Research Letters**, 18:75–79, 1995.
- C. Chaouiya. Petri net modelling of biological networks. **Briefings in Bioinformatics**, 8(4):210–219, 2007.
- H. Chen e X. Shen. Computing the stationary distribution of an SRBM in an orthant with applications to queueing networks. **Queueing Systems Theory and Applications**, 45(1):27–45, 2003.
- H. Chen e D.D. Yao. **Fundamentals of Queueing Networks**. Springer–Verlag, New York, 2001.
- L. Chen, G. Qi-Wei, M. Nakata, H. Matsuno, e S. Miyano. Modelling and simulation of signal transductions in an apoptosis pathway by using timed petri nets. **Journal of Biosciences**, 32(1):113–127, 2007.
- K.L. Chung. **A Course in Probability Theory, third edition**. Academic Press, 2001.
- J. Dai. **Steady-state analysis of reflected Brownian motions: characterization numerical methods and queueing applications**. Tese de Doutorado, Department of Mathematics, Stanford University, 1990.

- J. G. Dai e R. J. Williams. Existence and uniqueness of semimartingale reflecting brownian motions in convex polyhedrons. **Theory of Probability and its Applications**, 40(1):1–40, 1995.
- R.L. Disney. Queueing networks. **Amer. Math. Soc. Sump. Appl. Math.**, 25: 53–83, 1981.
- P. Dupuis e H. Ishii. On lipschitz continuity of the solution mapping to the skorohod problem, with applications. **Stochastics and Stochastics Reports**, 35:31–62, 1991.
- P. Dupuis e H. Ishii. SDEs with oblique reflections on nonsmooth domains. **The Annals of Probability**, 21(1):554–580, 1993.
- D.L. Eager, E.D. Lazowska, e J. Zahorja. A comparison of receiver-initiated and sender-initiated adaptive load sharing. **SIGMETRICS Perform. Eval. Rev.**, 13(2):1–3, 1985.
- P. Erdős e M. Kac. On certain limit theorems in the theory of probability. **Bull. Amer. Math. Soc.**, 52:292–302, 1946.
- A. K. Erlang. The theory of probabilities and telephone conversations. **Nyt Tidskrift for Matematik**, 20:33, 1909.
- A. K. Erlang. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. **The Post Office Electrical Engineer's Journal**, 10:189–197, 1917.
- S.N. Ethier e T.G. Kurtz. **Markov Processes**. John Wiley & Sons, New York, 1986.
- F.G. Foster. On the stochastic matrices associated with certain queuing processes. **The Annals of Mathematical Statistics**, 24(3):355–360, 1953.

- J. Fourneau, E. Gelenbe, e R. Suros. G-networks with multiple classes of negative and positive customers. **Theoretical Computer Science**, 155(1):141–156, 1996.
- J. Fourneau e D. Verchère. G-networks with triggered batch state-dependent movement. In: **MASCOTS '95: Proceedings of the 3rd International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems**, páginas 33–37, Washington, DC, USA, 1995. IEEE Computer Society.
- E. Gelenbe. Random neural networks with negative and positive signals and product form solution. **Neural Computation**, 1(4):502–511, 1989.
- E. Gelenbe. Product-form queueing networks with negative and positive customers. **Journal of Applied Probability**, 28:656–663, 1991.
- E. Gelenbe. G-networks with triggered customer movement. **Journal of Applied Probability**, 30:742–748, 1993.
- E. Gelenbe. G-networks: a unifying model for neural and queueing networks. **Annals of Operations Research**, 48(5), 1994.
- E. Gelenbe. Steady-state solution of probabilistic gene regulatory networks. **Physical Review E**, 76:031903–1–031903–8, 2007.
- E. Gelenbe e J. Fourneau. G-networks with resets. **Performance Evaluation**, 49(1-4):179–191, 2002.
- E. Gelenbe, P. Glynn, e K. Sigman. Queues with negative arrivals. **Journal of Applied Probability**, 28(1):245–250, 1991.
- E. Gelenbe e R. Schassberger. Stability of product form G-networks. **Probability in the Engineering and Informational Sciences**, 6:271–276, 1992.
- E. Gelenbe e A. Stafylopatis. Global behavior of homogeneous random neural systems. **Applied Mathematical Modelling**, 15:534–541, 1991.

- A. Gómez-Corral. On a tandem G-network with blocking. **Advances in Applied Probability**, 34:626–661, 2002.
- A. Gómez-Corral e M.E. Martos. Performance of two-stage tandem queues with blocking: the impact of several flows of signals. **Performance Evaluation**, 63(9):910–938, 2006.
- C.S. Gonçalves, J. Almeida, V. Almeida, R. Baeza-Yates, B. Ribeiro-Neto, A. Ziviani, e N. Ziviani. A capacity planning model for web search engines. Manuscript submitted for publication, 2007a.
- C.S. Gonçalves, R. Baeza-Yates, B. Ribeiro-Neto, A. Ziviani, e N. Ziviani. Analyzing imbalance among homogeneous index servers in a web search system. **Information Processing and Management**, 43:592–608, 2007b.
- D. Gross e C.M. Harris. **Fundamentals of queueing theory (3rd ed.)**. John Wiley & Sons, New York, 1998.
- V. Guffens. **Compartmental Fluid-Flow Modeling in Packet Switching Networks with Hop-by-Hop Control**. Tese de Doutorado, University of Louvain, Belgium, 2005.
- V. Guffens, E. Gelenbe, e G. Bastin. Qualitative dynamical analysis of queueing networks with inhibition. In: **interperf '06: Proceedings from the 2006 workshop on Interdisciplinary systems approach in performance evaluation and design of computer & communications systems**, 2006.
- S. Hardy e P.N. Robillard. Modeling and simulation of molecular biology systems using petri nets: Modeling goals of various approaches. **Journal of Bioinformatics and Computational Biology**, 2(4):619–637, 2004.
- S. Hardy e P.N. Robillard. Petri net-based method for the analysis of the dynamics of signal propagation in signaling pathways. **Bioinformatics**, 24(2):209–217, 2008.

- J. M. Harrison e M.I. Reiman. Reflected brownian motion in an orthant. **The Annals of Probability**, 9(2):302–308, 1981.
- J.M. Harrison. Heavy traffic analysis of a system with parallel servers: Asymptotic optimality of discrete-review policies. **The annals of Applied Probability**, 8(3):822–848, 1998.
- J.M. Harrison e M.J. López. Heavy traffic resource pooling in parallel-server systems. **Queueing Systems: Theory and Applications**, 33(4):339–368, 1999.
- P.G. Harrison. Compositional reversed markov processes, with applications to G-networks. **Performance Evaluation**, 57(3):379–408, 2004.
- P.G. Harrison e E. Pitel. The  $M/G/1$  queue with negative customers. **Advances in Applied Probability**, 28:540–566, 1996.
- J.R Jackson. Network of waiting lines. **Operations Research**, 5:674–683, 1957.
- J.R. Jackson. Jobshop-like queueing systems. **Manage. Sci.**, 10:131–142, 1963.
- J. Jacod e A.N. Shiryaev. **Limit Theorems for Stochastic Processes**. Springer, New York, 1987.
- D. Jarvis e H.J. Kushner. Codes for optimal stochastic control: documentation and users guide. Technical report, Brown University, Lefschetz Center for Dynamical Systems Report 96-3, 1996.
- O. Kallenberg. **Foundations of Modern Probability**. Springer-Verlag, New York, 2002.
- I. Karatzas e S.E. Shreve. **Brownian Motion and Stochastic Calculus**. Springer-Verlag, New York, 1988.
- D.G. Kendall. Some problems in the theory of queues. **Journal of the Royal Statistical Society**, 13(2):151–185, 1951.

- D.G. Kendall. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. **Ann. Math. Statist.**, 24(3):338–354, 1953.
- D.G. Kendall. Some recent work and further problems in the theory of queues. **Theor. Probability Appl.**, 9:1–13, 1964.
- J.F.C. Kingman. The single server queue in heavy traffic. **Proc. Camb. Phil. Soc.**, 57:902–904, 1961.
- J.F.C. Kingman. Some inequalities for the queue  $GI/G/1$ . **Biometrika**, 49(3/4): 315–324, 1962.
- J.F.C. Kingman. A martingale inequality in the theory of queues. **Proc. Camb. Phil. Soc.**, 59:359–362, 1963.
- J.F.C. Kingman. On the algebra of queues. **Jornal of Applied Probability**, 3 (2):285–326, 1966.
- J.F.C. Kingman. Inequalities in the theory of queues. **Journal of the Royal Statistical Society. Series B (Methodological)**, 32(1):102–110, 1970.
- L. Kleinrock. **Queueing Systems, Volume I: Theory**. John Wiley & Sons, New York, 1975.
- P.E. Kloeden, E. Platen, e H. Schurz. **Numerical Solution of SDE Through Computer Experiments**. Springer–Verlag, New York, 1994.
- T. G. Kurtz. Lectures on stochastic analysis. Lecture Notes, University of Wisconsin - Madison, 2001. URL <http://www.math.wisc.edu/~kurtz/m735.htm>.
- H.J. Kushner. **Heavy traffic Analysis of Controlled Queueing and Communication Networks**. Springer-Verlag, New York, 2001.
- H.J. Kushner e P.G. Dupuis. **Numerical Methods for Stochastic Control Problems in Continuous Time**. Springer–Verlag, New York, 1992.

- H.J. Kushner e L.F. Martins. Heavy traffic analysis of a data transmission system with many independent sources. **SIAM Journal on Applied Mathematics**, 53(4):1095–1122, 1993.
- H.J. Kushner, J. Yang, e D. Jarvis. Controlled and optimally controlled multiplexing systems: A numerical exploration. **Queueing Systems**, 20:255–291, 1995.
- S.C. Leite e M.D. Fragoso. G-queues under heavy traffic. In: **Anais do Congresso Nacional de Matemática Aplicada e Computacional**, Florianópolis, SC, 2007.
- S.C. Leite e M.D. Fragoso. Diffusion approximation of state dependent G-networks under heavy traffic. **Journal of Applied Probability**, 45:347–362, 2008a.
- S.C. Leite e M.D. Fragoso. Heavy traffic analysis of state-dependent fork-join queues with triggers. In: **Annals of the International Symposium on Performance Evaluation of Computer and Telecommunication Systems**, Edimburgo, Escócia, 2008b.
- S.C. Leite e M.D. Fragoso. Heavy traffic analysis of state-dependent fork-join queues with triggers and an application to web search systems. Submetido para publicação, 2008c.
- S.C. Leite e M.D. Fragoso. On the analysis of G-queues under heavy traffic. In: **Annals of the 47th IEEE Conference on Decision and Control**, Cancun, Mexico, 2008d.
- S.C. Leite e M.D. Fragoso. An unified approach to signaling stochastic networks and their heavy traffic approximations. Aceito para apresentação no congresso **European Control Conference**, 2009.
- Q. Li e Y.Q. Zhao. A  $MAP/G/1$  queue with negative customers. **Queueing Systems**, 47:5–43, 2004.

- D.V. Lindley. The theory of queues with a single server. **Proc. Camb. Phil. Soc.**, 48:277–289, 1952.
- J.D.C. Little. A proof of the queueing formula  $l = \lambda w$ . **Operations Research**, 9:383–387, 1961.
- D.M. Lucantoni. New results on the single server queue with a batch markovian arrival process. **Stochastic Models**, 7(1):1–46, 1991.
- D.M. Lucantoni. The *BMAP/G/1* queue: A tutorial. In: L. Donatiello e R. Nelson Editors (editores), **Models and Techniques for Performance Ecaluation of Computer and Communications Systems**, páginas 330–358. Springer Verlag, 1993.
- D.M. Lucantoni, K.S. Meier-Hellstern, e M.F. Neuts. A single-server queue with server vacations and a class of non-renewal arrival process. **Adv. Appl. Prob.**, 22:676–705, 1990.
- A. Mandelbaum e G. Pats. State-dependent stochastic network. part i: Approximations and applications with continuous diffusion limits. **The annals of Applied Probability**, 8:569–646, 1998.
- M. Ajmone Marsan. Stochastic petri nets: an elementary introduction. In: **Advances in Petri nets 1989**, páginas 1–29, New York, 1990. Springer-Verlag.
- M.A. Marsan, G. Conte, e G. Balbo. A class of generalized stochastic petri nets for the performance evaluation of multiprocessor systems. **ACM Transactions on Computer Systems**, 2(2):93–122, 1984.
- M. Molloy. **On the integration of delay and throughput measures in distributed processing models**. Tese de Doutorado, University of California, Los Angeles, 1981.
- M. Molloy. Performance analysis using stochastic petri nets. **IEEE Transactions on Computers**, c-31(9):913–917, 1982.

- T. Murata. Petri nets: Properties, analysis and applications. **Proceedings of the IEEE**, 77(4):541–580, 1989.
- S. Natkin. **Reseaux de Petri Stochastiques**. Tese de Doutorado, CNAM-PARIS, Paris, 1980.
- M.F. Neuts. Probability distributions of phase type. In Liber Amicorum Prof. Emeritus H. Florin, Department of Mathematics, University of Louvain, 1975.
- M.F. Neuts. A versatile markovian point process. **Journal of Applied Probability**, 16(4):764–779, 1979.
- M.F. Neuts. **Matrix-Geometric Solutions in Stochastic Models**. John Hopkins University Press, Baltimore, 1981.
- B. Oksendal. **Stochastic Differential Equations**. Springer-Verlag, New York, 2005.
- A. Pacheco e N.U. Prabhu. Markov-additive processes of arrivals. In: **Advances in Queueing: Theory, Methods and Open Problems**, páginas 167–194. CRC Press, Boca Raton, FL, 1995.
- C. A. Petri. Nets, time and space. **Theoretical Computer Science**, 153(1-2): 3–48, 1996.
- C.A. Petri. **Kommunikation mit Automaten**. Tese de Doutorado, Institut fur Instrumentelle Mathematik, Bonn, 1962.
- F.J. Piera, R.R. Mazumdar, e F.M. Guillemin. On the local times and boundary properties of reflected diffusions with jumps in the positive orthant. **Markov Processes and Related Fields**, 12:561–582, 2006.
- J.W. Pinney, D.R. Westhead, e G.A. McConkey. Petri net representations in systems biology. **Biochemical Society Transactions**, páginas 1513–1515, 2003.

- N.U. Prabhu. **Stochastic storage processes : queues, insurance risk, and dams**. Springer-Verlag, New York, 1980.
- Yu. Prohorov. Transient phenomena in process of mass service. **Litovsk. Mat. Sb.**, 3:199–205, 1963. (In Russian).
- P. Protter. **Stochastic Integration and Differential Equations: A New Approach**. Springer–Verlag, New York, 1995.
- V. Ramaswami. The  $N/G/1$  queue and its detailed analysis. **Adv. Appl. Prob.**, 12:222–261, 1980.
- C. Ramchandani. **Analysis of Asynchronous Concurrent Systems by Timed Petri Nets**. Tese de Doutorado, Massachusetts Institute of Technology, Massachusetts, 1973.
- R. Ratcliff e G. McKoon. The diffusion decision model: Theory and data for two-choice decision tasks. **Neural Computation**, 20:873–922, 2008.
- M.I. Reiman. Open queueing networks in heavy traffic. **Mathematics of Operations Research**, 9(3):441–458, 1984.
- M.I. Reiman e R.J. Williams. A boundary property of semimartingale reflecting brownian motions. **Probability Theory and Related Fields**, 77(1):87–97, 1988.
- W. Reisig. **Petri Nets: Introduction**. Springer–Verlag, New York, 1985.
- X. Shen, H. Chen, J.G. Dai, e W. Dai. The finite element method for computing the stationary distribution of and SRBM in a hypercube with applications to finite buffer queueing networks. **Queueing Systems**, 42:33–62, 2002.
- Y.W. Shin. Multi-server retrial queue with negative customers and disasters. **Queueing Systems**, 55:223–237, 2007.

- Y.W. Shin e B.D. Choi. A queue with positive and negative arrivals governed by a markov chain. **Probability in the Engineering and Informational Sciences**, 17(4):487–501, 2003.
- K. Sigman. Lecture notes on queueing theory. Columbia University, New York, 2004. URL <http://www.columbia.edu/~ks20/6704-04/6704-04.html>.
- W.L Smith. On the distribution of queueing times. **Proc. Camb. Phil. Soc.**, 49: 449–461, 1952.
- D.W. Stroock. **An Introduction to Markov Process**. Springer-Verlag, New York, 2005.
- F.J.W. Symons. Introduction to numerical petri nets, a general graphical model of concurrent processing systems. **Australian Telecommunications Research**, 14(1):28–33, 1980.
- D. Tipper e M.K. Sundareshan. Numerical methods for modeling computer networks under nonstationary conditions. **IEEE J. Select. Areas Commun**, 8(6):1682–1695, 1990.
- A. F. Veinott. Discrete dynamic programming with sensitive discount optimality criteria. **Annals of Mathematical Statistics**, 40:1635–1660, 1969.
- W. Whitt. Complements to heavy traffic limit theorems for the  $GI/G/1$  queue. **J. Appl. Prob.**, 9:185–191, 1972.
- W. Whitt. Heavy traffic limit theorems for queues: A survey. In: M. Beckmann e H.P. Kunzi (editores), **Lecture Notes in Economics and Mathematical Systems**. Springer-Verlag, New York, 1974.
- W. Whitt. The reflection map with discontinuities. **Mathematics of Operations Research**, 26(3):447–484, 2001.
- W. Whitt. **Stochastic-Process Limits**. Springer-Verlag, New York, 2002.

D.J. Wilkinson. **Stochastic Modelling for Systems Biology**. Chapman & Hall/CRC, Boca Raton, Florida, 2006.

R.J. Williams. On dynamic scheduling of a parallel server system with complete resource pooling. Proceedings of the Fields Institute Workshop on Analysis and Simulation of Communication Networks. American Mathematical Society, Providence, RI. 48, 1999.

# Apêndice A

## Processo de Difusão Refletido

O estudo de processos de difusão presos em um poliedro convexo com direções de reflexão oblíquas em relação às normais de cada face do poliedro é de fundamental importância para compreender aproximações de sistemas de fila sob tráfego pesado. Este apêndice reúne uma série de resultados importantes que contribuíram para o entendimento das propriedades de equações diferenciais estocásticas refletidas e, em especial, do termo de reflexão, o processo que mantém a difusão presa no domínio de interesse.

Na primeira parte deste apêndice serão apresentados resultados referentes à existência de solução para as equações diferenciais estocásticas refletidas. A razão principal de incluir este tema neste trabalho é que ele serve como um meio para introduzir alguns resultados importantes: o mapeamento de reflexão, o problema de Skorohod, e as condições sobre as direções de reflexão suficientes para garantir algumas propriedades importantes para o termo de reflexão.

Em seguida na Seção A.2, será apresentado um teorema que estabelece condições para a convergência fraca de processos de reflexão aproximados. Este resultado é de fundamental importância para o estudo de aproximações por difusão em sistemas de filas, e será usado em praticamente todos os teoremas referentes a este tema nesta tese.

Na Seção A.3, será apresentado uma propriedade dos processos de reflexão que será usado no Capítulo 7 desta tese.

## A.1 Existência e Unicidade de Soluções para Equações Diferenciais Estocásticas Refletidas

Nesta seção, a existência e unicidade de solução para equações diferenciais estocásticas refletidas (EDER) será estudada. As soluções destas equações tomam valor em um poliedro convexo em  $\mathbb{R}^d$ . Este poliedro é definido como a interseção finita de meio-espacos fechados dados por

$$G_i \triangleq \{\xi \in \mathbb{R}^d : \langle \xi, n_i \rangle \geq B_i\},$$

para  $i \in \{1, \dots, N\}$ , onde  $N \in \mathbb{N}$ ,  $n_i \in \mathbb{R}^d$  e  $B_i \in \mathbb{R}$  são dados. Desta forma, define-se  $G \triangleq \bigcap_{i \in N} G_i$ .

Junto com esse domínio  $G$ , é necessário definir um conjunto de “direções de reflexão”, que são direções que serão usadas para “empurrar” o processo de volta ao espaço de estado caso tente escapar. Define-se as direções de reflexão como vetores  $d_i \in \mathbb{R}^d$  definidos em cada face  $\partial G_i \triangleq \{\xi \in \mathbb{R}^d : \langle \xi, n_i \rangle = B_i\}$ ,  $i \in \{1, \dots, N\}$  (Casos mais gerais podem ser encontrados em (Dupuis e Ishii, 1993)). Defina para cada  $\xi \in \partial G$  o conjunto  $D(\xi) \triangleq \{d_i : \xi \in \partial G_i, i \in \{1, \dots, N\}\}$  que contém as direções de reflexão definidas no ponto  $\xi$ . De forma a condensar e facilitar referências futuras, a seguinte condição resume a discussão acima:

**Condição A.1.1.** *Seja  $G_i \triangleq \{\xi \in \mathbb{R}^d : \langle \xi, n_i \rangle \geq B_i\}$  para um dado vetor normal unitário  $n_i \in \mathbb{R}^d$  e uma constante  $B_i \in \mathbb{R}$ , para  $i \in \{1, \dots, N\}$ . Seja  $\partial G_i$  a fronteira  $\{\xi \in \mathbb{R}^d : \langle \xi, n_i \rangle = B_i\}$ . Em cada fronteira  $\partial G_i$  está definido um vetor unitário  $d_i \in \mathbb{R}^d$  tal que  $\langle n_i, d_i \rangle > 0$ , que é chamado de direção de reflexão. O espaço de estados  $G$  é dado por:  $G \triangleq \bigcap_{i=1}^n G_i$ .*

Diz-se que o processo estocástico  $x$  é solução de uma equação diferencial estocástica refletida se dado  $x_0 \in G$ ,  $x$  satisfaz:

$$x(t) = x_0 + \int_0^t b(x(s))ds + \int_0^t \sigma(x(s))dw(s) + z(t), \quad (\text{A.1})$$

onde  $x(t) \in G$  para todo  $t > 0$ ,  $b : G \rightarrow \mathbb{R}^d$  é uma função integrável,  $\sigma : G \rightarrow \mathbb{R}^{d \times d}$  é uma função integrável,  $w$  é um processo de Wiener  $\mathcal{F}_t$ -mensurável onde  $\mathcal{F}_t$  é a mínima  $\sigma$ -álgebra que mensura  $\{w(s), z(s); s \leq t\}$  e o processo  $z$  satisfaz:

- (i)  $|z|(t) < \infty$  para cada  $t \geq 0$ ,
- (ii)  $|z|(t) = \int_{[0,t]} \mathbb{I}\{x(s) \in \partial G\} d|z|(s)$ ,
- (iii) e existe uma função mensurável  $\gamma : [0, \infty) \rightarrow \mathbb{R}^d$  tal que  $\gamma(s) \in D(x(s))$  (d|z|-q.c.) e  $z(t) = \int_{[0,t]} \gamma(s) d|z|(s)$ , para todo  $t \geq 0$ .

O processo  $z$  é usualmente chamado de “processo de reflexão”.

### A.1.1 Existência e Unicidade Forte

O problema principal para a demonstração de existência e unicidade forte é estabelecer condições sobre o termo de reflexão, mais especificamente as direções de reflexão, de forma que propriedades do tipo dada pela Equação (A.3) sejam garantidas. Esse problema é tratado de duas formas diferentes na literatura. Uma delas é feita através do *mapeamento de reflexão* que será definido com mais detalhes abaixo. A outra forma, que surgiu posteriormente, é tratada através do chamado *problema de Skorohod*.

Além do interesse histórico, outras razões para apresentar estas duas formas diferentes são que ambas continuam sendo usadas na literatura e possuem vantagens e desvantagens.

#### A.1.1.1 Método do Mapeamento de Reflexão

O mapeamento de reflexão foi introduzido por (Harrison e Reiman, 1981) em problemas de fila. Geralmente, esse mapeamento é definido quando o espaço de estados  $G$  é o quadrante positivo de  $\mathbb{R}^d$ , e esse será o caso considerado aqui. Seja  $D_0(\mathbb{R}^d; 0, \infty) \triangleq \{\psi \in D(\mathbb{R}^d; 0, \infty) : \psi(0) \geq 0\}$ .

**Definição A.1.2.** (*Mapeamento de Reflexão*) Seja  $G \triangleq \{\xi : \xi_i \geq 0, i \in \{1, \dots, N\}\}$  e seja  $Q = \{q_{ij}; i, j \in \{1, \dots, N\}\}$  uma matriz com entradas reais. Para  $\psi \in$

$D_0(\mathbb{R}^d; 0, \infty)$ , o mapeamento de reflexão é definido como a função  $\Psi : D_0(\mathbb{R}^d; 0, \infty) \rightarrow D(\mathbb{R}^{2d}; 0, \infty)$  tal que se  $(\phi, \eta) = \Psi(\psi)$  então:

(i)  $\phi(t) = \psi(t) + (I - Q')\eta(t)$  para todo  $t > 0$ ,

(ii)  $\phi(0) = \psi(0)$ ,

(iii)  $\phi(t) \in G$ ,

(iv)  $\eta(0) = 0$ ,

(v)  $\eta$  é não-decrescente,

(vi)  $\eta_i$  satisfaz  $\int_0^t \phi_i(s) d\eta_i(s) = 0$ , para todo  $t > 0$ . Essa propriedade é geralmente chamada de “propriedade de complementação”.

Observe que neste caso a função  $\zeta(t) = (I - Q')\eta(t)$  faz o papel de manter  $\phi$  presa no quadrante positivo de  $\mathbb{R}^d$ . Cada direção de reflexão  $d_i$ , definida nas faces do espaço de estado  $\partial G_i \triangleq \{\xi \in \mathbb{R}^d : \xi_i \geq 0\}$ , é dada pela  $i$ -ésima coluna da matriz  $(I - Q')$ .

Será necessário introduzir seguinte condição sobre a matriz  $Q$ :

**Condição A.1.3.** (Condição de Harrison e Reiman) *A matriz  $Q$  tem raio espectral inferior à 1, ou em símbolos:  $\sigma(Q) < 1$ .*

O caso em uma dimensão será tratado primeiro. Este resultado será usado para obter o mesmo resultado para o caso multidimensional.

**Lema A.1.4.** *Para  $\psi \in D_0(\mathbb{R}; 0, \infty)$  existe um único par  $(\psi, \zeta)$  satisfazendo para todo  $t \geq 0$*

$$\phi(t) = \psi(t) + \zeta(t), \tag{A.2}$$

onde (i)  $\phi(t) \geq 0$  para todo  $t \geq 0$ , (ii)  $\zeta(0) = 0$ , (iii)  $\zeta$  é não-decrescente, e (iv)  $\zeta$  somente aumenta nos instantes  $t$  tais que  $\phi(t) = 0$ . Além disso,  $\zeta$  é dado por  $\zeta(t) = 0 \vee \sup_{s \leq t} (-\psi(s))$ .

*Demonstração.* Observe que  $\zeta$  dado por  $\zeta(t) = 0 \vee \sup_{s \leq t} (-\psi(s))$  e  $\phi$  dado como na Equação (A.2) é solução do problema acima. De fato, verificando os itens acima:

(i) Suponha que para algum  $t \geq 0$ , tem-se  $\phi(t) < 0$ , então

$$0 > \psi(t) + 0 \vee \sup_{s \leq t} (-\psi(s)) \Rightarrow -\psi(t) > 0 \vee \sup_{s \leq t} (-\psi(s)),$$

que leva à uma contradição.

(ii) Como  $\phi(0) \geq 0$ , então  $\zeta(0) = 0$ .

(iii) Claramente  $\zeta$  é não-decrescente pela definição.

(iv) Como  $\zeta$  é um càdlàg, basta considerar dois casos. Primeiramente, suponha que exista um intervalo  $[s, t]$  onde  $\zeta$  é estritamente crescente. Pela definição de  $\zeta$ ,  $\zeta(u) = -\psi(u)$  para todo  $u \in [s, t]$ , portanto  $\phi(u) = \psi(u) - \psi(u) = 0$  para todo  $u \in [s, t]$ . Agora suponha que existe um  $t > 0$  tal que  $\zeta(t) - \zeta(t-) > 0$ . Então  $\zeta(t) = -\psi(t)$  e portanto  $\phi(t) = 0$ .

Será mostrado agora que  $\zeta(t) = 0 \vee \sup_{s \leq t} (-\psi(s))$  é de fato a única solução (essa parte da prova foi retirada de (Kurtz, 2001)). Para  $t < \tau_0 \triangleq \inf\{s : \psi(s) \leq 0\}$  a condição (iv) implica que qualquer solução  $\zeta$  terá que satisfazer  $\zeta(t) = 0$ . Para  $t \geq \tau_0$ ,  $\zeta(t) \geq -\psi(t)$  pela a condição (i) e como  $\zeta$  é não decrescente,  $\zeta(t) \geq \sup_{s \leq t} (-\psi(s))$ .

Se  $t$  é um ponto onde  $\zeta$  aumenta então  $\phi(t) = 0$  e

$$\zeta(t) = -\psi(t) \leq \sup_{s \leq t} (-\psi(s)).$$

Como o lado esquerdo da inequação acima é não-decrescente,  $\zeta(t) \leq \sup_{s \leq t} (-\psi(s))$  para todo  $t > \tau_0$ . E portanto segue o resultado.  $\square$

**Teorema A.1.5.** *Seja  $\psi \in D_0(\mathbb{R}^d; 0, \infty)$  e suponha que a matriz  $Q$  satisfaz a Condição A.1.3. Então existe um único  $\phi$  e um único  $\eta$  pertencentes à  $D(\mathbb{R}^d; 0, \infty)$  tais que satisfazem as condições (i) – (vi) da Definição A.1.2 e portanto  $(\phi, \eta) =$*

$\Psi(\psi)$ . Além disso, seja  $(\phi^1, \eta^1) = \Psi(\psi^1)$  e  $(\phi^2, \eta^2) = \Psi(\psi^2)$ , para  $(\psi^1, \psi^2) \in D(\mathbb{R}^d; 0\infty)^2$  então para cada  $t > 0$

$$|\phi^1(t) - \phi^2(t)| + |\eta^1(t) - \eta^2(t)| \leq C \sup_{s \leq t} |\psi^1(s) - \psi^2(s)|. \quad (\text{A.3})$$

*Demonstração.* Essa demonstração segue as idéias em (Harrison e Reiman, 1981). Primeiramente, seja norma  $\|A\| = \max_j \sum_i A_{ij}$ . Como a matriz  $Q$  tem raio espectral menor que um, existe uma matriz diagonal  $\Lambda$  com elementos da diagonal positivos tal que a matriz não-negativa  $Q^* \triangleq \Lambda^{-1}Q\Lambda$  satisfaz  $\|Q^*\| < 1$  (veja (Veinott, 1969)). Portanto será assumido, sem perda de generalidade, que  $\|Q\| = \alpha < 1$ .

Para todo  $T > 0$  e  $\nu \in D(\mathbb{R}^d; 0, \infty)$  defina  $\|v\|_T = \sup_i \sup_{s \leq T} |v_i(s)|$  e

$$\pi_j(\psi, \nu)(t) = 0 \vee \sup_{s \leq t} ((Q\nu(s))_j - \psi_j(s)),$$

onde  $(Q\nu(s))_j$  denota a  $j$ -ésima componente do vetor  $Q\nu(s)$ . Então para  $\nu_1, \nu_2 \in D(\mathbb{R}^d; 0, \infty)$  quaisquer, tem-se que:

$$\begin{aligned} \|\pi_1(\psi, \nu_1) - \pi(\psi, \nu_2)\|_T &= \sup_i \sup_{s \leq T} |\pi_i(\psi, \nu_1)(t) - \pi_i(\psi, \nu_2)(t)| \\ &\leq \sup_i \sup_{s \leq T} |(Q\nu_1(s))_i - (Q\nu_2(s))_i| \\ &= \sup_i \sup_{s \leq T} |(Q(\nu_1(s) - \nu_2(s)))_i| \\ &\leq \alpha \|v_1 - v_2\|_T. \end{aligned}$$

Seja  $\eta^0 \triangleq 0$  e defina recursivamente para  $n \geq 1$

$$\phi^{n+1} = [\psi(t) - Q'\eta^n(t)] + \eta^{n+1}(t).$$

Pelo Lema A.1.4, cada componente  $\phi_i^{n+1}$  satisfaz:

$$\phi_i^{n+1} = [\psi_i(t) - (Q'\eta^n(t))_i] + \pi_i(\psi, \eta^n).$$

Portanto,  $\eta^{n+1}(t) = \pi(\psi, \eta^n)$  e desta forma:

$$\|\eta^{n+1} - \eta^n\|_T \leq \alpha \|\eta^n - \eta^{n-1}\|_T \leq \alpha^n.$$

Como  $\alpha < 1$ , existe um  $\eta \in D(\mathbb{R}^d; 0, \infty)$  tal que  $\|\eta^n - \eta\|_T \rightarrow 0$  quando  $n \rightarrow \infty$ . Além disso,  $\eta_i(t) = \pi_i(\psi, \eta)$ . Usando argumentos similares aos usados no caso unidimensional, verifica-se que  $\eta_i$  como dado acima e  $\phi_i$  correspondente satisfazem as condições da Definição A.1.2.

Para chegar à Inequação (A.3), suponha que  $(\phi^1, \eta^1)$  e  $(\phi^2, \eta^2)$  são soluções do problema dado pela Definição A.1.2 para  $(\psi^1, \psi^2) \in D(\mathbb{R}^d; 0, \infty)^2$ , respectivamente. Portanto,

$$\begin{aligned} \|\eta^1 - \eta^2\|_T &= \|\pi(\phi^1, \eta^1) - \pi(\phi^2, \eta^2)\|_T \\ &= \sup_i \sup_{s \leq T} |\pi_i(\phi^1, \eta^1)(s) - \pi_i(\phi^2, \eta^2)(s)| \\ &= \sup_i \sup_{s \leq T} |[(Q\eta^1(s))_i - (Q\eta^2(s))_i] + [\psi^2(s) - \psi^1(s)]| \\ &\leq \alpha \|\eta^1 - \eta^2\|_T + \|\psi^1 - \psi^2\|_T, \end{aligned}$$

e então

$$\|\eta^1 - \eta^2\|_T \leq \frac{1}{1 - \alpha} \|\psi^1 - \psi^2\|_T,$$

de onde segue o resultado. □

Apesar da definição do mapeamento de reflexão não cobrir todos os casos de interesse, a definição desse mapeamento é de extrema importância para teoria de filas. É possível mostrar que esse mapeamento é contínuo sobre o espaço  $D(\mathbb{R}^d; 0, \infty)$  usando a topologia  $M_1$  de Skorohod (para maior detalhes, veja (Whitt, 2001, 2002)).

### A.1.1.2 Problema de Skorohod

Na tentativa de generalizar o resultado acima para o caso onde o espaço de estado  $G$  é um poliedro, foi observado que a função  $\zeta(t) = (I - Q')\eta(t)$  como dada na Definição A.1.2 não será única caso as colunas da matriz  $(I - Q')$  não sejam linearmente independentes. Para evitar esse problema, foi introduzido uma nova forma de descrever o problema.

**Definição A.1.6.** (*Problema de Skorohod*) Para cada  $\psi \in D(\mathbb{R}^d; 0, \infty)$ ,  $\psi(0) \in G$ , diz-se que  $(\phi, \zeta) \in D(\mathbb{R}^d; 0, \infty)^2$  resolve o problema de Skorohod para  $\psi$  se:

(i)  $\phi(t) = \psi(t) + \zeta(t)$  para todo  $t \geq 0$ ,

(ii)  $\phi(t) \in G$  para todo  $t \geq 0$ ,

(iii)  $|\zeta|(t) < \infty$  para cada  $t$ ,

(iv)  $|\zeta|(t) = \int_0^t \mathbb{I}\{\phi(s) \in \partial G\} d|\zeta|(s)$ ,

(v) Existe um  $\gamma : [0, \infty) \rightarrow \mathbb{R}^d$  mensurável tal que para cada  $t \geq 0$   $\gamma(t) \in \text{Cone}(D(\phi(t)))$  e para  $d|\zeta|$ -quase todo  $t$ ,

$$\zeta(t) = \int_0^t \gamma(s) d|\zeta|(s).$$

Além de generalizar o espaço de estado  $G$ , as condições sobre as direções de reflexão também são relaxadas.

**Condição A.1.7.** (“Condição do Conjunto  $B$ ”) Existe um conjunto  $B$  convexo e compacto, onde  $0 \in B^0$ , tal que se  $\nu(z)$  denota o conjunto de normais internas à  $B$  definidas em  $\partial B$ , então existe um  $\delta > 0$  tal que:

$$\left\{ \begin{array}{l} z \in \partial B \\ |\langle z, n_i \rangle| < \delta \end{array} \right\} \Rightarrow \langle \nu, d_i \rangle = 0 \quad \forall \nu \in \nu(z),$$

para  $i \in \{1, \dots, N\}$ .

Em (Dupuis e Ishii, 1991) é mostrado que no caso considerado na Seção A.1.1.1, i.e.,  $G \triangleq \mathbb{R}_+^d$  e  $d_i = (I - Q')_i$ , o fato que  $\sigma(Q) < 1$  implica na existência de tal conjunto  $B$  como definido acima. Além disso Kushner (Kushner, 2001) apresenta um exemplo onde existe um conjunto  $B$  que satisfaz a Condição A.1.7 mas onde a condição do raio espectral não é atendida.

A prova de existência de solução para o problema de Skorohod pode ser encontrada em (Dupuis e Ishii, 1991), não há necessidade de demonstrá-la aqui. O principal objetivo aqui é em estabelecer uma relação como a Inequação (A.3). Para entender melhor a condição do conjunto  $B$  é importante entender como ela é usada para obter a propriedade desejada.

O seguinte lema será usado:

**Lema A.1.8.** *Se existe um conjunto  $B$  satisfazendo a Condição A.1.7, então  $B$  também satisfaz:*

$$z \in \partial B \text{ e } \nu \in \nu(z) \Rightarrow \langle z, n_i \rangle \langle \nu, d_i \rangle \leq 0,$$

para  $i \in \{1, \dots, N\}$ .

*Demonstração.* Veja Lema 2.1 de (Dupuis e Ishii, 1991). □

**Teorema A.1.9.** *Suponha a existência de um conjunto  $B$  que satisfaz a condição acima. Seja  $(\phi^1, \zeta^1)$  e  $(\phi^2, \zeta^2)$  soluções para o problema de Skorohod para  $\psi^1$  e  $\psi^2$ , respectivamente. Então existe um  $K < \infty$  tal que*

$$\begin{aligned} \sup_{s \leq t} |\zeta^1(s) - \zeta^2(s)| &\leq K \sup_{s \leq t} |\psi^1(s) - \psi^2(s)|, \\ \sup_{s \leq t} |\phi^1(s) - \phi^2(s)| &\leq K \sup_{s \leq t} |\psi^1(s) - \psi^2(s)| \end{aligned}$$

e portanto existe um  $C < \infty$

$$|\phi^1(t)| + |\zeta^1(t)| \leq C \sup_{s \leq t} |\psi^1(s)|. \tag{A.4}$$

*Demonstração.* A prova segue como em (Dupuis e Ishii, 1991). Tome um  $T < \infty$  e defina  $\Delta\psi = \psi^1 - \psi^2$ ,  $\Delta\phi = \phi^1 - \phi^2$ ,  $\Delta\zeta = \zeta^1 - \zeta^2$  e

$$C = \sup_{t \leq T} |\Delta\psi(t)|.$$

É suficiente mostrar que  $\Delta\zeta(t) \in aB$  para  $0 \leq t \leq T$  e  $a > C$ . A prova é feita por contradição. Suponha que para algum  $t$ ,  $\Delta\zeta(t) \notin aB$  e defina

$$\tau = \inf\{t \in [0, T] : \Delta\zeta(t) \notin aB^0\}.$$

Pela continuidade à direita de  $\Delta\zeta$ , apenas duas possibilidades são possíveis: ou  $\Delta\zeta(\tau-) \in \partial aB$  ou  $\Delta\zeta(\tau-) \in aB^0$ . Considere o primeiro caso onde  $\Delta\zeta(\tau-) \in \partial aB$ . Defina  $z = \Delta\zeta(\tau-)$  e tome  $\nu \in \nu(\delta z/a)$ . Para qualquer  $u \in [0, \tau)$  tem-se que  $\Delta\zeta(u) \in aB$  implica que

$$\langle z - \Delta\zeta(u), \nu \rangle < 0,$$

pela convexidade de  $B$ . Portanto

$$\int_{(u, \tau)} \langle \gamma^1(s), \nu \rangle d|\zeta^1|(s) < \int_{(u, \tau)} \langle \gamma^2(s), \nu \rangle d|\zeta^2|(s)$$

onde  $\gamma^1(s) \in D(\phi^1(s))$  e  $\gamma^2(s) \in D(\phi^2(s))$ . A inequação acima leva à duas possibilidades:

$$\int_{(u, \tau)} \langle \gamma^1(s), \nu \rangle d|\zeta^1|(s) < 0$$

ou

$$\int_{(u, \tau)} \langle \gamma^2(s), \nu \rangle d|\zeta^2|(s) > 0.$$

Portanto para algum  $s \in (u, \tau)$  tem-se que

$$\langle \gamma^1(s), \nu \rangle < 0 \text{ onde } \phi^1(s) \in \partial G$$

ou

$$\langle \gamma^2(s), \nu \rangle > 0 \text{ onde } \phi^2(s) \in \partial G,$$

já que  $|\zeta^1|$  (resp.,  $|\zeta^2|$ ) somente aumenta quando  $\phi^1$  (resp.,  $\phi^2$ ) está nas fronteiras. Como a escolha de  $u$  é arbitrária, toma-se uma sequência  $s_k \in [0, \tau)$  tal que  $s_k \uparrow \tau$  e  $s_k$  satisfaz uma das duas propriedades acima. Somente o primeiro caso será tratado já que o outro segue de forma análoga. Portanto, para algum  $i \in \{1, \dots, N\}$

$$\langle d_i, \nu \rangle < 0 \text{ e } \phi^1(s_k) \in \partial G_i.$$

Sem perda de generalidade, suponha que  $i$  é fixo para valores grandes de  $k$ . Portanto fazendo  $k \rightarrow \infty$

$$\langle d_i, \nu \rangle < 0 \text{ e } \phi^1(\tau-) \in \partial G_i.$$

Usando o Lema A.1.8 e a propriedade do conjunto  $B$ , tem-se que

$$\langle \delta z/a, n_i \rangle \geq \delta.$$

Desta forma,

$$a \leq \langle z, n_i \rangle = \langle (\phi^1 - \phi^2 - \psi^1 + \psi^2)(\tau-), n_i \rangle \leq \langle (\phi^1 - \phi^2)(\tau-), n_i \rangle + C$$

que implica que  $\langle (\phi^1 - \phi^2)(\tau-), n_i \rangle > 0$ . Mas como  $\phi^1(\tau-) \in \partial G_i$  e  $\phi^2(\tau-) \in G$  tem-se que

$$\langle (\phi^1 - \phi^2)(\tau-), n_i \rangle \leq 0$$

que gera uma contradição.

Fica faltando tratar o caso onde  $\Delta\zeta(\tau-) \in aB^0$ . Nesse caso escolha  $b \geq a$  tal que  $\Delta\zeta(\tau) \in \partial bB$  e a prova segue de forma similar ao que foi feito no caso anterior. Para maiores detalhes veja (Dupuis e Ishii, 1991).  $\square$

### A.1.1.3 Prova da Existência Forte

Com a propriedade estabelecida acima, é possível mostrar a existência e unicidade de solução da mesma forma que é geralmente provado para o caso sem o termo de reflexão. Considere a seguinte condição:

#### Condição A.1.10.

(i) Para todo  $\xi^1, \xi^2 \in \mathbb{R}^d$  existe um constante positiva  $K_1 < \infty$  tal que

$$|b(\xi^1) - b(\xi^2)| + \|\sigma(\xi^1) - \sigma(\xi^2)\| \leq K_1 |\xi^1 - \xi^2|,$$

onde  $\|A\|^2 \triangleq \text{tr}AA'$ , para uma matriz  $A \in \mathbb{R}^{d \times d}$  qualquer.

(ii) Para cada  $\xi \in \mathbb{R}^d$ , existe uma constante positiva  $K_2 < \infty$  tal que

$$|b(\xi)|^2 + \|\sigma(\xi)\|^2 \leq K_2 (1 + |\xi|^2).$$

**Teorema A.1.11.** *Considere a equação diferencial estocástica refletida como dada pela Equação (A.1). Suponha que as direções de reflexão satisfazem “a condição do conjunto B”, e as funções  $b$  e  $\sigma$  satisfazem a Condição A.1.10. Então para cada  $x_0 \in G$  existe uma única solução para a Equação (A.1).*

*Demonstração.* A prova começa com a demonstração de unicidade. Suponha que  $(x^1, z^1)$  e  $(x^2, z^2)$  são soluções para a Equação (A.1) dado a condição inicial  $x^1(0) =$

$x^2(0) = x_0$ . Então, para  $0 \leq t < T < \infty$ , tem-se:

$$\sup_{s \leq t} |x^1(t) - x^2(t)|^2 \leq K \sup_{s \leq t} \left( \left| \int_0^s b(x^1(u)) - b(x^2(u)) du \right. \right. \\ \left. \left. + \int_0^s [\sigma(x^1(u)) - \sigma(x^2(u))] dw(u) \right| \right)^2$$

onde foi usado o resultado do Teorema A.1.9. Portanto, chega-se à:

$$\sup_{s \leq t} |x^1(t) - x^2(t)|^2 \leq 2K \sup_{s \leq t} \left( \left| \int_0^s b(x^1(u)) - b(x^2(u)) du \right|^2 \right. \\ \left. + \left| \int_0^s [\sigma(x^1(u)) - \sigma(x^2(u))] dw(u) \right|^2 \right)$$

Aplicando o resultado dado pela Equação (5.1.5) de (Arnold, 1974) página 81, tem-se que

$$\mathbb{E} \left[ \sup_{s \leq t} |x^1(t) - x^2(t)|^2 \right] \leq 2K \mathbb{E} \left[ \sup_{s \leq t} s \int_0^s |b(x^1(u)) - b(x^2(u))|^2 du \right. \\ \left. + 4 \int_0^t \|\sigma(x^1(u)) - \sigma(x^2(u))\|^2 dw(u) \right].$$

Com a Condição A.1.10(i), sabe-se que existe um  $\tilde{K}(T) > 0$  tal que

$$\mathbb{E} \left[ \sup_{s \leq t} |x^1(t) - x^2(t)|^2 \right] \leq \tilde{K}(T) \mathbb{E} \left[ \sup_{s \leq t} \int_0^s |x^1(u) - x^2(u)|^2 du \right] \\ \leq \tilde{K}(T) \int_0^t \mathbb{E} \left[ \sup_{u \leq s} |x^1(u) - x^2(u)|^2 \right] ds$$

Portanto, com a desigualdade de Bellman & Gronwall (página 107 de (Arnold, 1974)), tem-se que

$$\mathbb{E} \left[ \sup_{s \leq t} |x^1(s) - x^2(s)|^2 \right] = 0.$$

Usando a desigualdade de Chebychev (página 192 de (Ash e Doléans-Dade, 2000))

tem-se que para  $0 < \epsilon < 1$ :

$$\mathbb{P} \left( \sup_{s \leq t} |x^1(s) - x^2(s)|^2 \geq \epsilon \right) \leq \frac{1}{\epsilon^2} \mathbb{E} \left[ \sup_{s \leq t} |x^1(s) - x^2(s)|^2 \right] = 0.$$

Logo,

$$\mathbb{P} \left( \sup_{t \leq T} |x^1(s) - x^2(s)|^2 > 0 \right) = 0$$

para qualquer  $T > 0$ .

Será tratado agora a existência usando o método clássico para a construção de soluções para EDE sem reflexão, isto é, através de uma iteração de Picard. Seja  $x^0(t) \triangleq x_0$  e para  $n \geq 1$ ,

$$x^{n+1}(t) \triangleq x_0 + \int_0^t b(x^n(s)) ds + \int_0^t \sigma(x^n(s)) dw(s) + z^{n+1}(t),$$

onde, para cada  $\omega \in \Omega$ ,  $(x^{n+1}(\omega), z^{n+1}(\omega))$  é solução do Problema de Skorohod para  $\psi^{n+1}(\omega; \cdot) \triangleq x_0 + \int_0^\cdot b(x^n(\omega; s)) ds + \int_0^\cdot \sigma(x^n(\omega; s)) dw(\omega; s)$ .

Observe que para cada  $t \geq 0$

$$\begin{aligned} \mathbb{E} [|x^{n+1}(t)|^2] &\leq \mathbb{E} K_1 \sup_{s \leq t} \left[ |x_0|^2 + s \int_0^s |b(x^n(u))|^2 du \right. \\ &\quad \left. + \left| \int_0^s \sigma(x^n(u)) dw(u) \right|^2 \right] \\ &\leq K_1 \mathbb{E} [|x_0|^2] + K_2(t^2) \left( 1 + \sup_{s \leq t} \mathbb{E} [|x^n(s)|^2] \right), \end{aligned}$$

onde foi usado a Condição A.1.10(ii) e o resultado dado pela Equação (5.1.5) em (Arnold, 1974) página 81.

Supondo que  $\mathbb{E} [|x_0|^2] < \infty$ , caso contrário usa-se a técnica de truncagem como é feito (Arnold, 1974) página 111, tem-se que  $\mathbb{E} [|x^n(t)|^2] < \infty$  para cada  $0 \leq t < \infty$ . Usando os argumentos empregados na prova de unicidade, é possível

escrever para cada  $0 \leq t < T < \infty$ :

$$\mathbb{E} \left[ \sup_{s \leq t} |x^{n+1}(s) - x^n(s)|^2 \right] \leq \tilde{K}(T) \int_0^t \mathbb{E} \left[ \sup_{u \leq s} |x^n(u) - x^{n-1}(u)|^2 \right] ds.$$

Usando a seguinte formula de Cauchy:

$$\int_{t_0}^t \int_{t_0}^{t_{n-1}} \cdots \int_{t_0}^{t_1} g(s) ds dt_1 \cdots dt_{n-1} = \int_{t_0}^t g(s) \frac{(t-s)^{n-1}}{(n-1)!} ds,$$

é possível escrever:

$$\mathbb{E} \left[ \sup_{s \leq t} |x^{n+1}(s) - x^n(s)|^2 \right] \leq \left( \tilde{K}(T) \right)^n \int_0^t \frac{(t-s)^{n-1}}{(n-1)!} \mathbb{E} \left[ \sup_{u \leq s} |x^1(u) - x^0(u)|^2 \right] ds.$$

Como

$$\begin{aligned} \mathbb{E} \left[ \sup_{u \leq s} |x^1(u) - x^0(u)|^2 \right] &\leq \mathbb{E} \left[ \sup_{u \leq s} |x^1(u)|^2 \right] + \mathbb{E} [|x_0|^2] \\ &\leq C(s^2) (1 + \mathbb{E} [|x_0|^2]) \triangleq \tilde{C}(s^2), \end{aligned}$$

onde  $C$  e  $\tilde{C}$  são constantes positivas para cada  $s$ , tem-se que

$$\mathbb{E} \left[ \sup_{s \leq t} |x^{n+1}(s) - x^n(s)|^2 \right] \leq \tilde{C}(T^2) \frac{(\tilde{K}(T)T)^n}{n!},$$

para  $n \geq 0$ . Com o Lema de Borel-Canteli (veja página 68 de (Ash e Doléans-Dade, 2000)) e a desigualdade de Chebychev, mostra-se que

$$\mathbb{P} \left( \limsup_n \left\{ \sup_{t \leq T} |x^{n+1}(t) - x^n(t)|^2 > 0 \right\} \right) = 0,$$

e portanto existe um  $x$  para o qual  $x^n$  converge quase em toda parte. Falta ainda verificar se esse  $x$  é de fato solução da EDER. Como o argumento usado para EDE sem reflexão é o mesmo que o necessário aqui, os detalhes são omitidos (para referência veja a página 110 de (Arnold, 1974)).  $\square$

### A.1.2 Existência de Solução Fraca

Nesta seção, somente será tratado a existência fraca. A prova de unicidade pode ser encontrada em (Dai e Williams, 1995) para movimentos Brownianos refletidos. Ela carrega detalhes desnecessários que não são relacionados com o tema principal deste capítulo. Como somente a existência será tratada, a seção se restringe a representação simplificada do processo de reflexão  $z$ . Isto é,  $x$  será a solução da EDER dada pela Equação (A.1) mas onde  $z$  satisfaz:

$$(iv) \quad z(t) = \sum_{i=1}^N d_i y_i(t), \text{ onde } y_i \text{ são processos com } y_i(0) = 0, \text{ não-decrescentes,} \\ \text{e que satisfazem } \int_0^t \mathbb{I}\{x(s) \in \partial G_i\} dy_i(s) = y_i(t) \text{ para qualquer } t > 0.$$

De forma semelhante às EDE usuais, as condições necessárias para a prova de existência e unicidade de soluções são relaxadas consideravelmente. Isto também vale para as condições impostas sobre as direções de reflexão.

**Condição A.1.12.** *Defina  $N(\xi) = \{n_i : \xi \in \partial G_i\}$  e  $D(\xi) = \{d_i : \xi \in \partial G_i\}$ . Para cada  $\xi \in \partial G$  existe um vetor  $\gamma \in \text{Cone}(N(\xi))$  tal que  $\langle \gamma, d \rangle > 0$  para todo  $d \in D(\xi)$ .*

Considere o caso onde  $G \triangleq \mathbb{R}_+^d$ , seja  $K \subseteq \{1, \dots, N\}$ , e defina a matriz  $D \in \mathbb{R}^{d \times |K|}$  como a matriz com colunas  $\{d_i, i \in K\}$ . De forma análoga, seja  $N \in \mathbb{R}^{d \times |K|}$  a matriz com colunas  $\{n_i, i \in K\}$ . Então a existência de tal  $\gamma$  descrito acima corresponde a existência de  $\alpha \in \mathbb{R}^{|K|} > 0$  tal que

$$(N\alpha) = \gamma$$

e portanto a condição acima implica que  $(N\alpha)'D > 0$ . Note que portanto a condição acima no caso  $G \triangleq \mathbb{R}_+^d$  é equivalente a condição de que a matriz  $D'N$  seja completamente- $\mathcal{S}$ , que é dada abaixo:

**Condição A.1.13** (Matriz Completamente- $\mathcal{S}$ ). *Uma matriz  $R \in \mathbb{R}^{a \times a}$  é chamada de completamente- $\mathcal{S}$  se para cada submatriz principal  $\tilde{R}$  de  $R$  (i.e., matriz construída removendo colunas e linhas de  $R$  com índices em um dado subconjunto de*

$\{1, \dots, a\}$ ) existe um vetor  $\nu$  tal que

$$\tilde{R}\nu > 0.$$

Portanto a Condição A.1.12 nada mais é do que uma generalização da condição completamente- $\mathcal{S}$ . Foi demonstrado em (Reiman e Williams, 1988) que esta condição é suficiente e necessária para a existência fraca de movimentos Brownianos presos no quadrante positivo de  $\mathbb{R}^d$ . Em palavras, pode-se interpretar a condição acima como a existência de uma direção que aponta para dentro do espaço de estados que pode ser usada para “empurrar” o processo de volta ao espaço de estados caso tente escapar.

Além da condição acima, é preciso garantir que essa direção que será usada para empurrar o processo possa ser de fato escrita como uma combinação linear positiva das direções de reflexão  $d_i$ . Isto é:

**Condição A.1.14.** *Para cada  $\xi \in \partial G$  existe um vetor  $\theta \in \text{Cone}(D(\xi))$  tal que  $\langle \theta, n \rangle > 0$  para todo  $n \in N(\xi)$ .*

Em (Dai e Williams, 1995) é discutido as situações em que a Condição A.1.12 é equivalente à Condição A.1.14. É verificado que quando o poliedro convexo  $G$  é simples as duas condições são equivalentes. Um poliedro convexo é simples quando para qualquer  $K \subseteq \{1, \dots, N\}$  tal que  $\bigcap_{i \in K} \partial G_i \equiv \{\xi_0\}$  ( $\xi_0$  sendo um vetor qualquer em  $\mathbb{R}^d$ ) exatamente  $d$  faces se encontram em  $\xi_0$ .

**Teorema A.1.15.** *Assuma as Condições A.1.1, A.1.12, A.1.14, e que as funções  $b$  e  $\sigma$  são limitadas e contínuas quando restritas à  $G$ . Então existe uma solução fraca para (A.1) com processo de reflexão satisfazendo o item (iv) acima. Isto é, existe um espaço de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$  e uma filtração  $\mathcal{F}_t$  onde são definidos processos  $x, z, e w$   $\mathcal{F}_t$ -mensuráveis que satisfazem a Equação (A.1).*

*Demonstração.* A prova deste teorema é baseada na prova do Teorema 5.4 de (Kushner, 2001). Será usado uma sequência de processos aproximantes  $(x^n, \psi^n, z^n)$  que converge fracamente para um processo satisfazendo a Equação (A.1).

Para cada  $n \in \mathbb{N}$ , defina a sequência  $x_k^n$  como  $x_0^n = x_0$  e

$$x_{k+1}^n = x_k^n + \delta\psi_k^n + \delta z_k^n$$

onde

$$\delta\psi_k^n \triangleq b(x_k^n) [(k+1)/n - k/n] + \sigma(x_k^n) [w((k+1)/n) - w(k/n)],$$

e  $w$  é um processo de Wiener padrão que toma valores em  $\mathbb{R}^d$ . Para definir  $\delta z_k^n$ , seja o conjunto  $K \triangleq \{i \in \{1, \dots, N\} : \langle x_k^n + \delta\psi_k^n, n_i \rangle < B_i\}$ . Se  $K = \emptyset$ , defina  $\delta z_k^n \triangleq 0$ . Caso contrário, seja  $d \in \text{Cone}\{d_i, i \in K\}$  tal que  $\|d\| = 1$ , e então neste caso

$$\delta z_k^n \triangleq \max_{i \in K} \left( \frac{B_i - \langle x_k^n + \delta\psi_k^n, n_i \rangle}{\langle d, n_i \rangle} \right) d.$$

É fácil observar que

$$\max_{i \in K} \left( \frac{B_i - \langle x_k^n + \delta\psi_k^n, n_i \rangle}{\langle d, n_i \rangle} \right) \leq \max_{i \in K} \left( \frac{1}{\langle d, n_i \rangle} \right) \|\delta\psi_k^n\|.$$

De fato, como para qualquer  $i \in K$ :

$$\begin{aligned} B_i - \langle x_k^n + \delta\psi_k^n, n_i \rangle &> 0 \Rightarrow \\ -\langle \delta\psi_k^n, n_i \rangle &> \langle x_k^n, n_i \rangle - B_i \geq 0. \end{aligned}$$

Logo  $-\langle \delta\psi_k^n, n_i \rangle = |\langle \delta\psi_k^n, n_i \rangle|$ . Portanto tem-se que

$$B_i - \langle x_k^n, n_i \rangle - \langle \delta\psi_k^n, n_i \rangle \leq |\langle \delta\psi_k^n, n_i \rangle| \leq \|\delta\psi_k^n\|,$$

e desta forma  $\|\delta z_k^n\| < C\|\delta\psi_k^n\|$ , para um  $C > 0$  que depende somente de  $d_i$  e  $n_i$ .

Também é possível verificar que para qualquer  $\mu, T > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \max_{k \leq Tn} |\delta\psi_k^n| \geq \mu \right) = 0,$$

e portanto para  $n$  grande  $\delta z_k^n$  está bem definido no sentido em que existirá um  $d \in \text{Cone}(D(\xi))$  tal que  $\langle d, n \rangle > 0$  para  $n \in N(\xi)$  onde  $\xi$  é o ponto em  $\partial G$  mais próximo de  $x_k^n + \delta\psi_k^n$  (a distância entre  $\xi$  e  $x_k^n + \delta\psi_k^n$  fica arbitrariamente pequena tomando  $n$  grande).

Defina agora  $z_0^n = \psi_0^n = 0$  e para  $k \geq 1$  faça

$$\begin{aligned} z_k^n &= \sum_{l=0}^{k-1} \delta z_l^n \\ \psi_k^n &= \sum_{l=0}^{k-1} \delta\psi_l^n. \end{aligned}$$

Sejam os processos a tempo contínuo  $z^n$ ,  $\psi^n$ , e  $x^n$  definidos como:

$$\begin{aligned} z^n(t) &= z_k^n \quad \text{para } t \in [k/n, (k+1)/n) \\ \psi^n(t) &= \psi_k^n \quad \text{para } t \in [k/n, (k+1)/n) \\ x^n(t) &= x_k^n \quad \text{para } t \in [k/n, (k+1)/n), \end{aligned}$$

para cada  $k \geq 0$ . Portanto:

$$x^n(t) = x_0 + \psi^n(t) + z^n(t). \tag{A.5}$$

Pelas propriedades das integrais estocásticas (veja a discussão do Capítulo 4 de (Arnold, 1974)) e pelo fato de que  $\sigma^n(\omega, t) \triangleq \sigma(x_k^n(\omega))$  para  $t \in [k/n, (k+1)/n)$  ser não-antecipativa e limitada, sabe-se que a soma dos termos  $\sigma(x_k^n) [w((k+1)/n) - w(k/n)]$  é limitada em probabilidade e possui como limite um processo com funções amostras quase certamente contínuas. Além disso, verifica-se que a soma dos termos  $b(x_k^n) [(k+1)/n - k/n]$  também é limitada em probabilidade usando o fato de  $b$  ser limitado em  $G$ . Portanto  $\{\psi^n\}$  é assintoticamente contínua.

Com a discussão acima, verifica-se que todas as condições do Teorema A.2.3 são atendidas e portanto  $\{z^n\}$  é assintoticamente contínua. Como o limite no sentido fraco de qualquer subsequência convergente de  $\{\psi^n\}$  e  $\{z^n\}$  possui funções amostras quase certamente contínuas, tem-se que  $\{x^n, \psi^n, z^n\}$  é assintoticamente contínua.

Agora somente falta verificar se o limite de alguma subsequência convergente de  $\{x^n, \psi^n, z^n\}$  satisfaz as condições estabelecidas para uma solução da EDER.

Pela definição de  $x^n$  e  $x_k^n$  não há dúvidas que a diferença

$$\int_0^t b(x^n(s))ds - \sum_{\{k:(k+1)/n < t\}} b(x_k^n) [(k+1)/n - k/n]$$

é zero em probabilidade quando  $n \rightarrow \infty$ . Para mostrar a convergência da integral estocástica satisfazendo a condição de que é  $\mathcal{F}_t$ -mensurável usa-se o Teorema 2.1.3 de (Kushner, 2001). Seja  $h(\cdot)$  uma função que toma valores reais limitada e contínua. Seja  $t, \tau > 0$ , e para qualquer  $p \in \mathbb{N}$ , seja  $0 \leq s_i \leq t$ , com  $i \in \{1, \dots, p\}$ . Todas estas variáveis são escolhidas de forma arbitrária. Então, pela definição de  $(x^n, z^n)$  e propriedades de  $w$  tem-se que

$$\begin{aligned} & \mathbb{E} \{h(x^n(s_i), \psi^n(s_i), z^n(s_i); i \leq p) \times [w(t + \tau + 1/n) - w(t + 1/n)]\} = 0 \\ & \mathbb{E} \{h(x^n(s_i), \psi^n(s_i), z^n(s_i); i \leq p) \\ & \quad \times [(w(t + \tau + 1/n) - w(t + 1/n))(w(t + \tau + 1/n) - w(t + 1/n))' - \tau I]\} = 0 \end{aligned}$$

Seja  $(x, \psi, z)$  o limite de uma subsequência convergente de  $\{x^n, \psi^n, z^n\}$ . Então, como os processos  $x, \psi, z$  tem funções amostras quase certamente contínuas e  $h$  é contínua, as equações acima valem para  $(x, \psi, z)$  em lugar de  $(x^n, \psi^n, z^n)$ . Desta forma, as condições do Teorema 2.1.3 de (Kushner, 2001) são satisfeitas e de fato a soma dos termos  $\sigma(x_k^n) [w((k+1)/n) - w(k/n)]$  converge para uma integral estocástica  $\mathcal{F}_t$ -mensurável onde  $\mathcal{F}_t$  é a mínima  $\sigma$ -álgebra que mensura  $\{w(s), z(s); s \leq t\}$ .

Da definição de  $z^n$ , pode-se verificar que  $z$  satisfaz as condições sobre o termo

de reflexão. □

## A.2 Convergência Fraca de Processos de Reflexão Aproximados

Considere processos estocásticos  $(x^n, \psi^n, z^n)$  e a variável aleatória  $x^n(0) \in G$  que satisfazem

$$x^n(t) = x^n(0) + \psi^n(t) + z^n(t) \quad (\text{A.6})$$

para todo  $n \in \mathbb{N}$  e  $t \geq 0$ . Além disso, para cada  $t > 0$ ,  $x^n(t)$  pertence à  $G^{\epsilon_n} \triangleq \{\xi \in \mathbb{R}^d : \text{dist}(\xi, G) \leq \epsilon_n\}$ . O processo  $z^n$  é encarregado de manter  $x^n$  dentro do espaço de estados  $G^{\epsilon_n}$ . A medida que  $n$  aumenta, deseja-se que  $\epsilon_n$  fique menor de forma que  $x^n$  fique restrito à um domínio mais próximo de  $G$ . Sob algumas hipóteses, será mostrado (no Teorema A.2.3 abaixo) que a sequência  $\{z^n\}$  é limitada em probabilidade. Este teorema é de fundamental importância o estudo de aproximações de filas sob tráfego pesado.

**Condição A.2.1.** *O processo  $z^n$  satisfaz as seguintes condições:*

- (i)  $z^n(t) = \sum_{i=1}^N d_i y_i^n(t)$  para todo  $t \geq 0$ , onde  $y_i^n(0) = 0$  e  $y_i^n$  são processos com função amostras não decrescentes e com variação total limitada.
- (ii) Existe uma sequência  $\mu_n$ , que converge para zero a medida que  $n \rightarrow \infty$ , tal que, se  $y_i^n(t) - y_i^n(s) > 0$ , para  $0 \leq s < t < \infty$ , então

$$\text{dist}(x^n(u), \partial G_i) \leq \mu_n \quad \forall u \in I,$$

para algum  $I \subseteq (s, t]$ ,  $I \neq \emptyset$ .

- (iii) Para cada  $T < \infty$ :

$$\sup_{t \leq T} |z^n(t) - z^n(t-)| \rightarrow 0 \quad (\text{A.7})$$

em probabilidade.

**Lema A.2.2.** *Assuma a Condição A.1.12 e tome  $\xi \in \partial G$ . Então nenhuma combinação linear positiva das direções de reflexão  $d_i \in D(\xi)$  pode ser zero.*

*Demonstração.* Assuma que existam  $\alpha_{i_k} > 0$ ,  $i_k \in \{i_1, \dots, i_K\}$ , tais que

$$\sum_{l=1}^K \alpha_{i_l} d_{i_l} = 0 \quad (\text{A.8})$$

para  $d_{i_l} \in D(\xi)$ . Para simplificar a notação seja  $\{i_1, \dots, i_K\} = \{1, \dots, K\}$ . Defina  $N = \{n_1, \dots, n_K\}$  como a matriz cujas colunas são direções normais definidas no ponto  $\xi$ . Seja também  $D = \{d_1, \dots, d_K\}$ . Portanto, pela Condição A.1.12, tem-se que existem um  $\gamma \in \mathbb{R}_{\geq 0}^K$  tal que

$$\gamma' N' D > 0.$$

Mas a Equação (A.8) afirma que  $D\alpha = 0$  portanto

$$\gamma' N' D\alpha = 0,$$

que gera uma contradição pois o vetor  $\alpha$  tem todas suas componentes positivas.  $\square$

**Teorema A.2.3.** *Assuma as Condições A.1.1, A.1.12, e A.2.1. Para cada  $n \in \mathbb{N}$  seja  $(x^n, \psi^n, z^n)$  processos estocásticos satisfazendo a Equação (A.6). Suponha que  $\{x^n(0)\}$  é limitada em probabilidade e  $\psi^n$  é assintoticamente contínuo. Então,  $z^n$  é assintoticamente contínuo.*

*Demonstração.* Esta prova usa as idéias presentes na prova do Teorema 3.6.1 de (Kushner, 2001). Suponha por contradição que existe um salto assintótico em algum  $y_i^n$ . Portanto, existe um conjunto  $M \subset \Omega$  com  $\mathbb{P}(M) > 0$ , onde para cada  $\omega \in M$  existe um  $t(\omega) > 0$ , um  $\nu_0(\omega) > 0$ , um  $S \in \mathbb{N}$ , e uma sequência  $\delta_n \rightarrow 0$  tais que

$$y_i^n(\omega; t(\omega) + \delta_n) - y_i^n(\omega; t(\omega)) \geq \nu_0(\omega) \quad \forall n \geq S.$$

Para  $\omega \in M$  considere o seguinte:

$$x^n(t + \delta_n) - x^n(t) = \psi^n(t + \delta_n) - \psi^n(t) + \sum_{i=1}^N (y_i^n(t + \delta_n) - y_i^n(t)) d_i,$$

onde foi omitido  $\omega$  para simplificar a notação. Defina, para qualquer função  $f : [0, \infty) \rightarrow \mathbb{R}^d$ ,  $d \geq 1$ , a função  $\tilde{f}(s) = f(t + s) - f(t)$ . Então

$$\tilde{x}^n(\delta_n) = \tilde{\psi}^n(\delta_n) + \sum_{i=1}^N d_i \tilde{y}_i^n(\delta_n).$$

O método da transformação do tempo (veja Seção 2.2.6 de (Kushner, 2001) para detalhes) será usado. Para isso, defina  $T^n(t) = t + \sum_i \tilde{y}_i^n(t)$  e  $\hat{T}^n(t) = \inf\{s : T^n(s) > t\}$ . Seja  $\hat{f}^n(t) \triangleq f^n(\hat{T}^n(t))$  para qualquer função  $f^n$ . Desta forma,

$$\hat{x}^n(\delta_n) = \hat{\psi}^n(\delta_n) + \sum_i \hat{y}_i^n(\delta_n) d_i.$$

É possível verificar facilmente que  $\hat{T}^n(\delta_n) \rightarrow 0$  com probabilidade um quando  $n \rightarrow \infty$ . Como  $\psi^n$  é assintoticamente contínuo e pelo Teorema de Representação de Skorohod (Billingsley, 1999), pode-se considerar que  $\psi^n \rightarrow \psi$  com probabilidade um, onde  $\psi$  é contínuo em  $\mathbb{P}$ -quase toda parte. Portanto,

$$\hat{\psi}^n(\delta_n) = \psi^n(t + \hat{T}^n(\delta_n)) - \psi^n(t) \rightarrow 0 \quad \mathbb{P} - \text{q.c.}$$

quando  $n \rightarrow \infty$ .

Por construção,  $\hat{y}_i$  é assintoticamente Lipschitz-contínua (veja o método de transformação de tempo Seção 2.2.6 de (Kushner, 2001) e Seção 11.1.2 de (Kushner e Dupuis, 1992) para detalhes), com constante de Lipschitz inferior à 1. Desta forma,  $\{\hat{y}^n\}$  é limitada em probabilidade.

Tome uma subsequência convergente de  $\{\hat{x}^n(\delta_n)\}$  e observe que seu limite

satisfaz

$$\Delta x(t) = \sum_{i=1}^N \Delta y_i(t) d_i,$$

onde  $\Delta f(t) = f(t+) - f(t)$  para qualquer função  $f$ . Pela Condição A.2.1(ii) pode-se supor que  $x(t) \in \partial G$ . Suponha que  $x(t) \in \partial G_i$  e  $x(t) \notin \partial G_j$ , para  $i \neq j$ . Portanto,  $\Delta y_i(t) > 0$  e  $\Delta y_j(t) = 0$ ,  $\forall j \neq i$ . Como  $\langle n_i, d_i \rangle > 0$ , um salto em  $y_i$  implica que  $x(t+) \in G^0$ , o que fere a Condição A.2.1(ii), já que  $y_i$  somente poderá aumentar nos instantes  $t$  em que  $x(t) \in \partial G$ .

Suponha agora que  $x(t) \in \partial G_i \cap \partial G_j$  e não pertença a mais nenhuma fronteira. Caso  $\Delta y_i(t) > 0$  e  $\Delta y_j(t) = 0$ , uma contradição aparece usando os mesmos argumentos do parágrafo acima. Portanto, considere o caso onde  $\Delta y_i(t) > 0$  e  $\Delta y_j(t) > 0$ . Do Lema A.2.2 nenhuma combinação linear positiva de  $d_i$  e  $d_j$  pode ser zero. Desta forma, um salto em  $\Delta z(t) = d_i \Delta y_i(t) + d_j \Delta y_j(t)$  implica que  $x(t+) \in G^0$ , ou seja, a Condição A.2.1(ii) é ferida novamente. O mesmo argumento é válido quando  $x(t)$  pertencer a mais de duas faces.  $\square$

### A.3 Propriedade do Termo de Reflexão nas Fronteiras

Nesta seção, será apresentado um resultado que afirma que o tempo total que um processo de difusão refletido (não-degenerado) passa próximo da intercessão entre duas fronteiras tem pouco efeito no processo de reflexão. Este resultado é importante pois garante uma representação única do termo de reflexão mesmo quando as direções de reflexão não são linearmente independentes.

Grande parte do estudo desta tese se concentrou em uma extensão deste resultado, mas ainda sem sucesso. A primeira tentativa foi a extensão do resultado publicado em (Piera et al., 2006), que prova o resultado de interesse com o uso de “*local times*”. Foi possível mostrar que a Condição A.1.12 é suficiente para a demonstração do Teorema 3.1 de (Piera et al., 2006) evitando uma condição mais restritiva imposta na referência. Contudo, foi descoberto em seguida a existência de um possível problema em uma passagem do Lema 2.2 de (Piera et al., 2006),

que foi usado para obter o resultado. A passagem do limite quando  $\epsilon$  tende à zero contém detalhes sutis que não possuem explicação clara. Além disso, a prova do lema não exclui o caso não-degenerado que pode ser usado como um contra exemplo para a validade do Teorema 3.1 de (Piera et al., 2006).

Em busca de uma possível solução, foi estudado a prova do Teorema 4.3.6 de (Kushner, 2001), que trata um caso menos geral mas que poderia auxiliar o estudo deste assunto. Contudo, a prova quando o coeficiente de difusão  $\sigma$  é dependente do estado também contém alguns possíveis problemas que não são tratados com detalhes. O autor do livro foi contactado, e ele confirmou o possível problema.

Portanto, como o resultado parece ainda estar em aberto para o caso mais geral (i.e., quando  $\sigma$  é dependente do estado e as direções de reflexão são variáveis aleatórias dependentes do tempo) será tratado o caso mais simples que foi retirado de (Dai e Williams, 1995) com algumas modificações sugeridas em (Kushner, 2001).

Seja  $x$  um processo estocástico dado por

$$x(t) = x_0 + \int_0^t b(x(s))ds + w(t) + \sum_{i=1}^N y_i(t)d_i, \quad (\text{A.9})$$

onde  $x_0 \in G$ ,  $x(t) \in G$  para todo  $t > 0$ ,  $w$  é um processo de Wiener  $\mathcal{F}_t$ -mensurável onde  $\mathcal{F}_t$  é a mínima  $\sigma$ -álgebra que mensura  $\{w(s), y_i(s); i \in \{1, \dots, N\}, s \leq t\}$  com matriz de covariância  $\Sigma$ , e  $y_i$  satisfaz:

- (i)  $y_i(0) = 0$ ;
- (ii)  $y_i$  tem quase certamente funções amostras não-decrescentes;
- (iii)  $y_i(t) = \int_0^t \mathbb{I}\{x(s) \in \partial G_i\} dy_i(s)$  para todo  $t \geq 0$ .

**Condição A.3.1.** *O coeficiente de deriva da Equação (A.9) é limitado quando restrito à  $G$ . A matriz de covariância  $\Sigma$  é definida positiva (i.e.,  $x$  é um processo não-degenerado).*

Será preciso ainda a seguinte condição no espaço de estados  $G$  e nas direções normais  $n_i$ .

**Definição A.3.2** (Conjunto Maximal (Dai e Williams, 1995)). *Chama-se um conjunto  $K \subseteq \{1, \dots, N\}$  de maximal se  $K \neq \emptyset$ ,  $\cap_{i \in K} \partial G_i \neq \emptyset$ , e  $\cap_{i \in K} \partial G_i \neq \cap_{i \in \tilde{K}} \partial G_i$  para qualquer  $\tilde{K} \subseteq \{1, \dots, N\}$  tal que  $K \subset \tilde{K}$ .*

**Condição A.3.3.** *Para qualquer conjunto maximal  $K \subseteq \{1, \dots, N\}$  existe um subconjunto  $L \subseteq K$  onde  $\{n_i; i \in L\}$  forma uma base linear para o espaço gerado por  $\{n_i; i \in K\}$ .*

O lema abaixo será usado na prova do Teorema A.3.5.

**Lema A.3.4.** *Suponha a Condição A.3.3 e seja  $K \subseteq \{1, \dots, N\}$  tal que  $\cap_{i \in K} \partial G_i \neq \emptyset$ . Então  $\cap_{i \in K} \partial G_i = \cap_{i \in L} \partial G_i$ , onde  $L \subseteq K$  é o conjunto da Condição A.3.3.*

*Demonstração.* Esta prova foi retirada do Lema 4.5 de (Dai e Williams, 1995). Pela propriedade de  $\{n_i; i \in L\}$ , existem constantes  $a_{ji}$  tais que

$$n_j = \sum_{i \in L} a_{ji} n_i \quad \forall j \in K.$$

Como  $\cap_{i \in K} \partial G_i \neq \emptyset$ , existe um  $\xi_0 \in \cap_{i \in K} \partial G_i$  e portanto

$$\langle n_j, \xi_0 \rangle = B_j \quad \forall j \in K.$$

Portanto, para  $\xi \in \cap_{i \in L} \partial G_i$ , tem-se que para  $j \in K$

$$\langle n_j, \xi \rangle = \sum_{i \in L} a_{ji} \langle n_i, \xi \rangle = \sum_{i \in L} a_{ji} B_i = \sum_{i \in L} a_{ji} \langle n_i, \xi_0 \rangle = \langle n_j, \xi_0 \rangle = B_j.$$

Logo  $\xi \in \cap_{i \in K} \partial G_i$ . □

**Teorema A.3.5.** *Seja  $x$  um processo estocástico como dado acima pela Equação (A.9). Assuma as Condições A.1.1, A.1.12, A.3.3, e A.3.1. Seja também  $T$  um tempo de parada (stopping time). Então para qualquer  $K \subseteq \{1, \dots, N\}$  com  $|K| \geq 2$ ,*

$$\int_0^T \mathbb{I}\{x(s) \in \cap_{i \in K} \partial G_i\} dy_i(s) = 0 \quad \mathbb{P} - q.c.,$$

para cada  $i \in \{1, \dots, N\}$ .

*Demonstração.* A prova deste teorema foi retirada de (Dai e Williams, 1995), e é uma extensão do resultado de (Reiman e Williams, 1988).

Primeiramente, observe que através do uso da transformação de Girsanov (e a condição de Novikov, veja Teorema de Girsanov (Oksendal, 2005), página 162) existe uma medida de probabilidade  $\tilde{\mathbb{P}}$  que é equivalente à  $\mathbb{P}$  onde  $x$  satisfaz

$$x(t) = x_0 + \tilde{w}(t) + \sum_{i=1}^N y_i(t) d_i,$$

para todo  $0 < t < \infty$ , em  $(\Omega, \mathcal{F}, \tilde{\mathbb{P}})$ , onde  $\tilde{w}$  é um processo de Wiener. Portanto, basta mostrar o resultado para  $b \equiv 0$ , sem perda de generalidade. Além disso, pode-se supor que  $x_0 \in \partial G$  sem perda de generalidade.

A prova é dividida em dois casos:

(i) Considere o caso onde  $K = \{1, \dots, N\}$  e  $\cap_{i \in K} \partial G_i \neq \emptyset$ . Da hipótese, tem-se a existência de  $L \subseteq K$  onde  $\{n_i; i \in L\}$  forma uma base linear para o espaço gerado por  $\{n_i; i \in K\}$ . Para simplificar a notação e sem perda de generalidade suponha que  $L = \{1, \dots, l\}$ . Seja  $A \in \mathbb{R}^{l \times d}$  a matriz cujas linhas são os vetores  $n_i$ , para  $i \in L$ . A Condição A.1.12 garante a existência de um vetor  $\eta = \sum_{i \in K} \gamma_i n_i$ , onde  $\gamma_i > 0$ , tal que  $\langle \eta, d_i \rangle > 0$  para todo  $i \in K$ . Como  $\{n_i; i \in L\}$  é uma base linear para o espaço gerado por  $\{n_i; i \in K\}$ , pode-se escrever  $\eta = A' \lambda$ , onde  $\lambda$  é um vetor coluna dado por  $(\lambda_1, \dots, \lambda_l)'$ .

Defina  $\Gamma = A \Sigma A'$  e observe que  $\Gamma$  é definida positiva já que  $A$  tem linhas linearmente independentes. Além disso,  $\cap_{i \in L} \partial G_i = \cap_{i \in K} \partial G_i$  pelo Lema A.3.4. Seja  $\xi_0 \in \cap_{i \in L} \partial G_i$  e observe que para qualquer  $\xi \in G$ , tem-se que

$$\langle \eta, (\xi - \xi_0) \rangle = \sum_{i \in K} \gamma_i \langle n_i, (\xi - \xi_0) \rangle = \sum_{i \in K} \gamma_i (\langle n_i, \xi \rangle - B_i) \geq 0.$$

Para cada  $\xi \in G$  e  $r \in (0, 1)$  defina

$$\begin{aligned} d^2(\xi, r) &= (A(\xi_0 - \xi) + r\Gamma\lambda)' \Gamma^{-1} (A(\xi_0 - \xi) + r\Gamma\lambda) \\ &\geq r^2 (\lambda' \Gamma \lambda) > 0. \end{aligned}$$

Para cada  $\epsilon \in (0, 1)$ , seja

$$\phi_\epsilon(\xi) \triangleq \begin{cases} \frac{1}{2-d} \int_\epsilon^1 r^{d-2} (d^2(\xi, r))^{(2-d)/2} dr & \text{se } d \geq 3, \\ \frac{1}{2} \int_\epsilon^1 \ln(d^2(\xi, r)) dr & \text{se } d = 2. \end{cases} \quad (\text{A.10})$$

Observe que, se  $\mathfrak{L}$  é definido como

$$\mathfrak{L}f = \frac{1}{2} \sum_{i,j=1}^d \left( \Sigma_{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} \right),$$

para qualquer função  $f \in C^2(G)$ , o termo dentro da integral da Equação (A.10) é  $\mathfrak{L}$ -harmônico em um domínio contendo  $G$ , já que  $\text{tr}(\Sigma A' \Gamma^{-1} A) = d$ . Além disso, tem-se que

$$\nabla \phi_\epsilon(\xi) = \int_\epsilon^1 r^{d-2} A' \Gamma^{-1} (A(\xi - \xi_0) + r\Gamma\lambda) (d^2(\xi, r))^{-d/2} dr,$$

e portanto

$$\langle d_i, \nabla \phi_\epsilon(\xi) \rangle = \int_\epsilon^1 r^{d-2} (\langle d_i, A' \Gamma^{-1} A(\xi - \xi_0) \rangle + r \langle d_i, \eta \rangle) (d^2(\xi, r))^{-d/2} dr,$$

para  $i \in K$ .

Seja  $u_i \triangleq \Gamma^{-1} A d_i$ . Observe que  $u_i \neq 0$  já que  $\langle u_i, \Gamma \lambda \rangle = \langle d_i, \eta \rangle > 0$ . Para  $\xi$  tal que  $\|A(\xi - \xi_0)\| < \epsilon \beta_i$ , onde  $\beta_i \triangleq \langle d_i, \eta \rangle / \|u_i\|$ , tem-se que para  $r > \epsilon$

$$\begin{aligned} \|d^2(\xi, r)\| &\leq \|\Gamma^{-1}\| (\|A(\xi - \xi_0)\| + \|r\Gamma\lambda\|)^2 \\ &\leq \|\Gamma^{-1}\| (\epsilon \beta_i + \|\Gamma\lambda\|)^2 r^2, \end{aligned}$$

e portanto,

$$\begin{aligned} \langle d_i, \nabla \phi_\epsilon(\xi) \rangle &\geq (\|\Gamma^{-1}\|(\epsilon\beta_i + \|\Gamma\lambda\|)^2)^{-d/2} \int_\epsilon^1 r^{d-2} (-\epsilon \langle d_i, \eta \rangle + r \langle d_i, \eta \rangle) r^{-d} dr \\ &\geq -c_i(\ln \epsilon + 1), \end{aligned}$$

para  $i \in K$ , onde  $c_i \triangleq (\|\Gamma^{-1}\|(\epsilon\beta_i + \|\Gamma\lambda\|)^2)^{-d/2} \langle d_i, \eta \rangle$ .

Seja  $\rho_i(\xi) \triangleq -\langle u_i, A(\xi - \xi_0) \rangle / \langle d_i, \eta \rangle$ . Então, seguindo os mesmos passos usados em (Reiman e Williams, 1988) para derivar as Inequações (20) - (24), é possível mostrar que

$$\langle d_i, \nabla \phi_\epsilon(\xi) \rangle \geq -\hat{c}_i,$$

para  $i \in K$  e qualquer  $\xi \in G$ , onde  $\hat{c}_i$  é uma constante não-negativa.

Para cada  $k \in \mathbb{N}$  defina os tempos de parada

$$\tau_k \triangleq \inf \{t \geq 0 : \|A(x(t) - x_0)\| \geq k \text{ ou } y_i(t) \geq k, \text{ para algum } i \in \{1, \dots, N\}\} \wedge T.$$

Observe que  $\phi_\epsilon$  e  $\nabla \phi_\epsilon$  são limitados em  $\{\xi \in G : \|A(\xi - x_0)\| \leq k\}$  e então aplicando a formula de Itô em  $\phi_\epsilon(x(\cdot \wedge \tau_k))$  e tomando a esperança tem-se

$$\begin{aligned} \mathbb{E} [\phi_\epsilon(x(t \wedge \tau_k)) - \phi_\epsilon(x_0)] &= \sum_{i \in K} \mathbb{E} \left[ \int_0^{t \wedge \tau_k} \langle d_i, \nabla \psi_\epsilon(x(s)) \rangle dy_i(s) \right] \\ &\geq -(\ln \epsilon + 1) \sum_{i \in K} c_i \mathbb{E} \left[ \int_0^{t \wedge \tau_k} \mathbb{I}\{\|A(x(s) - x_0)\| < \epsilon\beta_i\} dy_i(s) \right] \\ &\quad - \sum_{i \in K} \hat{c}_i \mathbb{E} [y_i(t \wedge \tau_k)]. \end{aligned}$$

Dividindo a equação acima por  $-(\ln \epsilon + 1)$  e fazendo com que  $\epsilon \rightarrow 0$ , tem-se que

$$\limsup_{\epsilon \rightarrow 0} \sum_{i \in K} c_i \mathbb{E} \left[ \int_0^{t \wedge \tau_k} \mathbb{I}\{\|A(x(s) - x_0)\| < \epsilon\beta_i\} dy_i(s) \right] \equiv 0.$$

Como  $\cap_{i \in K} \partial G_i = \cap_{i \in L} \partial G_i \subset \{\xi \in G : A(\xi - x_0) \leq 0\}$ , tem-se que

$$\int_0^{t \wedge \tau_k} \mathbb{I}\{x(s) \in \cap_{i \in K} \partial G_i\} dy_i(s) = 0 \quad \mathbb{P} - q.c..$$

O resultado segue fazendo  $k \rightarrow \infty$ ,  $t \rightarrow \infty$ , e observando que  $(t \wedge \tau_k) \uparrow T$ .

(ii) O caso geral é feito através de uma indução reversa junto com um argumento de localização próximo da fronteira de interesse. Observe que é suficiente mostrar o teorema para  $K$  maximal.

Usa-se como base para esta indução a prova de (i) acima. Caso  $\{1, \dots, N\}$  não seja tal que  $\cap_{i \in \{1, \dots, N\}} \partial G_i \neq \emptyset$ , o argumento de localização do processo próximo à fronteira de interesse pode ser usado em combinação com (i) para obter a base da indução. Os detalhes não serão mostrados mas ficarão claro com os argumentos a seguir.

Suponha que para  $2 \leq j \leq N$ , o teorema é verdadeiro para todo  $K$  maximal com  $j < |K| \leq N$ . Pela hipótese, indutiva tem-se que

$$\begin{aligned} & \int_0^T \mathbb{I}\{x(s) \in \cap_{i \in K} \partial G_i\} dy_i(s) \\ &= \int_0^T \mathbb{I}\{x(s) \in \cap_{i \in K} \partial G_i\} \mathbb{I}\{\langle n_l, x(s) \rangle > B_l, \quad \forall l \in \{1, \dots, N\} \setminus K\} dy_i(s). \end{aligned}$$

Agora defina a sequência de tempos de parada  $\{\sigma_k\}_{k=0}^\infty$  e  $\{\tau_k\}_{k=1}^\infty$  para  $\epsilon > 0$

$$\tau_k = \inf\{s \geq \sigma_{k-1} : \langle n_l, x(s) \rangle - B_l \leq \epsilon/2 \text{ para algum } l \in \{1, \dots, N\} \setminus K\} \wedge T$$

$$\sigma_k = \inf\{s \geq \tau_k : \langle n_l, x(s) \rangle - B_l \geq \epsilon \text{ para todo } l \in \{1, \dots, N\} \setminus K\} \wedge T.$$

Portanto, para  $i \in K$  tem-se que

$$\begin{aligned} & \int_0^T \mathbb{I}\{x(s) \in \cap_{i \in K} \partial G_i\} \mathbb{I}\{\langle n_l, x(s) \rangle - B_l \geq \epsilon \quad \forall l \in \{1, \dots, N\} \setminus K\} dy_i(s) \\ & \leq \sum_{k=0}^\infty \int_{\sigma_k}^{\tau_{k+1}} \mathbb{I}\{x(s) \in \cap_{i \in K} \partial G_i\} dy_i(s). \end{aligned} \tag{A.11}$$

Observe que  $y_i$  não aumenta em  $[\sigma_k, \tau_{k+1}]$  para  $l \in \{1, \dots, N\} \setminus K$ . Portanto, pelo

caso (i) tem-se que

$$0 = \int_0^\tau \mathbb{I}\{x(s + \sigma_k) \in \cap_{i \in K} \partial G_i\} dy_i(s + \sigma_k) = \int_{\sigma_k}^{\tau_{k+1}} \mathbb{I}\{x(s) \in \cap_{i \in K} \partial G_i\} dy_i(s)$$

$\mathbb{P} - q.c.$  em  $\{\sigma_k < \infty\}$ , onde  $\tau \triangleq (\tau_{k+1} - \sigma_k) \mathbb{I}\{\sigma_k < \infty\}$ . Da Inequação (A.11) e da observação acima o teorema está provado.  $\square$

# Apêndice B

## Resultados Auxiliares e Definições

Neste apêndice reuni-se uma série de resultados e definições que são usados neste trabalho.

### B.1 Convergência Fraca

Seja  $S$  um espaço métrico e  $\mathcal{S}$  a  $\sigma$ -álgebra de Borel (gerada por conjuntos abertos de  $S$ ).

**Definição B.1.1.** *Seja  $\{P_n\}$  uma sequência de medidas de probabilidade definidas em  $(S, \mathcal{S})$ . Diz-se que  $\{P_n\}$  converge fracamente para a medida de probabilidade  $P$  definida em  $(S, \mathcal{S})$  se para cada função  $f : S \rightarrow \mathbb{R}$  uniformemente contínua e limitada tem-se:*

$$\int_S f dP_n \rightarrow \int_S f dP.$$

*Denota-se  $P_n \Rightarrow P$  quando  $\{P_n\}$  converge fracamente para  $P$ . Veja também o Teorema de Portmanteau para condições equivalentes (e.g., Teorema 2.1 página 16 de (Billingsley, 1999)).*

**Definição B.1.2** (Família de Medidas de Probabilidade Limitada). *Uma família  $\Pi$  de medidas de probabilidade é dita limitada se para cada  $\epsilon > 0$  existe um conjunto compacto  $K$  tal que*

$$P(K) > 1 - \epsilon$$

para cada  $P$  em  $\Pi$ .

Considere uma sequência de elementos aleatórios  $\{X_n\}$  definidos no espaço de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$ , onde, para cada  $n \in \mathbb{N}$ ,  $X_n$  é  $(\Omega, \mathcal{F})$ - $(S, \mathcal{S})$  mensurável. Seja  $P_n$  a medida de probabilidade induzida por  $X_n$  e  $\mathbb{P}$  em  $(S, \mathcal{S})$ . Seja também  $X$  um elemento aleatório  $(\Omega, \mathcal{F})$ - $(S, \mathcal{S})$  mensurável e que induz a medida de probabilidade  $P$  em  $(S, \mathcal{S})$ . Diz-se que  $X_n$  converge fracamente para  $X$  se  $P_n \Rightarrow P$ , e denota-se  $X_n \Rightarrow X$ . Ou seja, para cada função  $f : S \rightarrow \mathbb{R}$  uniformemente contínua e limitada

$$\mathbb{E}[f(X_n)] = \int_{\Omega} f(X_n) d\mathbb{P} \rightarrow \int_{\Omega} f(X) d\mathbb{P} = \mathbb{E}[f(X)].$$

**Definição B.1.3** (Sequência de Elementos Aleatórios Limitada em Probabilidade).

Seja  $\{X_n\}$  como definido no parágrafo acima. Diz-se que  $\{X_n\}$  é limitada em probabilidade se a família  $\{P_n\}$  é limitada.

O teorema abaixo é de fundamental importância para a teoria de convergência fraca. O resultado possibilita criar critérios para a verificação da existência de convergência fraca.

**Teorema B.1.4** (Teorema de Prohorov). *Se a família  $\Pi$  é limitada, então para qualquer sequência  $\{P_n\}$  de  $\Pi$  existe uma subsequência  $\{P_{n_i}\}$  que converge fracamente.*

*Demonstração.* Veja seção 5 página 60 de (Billingsley, 1999). □

Portanto, se  $\{P_n\}$  é limitada e qualquer subsequência de  $\{P_n\}$  tem limite  $P$ , tem-se que  $P_n \Rightarrow P$  (veja o corolário da página 59 de (Billingsley, 1999)).

### B.1.1 Espaço de Skorohod

Geralmente o espaço métrico  $S$  tratado na seção anterior é  $C(\mathbb{R}^d; 0, \infty)$ , aquele constituído de funções  $f : [0, \infty) \rightarrow \mathbb{R}^d$  contínuas, e com topologia uniforme.

Isto é, a medida de distância entre dois pontos  $x, y$  é dada por

$$\rho(x, y) = \sup_t |x(t) - y(t)|.$$

Também é de interesse tratar funções com descontinuidades. Neste caso  $S$  será o espaço  $D(\mathbb{R}^d; 0, \infty)$  de funções  $f : [0, \infty) \rightarrow \mathbb{R}^d$  contínuas à direita e como limites à esquerda que são usualmente chamadas de *Càdlàg* (do Francês: “*continu à droite, limites à gauche*”). A topologia de Skorohod ( $J_1$  usual) é usada neste caso (veja página 123 de (Billingsley, 1999)). Geralmente o espaço  $D(\mathbb{R}^d; 0, \infty)$  com topologia de Skorohod é chamado simplesmente de espaço de Skorohod.

### B.1.2 Critérios

Abaixo estão alguns critérios usados para mostrar que uma sequência  $\{X_n\}$  de processos estocásticos é limitada em probabilidade.

**Proposição B.1.5.** *Suponha que para cada  $n \in \mathbb{N}$ ,  $X_n$  é um processo definido em  $(\Omega, \mathcal{F}, \mathbb{P})$  com funções amostras em  $C(\mathbb{R}^d; 0, \infty)$ . A sequência  $\{X_n\}$  é limitada em probabilidade se e somente se satisfaz condições (i) e (ii) abaixo:*

(i) Para cada  $\nu > 0$

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|X_n(0)| \geq \nu) = 0$$

(ii) Para cada  $\nu > 0$  e  $T > 0$ :

$$\lim_{\delta \rightarrow 0} \limsup_n \mathbb{P} \left( \sup_{t \leq T} \sup_{s \leq \delta} |X_n(t+s) - X_n(t)| \geq \nu \right) = 0.$$

*Demonstração.* Teorema 7.3 de (Billingsley, 1999) página 82. □

**Definição B.1.6** (Assintoticamente Contínuo). *Seja  $X_n$  um processo definido em  $(\Omega, \mathcal{F}, \mathbb{P})$  com funções amostras em  $D(\mathbb{R}^d; 0, \infty)$  para cada  $n \in \mathbb{N}$ . Diz-se que a*

sequência  $\{X_n\}$  é assintoticamente contínua se é limitada em probabilidade e qualquer subsequência convergente tem como limite um processo com funções amostras contínuas com probabilidade um. Abusando um pouco da nomenclatura, diz-se que  $X_n$  é assintoticamente contínuo quando a sequência  $\{X_n\}$  é assintoticamente contínua.

**Proposição B.1.7.** *Suponha que para cada  $n \in \mathbb{N}$ ,  $X_n$  é um processo definido em  $(\Omega, \mathcal{F}, \mathbb{P})$  com funções amostras em  $D(\mathbb{R}^d; 0, \infty)$ . A sequência  $\{X_n\}$  é assintoticamente contínua se e somente se satisfaz condições (i) e (ii) abaixo:*

(i) *Para cada  $\eta > 0$  e  $T \geq 0$ , existe uma constante  $M_{\eta, T} > 0$  tal que*

$$\liminf_{n \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq t \leq T} |X_n(t)| \leq M_{\eta, T} \right) \geq 1 - \eta.$$

(ii) *Para cada  $\epsilon > 0$ ,  $\eta > 0$  e  $T > 0$ , existe uma  $\lambda \in (0, T)$  tal que*

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{t \in [0, T-\lambda]} \sup_{u, v \in [t, t+\lambda]} |X_n(u) - X_n(v)| \geq \epsilon \right) \leq \eta.$$

*Demonstração.* Veja Proposição VI.3.26 de (Jacod e Shiryaev, 1987). □

### B.1.3 Convergência Fraca para Processos de Wiener

Nesta seção são listados alguns resultados referentes a convergência fraca de processos para o processos de Wiener.

**Teorema B.1.8** (Teorema de Donsker). *Suponha que as variáveis aleatórias  $\xi_i$ ,  $i \geq 1$ , são independentes e identicamente distribuídas com média zero e variância  $\sigma^2$ . Então o processo estocástico  $w^n(\cdot)$  dado por*

$$w^n(\cdot) = \frac{1}{\sqrt{n}} \sum_{i=1}^{nt} \xi_i,$$

*converge fracamente para um processo de Wiener com matriz de covariância  $\Sigma$ .*

*Demonstração.* Veja a prova do Teorema 14.1 em (Billingsley, 1999). □

Os Teoremas B.1.10, B.1.9, e B.1.12 seguintes foram retirados de (Kushner, 2001) para facilitar a referência. Eles são respectivamente os Teoremas 2.8.6, 2.8.9, e 2.8.2 de (Kushner, 2001).

**Teorema B.1.9.** *Sejam  $\xi_i^n$ , para  $i, n \geq 1$ , vetores aleatórios mutuamente independentes (em  $i$ ) para cada  $n$  e com média zero e matriz de covariância dada por  $\Sigma_{n,i}$ . Suponha que existe uma matrix  $\Sigma$  tal que  $\Sigma_{n,i} \rightarrow \Sigma$ , quando  $n \rightarrow \infty$ , uniformemente em  $i$ . Seja  $\{|\xi_l^n|^2; l, n\}$  uniformemente integrável. Então o processo  $w^n(\cdot)$  dado por*

$$w^n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{nt} \xi_i^n$$

*converge fracamente para um processo de Wiener com matriz de covariância  $\Sigma$ .*

**Teorema B.1.10.** *Defina  $w^n(\cdot)$  como*

$$w^n(\cdot) = \frac{1}{\sqrt{n}} \sum_{l=1}^{nt} \xi_l^n.$$

*Suponha que para  $T < \infty$ , o seguintes conjuntos são uniformemente integráveis:*

$$\{|\xi_i^n|^2; n, i\},$$

$$\left\{ \left| \sum_{j=i+1}^{nT} \mathbb{E} [\xi_j^n | \xi_k^n, k \leq i] \right|^2; i < nT, n > 0 \right\}.$$

*Então  $\{w^n(\cdot)\}$  é limitado em probabilidade, e todo limite no sentido fraco tem funções amostras contínuas com probabilidade um. Suponha que existam matrizes  $\Sigma_0$  and  $\Sigma_1$  e números reais  $N_n \rightarrow \infty$  quando  $n \rightarrow \infty$  tal que  $N_n/n \rightarrow 0$  e*

$$\frac{1}{N_n} \sum_{l=i+1}^{i+N_n} \mathbb{E} [\xi_l^n (\xi_l^n)' | \xi_k, k \leq i] \rightarrow \Sigma_0$$

em probabilidade quando  $n, i \rightarrow \infty$ . Além disso, suponha que o seguinte é verdade quando  $j - a \rightarrow \infty$ ,  $m - j \rightarrow \infty$ , e  $a, n, j \rightarrow \infty$ :

$$\sum_{k=j+1}^m \mathbb{E} [\xi_k^n (\xi_j^n)'] \rightarrow \Sigma_1,$$

$$\mathbb{E} \left[ \left| \sum_{k=j+1}^m \mathbb{E} [\xi_k^n (\xi_j^n)' | \xi_i, i \leq a] - \Sigma_1 \right| \right] \rightarrow 0.$$

Então  $w^n(\cdot)$  converge fracamente para um processo de Wiener com matriz de covariância  $\Sigma = \Sigma_0 + \Sigma_1 + \Sigma_1'$ .

No seguinte teorema será necessário usar a seguinte definição:

**Definição B.1.11** (Processo de Doob-Meyer Associado). *Seja  $M$  um  $\mathcal{F}_t$ -martingale com funções amostras em  $D(\mathbb{R}^d; 0, \infty)$ . O processo de Doob-Meyer do submartingale  $M^2$  é o processo  $\langle M \rangle$   $\mathcal{F}_t$ -previsível, localmente integrável, e não-decrescente tal que  $M^2 - \langle M \rangle$  é um  $\mathcal{F}_t$ -martingale (a existência e unicidade de tal processo é garantida pelo Teorema de Doob-Meyer, veja por exemplo o Teorema 25.5 na página 493 de (Kallenberg, 2002)). Diz-se então que  $\langle M \rangle$  é o processo de Doob-Meyer associado ao martingale  $M$ .*

**Teorema B.1.12.** *Sejam  $V^n$ ,  $M^n$ , e  $\Sigma^n$  processos estocásticos, onde  $M^n$  tem funções amostras em  $D(\mathbb{R}^d; 0, \infty)$ ,  $\Sigma^n$  tem funções amostras em  $D(\mathbb{R}^{d \times d}; 0, \infty)$ , e  $V^n$  possui funções amostras em  $D(S; 0, \infty)$ , onde  $S$  é qualquer espaço métrico separável e completo. Seja  $\mathcal{F}_t^n$  a mínima  $\sigma$ -álgebra que mensura  $\{V^n(s), M^n(s), \Sigma^n(s); s \leq t\}$ . Suponha que  $M^n$  é um  $\mathcal{F}_t^n$ -martingale com processo de Doob-Meyer associado dado por  $\langle M^n \rangle(t) = \int_0^t \Sigma^n(s) ds$ , onde  $\{\Sigma^n\}$  é limitado. Além disso, suponha que a maior descontinuidade de  $M^n$  seja menor do que  $\delta_n$  onde  $\delta_n \rightarrow \infty$  quando  $n \rightarrow \infty$ . Então,  $\{M^n\}$  é limitado em probabilidade e qualquer limite no sentido fraco será contínuo.*

*Se além disso  $(V^n, M^n, \Sigma^n)$  convergir fracamente para  $(V, M, \Sigma)$ , então  $M$  é um  $\mathcal{F}_t$ -martingale com processo de Doob-Meyer associado dado por  $\langle M \rangle(t) =$*

$\int_0^t \Sigma(s)ds$ , onde  $\mathcal{F}_t$  é a mínima  $\sigma$ -álgebra gerada por  $\{M(s), \Sigma(s), V(s); s \leq t\}$ . Caso  $\Sigma$  não seja aleatório, então  $M$  é um processo de Wiener  $\mathcal{F}_t$ -mensurável com matriz de covariância dada por:  $\mathbb{E}[M(1)M(1)'] = \int_0^1 \Sigma(s)ds$ .

**Lema B.1.13.** Para cada  $n \in \mathbb{N}$ , seja  $M^n$  um  $\mathcal{F}_t^n$ -martingale com processo de Doob-Meyer associado dado por

$$\langle M \rangle(t) = \int_0^t \Sigma^n(s)ds,$$

onde  $\Sigma^n$  é um processo  $\mathcal{F}_t$ -mensurável com funções amostras em  $D(\mathbb{R}; 0, \infty)$  limitadas. Além disso, suponha que  $M^n(0) \equiv 0$  e que a maior descontinuidade de  $M^n$  seja menor do que  $\delta_n$ , onde  $\delta_n \rightarrow 0$  quando  $n \rightarrow \infty$ . Seja  $H^n$  um processo  $\mathcal{F}_t$ -adaptado com funções amostras em  $D(\mathbb{R}; 0, \infty)$  limitadas para cada  $n \in \mathbb{N}$ . Então se o processo  $G^n$  tomando valores

$$G^n(t) \triangleq \int_0^t H^n(s)\Sigma^n(s)ds$$

converge fracamente para o processo zero, então o processo tomando valores  $\int_0^t H^n(s)dM^n(s)$  converge fracamente para o processo zero.

*Demonstração.* Do Teorema B.1.12, sabe-se que  $\{M^n\}$  é assintoticamente contínua. Do Corolário 3 na página 66 e Teorema 29 na página 68 de (Protter, 1995), tem-se que

$$\begin{aligned} \mathbb{E} \left[ \left( \int_0^t H^n(s)dM^n(s) \right)^2 \right] &= \mathbb{E} \left[ \int_0^t H^n(s)d[M^n](s) \right] \\ &= \mathbb{E} \left[ \int_0^t H^n(s)\Sigma^n(s)ds \right] + \Delta_n, \end{aligned}$$

onde foi usado a definição da página 62 de (Protter, 1995) (veja também comentários na página 63). O termo  $\Delta_n$  vai para zero a medida que  $n \rightarrow \infty$ , já que  $\{M^n\}$  é assintoticamente contínua.

Chega-se ao resultado desejado usando a Inequação de Quadrática Maxi-

mal de Doob (veja (Protter, 1995) página 12) em conjunto com a inequação de Chebychev (veja página 192 de (Ash e Doléans-Dade, 2000)).  $\square$

## B.2 Processos de Salto

Nesta seção, será listado alguns resultados referentes a processos de saltos.

**Definição B.2.1** ( $\mathcal{F}_t$ -Intensidade de um Processo de Salto). *Seja  $N$  um processo de salto adaptado à uma história  $\mathcal{F}_t$ . Seja  $\lambda$  um processo não negativo e  $\mathcal{F}_t$ -mensurável tal que para todo  $t \geq 0$*

$$\int_0^t \lambda(s) ds < \infty \quad \mathbb{P} - q.c..$$

*Se para todo processo  $C$  não-negativo e  $\mathcal{F}_t$ -previsível tem-se*

$$\mathbb{E} \left[ \int_0^\infty C(s) dN(s) \right] = \mathbb{E} \left[ \int_0^\infty C(s) \lambda(s) ds \right],$$

*então é dito que  $N$  admite a  $\mathcal{F}_t$ -intensidade  $\lambda$ .*

**Teorema B.2.2.** *Se  $N$  admite a  $\mathcal{F}_t$ -intensidade  $\lambda$ , então  $N$  é  $\mathbb{P}$ -não-explosivo e*

- (i)  $M(t) = N(t) - \int_0^t \lambda(s) ds$  é um  $\mathcal{F}_t$ -martingale local;
- (ii) Se  $X$  é um processo  $\mathcal{F}_t$ -previsível tal que  $\mathbb{E} \left[ \int_0^t |X(s)| \lambda(s) ds \right] < \infty$ , para todo  $t \geq 0$ , então  $\int_0^t X(s) dM(s)$  é um  $\mathcal{F}_t$ -martingale.
- (iii) Se  $X$  é um processo  $\mathcal{F}_t$ -previsível tal que  $\int_0^t |X(s)| \lambda(s) ds < \infty$ ,  $\mathbb{P} - q.c.$ , então  $\int_0^t X(s) dM(s)$  é um  $\mathcal{F}_t$ -martingale local.

*Demonstração.* Veja Teorema T8 na página 27 de (Brémaud, 1981).  $\square$

**Lema B.2.3.** *Seja  $A$  um processo  $\mathcal{F}_t$ -adaptado definido como  $A(t) \triangleq N(\xi(t))$ , para todo  $t \geq 0$ , onde  $\xi(t) \triangleq \int_0^t \lambda(s) ds$ ,  $\lambda$  é um processo  $\mathcal{F}_t$ -mensurável e tomando valores em  $\mathbb{R}_{\geq 0}$ , e  $N$  é um processo de Poisson padrão (i.e., com taxa igual à um). Então  $\lambda$  é a  $\mathcal{F}_t$ -intensidade de  $A$ .*

*Demonstração.* A prova é uma adaptação das idéias do Teorema T16 na página 41 de (Brémaud, 1981). É suficiente mostrar que para todos os processos  $C$  não-negativos e  $\mathcal{F}_t$ -previsíveis que podem ser escritos como

$$C(\omega; t) = \mathbb{I}\{\omega \in B\}\mathbb{I}\{t \in (a, b]\},$$

onde  $0 \leq a \leq b \leq \infty$ ,  $B \in \mathcal{F}_a$ , e  $(\omega, t) \in \Omega \times [0, \infty)$ , tem-se

$$\mathbb{E} \left[ \int_0^\infty C(s) dA(s) \right] = \mathbb{E} \left[ \int_0^\infty C(s) \lambda(s) ds \right].$$

Para simplificar a notação,  $\omega$  é omitido e  $C$  é escrito como

$$C(t) = \mathbb{I}_B \mathbb{I}\{t \in (a, b]\},$$

para todo  $t \geq 0$ .

Defina o tempo de parada  $\mathcal{F}_t$ -mensurável  $\tau(t)$  como

$$\int_0^{\tau(t)} \lambda(s) ds = t \wedge \left| \int_0^\infty \lambda(s) ds \right|.$$

Observe que

$$\begin{aligned} \mathbb{E} \left[ \int_0^\infty C(s) dA(s) \right] &= \mathbb{E} [\mathbb{I}_B (A(b) - A(a))] \\ &= \mathbb{E} [\mathbb{I}_B (N(\xi(b)) - N(\xi(a)))] \\ &= \mathbb{E} \left[ \int_0^\infty \mathbb{I}_B \mathbb{I}\{s \in (\xi(a), \xi(b)]\} dN(s) \right]. \end{aligned}$$

Seja  $\tilde{C}(t) \triangleq \mathbb{I}_B \mathbb{I}\{t \in (\xi(a), \xi(b)]\}$  para todo  $t \geq 0$ . Se for estabelecido que  $\tilde{C}$  é  $\mathcal{F}_{\tau(t)}$ -previsível e que  $N(\cdot \wedge \xi(\infty))$  é  $\mathcal{F}_{\tau(t)}$ -adaptado tem-se

$$\begin{aligned} \mathbb{E} \left[ \int_0^\infty \tilde{C}(s) dN(s) \right] &= \mathbb{E} [\mathbb{I}_B (\xi(b) - \xi(a))] = \mathbb{E} \left[ \int_0^\infty \mathbb{I}_B \mathbb{I}\{s \in (a, b]\} \lambda(s) ds \right] \\ &= \mathbb{E} \left[ \int_0^\infty C(s) \lambda(s) ds \right], \end{aligned}$$

onde foi usado a definição de  $\mathcal{F}_t$ -intensidade para o processo de Poisson padrão  $N$ .

Seja  $\mathcal{F}_t^A$  é a mínima  $\sigma$ -álgebra que mensura  $\{A(s); s \leq t\}$ , logo o processo  $A(\tau(t)) = N(t \wedge |\xi(\infty)|)$  é adaptado à  $\mathcal{F}_{\tau(t)}^A$ . Como  $\mathcal{F}_{\tau(t)}^A \subseteq \mathcal{F}_{\tau(t)}$ , tem-se que  $N(t \wedge |\xi(\infty)|)$  é  $\mathcal{F}_{\tau(t)}$ -adaptado.

Para verificar que  $\tilde{C}$  é  $\mathcal{F}_{\tau(t)}$ -previsível, somente é necessário verificar se o processo é  $\mathcal{F}_{\tau(t)}$ -mensurável, já que  $\tilde{C}$  é contínuo à esquerda. Observe que

$$\begin{aligned} \{t \leq \xi(b)\} &\equiv \{b \geq \tau(t)\} \\ \{t \geq \xi(a)\} &\equiv \{a \leq \tau(t)\}. \end{aligned}$$

Portanto,  $\{t \in (\xi(a), \xi(b))\} \equiv \{a \leq \tau(t) \leq b\}$ , e então

$$\tilde{C}(t) = \mathbb{I}_B \mathbb{I}\{t \in (\xi(a), \xi(b))\} = \mathbb{I}_B \mathbb{I}\{a \leq \tau(t) \leq b\}$$

é  $\mathcal{F}_{\tau(t)}$ -mensurável já que  $B \in \mathcal{F}_a$ . □

**Lema B.2.4.** *Seja  $N$  um processo de salto  $\mathcal{F}_t$ -mensurável com  $\mathcal{F}_t$ -intensidade dada por  $\lambda$ . Suponha que o tamanho de qualquer salto de  $N$  seja igual à um, quase certamente. Seja  $M$  o martingale garantido pelo Teorema B.2.2. Então, o processo de Doob-Meyer associado à  $M$  é dado por*

$$\langle M \rangle(t) = \int_0^t \lambda(s) ds.$$

*Demonstração.* Veja a discussão na página 62 de (Kushner, 2001). □

**Lema B.2.5.** *Seja  $N$  um processo de salto  $\mathcal{F}_t$ -mensurável com  $\mathcal{F}_t$ -intensidade dada por  $\lambda$ . Seja  $I_i$ , com  $i \in \{1, \dots, q\}$  e  $q \in \mathbb{N}$ , um processo estocástico  $\mathcal{F}_t$ -mensurável, onde para cada  $t \geq 0$ ,  $I_i(t)$  é uma função indicadora de eventos disjuntos para cada  $i \in \{1, \dots, q\}$ . Além disso, suponha que  $\sum_{i=1}^q I_i(t) \leq 1$ , se houver um salto em  $N$  no instante  $t$ , e igual à zero caso contrário. Defina*

$$Q_i(t) = \mathbb{E}[I_i(t) | \mathcal{F}_{t-}].$$

Seja  $F$  um processo  $\mathcal{F}_t$ -previsível tomando valores em  $\mathbb{R}^q$  e defina o  $\mathcal{F}_t$ -martingale  $M = (M_1, \dots, M_q)'$  como:

$$M_i(t) = \int_0^t [I_i(s) - Q_i(s)] F_i(s) dN(s).$$

Então o processo de Doob-Meyer associado à  $M$  é dado por

$$\langle M \rangle(\cdot) = \int_0^\cdot \Sigma(s) F(s) F(s)' \lambda(s) ds,$$

onde

$$\langle M \rangle_{ij} \triangleq \begin{cases} \langle M_i \rangle & \text{se } i = j \\ \langle M_i, M_j \rangle & \text{se } i \neq j, \end{cases}$$

e  $(\Sigma)_{ij}(\cdot) \triangleq \delta_{ij} Q_i(\cdot) - Q_i(\cdot) Q_j(\cdot)$ , para  $\delta_{ij} \triangleq \mathbb{I}\{i = j\}$ , e  $(i, j) \in \{1, \dots, q\}^2$ .

*Demonstração.* Da identidade de polarização (veja (Kallenberg, 2002) página 516), tem-se:

$$\langle M_i, M_j \rangle = \frac{1}{4} [\langle M_i + M_j \rangle - \langle M_i - M_j \rangle].$$

Portanto, será considerado primeiramente o compto de  $\langle M_i + M_j \rangle$  o qual será denotado por  $m_{ij}^+$  para simplificar a notação. Pela definição de processos de Doob-Meyer associado, tem-se que para  $t$  e  $\tau \geq 0$ :

$$\mathbb{E} [[M_i(t + \tau) + M_j(t + \tau)]^2 - m_{ij}^+(t + \tau) | \mathcal{F}_t] = [M_i(t) + M_j(t)]^2 - m_{ij}^+(t).$$

Seja  $\Delta(f) \triangleq f(t + \tau) - f(t)$ , para qualquer função  $f$  definida em  $[0, \infty)$ . Observe

que:

$$\begin{aligned}
& \mathbb{E} \left[ [M_i(t + \tau) + M_j(t + \tau)]^2 - [M_i(t) + M_j(t)]^2 \mid \mathcal{F}_t \right] \\
&= \mathbb{E} \left[ [\Delta(M_i + M_j) + 2(M_i + M_j)(t)] \Delta(M_i + M_j) \mid \mathcal{F}_t \right] \\
&= \mathbb{E} \left[ \Delta(M_i + M_j)^2 \mid \mathcal{F}_t \right] + 2(M_i + M_j)(t) \mathbb{E} \left[ \Delta(M_i + M_j) \mid \mathcal{F}_t \right] \\
&= \mathbb{E} \left[ \Delta(M_i + M_j)^2 \mid \mathcal{F}_t \right],
\end{aligned}$$

onde foi usado o fato que  $M_j$  e  $M_i$  são martingales na última passagem. Desta forma,

$$\mathbb{E} \left[ m_{ij}^+(t + \tau) - m_{ij}^+(t) \mid \mathcal{F}_t \right] = \mathbb{E} \left[ \Delta(M_i + M_j)^2 \mid \mathcal{F}_t \right]. \quad (\text{B.1})$$

Observe que é possível escrever o seguinte:

$$\begin{aligned}
& \mathbb{E} \left[ \Delta(M_i + M_j)^2 \mid \mathcal{F}_t \right] = \\
& \mathbb{E} \left[ \left( \sum_{l=1}^{\infty} [(I_i(s_l) - Q_i(s_l)) F_i(s_l) + (I_j(s_l) - Q_j(s_l)) F_j(s_l)] \mathbb{I}\{s_l \leq t + \tau\} \right)^2 \mid \mathcal{F}_t \right],
\end{aligned}$$

onde  $s_l$  denota o instante do  $l$ -ésimo salto de  $N$  após ou em  $t$ . Como  $s_m$  e  $s_n$  são diferentes quando  $m \neq n$ , e usando o fato de que  $\mathbb{E} [I_i(t) \mid \mathcal{F}_{t-}] = Q_i(t)$ , tem-se

$$\mathbb{E} [(I_i(s_m) - Q_i(s_m))(I_j(s_n) - Q_j(s_n)) F_i(s_m) F_j(s_n) \mathbb{I}\{s_n, s_m \leq t + \tau\} \mid \mathcal{F}_t] = 0,$$

onde foi usada a “regra da torre” (i.e.,  $\mathbb{E} [\dots \mid \mathcal{F}_t] = \mathbb{E} [\mathbb{E} [\dots \mid \mathcal{F}_{s_{h-}}] \mid \mathcal{F}_t]$ , para  $h = \max(m, n)$ ). Seja

$$a_r(\cdot) \triangleq \int_0^\cdot (I_r(s) - Q_r(s-))^2 F_r(s) dN(s),$$

para  $r = i, j$ . Portanto,

$$\begin{aligned}
\mathbb{E} [\Delta(M_i + M_j)^2 | \mathcal{F}_t] &= \mathbb{E} [\Delta(a_i) - \Delta(a_j) | \mathcal{F}_t] \\
&+ 2 \sum_{l=1}^{\infty} \mathbb{E} [(I_i(s_l) - Q_i(s_l))(I_j(s_l) - Q_j(s_l))F_i(s_l)F_j(s_l)\mathbb{I}\{s_l \leq t + \tau\} | \mathcal{F}_t] \\
&= \mathbb{E} [\Delta(a_i) + \Delta(a_j) | \mathcal{F}_t] \\
&+ 2 \sum_{l=1}^{\infty} \mathbb{E} [(\delta_{ij}Q_i(s_l) - Q_i(s_l)Q_j(s_l)) F_i(s_l)F_j(s_l)\mathbb{I}\{s_l \leq t + \tau\} | \mathcal{F}_t],
\end{aligned}$$

onde foi usado a regra da torre novamente. Usando a definição de intensidade estocástica (Definição B.2.1), tem-se

$$\begin{aligned}
&\mathbb{E} \left[ \int_t^{t+\tau} (\delta_{ij}Q_i(s) - Q_i(s)Q_j(s)) F_i(s-)F_j(s-)dN(s) \middle| \mathcal{F}_t \right] \\
&= \mathbb{E} \left[ \int_t^{t+\tau} (\delta_{ij}Q_i(s) - Q_i(s)Q_j(s)) F_i(s)F_j(s)\lambda(s)ds \middle| \mathcal{F}_t \right],
\end{aligned}$$

e para  $r = i, j$ ,

$$\mathbb{E} [\Delta(a_r) | \mathcal{F}_t] = \mathbb{E} \left[ \int_t^{t+\tau} (I_r(s) - Q_r(s))^2 F_r(s)\lambda(s)ds \middle| \mathcal{F}_t \right].$$

Defina  $\tilde{a}_r(\cdot) \triangleq \int_0^\cdot (I_r(s) - Q_r(s))^2 F_r(s)\lambda(s)ds$ . Usando a Equação (B.1) e como o processo de Doob-Meyer é único (indistinguível), tem-se

$$m_{ij}^+(t) = \tilde{a}_i(t) + \tilde{a}_j(t) + 2 \int_0^t (\delta_{ij}Q_i(s) - Q_i(s)Q_j(s)) F_i(s)F_j(s)\lambda(s)ds.$$

Usando o mesmo argumento para  $\langle M_i - M_j \rangle \triangleq m_{ij}^-$ , tem-se

$$m_{ij}^-(t) = \tilde{a}_i(t) + \tilde{a}_j(t) - 2 \int_0^t (\delta_{ij}Q_i(s) - Q_i(s)Q_j(s)) F_i(s)F_j(s)\lambda(s)ds.$$

Portanto,

$$\langle M_i, M_j \rangle(t) = \frac{1}{4} [m_{ij}^+(t) - m_{ij}^-(t)] = \int_0^t (\delta_{ij}Q_i(s) - Q_i(s)Q_j(s)) F_i(s)F_j(s)\lambda(s)ds.$$

□

**Teorema B.2.6.** *Considere  $\{\xi_l^n, l < \infty\}$ , onde para cada  $l$  e  $n$ ,  $\xi_l^n$  são variáveis aleatórias que tomam valores positivos com probabilidade um. Seja  $\{h^n\}$  limitada em probabilidade, onde para cada  $n$ ,  $h^n$  é um processo definido como:*

$$h^n(t) \triangleq \frac{1}{\sqrt{n}} \sum_{l=1}^{nt} (\xi_l^n - \bar{\xi}^n).$$

Além disso suponha que, para cada  $l$ ,  $\mathbb{E}[\xi_l^n] \triangleq \bar{\xi}^n \rightarrow \bar{\xi} \in \mathbb{R}_{>0}$  quando  $n \rightarrow \infty$ . Seja  $\mathcal{J}^n$  um processo não decrescente tal que  $\mathcal{J}^n(0) = 0$  e para cada  $t \in \mathbb{R}_{\geq 0}$ ,  $\mathcal{J}^n(t) \leq nt$ . Então  $\{N^n\}$ , com

$$N^n(t) \triangleq \frac{1}{n} \max \left\{ m \in \mathbb{N}_0 : \sum_{l=1}^m \xi_l^n \leq nt - \mathcal{J}^n(t) \right\},$$

é limitada em probabilidade e qualquer limite no sentido fraco tem funções amostras Lipschitz-contínuas, com constante de Lipschitz inferior à  $1/\bar{\xi}$ . Se além disso  $\{\mathcal{J}^n/\sqrt{n}\}$  for limitada em probabilidade, o processo  $N^n$  converge no sentido fraco para o processo  $N$  que toma valores  $N(t) \triangleq t/\bar{\xi}$ .

*Demonstração.* Esta prova é uma modificação de um resultado contido na prova do Teorema 5.1.1 de (Kushner, 2001). Seja

$$\mathcal{T}^n(t) \triangleq \frac{1}{n} \sum_{l=1}^{nt} \xi_l^n$$

Observe que

$$N^n(\mathcal{T}^n(t)) = \frac{1}{n} \max \left\{ m \in \mathbb{N}_0 : \sum_{l=1}^m \xi_l^n \leq \sum_{l=1}^{nt} \xi_l^n - \mathcal{J}^n(\mathcal{T}^n(t)) \right\} \leq t + \epsilon_{n,1} \quad (\text{B.2})$$

onde  $|\epsilon_{n,1}| < 1/n$ . Além disso,

$$\mathcal{T}^n(N^n(t)) = \frac{1}{n} \sum_{l=1}^{\max\{m: \sum_{l=1}^m \xi_l^n \leq nt - \mathcal{J}^n(t)\}} \xi_l^n = t - \frac{\mathcal{J}^n(t)}{n} + \epsilon_{n,2} \quad (\text{B.3})$$

onde  $|\epsilon_{n,2}| < (1/n)\xi_{N^n(t)+1}^n$  é um erro negligenciável e que converge para o processo “zero” quando  $n \rightarrow \infty$  já que  $\{h^n\}$  é limitada em probabilidade.

O processo  $\mathcal{T}^n(t)$  pode ser reescrito como:

$$\mathcal{T}^n(t) \triangleq \frac{1}{n} \sum_{l=1}^{nt} (\xi_l^n - \bar{\xi}^n) + \bar{\xi}^n t.$$

Já que  $\{h^n\}$  é limitada em probabilidade, o primeiro termo converge no sentido fraco para o processo zero, e  $\mathcal{T}^n(\cdot) \Rightarrow \mathcal{T}(\cdot)$  onde  $\mathcal{T}(t) \triangleq \bar{\xi} t$ , pela hipótese  $\bar{\xi}^n \rightarrow \bar{\xi}$ . Portanto, para quaisquer  $\epsilon > 0$  e  $t > 0$ :

$$\lim_n \mathbb{P} \left( \sup_{s \leq t} N^n(s) < \frac{t}{\bar{\xi}} + \epsilon \right) = 1 \quad (\text{B.4})$$

pela Equação (B.2).

Usando a Equação (B.3), observe que para qualquer  $\tau > 0$

$$\mathcal{T}^n(N^n(t + \tau)) - \mathcal{T}^n(N^n(t)) = \tau - \frac{\mathcal{J}^n(t + \tau) - \mathcal{J}^n(t)}{n} + \tilde{\epsilon}_{n,2} \leq \tau + \tilde{\epsilon}_{n,2}$$

onde a ultima passagem é possível já que  $\mathcal{J}^n(\cdot)$  é não-negativo e não-decrescente. Portanto, para quaisquer constantes  $\epsilon > 0$ ,  $\tau > 0$  e  $t \in \mathbb{R}_{\geq 0}$

$$\lim_n \mathbb{P} (|N^n(t + \tau) - N^n(t)| < \tau/\bar{\xi} + \epsilon) = 1. \quad (\text{B.5})$$

Observe que usando as Equações (B.4) e (B.5) as condições da Proposição B.1.7 são satisfeitas e, portanto,  $\{N^n(\cdot)\}$  satisfaz a primeira parte do teorema.

Assuma agora que  $\{\mathcal{J}^n/\sqrt{n}\}$  é limitada em probabilidade. Da Equação (B.3) e deixando que  $n \rightarrow \infty$  tem-se  $\mathcal{T}(N(t)) = t$ , e portanto,

$$N(t) = t/\bar{\xi},$$

já que  $\{\mathcal{J}^n/n\}$  converge no sentido fraco para o processo “zero”. □

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)