

# Analysis of Dependence Metrics for Queueing Processes

Abdelghafour Es-Saghouani

Analysis of Dependence Metrics for Queueing Processes

Abdelghafour Es-Saghouani

# **Analysis of Dependence Metrics for Queueing Processes**

Analysis of Dependence Metrics for Queueing Processes  
Abdelghafour Es-Saghouani  
Proefschrift Universiteit van Amsterdam  
Met literatuur opgave en samenvatting in het Nederlands  
ISBN: 978-90-9024-719-9

Dit proefschrift werd mogelijk gemaakt door



Nederlandse Organisatie voor Wetenschappelijk Onderzoek

Dit proefschrift werd medemogelijk gemaakt door

THOMAS STELTJES INSTITUTE  
FOR MATHEMATICS



# Analysis of Dependence Metrics for Queueing Processes

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor  
aan de Universiteit van Amsterdam  
op gezag van de Rector Magnificus  
prof. dr. D.C. van den Boom  
ten overstaan van een door het college voor promoties ingestelde  
commissie, in het openbaar te verdedigen in de Agnietenkapel  
op dinsdag 17 november 2009, te 12:00 uur

door

**Abdelghafour Es-Saghouani**

geboren te Tazourakht, Marokko

## **Promotiecommissie**

Promotor: prof. dr. M.R.H. Mandjes

Overige leden: prof. dr. O.J. Boxma  
prof. dr. N.M. van Dijk  
prof. dr. C.A.J. Klaassen  
dr. K.G. Dębicki  
dr. R. Núñez Queija  
dr. ir. W.R.W. Scheinhardt

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

---

## Voorwoord

Het proefschrift waarvan u nu het voorwoord leest, is het resultaat van het promotieonderzoek dat ik de afgelopen vier jaar heb uitgevoerd aan het Korteweg-de Vries Instituut onder begeleiding van mijn promotor Michel Mandjes. Michel, ik ben je zeer erkentelijk voor de heel prettige samenwerking en het vertrouwen dat je altijd in mij hebt gesteld. Welke vraag ik ook had, jouw deur stond altijd open. Zonder jou had dit proefschrift er niet gelegen. Mijn meest oprechte dank hiervoor. Aan mijn bezoek aan jou in Stanford koester ik nog steeds mooie herinneringen. Ik wil jou en Miranda maar ook de kleine Chloe bedanken voor jullie gastvrijheid en hartelijkheid.

Naast Michel wil ik ook voor Peter Spreij mijn waardering uitspreken. Elke keer als ik met een vraag zat, nam hij de tijd om met mij hierover te discussiëren, wat vaak eindigde met een verwijzing naar een boek dat ik dan van hem mocht lenen. Heel veel dank hiervoor. Daarnaast wil ik graag alle commissieleden bedanken voor hun bereidheid om mijn dissertatie te beoordelen en zitting te nemen in mijn promotiecommissie.

Chapters 3 and 4 of this thesis are the result of joint work with Krzys Dębicki from Wrocław University. I would like to express my gratitude to Krzys and thank him for the opportunity he gave me to visit him for a week to finish our first joint paper (Chapter 3). By the end of my stay we had a discussion on another topic which resulted in a second paper (Chapter 5). I also would like to thank his wife Asia for her kindness and hospitality.

Verder wil ik graag alle mensen van het KdV-instituut bedanken voor de prettige sfeer die er altijd heerst. In het bijzonder wil ik mijn kamergenoten Said en Sjors bedanken voor de mooie tijden in het Euclidesgebouw, maar ook andere collega's Benjamin, Enno, Frank, Jevgenijs, Kamil, Michel, Pascal, Phyllis en Ramon. Mijn oprechte dank gaat ook uit naar het secretariaat van het KdV in de personen van Evelien en Hanneke. René, bedankt voor alles.

Verder wil ik nog mijn dank uitspreken aan mijn vriend Hicham die mij in mijn

eerste jaren in Nederland zowel binnen als buiten de UvA de weg wees.

Tenslotte wil ik mijn broers en zussen Naziha, Abdelmalek, Nadia, Najib, Amal, Samir en Amine bedanken voor hun steun, vooral sinds ik in Nederland ben. Mes plus sincères remerciements vont à mon grand frère Nourddin, je te suis reconnaissant pour tout ce que tu as signifié dans ma vie. Als laatste, maar niet in het minst, wil ik mijn meest oprechte dank betuigen aan mijn ouders, mijn vrouw Hayat en mijn dochter Manal. Ik ben jullie zeer erkentelijk voor jullie onvoorwaardelijke liefde en onmisbare steun en vertrouwen gedurende de afgelopen jaren. Daarom draag ik het proefschrift aan jullie op.

إِلَى أَبِي، أُمِّي، زَوْجَتِي حَيَاةَ وَابْنَتِي مَنَال  
عَبْدُالْعَفُورِ السَّغْوَانِي

Amsterdam, september 2009

Abdelghafour Es-Saghouani

---

# Contents

<b>Voorwoord</b>	<b>i</b>
<b>Contents</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Workload process . . . . .	3
1.2 Dependence metrics . . . . .	4
1.3 Gaussian processes . . . . .	6
1.4 Lévy processes . . . . .	7
1.5 Markov modulated fluid input processes . . . . .	9
1.6 Overview and contributions . . . . .	10
<b>2 Gaussian queue under many-sources scaling</b>	<b>13</b>
2.1 Introduction . . . . .	13
2.2 Preliminaries . . . . .	16
2.3 Main results . . . . .	20
2.4 Proofs . . . . .	23
2.5 Discussion and concluding remarks . . . . .	31
Appendix	
2.A Proof of an auxiliary result . . . . .	34
<b>3 Gaussian queue under large buffer scaling</b>	<b>37</b>
3.1 Introduction . . . . .	37
3.2 Model and problem description . . . . .	38
3.3 Notation and preliminaries . . . . .	40
3.4 General results . . . . .	42
3.5 Special cases . . . . .	51
3.6 Discussion and concluding remarks . . . . .	61

<b>4</b>	<b>Correlation structure of Lévy-driven queues</b>	<b>63</b>
4.1	Introduction . . . . .	63
4.2	Laplace transform of the correlation function . . . . .	64
4.3	Structural properties of the correlation function . . . . .	67
4.4	Correlation asymptotics for light-tailed input . . . . .	70
4.5	Correlation asymptotics for heavy-tailed input . . . . .	73
4.6	Concluding remarks . . . . .	75
	Appendix	
4.A	Complete monotonicity . . . . .	76
<b>5</b>	<b>Transient asymptotics of Lévy-driven queues</b>	<b>79</b>
5.1	Introduction . . . . .	79
5.2	Notation and preliminaries . . . . .	81
5.3	General results . . . . .	82
5.4	Heavy-tailed Lévy input . . . . .	87
5.5	Light-tailed input . . . . .	94
5.6	Discussion and concluding remarks . . . . .	103
<b>6</b>	<b>Markov fluid-driven queues</b>	<b>105</b>
6.1	Introduction . . . . .	105
6.2	Model and preliminaries . . . . .	107
6.3	Analysis of the busy period . . . . .	109
6.4	Busy period asymptotics . . . . .	112
6.5	The covariance function of the workload process . . . . .	115
6.6	Example . . . . .	119
6.7	Concluding remarks . . . . .	123
<b>7</b>	<b>Exact multivariate workload asymptotics</b>	<b>125</b>
7.1	Introduction . . . . .	125
7.2	Model, objective, and preliminaries . . . . .	126
7.3	Exact workload asymptotics . . . . .	129
	<b>Bibliography</b>	<b>133</b>
	<b>Samenvatting</b>	<b>141</b>
	<b>About the author</b>	<b>143</b>

## Chapter 1

---

### Introduction

Having applications in everyday situations that we face, queueing theory is an appealing part of applied probability. Queueing theory started to attract interest from mathematicians at the beginning of the last century, when a paper of Erlang [49] on the theory of telephone traffic appeared. In this pioneering work, Erlang showed that the number of telephone calls made during a given interval of time, assuming random origination of the calls, follows a Poisson distribution. In a second paper, which is generally considered as the more important one, Erlang [50] studied blocking probabilities when a fixed number of telephone lines are available, and proved what has become the famous Erlang loss formula.

Nowadays, queueing theory has turned into an indispensable tool for modeling various real life phenomena. For instance, it is used in the study of production and storage systems, where the objective is to regulate the demand so as to achieve a storage of desirable level, but also in communication networks, e.g., telephone networks, Internet and computer systems where one aims at guaranteeing a given level of service. In other application domains models are used that are closely related to queueing systems, for instance in insurance risk. There one of the objectives is the study of the ruin probability of an insurance company, given some initial capital.

A description of a simple queueing system is as follows. Suppose that we have a service station that processes work fed into the system. If traffic arrives at a faster rate than it can be served, then the unfinished work can be stored in a buffer, which, for ease, is assumed to have infinite storage capacity. Mathematically such a system can be modeled as follows. Let  $X \equiv \{X(t) : t \in \mathbb{R}\}$  be the stochastic process describing the way the input arrives to the service station, and let  $A(s, t) = X(t) - X(s)$  denote the amount of traffic entering between time epochs  $s$  and  $t$  ( $s \leq t$ ). Furthermore, the system is emptied at a constant rate  $c > 0$ , meaning that  $ct$  units of work can be potentially processed in any time window of length  $t$ . This naturally defines a workload process, denoted henceforth by  $Q \equiv \{Q(t) : t \in \mathbb{R}\}$ , representing the amount of work in storage at time  $t$ . In the literature, the process  $Q$  is also known as the buffer content process or the virtual waiting time process. Throughout this thesis we will use these terms interchangeably.

Obviously, the most interesting situation is the one where the state of the system alternates between busy periods and idle periods, rather than being essentially

continuously busy. To make sure that the system alternates between busy and idle periods, we should require that the average rate at which the work arrives at the system is less than the rate at which the system is emptied. Assuming that the input process  $X$  has stationary increments,  $A(s, t) := X(t) - X(s)$  for  $s \leq t$ , by which we mean that  $A(s, t)$  and  $A(s+u, t+u)$  have the same probability distribution for any  $u$ , then it can easily be shown that the mean of the increments  $\mathbb{E}A(s, t)$  is linear in  $t - s$ . In this case we can write  $\mathbb{E}A(s, t) = \varpi \cdot (t - s)$ , with  $\varpi = \mathbb{E}A(0, 1)$ . Then the criterion just described above reduces to requiring that

$$\varpi < c. \tag{1.1}$$

If Condition (1.1) is satisfied, then the queueing system is said to be *stable* and we will refer to (1.1) as the stability condition. Indeed, if  $\varpi > c$ , then the content of the queue will eventually grow beyond any bound.

In this monograph the main stochastic process that we will be interested in, is the content process  $\{Q(t) : t \in \mathbb{R}\}$ . In Section 1.1, we will see that the workload  $Q(t)$  of the queue at time  $t$  can be formulated as a functional of the *net input process*  $\{A(s, t) - c(t - s) : s \leq t\}$ . Assuming stationarity of the increments  $A(s, t)$  of the input process, and imposing condition (1.1), it can be shown that the distributions of  $Q(t)$  have a weak limit as  $t \rightarrow \infty$ , and we let  $Q_e$  denote the stationary (or steady-state, or equilibrium) workload. Most of the research concerning the process  $Q$  has dealt with the stochastic properties of the steady-state workload  $Q_e$ . For special input processes explicit results are available for the steady-state distribution  $\mathbb{P}(Q_e \leq q)$ , or sometimes for the Laplace-Stieltjes transform  $\mathbb{E}e^{-sQ_e}$ ,  $s \geq 0$ , of the stationary workload  $Q_e$ . In contrast, transient behavior of the system (i.e., the time-dependent case where one studies  $\mathbb{P}(Q(t) \leq q)$  given the state of the system at time 0) is considerably less explored, and only in relatively few situations one has succeeded in deriving explicit results.

The main objective of this thesis, as its title indicates, is to investigate the dependence structure of the workload process  $Q$ . Specifically, we will investigate to what extent the dependence structure of the input process is inherited by the workload process. Throughout this thesis we will assume that the system is already in stationarity at time  $t = 0$ . Under this assumption we will study two types of metrics capturing the dependence structure of the workload process  $Q$ .

- The covariance function  $R(t)$ , to be defined in (1.9), is a measure that gives insight in the speed at which the correlation between  $Q(0)$  and  $Q(t)$  for  $t \geq 0$  vanishes as  $t$  grows large.
- A second measure that will be extensively studied, is the measure  $R(T|p, q)$ , to be defined in (1.13), featuring the joint probability of the workload exceeding some thresholds  $p$  at time 0 and  $q$  at time  $T$ .

The remainder of this introductory chapter is organized as follows. Section 1.1 describes the workload process in both discrete-time and continuous-time queueing systems. In Section 1.2, we recall some basic properties of the covariance function and give some motivations for its study. Then, in the same section we will introduce and motivate the choice of the alternative dependence metric  $\mathbb{R}(T|p, q)$ . Sections 1.3, 1.4 and 1.5 present basic background results on the stochastic processes considered as input processes, namely Gaussian, Lévy and Markov modulated fluid processes, respectively. The overview and contributions of the thesis are given in Section 1.6.

## 1.1 Workload process

For ease of exposition, let us begin by considering a *discrete-time* queueing system. Suppose that at discrete time epochs  $t = 1, 2, \dots$ , the amount of traffic entering the system at time  $t$  is given by the random quantity  $Y(t)$ . Define the stochastic process  $\{X(t) : t = 0, 1, \dots\}$  as follows:  $X(0) = 0$  and for  $t \geq 1$ ,  $X(t) = Y(1) + Y(2) + \dots + Y(t)$ . Thus  $X(t)$  is the amount of traffic arrived to the system in time epochs 1 up to  $t$ . Let the queue be emptied at a constant rate  $c > 0$ , and assume in addition that the queue was empty at time epoch  $t = 0$ . The workload process  $\{Q(t) : t \geq 0\}$  induced by the process  $\{Y(t) : t \geq 1\}$  is defined by  $Q(0) = 0$ , and for  $t \geq 0$

$$Q(t+1) = \begin{cases} Q(t) + Y(t+1) - c & \text{if } Q(t) + Y(t+1) - c \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (1.2)$$

In other words: the process  $\{Q(t) : t \geq 0\}$  is given by the following recurrence relation, which is known as the Lindley recursion:

$$Q(0) = 0, \quad Q(t+1) = \max\{Q(t) + Y(t+1) - c, 0\}, \quad t \geq 0. \quad (1.3)$$

Solving the recurrence relation (1.3) shows that

$$Q(t) = \max_{s=0,1,\dots,t} (X(t) - X(s) - c(t-s)) = \max_{s=0,1,\dots,t} (A(s,t) - c(t-s)).$$

Since we assumed that the increments  $A(s, t)$  of the process  $X$  are stationary and that the stability condition (1.1) is satisfied, it can be shown that

$$Q_e \stackrel{d}{=} \sup_{s \in \mathbb{N}_0} (A(-s, 0) - cs), \quad (1.4)$$

where  $\stackrel{d}{=}$  denotes equality in distribution, and  $A(-s, 0)$  is to be interpreted as the amount of traffic entering the system in time epochs  $-s$  up to  $-1$ . This result is often attributed to Reich [100].

Now consider the *continuous-time* case. Suppose that the input of a queueing system is a stochastic process  $\{X(t) : t \in \mathbb{R}\}$  with stationary increments  $A(s, t)$ . As before, we assume that the queue is emptied at a positive deterministic rate  $c > 0$ . With  $Q(0) \geq 0$ , the continuous-time analogue of (1.3) is obtained by reflecting  $\{Q(0) + A(s, t) - c(t - s) : s \leq t\}$  at zero. More specifically, for  $t \geq 0$ , the workload  $Q(t)$  at time  $t$  is given by

$$Q(t) = A(0, t) - ct + \max\left(Q(0), -\inf_{0 \leq s \leq t} (A(0, s) - cs)\right), \quad (1.5)$$

which, using straightforward calculus, can be rewritten as

$$Q(t) = \max\left(\sup_{0 \leq s \leq t} (A(s, t) - c(t - s)), Q(0) + A(0, t) - ct\right). \quad (1.6)$$

Under the stability condition (1.1), it can be shown, see for instance [69], that there exists a stationary stochastic process defined by

$$Q(t) = \sup_{-\infty < s \leq t} (A(s, t) - c(t - s)), \quad -\infty < t < \infty, \quad (1.7)$$

satisfying (1.6). The steady-state workload is then given by

$$Q_e \stackrel{d}{=} \sup_{s \geq 0} (A(-s, 0) - cs), \quad (1.8)$$

where  $A(-s, 0)$  is to be interpreted as the amount of traffic generated during the time interval  $[-s, 0)$ . Observe that (1.8) is the continuous counterpart of (1.4).

## 1.2 Dependence metrics

Consider the stationary workload process  $\{Q(t) : t \geq 0\}$  defined by (1.7). By stationarity, its mean  $\mathbb{E}Q(t)$  and variance  $\text{Var}Q(t)$  are constant in  $t$ , and will be denoted henceforth by  $\mu$  and  $v$  respectively. Moreover, by stationarity  $\mathbb{E}Q(s)Q(t)$  depends only on the difference  $t - s$ . Furthermore, assuming the variance  $v$  is finite, it can be shown, using the Cauchy-Schwarz inequality, that also for every  $s$  and  $t$ ,  $\mathbb{E}Q(s)Q(t)$  is finite. Therefore, the covariance function  $R(\cdot)$  and the correlation function  $r(\cdot)$  are well-defined, and are given by, respectively,

$$R(t) := \text{Cov}(Q(0), Q(t)) = \mathbb{E}Q(0)Q(t) - \mathbb{E}Q(0)\mathbb{E}Q(t) = \mathbb{E}Q(0)Q(t) - \mu^2 \quad (1.9)$$

$$r(t) := \text{Corr}(Q(0), Q(t)) = \frac{\text{Cov}(Q(0), Q(t))}{\sqrt{\text{Var}Q(0)}\sqrt{\text{Var}Q(t)}} = \frac{R(t)}{v} \quad (1.10)$$

There are several reasons for studying the covariance function  $R(\cdot)$  of the workload process  $Q$ , see for instance [102].

- In the first place, interesting stochastic properties of the process  $Q$  can be captured via its covariance function. Indeed, for instance  $L^2$ -continuity (that is, quadratic mean continuity), -differentiability and -integrability of the workload process  $Q$  depend upon similar properties of its covariance function.
- Moreover, knowledge of the covariance function  $R(\cdot)$  is important to assess the variance of estimators for the mean workload  $\mu = \mathbb{E}Q_e$ , for instance when one considers estimators of the type

$$\hat{\mu} = \frac{1}{T} \int_0^T Q(t) dt.$$

This can be seen as follows. First notice that  $\hat{\mu}$  is an unbiased estimator of  $\mu$  ( $\mathbb{E}\hat{\mu} = \mu$ ) if the queue was in stationarity at time 0. To assess its precision we have to evaluate its standard error, which can be calculated from

$$\text{Var}\hat{\mu} = \frac{1}{T^2} \int_0^T \int_0^T \text{Cov}(Q(s), Q(t)) ds dt = \frac{2}{T^2} \int_0^T \int_0^t R(u) du dt.$$

This explains the interest in studying the covariance function.

- Furthermore, application of spectral analysis to queueing theory requires the determination of the correlation function as the Fourier transform of the spectral density. Indeed, if the correlation function  $r(\cdot)$  is continuous at  $t = 0$ , and hence at every point  $t \in [0, \infty)$ , then it can be represented as

$$r(t) = \int_{-\infty}^{\infty} e^{itx} dF(x), \quad (1.11)$$

where  $F(\cdot)$  is real, strictly increasing and bounded.  $F(\cdot)$  is called the spectral distribution of the process  $Q$ . Since  $Q$  is a real-valued process,  $r(t)$  will then be real and (1.11) yields

$$r(t) = \int_{-\infty}^{\infty} \cos(tx) dF(x) = F(0) + 2 \int_0^{\infty} \cos(tx) dF(x). \quad (1.12)$$

For further reading on the covariance function of stationary stochastic processes, we refer the reader to Cramér and Leadbetter [34].

In this thesis we also consider an alternative measure for the study of the dependence structure of the workload process. Assuming the queue is already in stationarity at time 0 and with  $Q_e$  denoting the stationary workload, we define for  $p, q$  and  $T > 0$

$$\mathbb{R}(T|p, q) := \frac{\mathbb{P}(Q(0) > p, Q(T) > q)}{\mathbb{P}(Q_e > p)\mathbb{P}(Q_e > q)}. \quad (1.13)$$

Observe that ‘the more independent’ the events  $\{Q(0) > p\}$  and  $\{Q(T) > q\}$  are, the more  $R(T|p, q)$  approaches the value 1. In this sense, the metric  $R(T|p, q)$  can be seen as a measure that describes the dependence of the events  $\{Q(0) > p\}$  and  $\{Q(T) > q\}$ . Moreover, the measure  $R(T|p, q)$  can be related to the covariance of the corresponding indicator functions. Indeed, considering the indicator functions  $\mathbf{1}_{\{Q(0) > p\}}$  and  $\mathbf{1}_{\{Q(T) > q\}}$  of the events  $\{Q(0) > p\}$  and  $\{Q(T) > q\}$ , respectively, it is easily seen that

$$\text{Cov}(\mathbf{1}_{\{Q(0) > p\}}, \mathbf{1}_{\{Q(T) > q\}}) = \mathbb{P}(Q_e > p)\mathbb{P}(Q_e > q)(R(T|p, q) - 1).$$

A further motivation for the choice of this dependence metric and its relation to mixing properties will be given in Chapter 2.

Now that we have defined the performance metrics of our interest, we proceed by briefly describing the three classes of input models that we consider in this monograph.

### 1.3 Gaussian processes

A real-valued stochastic process  $\{X(t) : t \in \mathbb{R}\}$  is called Gaussian, if all its finite dimensional distributions are multivariate normal distributions, i.e., for all  $1 \leq n < \infty$ , and  $t_1 < t_2 < \dots < t_n$ , the random vector  $(X(t_1), X(t_2), \dots, X(t_n))$  has a multivariate normal distribution. In particular, for each  $t$ , the random variable  $X(t)$  has a normal distribution with some mean  $m(t)$  and variance  $v(t)$ .

Furthermore, a Gaussian process is completely characterized by its mean function  $m(t) = \mathbb{E}X(t)$  and covariance function  $\Gamma(s, t) = \text{Cov}(X(s), X(t))$ . Notice that every covariance function is nonnegative definite in the sense that, for any finite set of indices  $\{t_1, \dots, t_k\}$ , and  $\alpha_1, \dots, \alpha_k$  arbitrary numbers,  $\Gamma(s, t)$  satisfies

$$\sum_{i,j=1}^k \alpha_i \Gamma(t_i, t_j) \alpha_j \geq 0. \quad (1.14)$$

It follows that, given any function  $\Gamma(s, t)$  satisfying (1.14), we can always construct a Gaussian process having  $\Gamma(s, t)$  as its covariance function, see for instance [34, pp. 80-82].

In this thesis, we only consider Gaussian processes with stationary increments  $A(s, t) = X(t) - X(s)$  for  $s \leq t$ . For this class of Gaussian processes it can be easily verified that

$$\Gamma(s, t) = \frac{1}{2}(v(t) + v(s) - v(t - s)). \quad (1.15)$$

As indicated by (1.15), the class of Gaussian processes with stationary increments is extremely rich. Indeed, if we define  $\Gamma(s, t)$  with  $v(t)$  a continuous positive function, such that  $\Gamma(s, t)$  is nonnegative definite, then we can construct a corresponding Gaussian process  $X$  with stationary increments.

Some examples of Gaussian processes, that will be studied in this thesis, are:

- *Fractional Brownian motion*, characterized by the following variance function

$$v(t) = |t|^{2H}, \quad H \in [0, 1] \text{ and } t \in \mathbb{R}; \quad (1.16)$$

where  $H$  is the so-called *Hurst parameter*. Observe that for  $H = 1/2$  we have Brownian motion.

- *The integrated Ornstein-Uhlenbeck process* having the variance function

$$v(t) = |t| - 1 + e^{-|t|}, \quad t \in \mathbb{R}. \quad (1.17)$$

For this class of input processes with  $\varpi < c$ , there exist no explicit formulae for the steady-state probability distribution of the stationary workload  $Q_e$ , except for Brownian motion input. For Brownian motion input, where  $v(t) = |t|$ , it is well-known that the stationary workload  $Q_e$  is exponentially distributed with parameter  $2(c - \varpi)$ , see for instance [61]. For other Gaussian processes, one has found approximations of  $\mathbb{P}(Q_e > B)$ , the probability that the workload exceeds some threshold  $B > 0$ , in different asymptotic regimes.

A simple (and remarkably good) approximation of  $\mathbb{P}(Q_e > B)$  is given by

$$\mathbb{P}(Q_e > B) \approx \exp \left\{ - \inf_{s \geq 0} I(s, B|c, \varpi) \right\}; \quad (1.18)$$

where

$$I(s, B|c, \varpi) := \frac{1}{2} \frac{(B + (c - \varpi)s)^2}{v(s)}, \quad B, s > 0. \quad (1.19)$$

It can be checked that for Brownian motion input, Relation (1.18) is exact. More details can be found in [57, 80] and the references therein.

For further reading on Gaussian processes, see [5, 6, 64], and for the use of Gaussian processes as input in queueing models, we refer to [4, 80] and the references therein.

## 1.4 Lévy processes

A Lévy process  $X \equiv \{X(t) : t \in \mathbb{R}\}$  is a stochastic process possessing the following properties:

- (i)  $X(0) = 0$  with probability one,
- (ii)  $X$  has independent increments, i.e., for  $s \leq t$ ,  $X(t) - X(s)$  is independent of  $\{X(u) : u \leq s\}$ ,
- (iii)  $X$  has stationary increments, i.e., for  $s \leq t$ ,  $X(t) - X(s)$  is equal in distribution to  $X(t - s)$ ,
- (iv) the sample-paths of  $X$  are almost surely right continuous with left limits.

Lévy processes are intimately related to the class of infinitely divisible distributions, which are completely characterized by the Lévy-Khintchine formula: a probability distribution  $F$  is infinitely divisible if and only if its characteristic exponent

$$\Psi(\vartheta) := \log \int_{\mathbb{R}} e^{i\vartheta x} F(dx), \quad \vartheta \in \mathbb{R},$$

takes the following expression, known as the Lévy-Khintchine representation,

$$\Psi(\vartheta) = i\delta\vartheta - \frac{1}{2}\sigma^2\vartheta^2 + \int_{\mathbb{R}} (e^{i\vartheta x} - 1 - i\vartheta x\mathbf{1}_{|x|<1}) \Pi(dx). \quad (1.20)$$

Here  $\delta \in \mathbb{R}$ ,  $\sigma^2 \geq 0$  and  $\Pi$  is a Lévy measure, i.e., a measure concentrated on  $\mathbb{R} \setminus \{0\}$  satisfying

$$\int_{\mathbb{R}} \min(1, x^2) \Pi(dx) < \infty. \quad (1.21)$$

A Lévy process  $X$  has the property that for all  $t$ ,  $\mathbb{E}e^{i\vartheta X(t)} = e^{t\Psi(\vartheta)}$ , where  $\Psi(\cdot)$  is the characteristic exponent of  $X(1)$ , which has an infinitely divisible distribution. Within the class of Lévy processes we distinguish the following three classes. If the measure  $\Pi$  gives no mass to the negative half line, i.e.,  $\Pi(-\infty, 0) = 0$ , which means that the process has no negative jumps, we say that the Lévy process  $X$  is *spectrally positive*. On the other hand, if the measure  $\Pi$  is concentrated on  $(-\infty, 0)$ , i.e.,  $\Pi(0, \infty) = 0$ , which means that the process has no positive jumps, we say that the Lévy process  $X$  is *spectrally negative*. Lévy processes which have monotone sample-paths almost surely are called subordinators.

In this thesis special attention is paid to the class of spectrally-positive Lévy processes. For this class we usually use Laplace transforms instead of characteristic functions. We define the Laplace exponent

$$\varphi(\vartheta) := \log \mathbb{E}e^{-\vartheta X(1)}, \quad \vartheta \geq 0,$$

which is given by

$$\varphi(\vartheta) = \delta\vartheta + \frac{1}{2}\sigma^2\vartheta^2 + \int_{(0, \infty)} (e^{-\vartheta x} - 1 + \vartheta x\mathbf{1}_{|x|<1}) \Pi(dx), \quad (1.22)$$

with  $\delta$ ,  $\sigma^2$  and  $\Pi$  as in (1.20).

Examples of Lévy processes are Brownian motion (for which  $\Pi \equiv 0$ ), the compound Poisson process (for which  $\delta = \sigma^2 = 0$ ), and  $\alpha$ -stable Lévy processes with characteristic exponent given by

$$\Psi(\vartheta) = -|\vartheta|^\alpha \left( 1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right) \operatorname{sgn}(\vartheta) \right), \quad (1.23)$$

where  $\alpha \in (0, 1) \cup (1, 2)$  and  $\beta \in [-1, 1]$ . For further reading on Lévy processes, we refer to [20, 22, 55, 74, 106].

Results on the characterization of the Laplace transform of the stationary workload  $Q_e$  for queues fed by spectrally one-sided input processes are due to Zolotarev [113]. However, in general it is not feasible to determine the steady-state distribution by explicit inversion of the Laplace transform. Therefore, one may then resort to study the asymptotics of  $\mathbb{P}(Q_e > B)$  as  $B \rightarrow \infty$ . Under different assumptions on the moment generating function  $\mathbb{E}e^{\vartheta X(1)}$  of  $X(1)$  asymptotics of  $\mathbb{P}(Q_e > B)$  as  $B \rightarrow \infty$  have been determined. For heavy-tailed input, i.e.,  $\mathbb{E}e^{\vartheta X(1)} = \infty$  for all  $\vartheta > 0$ , we refer to for instance [13, 47]. For light-tailed input, i.e.,  $\mathbb{E}e^{\vartheta X(1)} < \infty$  for some  $\vartheta > 0$ , asymptotic results can be found in e.g. [21, 59].

## 1.5 Markov modulated fluid input processes

In this section we describe a Markov modulated fluid input process. Let  $J \equiv \{J(t) : t \geq 0\}$  be an irreducible continuous-time Markov chain defined on a finite state space  $\mathcal{E} = \{1, 2, \dots, N\}$ . Let  $(\pi_i, i \in \mathcal{E})$  denote the stationary distribution of the Markov chain  $J$ . Further, let  $\{r_i : i \in \mathcal{E}\}$  be a finite set of real numbers. The process  $\{A(s, t) : 0 \leq s \leq t\}$  defined by

$$A(s, t) = \int_s^t r_{J(u)} du, \quad (1.24)$$

is called a Markov modulated process. This type of input processes is widely used in, e.g., manufacturing and communication networks.

Under the stability condition, which for this class of input processes is given by

$$\sum_{i=1}^N r_i \pi_i < 0,$$

the steady-state distribution of  $(Q_e, J_e)$  exists and can be explicitly derived. For more background on this model (and several variants) we refer to for instance [72, 103, 107].

## 1.6 Overview and contributions

In this section we give a short overview of the remainder of this thesis. In the next two chapters of this thesis we consider Gaussian queues, that is, queues fed by Gaussian processes, such as fractional Brownian motion (fBm) and the integrated Ornstein-Uhlenbeck (iOU) process. As already mentioned in Section 1.3, for these input processes no explicit expression for the distribution of the stationary workload is known. Hence it is unfeasible to analyze the covariance function  $R(t)$  of the stationary version of the workload process (as this would even require the knowledge of  $\mathbb{E}Q(0)Q(t)$ , the *joint* moment function). Therefore, we will analyze the dependence structure of the workload process by considering the metric  $R(T|p, q)$  defined in (1.13). In Chapter 2, based on [53], we analyze the behavior of the metric  $R(T|p, q)$  for  $T \rightarrow \infty$  under the so-called many-sources scaling. That is, we assume that the input of the queue is an aggregation of  $n$  i.i.d. Gaussian processes. To keep the queue stable we scale the service rate by  $n$  and in addition we scale also the buffer with  $n$ . We denote the resulting workload process by  $Q^n$ . We focus on two special cases, viz. fBm and iOU. For large values of  $T$ , we study rough, logarithmic asymptotics of our dependence metric associated with  $Q^n$  as  $n \rightarrow \infty$ . Relying on (the generalized version of) Schilder's theorem, we are able to characterize its decay. The main result of this chapter is that, at least for the special cases considered, the dependence structure of the input process essentially carries over to the workload process (in the asymptotic regime that we have chosen and in terms of our specific notion of dependence, viz., the metric  $R(T|p, q)$ ).

Chapter 3, based on [38], is devoted to the analysis of transient characteristics of Gaussian queues under the so-called large buffer regime. More specifically, we determine the logarithmic asymptotics of  $\mathbb{P}(Q(0) > pB, Q(TB) > qB)$  as  $B \rightarrow \infty$  and hence the logarithmic asymptotics of  $R(TB|pB, qB)$  as  $B \rightarrow \infty$  can also be determined. For any pair  $(p, q)$  three regimes can be distinguished:

- (A) For small values of  $T$ , either of the events  $\{Q(0) > pB\}$  and  $\{Q(TB) > qB\}$  will essentially imply the other. More specifically: if  $p > q$  then the event  $\{Q(TB) > qB\}$  'comes for free', and if  $q > p$ , then the event  $\{Q(0) > pB\}$  'comes for free'.
- (B) There is an intermediate range of values of  $T$  for which it is to be expected that both  $\{Q(0) > pB\}$  and  $\{Q(TB) > qB\}$  are 'tight' (in the sense that none of them implies the other with overwhelming probability), but that the time epochs 0 and  $T$  tend to lie in the same busy period.
- (C) Finally, for large values of  $T$  still both events are 'tight', but now they occur in different busy periods with overwhelming probability.

For the short-range dependent case explicit calculations are presented, whereas for the long-range dependent case structural results are proven.

The Lévy-driven queue, that is, a queue with Lévy input, is studied in Chapters 4 and 5. In Chapter 4, which is based on [51], we consider a queue fed by a spectrally positive Lévy process. For this class of Lévy processes the Laplace transform of the stationary workload is known, see Zolotarev [113]. Using this result and a result taken from [65], we are able to derive the Laplace transform  $\rho(\vartheta) = \int_0^\infty r(t)e^{-\vartheta t} dt$  of the correlation function  $r(t)$  of the stationary workload process  $\{Q(t) : t \geq 0\}$ . This expression allows us to prove structural properties of  $r(\cdot)$ . More specifically, we prove that the correlation function is positive, decreasing, and convex, relying on the machinery of completely monotone functions. We also show that  $r(\cdot)$  can be represented as the complementary distribution function of a specific random variable. These results are used to compute the asymptotics of  $r(t)$ , for  $t$  large, for the cases of light-tailed and heavy-tailed Lévy input.

In Chapter 5, based on [37], we consider a queue fed by a general Lévy process. In this case we will study the metric  $R(T_B|pB, qB)$  for various types of functions  $T_B$ . The main focus will be on refined exact asymptotics (rather than rough logarithmic asymptotics) of rare event probabilities of the type  $\mathbb{P}(Q(0) > pB, Q(T_B) > qB)$ , for given positive numbers  $p, q$ , and a positive deterministic function  $T_B$ . The following contributions are then made.

- We first identify conditions on the function  $T_B$  under which the probability of interest is dominated by the ‘most demanding event’, in the sense that it is asymptotically equivalent to  $\mathbb{P}(Q_e > \max\{p, q\}B)$  for  $B$  large. These conditions essentially reduce to  $T_B$  being sublinear (i.e.,  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ ).
- A second condition on  $T_B$  is derived under which the probability of interest essentially ‘decouples’, in that  $\mathbb{P}(Q(0) > pB, Q(T_B) > qB)$  is asymptotically equivalent to  $\mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB)$  for  $B$  large. For various models considered in the literature this ‘decoupling condition’ reduces to requiring that  $T_B$  is superlinear (i.e.,  $T_B/B \rightarrow \infty$  as  $B \rightarrow \infty$ ). This is not true for certain ‘heavy-tailed’ cases, for instance the situations in which the Lévy input process corresponds to an  $\alpha$ -stable process, or to a compound Poisson process with regularly varying job sizes, in which the ‘decoupling condition’ reduces to  $T_B/B^2 \rightarrow \infty$ . For these input processes we also establish the asymptotics of the probability under consideration for  $T_B$  increasing superlinearly but subquadratically.

Moreover, for light-tailed input, special attention is paid to the case where  $T_B$  is a linear function in  $B$ , that is we suppose that  $T_B = RB$ , for some  $R > 0$ . We derive intuitively appealing asymptotics, by intensively relying on sample-path large deviations results. The regimes obtained in this case, can be interpreted in terms of most

likely paths to overflow.

Chapter 6 is based on [52] and deals with a Markov-fluid-driven queue, that is, a queue fed by a Markov modulated process. For this model we consider two important metrics, viz., the busy period of the system and the covariance function of the stationary workload process, which we capture in terms of their Laplace transforms. Relying on sample-path large deviations, we identify the logarithmic asymptotics of the probability that the busy period lasts longer than  $t$ , as  $t \rightarrow \infty$ .

In Chapter 7 we consider a discrete-time model with a general input process  $X$  having stationary increments. More specifically, we consider the workload process of a queue operating in discrete time, focusing on the (multivariate) distribution of the workloads at different points in time. In a many-sources framework exact asymptotics are determined, relying on large deviations results for the sample means of multivariate random variables.

Each chapter of this monograph can be read independently of the other chapters, that is, all chapters are essentially self-contained. Furthermore, every chapter starts with an introduction where a historical account of the literature is given. For the key quantities we have used a consistent notation.

## Chapter 2

---

# Gaussian queue under many-sources scaling

In this chapter we consider a single-server queue fed by Gaussian processes. Under the many-sources scaling we will study the logarithmic asymptotics of the metric  $R(T|p, q)$ . Explicit results will be given for the special cases of fractional Brownian motion and integrated Ornstein-Uhlenbeck inputs. In the following section we give an overview of the literature on queueing systems with Gaussian input processes, and give further background on the problem dealt with in this chapter.

### 2.1 Introduction

Traffic measurement studies have provided convincing statistical evidence that in various networking environments traffic exhibits strong dependence over a wide range of time-scales. These studies, starting off in the early 1990s with the famous article by Leland *et al.* [76] on Ethernet traffic, showed that the traffic rate process was *long-range dependent*: with  $Z(t)$  the traffic rate at time  $t$ , the autocorrelation function  $c(T)$  of the traffic rate (i.e., the correlation coefficient between  $Z(0)$  and  $Z(T)$ ) vanishes extremely slowly as a function of the lag  $T$  – more precisely:  $c(T)$  decays so slowly that  $\sum_{T \in \mathbb{N}} c(T) = \infty$ .

This explains the interest in the performance evaluation of queues fed by long-range dependent traffic. Notably, the traffic models that were predominantly used till the mid-1990s did not allow for any long-range dependence: they usually corresponded to short-range dependent traffic processes (such as Poisson processes, Markov-modulated Poisson processes, or exponential on-off sources). In the late 1990s, Gaussian traffic models have gained more interest and popularity for modeling network traffic. One of their attractive features is that they cover a broad variety of dependence structures, ranging from short-range (the integrated Ornstein-Uhlenbeck process, Brownian motion) to long-range dependent (fractional Brownian motion with *Hurst parameter*  $H > \frac{1}{2}$ , see [76]). In [67] it is argued that the use of Gaussian traffic models is justified as long as the aggregation is sufficiently large, both in number of flows and time. We refer to [62, 90, 104] for excellent studies on network traffic modeling.

The fact that network traffic is long-range dependent is of crucial importance from the perspective of traffic engineering in communication networks. Where the

short-range dependent models usually lead to buffer overflow probabilities that decay exponentially in the buffer size, long-range dependent models are considerably less benign: in case of fractional Brownian motion input with Hurst parameter  $H$  this decay is ‘Weibullian’ [48, 93] (that is, roughly like  $\exp(-\alpha B^{2-2H})$ , for some  $\alpha > 0$ , and  $B$  denoting the buffer size, which is slower than exponential for  $H > \frac{1}{2}$ ), or even polynomial [96, 114] (e.g., for on-off sources with regularly varying on-times). In other words: modeling traffic by short-range dependent process would lead to estimates of the overflow probability that are considerably too optimistic.

For Gaussian queues (that is, queues fed by Gaussian processes), so far primary interest was in the characterization of the buffer overflow probability. Notably, in two limiting regimes asymptotic results were obtained: in the large-buffer regime (where the buffer threshold grows large), and in the many-sources regime (in which the number of Gaussian inputs grows large, and the buffer and service speed are scaled accordingly [111]). Without exhaustively mentioning all relevant contributions, logarithmic asymptotics for the large-buffer case are due to [36, 48], whereas exact asymptotics can be found in, e.g., [46, 92], and in the many-sources regime logarithmic asymptotics are in [4, 28] and the exact asymptotics in [39].

To the best of our knowledge, hardly any attention has been paid to the characterization of the *dependence structure of the workload process* of Gaussian queues. This is remarkable, as, from an engineering standpoint, knowledge of the dependence structure is clearly quite relevant. Most importantly, it would give us a handle on the timescale after which it is justified to approximate transient probabilities by their steady-state counterpart. Also procedures that ‘learn’ the characteristics of the input process by observing the workload process [85] would greatly benefit from insights into the degree of dependence between two subsequent observations (more precisely, it can be determined from what timescale on one could safely neglect the dependence between the observations).

Seen from a more mathematical angle, an interesting fundamental question is: to what extent the dependence structure of the input process is inherited by the workload process? Or, put differently, does long-range dependent input give rise to a long-range dependent workload process? The results obtained in this chapter show that indeed for fractional Brownian motion (in the sequel abbreviated to fBm) and integrated Ornstein-Uhlenbeck (iOU) the dependence structure of the workload process strongly resembles that of the input process: Weibullian decay for fBm, and exponential decay for iOU.

A first aim would be to analyze the covariance of  $Q(0)$  and  $Q(T)$ , which we denote by  $R(T)$  or the corresponding correlation coefficient denoted by  $r(T)$ , cf. (1.9) and (1.10). However, it is not clear what methodology can be used to analyze these covariances. It is noted, for instance, that large deviations type of results are not of any help here, as covariances are quantities related to expected values, which cannot

be represented as rare-event probabilities (where we also recall that in the setting of queues with Gaussian input, apart from a few special cases, one has not even succeeded so far to compute the *mean* workload  $\mu = \mathbb{E}Q_e$ ).

To overcome this problem we have chosen the following solution:

- We choose a measure for dependence that is more tractable than the covariance  $R(T)$ . This new metric measures the difference between  $\log \mathbb{P}(Q(0) > p, Q(T) > q)$  and  $\log (\mathbb{P}(Q(0) > p)\mathbb{P}(Q(T) > q))$ , for given  $p, q > 0$ ; popularly speaking, the more independent  $\{Q(0) > p\}$  and  $\{Q(T) > q\}$  are, the smaller the difference. A more specific goal is to characterize for fBm and iOU how our metric decays to 0 when  $T$  grows to infinity.
- We work in the many-sources asymptotic regime [111] that was mentioned above. As a consequence, we can use an extensive set of useful techniques, most notably (sample-path) large deviations results, in particular (the generalized version of) Schilder's theorem.

More precisely, the setting we consider is as follows: we let  $n$  i.i.d. Gaussian processes  $X_1(\cdot), \dots, X_n(\cdot)$  feed into a queue in which both the service speed and the buffer content are scaled by  $n$ ; we denote the workload of the resulting queueing system at time  $t$  by  $Q^n(t)$ . The results presented in this chapter are asymptotic in  $n$ .

As mentioned above, we specialize to the important cases of fBm and iOU input. Our main conclusion is that, using the metric introduced above and considering the many-sources regime, the dependence structure of the input process essentially carries over to the workload process.

Above we argued that Gaussian models (and in particular fBm) are good traffic descriptors in the setting of communication networks as long as there is sufficient aggregation [67]. We stress, however, that this is an issue that should be handled with care, as it depends very much on the situation at hand whether this is the case. [90] presents a systematic assessment of this issue. There the stochastic properties of a superposition of  $n$  sources with heavy-tailed on-times (or bursts), and alternatively a corresponding  $M/G/\infty$  input model, is considered, after rescaling time with  $T$ . Conditions are discussed under which the limiting process indeed looks like fBm, while in other situations  $\alpha$ -stable Lévy motion is more suitable. More specifically, if the situation at hand is such that the rate at which bursts are generated is large in relation to the tail of the distribution of the burst duration, fBm is an appropriate approximation. For a more detailed discussion we refer to e.g. [90, 56].

The remainder of this chapter is organized as follows. In Section 2.2 we recall some important results about Gaussian processes and the large-deviations theorems that will be needed in the sequel. In Section 2.3 we give the main results of this

chapter while their proofs are given in Section 2.4. In the last section we give a heuristic approach that extends our results to queues fed by more general Gaussian processes, and a number of concluding remarks.

## 2.2 Preliminaries

This section consists of two subsections. In Section 2.2.1 we recall basic properties of Gaussian processes, while in Section 2.2.2 we state two important tools from large deviations theory: (the multivariate version of) Cramér's theorem and (the generalized version of) Schilder's theorem.

### 2.2.1 Gaussian processes

Let  $X_i(\cdot)$  denote a sequence of i.i.d. centered Gaussian processes with continuous sample paths and stationary increments  $A_i(\cdot, \cdot)$ ,  $i = 1, \dots, n$ ; it is assumed that  $X_i(0) \equiv 0$  for all  $i$ . As in Chapter 1, for  $s < t$ , we interpret  $A_i(s, t)$  as the amount of the traffic generated by the  $i$ -th Gaussian source  $X_i(\cdot)$  in the time interval  $(s, t]$ . Moreover, we let the  $X_i(t)$  be 'two-sided', that is, defined for all  $t \in \mathbb{R}$ .

We denote by  $X(\cdot)$  the generic Gaussian process corresponding to a single source, and  $A(s, t) := X(t) - X(s)$  denote its increments, i.e., for all  $i$ ,  $X_i(\cdot)$  and  $A_i(\cdot, \cdot)$  have the same distribution as  $X(\cdot)$  and  $A(\cdot, \cdot)$ , respectively. A (centered) Gaussian process is characterized by its variance function  $v(\cdot)$  (which is necessarily continuous); because of the stationarity of the increments of our process, we have for  $s < t$ ,  $\text{Var}A(s, t) = v(t - s)$ .

In the sequel the bivariate random variable  $(A(-s, 0), A(T - t, T))$  (for large values of  $T$ ) is frequently used. Its distribution is a bivariate Normal distribution with zero mean vector and covariance matrix  $\Sigma_T(s, t)$  given by

$$\Sigma_T(s, t) := \begin{pmatrix} v(s) & \Gamma_T(s, t) \\ \Gamma_T(s, t) & v(t) \end{pmatrix},$$

with  $\Gamma_T(s, t) := \text{Cov}(A(-s, 0), A(T - t, T))$ . For  $s > 0$  and  $0 < t < T$ , this covariance reduces to

$$\Gamma_T(s, t) = \frac{v(T + s) - v(T) - v(T - t + s) + v(T - t)}{2};$$

for other ranges of  $s$  and  $t$  similar expressions can be given.

Gaussian sources have the intrinsic inconvenience that in principle negative traffic can be generated:  $A(s, t)$  (with  $t > s$ ) is not necessarily non-negative. When using

the representation for the workload at time  $t$  (take for ease a queue fed by a single Gaussian source, with service rate  $c > 0$ )

$$Q(t) := \sup_{s \geq 0} \{A(t-s, t) - cs\},$$

this turns out to be not an issue: the probabilistic properties of the above functional of the Gaussian process  $X(\cdot)$  can be evaluated, irrespective of whether the input process allows negative increments.

In our study we focus, without loss of generality, on centered Gaussian processes, but it is straightforward to adapt the results to the case of non-centered Gaussian processes, as the queueing system in which the input has mean rate  $m \neq 0$  and service rate  $c$  (larger than  $m$  to ensure stability) coincides with the system with centered input and service rate  $c - m$ .

In this chapter we focus on two special Gaussian processes: (standard) fractional Brownian motion (or *fBm*;  $v(t) = t^{2H}$ , with  $H \in (0, 1)$ ), and the integrated Ornstein-Uhlenbeck process (or *iOU*;  $v(t) = t - 1 + e^{-t}$ ).

**Lemma 2.2.1.** *Fix  $s, t > 0$ ; let  $t < T$ .*

- *fBm.* For  $H > \frac{1}{2}$ ,  $\Gamma_T(s, t)$  is positive, and decreases to 0 when  $T \rightarrow \infty$ . For  $H < \frac{1}{2}$ ,  $\Gamma_T(s, t)$  is negative, and increases to 0 when  $T \rightarrow \infty$ .
- *iOU.*  $\Gamma_T(s, t)$  is positive, and decreases to 0 when  $T \rightarrow \infty$ .

*Proof.* First focus on *fBm*. It is immediate that

$$\Gamma_T(s, t) = \gamma_T^{(\text{fBm})}(s, t) := \frac{1}{2} ((T+s)^{2H} - T^{2H} - (T-t+s)^{2H} + (T-t)^{2H}).$$

Consider  $H > \frac{1}{2}$ . It is readily checked that in order to show that  $\Gamma_T(s, t)$  is positive, we have to prove that

$$1 - (1-t)^{2H} < (1+s)^{2H} - (1-t+s)^{2H},$$

or, equivalently, that  $(1+s)^{2H} - (1-t+s)^{2H}$  increases in  $s > 0$  (for all  $t \in (0, 1)$ ). Differentiation with respect to  $s$  leads to the claim  $(1+s)^{2H-1} > (1+s-t)^{2H-1}$ , which is indeed true for  $H > \frac{1}{2}$ . The fact that  $\Gamma_T(s, t)$  is decreasing in  $T$  (with limit 0) is proven in the same way. The case  $H < \frac{1}{2}$  can be dealt with similarly.

For *iOU*,

$$\begin{aligned} \Gamma_T(s, t) &= \gamma_T^{(\text{iOU})}(s, t) := \frac{1}{2} (e^{-T-s} - e^{-T} - e^{-T+t-s} + e^{-T+t}) \\ &= \frac{1}{2} (1 - e^{-s})(e^t - 1)e^{-T}, \end{aligned}$$

which is indeed positive and decreasing in  $T$ . □

### 2.2.2 Large deviations results

In this subsection we give a brief description of the main results from the large deviations theory for Gaussian processes. The proofs of the theorems presented here can be found in [44, 45]; see for more background [80]. We first state Cramér's theorem, that relates to  $d$ -dimensional random variables, and then Schilder's theorem, that describes the sample-path large deviations of Gaussian processes.

Let  $X \in \mathbb{R}^d$  be a  $d$ -dimensional random vector. We denote the moment generating function of  $X$  by  $M(\theta) := \mathbb{E} \exp(\langle \theta, X \rangle)$  and its logarithm by  $\Lambda(\theta) := \log M(\theta)$ . Its convex conjugate  $\Lambda^*$  is defined by  $\Lambda^*(x) := \sup_{\theta \in \mathbb{R}^d} (\langle \theta, x \rangle - \Lambda(\theta))$ , with  $\langle \cdot, \cdot \rangle$  denoting the usual inner product:  $\langle \theta, x \rangle := \theta^T x = \sum_{i=1}^d \theta_i x_i$ . We first state (the multivariate version of) Cramér's theorem which characterizes the logarithmic rate of the convergence of the empirical mean of i.i.d. random vectors in  $\mathbb{R}^d$ .

**Theorem 2.2.2** (Multivariate Cramér). *Let  $X_i \in \mathbb{R}^d$  be i.i.d.  $d$ -dimensional random vectors, distributed as a random vector  $X$ . Then the following LDP applies [44, 45]:*

(a) *For any closed set  $F \subset \mathbb{R}^d$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n X_i \in F \right) \leq - \inf_{y \in F} \Lambda^*(y);$$

(b) *For any open set  $G \subset \mathbb{R}^d$ ,*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n X_i \in G \right) \geq - \inf_{y \in G} \Lambda^*(y),$$

where the large deviations rate function  $\Lambda^*(\cdot)$  is as given above.

**Remark 2.2.3.** Consider the case that  $X$  has a multivariate Normal distribution with mean vector 0 and  $d \times d$ -non-singular matrix covariance matrix  $\Sigma$ . Then using that  $\Lambda(\theta) = \frac{1}{2} \theta^T \Sigma \theta$  we obtain

$$\theta^* = \Sigma^{-1} x \quad \text{and} \quad \Lambda^*(x) = \frac{1}{2} x^T \Sigma^{-1} x, \tag{2.1}$$

where  $\theta^*$  is the optimizer in the definition of  $\Lambda^*$ . ♠

Before stating the generalized Schilder's theorem we sketch the framework of the Schilder's sample path large deviations principle as established in [16], see also [45]. We use the same setup and notation as in [80, 81]. We consider  $n$  i.i.d. centered Gaussian processes  $A_i(\cdot)$  and define the path space  $\Omega$  as

$$\Omega := \left\{ \omega : \mathbb{R} \rightarrow \mathbb{R}, \text{ continuous, } \omega(0) = 0, \lim_{|t| \rightarrow \infty} \frac{\omega(t)}{1 + |t|} = 0 \right\}$$

which becomes a Banach space by equipping it with the norm

$$\|\omega\|_{\Omega} := \sup_{t \in \mathbb{R}} \frac{|\omega(t)|}{1 + |t|}.$$

In Addie *et al.* [4] it is shown that  $X(\cdot)$  can be realized in  $\Omega$  under Assumption 2.2.4; it is clear that both fBm and iOU satisfy this requirement.

**Assumption 2.2.4.** *The variance function  $v(\cdot)$  of the process  $X(\cdot)$  is continuous and it satisfies*

$$\lim_{t \rightarrow \infty} \frac{v(t)}{t^{\alpha}} = 0 \quad (2.2)$$

for some  $\alpha \in (0, 2)$ .

Next we introduce the *reproducing kernel Hilbert space*  $\mathbb{R} \subset \Omega$ , with the property that its elements are roughly as smooth as the covariance function  $\Gamma(s, \cdot)$ , see Adler [5] for more details. We start from a subspace  $\mathbb{R}^* \subset \Omega$ , defined by

$$\mathbb{R}^* := \left\{ \omega \in \Omega, \omega(\cdot) = \sum_{i=1}^n a_i \Gamma(s_i, \cdot), a_i, s_i \in \mathbb{R}, n \in \mathbb{N} \right\},$$

with  $\Gamma(s, t) := \text{Cov}(A(0, s), A(0, t))$ . The inner product on this space  $\mathbb{R}^*$  is defined as follows, for  $\omega_a, \omega_b \in \mathbb{R}^*$

$$\langle \omega_a, \omega_b \rangle_{\mathbb{R}} := \left\langle \sum_{i=1}^n a_i \Gamma(s_i, \cdot), \sum_{j=1}^n b_j \Gamma(s_j, \cdot) \right\rangle_{\mathbb{R}} = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \Gamma(s_i, s_j). \quad (2.3)$$

Now we can introduce the norm  $\|\omega\|_{\mathbb{R}} := \sqrt{\langle \omega, \omega \rangle_{\mathbb{R}}}$ . The closure of  $\mathbb{R}^*$  under this norm is defined as the space  $\mathbb{R}$ . Then the rate function of the sample-path large deviations principle (LDP) is defined as follows:

$$I(\omega) := \begin{cases} \frac{1}{2} \|\omega\|_{\mathbb{R}}^2 & \text{if } \omega \in \mathbb{R}; \\ \infty & \text{otherwise.} \end{cases} \quad (2.4)$$

For a sequence of  $n$  i.i.d. centered Gaussian processes  $X_i(\cdot)$  the following sample-path LDP holds [16, 45].

**Theorem 2.2.5** (Generalized Schilder). *The following sample-path LDP applies:*

(a) *For any closed set  $F \subset \Omega$ ,*

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n X_i(\cdot) \in F \right) \leq - \inf_{\omega \in F} I(\omega);$$

(b) *For any open set  $G \subset \Omega$ ,*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n X_i(\cdot) \in G \right) \geq - \inf_{\omega \in G} I(\omega),$$

### 2.3 Main results

As mentioned in the introduction of this chapter, our main interest lies in the investigation of the dependence structure of the workload process. Since the distribution of the steady-state workload has been found explicitly only for the case of Brownian motion input, we resort to an asymptotic framework, viz. the so-called many-sources regime. In this regime the number of Gaussian inputs, say  $n$ , grows large, and the service rate is scaled accordingly. In this framework, the stationary workload process is given by

$$Q^n(t) := \sup_{s \leq t} \sum_{i=1}^n A_i(s, t) - nc(t-s) \stackrel{d}{=} Q_e^n = \sup_{s \geq 0} \sum_{i=1}^n A_i(-s, 0) - ncs. \quad (2.5)$$

As we wish to investigate the dependence structure of the workload process, we could try to characterize the autocorrelation

$$\delta_n(T) := \frac{\mathbb{E}Q^n(0)Q^n(T) - \mathbb{E}Q^n(0)\mathbb{E}Q^n(T)}{\sqrt{\text{Var}Q^n(0)}\sqrt{\text{Var}Q^n(T)}} = \frac{\mathbb{E}Q^n(0)Q^n(T) - \mathbb{E}Q_e^n\mathbb{E}Q_e^n}{\text{Var}Q_e^n}.$$

It is evident that  $\delta_n(T) \downarrow 0$  as  $T \uparrow \infty$ , but the question is how fast it vanishes.

Unfortunately, this notion of dependence is hard to handle — not even an explicit expression for  $\mathbb{E}Q_e^n$  is known for non-Brownian Gaussian input processes. We therefore introduce an alternative notion of dependence. The following metric describes the degree of dependence between the events  $\{Q^n(0) > np\}$  and  $\{Q^n(T) > nq\}$  for positive  $p, q$ .

**Definition 2.3.1.** For given positive numbers  $p, q$  define

$$\kappa_n(T) := \frac{\mathbb{P}(Q^n(0) > np, Q^n(T) > nq)}{\mathbb{P}(Q^n(0) > np)\mathbb{P}(Q^n(T) > nq)} = \frac{\mathbb{P}(Q^n(0) > np, Q^n(T) > nq)}{\mathbb{P}(Q_e^n > np)\mathbb{P}(Q_e^n > nq)}. \quad (2.6)$$

Furthermore, let  $\kappa(T)$  be the limit of  $\log \kappa_n(T)/n$  as  $n \rightarrow \infty$ .

**Remark 2.3.2.** It should be noticed that  $\kappa_n(T)$  relates directly to the metric  $R(T|p, q)$  given by (1.13). Indeed, one only needs to replace  $Q(0), Q(T), c, p$  and  $q$  by  $Q^n(0), Q^n(T), nc, np$  and  $nq$ , respectively, in the definition of  $R(T|p, q)$ . ♠

It is evident that various other dependence measures could be thought of. Our measure is reminiscent of quantities used when defining mixing conditions, see e.g. [29]. For instance, with  $\mathcal{A}_s^t$  defining the  $\sigma$ -field  $\sigma(Q(u) : s \leq u \leq t)$ , and

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$

we say that  $Q(t)$  is strongly mixing if  $\alpha(T) := \sup_s \alpha(\mathcal{A}_{-\infty}^s, \mathcal{A}_{s+T}^\infty) \rightarrow 0$  as  $T \rightarrow \infty$ . The relation between the decay of  $\kappa(T)$  and mixing conditions is not *a priori* clear;

also due to the fact that a supremum over  $A \in \mathcal{A}_{-\infty}^s$  and  $B \in \mathcal{A}_{s+T}^{\infty}$  needs to be computed, it is typically hard to characterize the decay of  $\alpha(T)$  and related quantities.

Before stating the main theorems of this section we first give the logarithmic asymptotics of the marginal probabilities involved in Definition 2.3.1. They are given by

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(Q^n(0) > np) = - \inf_{s>0} \frac{(p+cs)^2}{2v(s)} \quad (2.7)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(Q^n(T) > nq) = - \inf_{t>0} \frac{(q+ct)^2}{2v(t)}, \quad (2.8)$$

using that the queue is in stationarity at both epochs; see for instance [4]. In [39] the following lemma was proved; it entails that the infima over  $s$  and  $t$  are attained and are unique under a specific assumption on the variance function.

**Lemma 2.3.3.** *Suppose that the standard deviation function  $\sigma(t) := \sqrt{v(t)}$  of the generic input process  $X(\cdot)$  is such that  $\sigma(t) \in \mathcal{C}^2([0, \infty))$  is strictly increasing and strictly concave. Then the right-hand sides of (2.7) and (2.8) have unique minimizers. Concavity of  $\sigma(t)$  is equivalent to requiring that*

$$2v(t)v''(t) - (v'(t))^2 \leq 0. \quad (2.9)$$

We denote the minimizers by  $s^*$  and  $t^*$ . It is readily checked that they solve

$$\begin{cases} 2cv(s) = (p+cs)v'(s); \\ 2cv(t) = (q+ct)v'(t). \end{cases} \quad (2.10)$$

Now we give the main results of this chapter. Theorem 2.3.4 states that for fBm input  $\kappa(T)$  decays to zero and its decay rate is  $T^{2H-2}$  as  $T \rightarrow \infty$ , which indicates that the workload process has essentially the same dependence structure as the input process. As will be discussed in more detail in Section 2.5, this means that the workload process is (in our metric) long-range dependent if the *Hurst parameter*  $H$  is greater than  $\frac{1}{2}$ . For fBm,  $\sigma(t) = t^H$  is concave, so Lemma 2.3.3 applies, and (2.10) has a unique solution; in fact,  $s^*$  and  $t^*$  can be explicitly calculated, and are given through

$$s^* := \frac{p}{c} \frac{H}{1-H}; \quad t^* := \frac{q}{c} \frac{H}{1-H}. \quad (2.11)$$

**Theorem 2.3.4** (fBm input). *If the input process is fBm we have the following logarithmic asymptotics for  $\kappa_n(T)$ :*

$$\begin{aligned} \kappa(T) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \kappa_n(T) \\ &= \frac{(p+cs^*)(q+ct^*)}{v(s^*)v(t^*)} \cdot \frac{1}{2} s^* t^* (2H)(2H-1) T^{2H-2} + o(T^{2H-2}) \\ &= \frac{(2H-1)c^2}{H} s^{*2-2H} t^{*2-2H} T^{2H-2} + o(T^{2H-2}). \end{aligned} \quad (2.12)$$

It is interesting to compare this result to the dependence structure of the input process. We could look at a counterpart of  $\kappa(T)$ , for instance

$$\lambda(T) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \frac{\mathbb{P}(\sum_{i=1}^n A_i(0, 1) > np, \sum_{i=1}^n A_i(T, T+1) > nq)}{\mathbb{P}(\sum_{i=1}^n A_i(0, 1) > np) \mathbb{P}(\sum_{i=1}^n A_i(T, T+1) > nq)} \right),$$

and consider its decay for  $T$  large. Denoting  $\gamma_T^{(\text{fBm})}(1, 1)$  by  $\gamma_T^{(\text{fBm})}$  (see Lemma 2.2.1), we have that  $\gamma_T^{(\text{fBm})} \sim \frac{1}{2} T^{2H-2} \rightarrow 0$  as  $T \rightarrow \infty$ . Then straightforward computations show that the bivariate version of Cramér's theorem implies that

$$\lambda(T) = -\frac{1}{2}(p, q) \begin{pmatrix} v(1) & \gamma_T^{(\text{fBm})} \\ \gamma_T^{(\text{fBm})} & v(1) \end{pmatrix}^{-1} \begin{pmatrix} p \\ q \end{pmatrix} + \frac{p^2}{2v(1)} + \frac{q^2}{2v(1)} \sim pq \gamma_T^{(\text{fBm})},$$

for  $T$  large, i.e., *also* decaying as  $T^{2H-2}$ ! The above arguments provide support for the claim that, in this metric, the workload process has essentially the same dependence structure as the input process.

**Remark 2.3.5.** For  $H = \frac{1}{2}$  we can explicitly calculate  $\kappa(T)$  for any  $T$ , relying on the formulas for the transient behavior of reflected Brownian motion, see e.g. [61, p. 49]. It turns out that for all  $T > c^{-1} \cdot (\sqrt{p} + \sqrt{q})^2$  it holds that  $\kappa(T) = 0$ ; observe that the existence of such a threshold value could be anticipated due to the independent increments. Also note that this result is in line with Theorem 2.3.4. ♠

Now consider the case of iOU input. In this case  $s^*$  and  $t^*$  cannot be explicitly calculated. They are uniquely determined though, as can be seen as follows. Criterion (2.9) reduces to

$$\varphi(t) := 2te^{-t} + e^{-2t} - 1 \leq 0,$$

which is true because  $\varphi(0) = 0$  and  $\varphi'(t) = e^{-t}(2 - 2t - 2e^{-t}) \leq 0$ .

Theorem 2.3.6 states that for iOU input the speed of convergence of  $\kappa(T)$  to 0 as  $T \rightarrow \infty$  is  $e^{-T}$ .

**Theorem 2.3.6** (iOU input). *If the input process is iOU we have the following logarithmic asymptotics for  $\kappa_n(T)$ :*

$$\begin{aligned} \kappa(T) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \kappa_n(T) \\ &= \frac{(p + cs^*)(q + ct^*)}{v(s^*)v(t^*)} \cdot \frac{1}{2} (1 - e^{-s^*})(e^{t^*} - 1)e^{-T} + o(e^{-T}) \\ &= 2c^2 e^{-(T-t^*)} + o(e^{-T}). \end{aligned} \tag{2.13}$$

In this case, it can be verified that

$$\gamma_T^{(\text{iOU})} := \gamma_T^{(\text{iOU})}(1, 1) \sim \frac{1}{2} e^{-T} \cdot \left( e - 2 + \frac{1}{e} \right).$$

As we have  $\lambda(T) \sim pq\gamma_T^{(\text{iOU})}$ , it again holds that the dependence structure of the workload process essentially coincides with that of the input process (i.e., both  $\kappa(T)$  and  $\lambda(T)$  are roughly proportional to  $e^{-T}$ ).

## 2.4 Proofs

In this section we give the proofs of the results we stated in the previous section. In the first subsection we derive a number of generic results, while we specialize to fBm and iOU in the last part of the section.

### 2.4.1 General results

The results of this subsection hold for any type of Gaussian sources (i.e., we do not restrict ourselves to fBm and iOU), the only exception being Proposition 2.4.4. We first define two sets of paths in  $\Omega$  that play a crucial role in our analysis.

$$\mathcal{S}_T := \{f \in \Omega : \exists s > 0, \exists t > 0 : -f(-s) > p + cs, f(T) - f(T-t) > q + ct\}; \quad (2.14)$$

$$\mathcal{S}_T(s, t) := \{f \in \Omega : -f(-s) > p + cs, f(T) - f(T-t) > q + ct\}. \quad (2.15)$$

Observe that  $\mathcal{S}_T$  is the union (over all  $s, t > 0$ ) of the  $\mathcal{S}_T(s, t)$ . Interestingly, the set of paths  $\mathcal{S}_T$  directly relates to the ‘joint overflow event’  $\{Q^n(0) > np, Q^n(T) > nq\}$ , as follows from the next lemma.

**Lemma 2.4.1.** *For any  $p, q > 0$ ,*

$$\mathbb{P}(Q^n(0) > np, Q^n(T) > nq) = \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n A_i(\cdot) \in \mathcal{S}_T\right).$$

*Proof.* This follows by applying (2.5):

$$\begin{aligned} & \mathbb{P}(Q^n(0) > np, Q^n(T) > nq) \\ &= \mathbb{P}\left(\sup_{s>0} \left\{ \sum_{i=1}^n A_i(-s, 0) - ncs \right\} > np, \sup_{t>0} \left\{ \sum_{i=1}^n A_i(T-t, T) - nct \right\} > nq\right) \\ &= \mathbb{P}\left(\exists s > 0 : \sum_{i=1}^n A_i(-s, 0) - ncs > np, \exists t > 0 : \sum_{i=1}^n A_i(T-t, T) - nct > nq\right) \\ &= \mathbb{P}\left(\exists s > 0 : \sum_{i=1}^n \frac{A_i(-s, 0)}{n} > p + cs, \exists t > 0 : \sum_{i=1}^n \frac{A_i(T-t, T)}{n} > q + ct\right) \\ &= \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n A_i(\cdot) \in \mathcal{S}_T\right), \end{aligned}$$

which proves the claimed.  $\square$

In the sequel we frequently use the following bivariate Normal large deviations rate function:

$$\Lambda_T^*(p + cs, q + ct) := \frac{1}{2}(p + cs, q + ct) (\Sigma_T(s, t))^{-1} \begin{pmatrix} p + cs \\ q + ct \end{pmatrix}.$$

By explicitly calculating the matrix inverse, we obtain that  $\Lambda_T^*(p + cs, q + ct)$  can be written in the following alternative form:

$$\begin{aligned} & \frac{1}{2} \frac{v(s)v(t)}{v(s)v(t) - \Gamma_T(s, t)^2} \\ & \cdot \left( \frac{(p + cs)^2}{v(s)} + \frac{(q + ct)^2}{v(t)} - 2 \frac{(p + cs)(q + ct)\Gamma_T(s, t)}{v(s)v(t)} \right). \end{aligned} \quad (2.16)$$

The next lemma determines the decay rate of the most likely path in  $\mathcal{S}_T(s, t)$ , for fixed values of  $s$  and  $t$ . It turns out that there are three different regimes.

**Lemma 2.4.2.** *For any  $p, q > 0$ ,*

$$\inf_{f \in \mathcal{S}_T(s, t)} I(f) = \bar{\Lambda}_T^*(p + cs, q + ct),$$

where  $\bar{\Lambda}_T^*(p + cs, q + ct)$  equals

$$\frac{(p + cs)^2}{2v(s)} \quad \text{if} \quad \frac{\Gamma_T(s, t)}{v(s)}(p + cs) > q + ct; \quad (2.17)$$

$$\frac{(q + ct)^2}{2v(t)} \quad \text{if} \quad \frac{\Gamma_T(s, t)}{v(t)}(q + ct) > p + cs; \quad (2.18)$$

$$\Lambda_T^*(p + cs, q + ct) \quad \text{otherwise.} \quad (2.19)$$

*Proof.* Multiplication of (2.17) and (2.18) would lead to

$$\Gamma_T^2(s, t) > v(s)v(t),$$

and hence Cauchy-Schwarz implies that the conditions in (2.17) and (2.18) cannot apply simultaneously.

Then recognize

$$\frac{\Gamma_T(s, t)}{v(s)}(p + cs) = \mathbb{E}(A(T - t, T) \mid A(-s, 0) = p + cs);$$

$$\frac{\Gamma_T(s, t)}{v(t)}(q + ct) = \mathbb{E}(A(-s, 0) \mid A(T - t, T) = q + ct).$$

The stated now follows immediately from the bivariate version of Cramér's theorem; see the solution of Exercise 4.1.9 as given on p. 42 of [80].  $\square$

The proof of the next proposition relies on Lemma 2.A.1, that is stated and proven in the appendix.

**Proposition 2.4.3.** *For any  $p, q > 0$ ,*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(Q^n(0) > np, Q^n(T) > nq) &= - \inf_{f \in \mathcal{S}_T} I(f) \\ &= - \inf_{s, t > 0} \bar{\Lambda}_T^*(p + cs, q + ct). \end{aligned}$$

*Proof.* From ‘Schilder’ and Lemma 2.4.1 we have

$$- \inf_{f \in \mathcal{S}_T} I(f) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(Q^n(0) > np, Q^n(T) > nq) \leq - \inf_{f \in \bar{\mathcal{S}}_T} I(f).$$

We first show that the above inequalities are actually equalities, by establishing that  $\mathcal{S}_T$  is an  $I$ -continuity set, that is,

$$\inf_{f \in \mathcal{S}_T} I(f) = \inf_{f \in \bar{\mathcal{S}}_T} I(f), \quad (2.20)$$

where the  $\bar{\mathcal{S}}_T$  denotes the closure of  $\mathcal{S}_T$ , and is given in Lemma 2.A.1.

This can be done in the same way as in the Appendix of [86]. Choose an arbitrary path  $f$  in  $\bar{\mathcal{S}}_T \cap \mathbb{R}$ , and approximate it by a path in  $\mathcal{S}_T$ , as follows. We use the sets  $\mathcal{S}(s), \mathcal{S}_T(t), \bar{\mathcal{S}}(s)$ , and  $\bar{\mathcal{S}}_T(t)$  as defined in the appendix 2.A. Due to Lemma 2.A.1 we have that  $f \in \bar{\mathcal{S}}(s) \cap \bar{\mathcal{S}}_T(t)$  for some  $s, t > 0$ . Let  $\eta(\cdot)$  be a path in  $\mathbb{R}$  that is strictly increasing and taking negative values for  $u \in (-\infty, 0)$  and positive values for  $u \in (0, \infty)$  (for instance  $\eta(u) := \text{sgn}(u)\sqrt{|u|}$  or  $\arctan u$ ). Define

$$f_n(u) := f(u) + \frac{\eta(u)}{n}.$$

Then  $f_n \in \mathcal{S}(s) \cap \mathcal{S}_T(t)$  as, for any  $s > 0$ , it holds that

$$-f_n(-s) = -f(-s) - \frac{\eta(-s)}{n} \geq p + cs - \frac{\eta(-s)}{n} > p + cs$$

and, for any  $t > 0$ ,

$$\begin{aligned} f_n(T) - f_n(T-t) &= f(T) - f(T-t) + \frac{\eta(T) - \eta(T-t)}{n} \\ &\geq q + ct + \frac{\eta(T) - \eta(T-t)}{n} > q + ct. \end{aligned}$$

Moreover, we have, for  $n \rightarrow \infty$ ,

$$\|f_n\|_{\mathbb{R}}^2 = \|f + \frac{1}{n}\eta\|_{\mathbb{R}}^2 \rightarrow \|f\|_{\mathbb{R}}^2,$$

which proves (2.20) and therefore also the first equality of the proposition.

The above entails that the decay rate of our interest equals

$$\inf_{s,t>0} \inf_{f \in (\mathcal{S}(s) \cap \mathcal{S}_T(t))} I(f).$$

Recall from (2.14) and (2.15) that  $\mathcal{S}_T$  is the union over all  $s \geq 0$  and  $t \geq 0$  of the  $\mathcal{S}_T(s, t)$ , and observe that  $\mathcal{S}_T(s, t) = \mathcal{S}(s) \cap \mathcal{S}_T(t)$ . The second equality of the proposition now follows directly from Lemma 2.4.2.  $\square$

The following proposition indicates, for fBm and iOU inputs, that for  $T$  large necessarily (2.19) is satisfied.

**Proposition 2.4.4.** *Consider fBm or iOU. For any  $p, q > 0$ , and  $T$  large enough*

$$\inf_{s,t>0} \bar{\Lambda}_T^*(p + cs, q + ct) = \inf_{s,t>0} \Lambda_T^*(p + cs, q + ct). \quad (2.21)$$

*Proof.* As, for any  $s, t > 0$  and any  $T > 0$ , it holds that  $\bar{\Lambda}_T^*(s, t) \leq \Lambda_T^*(s, t)$ , it suffices to prove that, for  $T$  sufficiently large,

$$\inf_{s,t>0} \bar{\Lambda}_T^*(p + cs, q + ct) \geq \inf_{s,t>0} \Lambda_T^*(p + cs, q + ct).$$

We prove this property in a number of steps.

- *Step 1.* As, evidently,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i(-s, 0) > p + cs, \frac{1}{n} \sum_{i=1}^n A_i(T - t, 0) > q + ct \right) \\ & \leq \min \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i(-s, 0) > p + cs \right), \right. \\ & \quad \left. \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i(T - t, 0) > q + ct \right) \right\}, \end{aligned}$$

we have that

$$\bar{\Lambda}_T(p + cs, q + ct) \geq \max \left\{ \frac{(p + cs)^2}{2v(s)}, \frac{(q + ct)^2}{2v(t)} \right\}.$$

- *Step 2.* Lemma 2.2.1 states that, for any fixed  $s, t$ ,  $\Gamma_T(s, t) \rightarrow 0$  as  $T \rightarrow \infty$ . It can be checked that this implies that also  $\bar{\Lambda}_T^*(p + cs, q + ct) \rightarrow \Lambda_\infty^*(p + cs, q + ct)$  as  $T \rightarrow \infty$ , where

$$\Lambda_\infty^*(p + cs, q + ct) = \frac{(p + cs)^2}{v(s)} + \frac{(q + ct)^2}{v(t)}$$

(to this end, observe that, for any fixed  $s, t$  the conditions in (2.17) and (2.18) are not fulfilled for  $T$  sufficiently large). It is clear that, when taking the infimum of  $\Lambda_\infty^*(p + cs, q + ct)$  over  $s, t > 0$ , the expression decouples into the sum of an infimum over  $s$  and an infimum over  $t$ . Both individual infima have a unique minimizer, namely  $s^*$  and  $t^*$  as introduced earlier. In the remainder of the proof we use the notation  $\ell := \Lambda_\infty^*(p + cs^*, q + ct^*)$ . It is clear that the above implies that for  $T$  sufficiently large

$$\inf_{s, t > 0} \bar{\Lambda}_T^*(p + cs, q + ct) \leq \bar{\Lambda}_T^*(p + cs^*, q + ct^*) \leq 2\ell. \quad (2.22)$$

- *Step 3.* Using Step 1, for any  $t > 0$ , both as  $s \downarrow 0$  and as  $s \rightarrow \infty$ , uniformly in  $T$ ,

$$\bar{\Lambda}_T^*(p + cs, q + ct) \geq \frac{(p + cs)^2}{2v(s)} \rightarrow \infty;$$

likewise, for any  $s > 0$ , both as  $t \downarrow 0$  and as  $t \rightarrow \infty$ , we have that

$$\bar{\Lambda}_T^*(p + cs, q + ct) \rightarrow \infty.$$

It implies that we can find  $\underline{\varepsilon}, \bar{\varepsilon} \in (0, \infty)$ , independent of  $T$ , such that for all  $s, t \notin [\underline{\varepsilon}, \bar{\varepsilon}]$  it holds that  $\bar{\Lambda}_T^*(p + cs, q + ct) \geq 3\ell$ .

- *Step 4.* Using (2.22) and Step 3, we conclude that we can restrict ourselves, for  $T$  sufficiently large, to  $s, t \in [\underline{\varepsilon}, \bar{\varepsilon}]$ . Again using that  $\Gamma_T(s, t) \rightarrow 0$  as  $T \rightarrow \infty$  (by virtue of Lemma 2.2.1), it is seen that for  $T$  large enough, for all  $s, t \in [\underline{\varepsilon}, \bar{\varepsilon}]$  the conditions in (2.17) and (2.18) are not satisfied, and therefore we have that  $\bar{\Lambda}_T^*(p + cs, q + ct) = \Lambda_T^*(p + cs, q + ct)$ . This entails that, for  $T$  sufficiently large,

$$\begin{aligned} \inf_{s, t > 0} \bar{\Lambda}_T^*(p + cs, q + ct) &= \inf_{s, t \in [\underline{\varepsilon}, \bar{\varepsilon}]} \bar{\Lambda}_T^*(p + cs, q + ct) \\ &= \inf_{s, t \in [\underline{\varepsilon}, \bar{\varepsilon}]} \Lambda_T^*(p + cs, q + ct) \\ &\geq \inf_{s, t > 0} \Lambda_T^*(p + cs, q + ct). \end{aligned}$$

This concludes the proof.  $\square$

In view of the fact that  $\Lambda_T^*(p + cs, q + ct) \rightarrow \Lambda_\infty^*(p + cs, q + ct)$ , we now also have that a sequence of local optimizers of the right-hand side of (2.21), say  $(s_T^*, t_T^*)$ , converges to  $(s^*, t^*)$  as  $T \rightarrow \infty$ . Relying on Taylor expansions around  $(s^*, t^*)$ , the vector  $(s_T^*, t_T^*)$  at which the function  $\Lambda_T^*(p + cs, q + ct)$  is minimal solves the following system:

$$\begin{aligned} &(p + cs)(2cv(s) - (p + cs)v'(s)) \\ &= 2 \left( \frac{q + ct}{v(t)} \right) \left( (cv(s) - (p + cs)v'(s)) \Gamma_T(s, t) + (p + cs)v(s) \frac{\partial \Gamma_T}{\partial s}(s, t) \right); \end{aligned} \quad (2.23)$$

$$\begin{aligned}
& (q + ct)(2cv(t) - (q + ct)v'(t)) \\
&= 2 \left( \frac{p + cs}{v(s)} \right) \left( (cv(t) - (q + ct)v'(t)) \Gamma_T(s, t) + (q + ct)v(t) \frac{\partial \Gamma_T}{\partial t}(s, t) \right)
\end{aligned} \tag{2.24}$$

where the partial derivatives of  $\Gamma_T(s, t)$  with respect to  $s$  and  $t$  are given by

$$\begin{aligned}
\frac{\partial \Gamma_T}{\partial s}(s, t) &= \frac{1}{2} (v'(T + s) - v'(T - t + s)); \\
\frac{\partial \Gamma_T}{\partial t}(s, t) &= \frac{1}{2} (v'(T - t + s) - v'(T - t)).
\end{aligned}$$

In the next two subsections we study the system (2.23)-(2.24), for both fBm and iOU inputs, by analyzing the behavior of  $s_T^*$ ,  $t_T^*$  in detail. This yields the desired information, needed in order to characterize the decay rate  $\kappa(T)$  for  $T$  large.

## 2.4.2 Proof for fBm input

As we have seen in the proof of Lemma 2.2.1, for  $T \rightarrow \infty$ ,

$$\gamma_T^{(\text{fBm})}(s, t) = st \cdot H(2H - 1) \cdot T^{2H-2} + o(T^{2H-2}).$$

For large  $T$  we obtain in the same way

$$\begin{aligned}
\frac{\partial \gamma_T^{(\text{fBm})}}{\partial s} &= t \cdot H(2H - 1) \cdot T^{2H-2} + o(T^{2H-2}); \\
\frac{\partial \gamma_T^{(\text{fBm})}}{\partial t} &= s \cdot H(2H - 1) \cdot T^{2H-2} + o(T^{2H-2}).
\end{aligned} \tag{2.25}$$

Inserting these into (2.23)-(2.24) we obtain

$$\begin{aligned}
& (2cs - 2H(p + cs)) \\
&= \frac{2H(2H - 1)(q + ct)(cs - (2H - 1)(p + cs))st}{t^{2H}(p + cs)} T^{2H-2} + o(T^{2H-2});
\end{aligned} \tag{2.26}$$

$$\begin{aligned}
& (2ct - 2H(q + ct)) \\
&= \frac{2H(2H - 1)(p + cs)(ct - (2H - 1)(q + ct))st}{s^{2H}(q + ct)} T^{2H-2} + o(T^{2H-2}).
\end{aligned} \tag{2.27}$$

Note that if we let  $T \rightarrow \infty$  in the last system, we retrieve (2.10), which has unique solution (2.11). Observe that in the system of equations (2.26)-(2.27), the right-hand side of the equations decays to 0 with speed  $T^{2H-2}$  as  $T$  grows to infinity. This observation, in conjunction with  $(s_T^*, t_T^*)$  converging to  $(s^*, t^*)$ , entails that we can express  $s_T^*$ ,  $t_T^*$  as follows:

$$\begin{cases} s_T^* = s^* + f(s^*, t^*)T^{2H-2} + o(T^{2H-2}); \\ t_T^* = t^* + g(s^*, t^*)T^{2H-2} + o(T^{2H-2}). \end{cases}$$

To determine the values of  $f(s^*, t^*)$  and  $g(s^*, t^*)$ , we proceed as follows. Using Taylor expansions we obtain for the left-hand side of (2.26), after tedious calculus,

$$\begin{aligned} & (p + cs) (2cv(s) - (p + cs)v'(s)) \\ &= 2H(p + cs^*)s^{*2H-2} (cs^* - (2H - 1)(p + cs^*)) f(s^*, t^*)T^{2H-2} + o(T^{2H-2}), \end{aligned}$$

and for the right-hand side

$$\begin{aligned} & 2 \left( \frac{q + ct}{v(t)} \right) \left( (cv(s) - (p + cs)v'(s)) \Gamma_T(s, t) + (p + cs)v(s) \frac{\partial \Gamma_T}{\partial s}(s, t) \right) \\ &= 2H(2H - 1)(q + ct^*)s^{*2H}t^{*1-2H} (cs^* - (2H - 1)(p + cs^*)) T^{2H-2} + o(T^{2H-2}) \end{aligned}$$

Doing the same for (2.27), and inserting (2.11), we find the following expressions for  $f$  and  $g$  at  $s^*, t^*$ :

$$\begin{aligned} f(s^*, t^*) &= (2H - 1) \frac{q}{p} s^{*2} t^{*1-2H} = (2H - 1) s^* t^{*2-2H}; \\ g(s^*, t^*) &= (2H - 1) \frac{p}{q} t^{*2} s^{*1-2H} = (2H - 1) t^* s^{*2-2H}. \end{aligned}$$

Inserting these expressions into  $\Lambda_T^*(p + cs, q + ct)$ , we can evaluate the components of (2.16):

$$\begin{aligned} \frac{(p + cs)^2}{s^{2H}} &= \frac{(p + cs^*)^2}{s^{*2H}} \left( 1 + 2 \left( \frac{c}{(p + cs^*)} - \frac{H}{s^*} \right) f(s^*, t^*)T^{2H-2} + o(T^{2H-2}) \right) \\ &= \frac{(p + cs^*)^2}{s^{*2H}} + o(T^{2H-2}); \end{aligned}$$

$$\begin{aligned} \frac{(q + ct)^2}{t^{2H}} &= \frac{(q + ct^*)^2}{t^{*2H}} \left( 1 + 2 \left( \frac{c}{(q + ct^*)} - \frac{H}{t^*} \right) g(s^*, t^*)T^{2H-2} + o(T^{2H-2}) \right) \\ &= \frac{(q + ct^*)^2}{t^{*2H}} + o(T^{2H-2}); \end{aligned}$$

$$\begin{aligned} & 2H(2H - 1) \cdot (p + cs)(q + ct)(st)^{1-2H} \cdot T^{2H-2} + o(T^{2H-2}) \\ &= 2H(2H - 1)(p + cs^*)(q + ct^*)s^{*1-2H}t^{*1-2H}T^{2H-2} + o(T^{2H-2}) \\ &= 2 \frac{(2H - 1)c^2}{H} s^{*2-2H}t^{*2-2H}T^{2H-2} + o(T^{2H-2}). \end{aligned}$$

We thus obtain the desired result, i.e.,

$$\kappa(T) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \kappa_n(T) = \frac{(2H - 1)c^2}{H} s^{*2-2H}t^{*2-2H}T^{2H-2} + o(T^{2H-2}).$$

### 2.4.3 Proof for iOU input

As in the fBm case, denote by  $s^*, t^*$  the minimizing point when there is independence, i.e., the solution of (2.10). We follow the same arguments as in the case of fBm. For fixed  $(s, t)$  the covariance  $\Gamma_T(s, t)$  is decreasing exponentially in  $T$ . The solution of system (2.23)-(2.24), say  $s_T^*, t_T^*$ , converges to  $s^*, t^*$ , and its convergence speed is of the order  $e^{-T}$  for large  $T$ . These observations entail that

$$\begin{cases} s_T^* = s^* + k(s^*, t^*)e^{-T} + o(e^{-T}); \\ t_T^* = t^* + \ell(s^*, t^*)e^{-T} + o(e^{-T}). \end{cases}$$

To determine  $k$  and  $\ell$  at  $s^*, t^*$ , we proceed as in the above subsection. We find

$$\begin{aligned} k(s^*, t^*) &= \\ &= \frac{q + ct^*}{p + cs^*} (e^{t^*} - 1) \cdot \frac{cv(s^*)(1 - e^{-s^*}) - (p + cs^*)(v'(s^*)(1 - e^{-s^*}) - v(s^*)e^{-s^*})}{(cv'(s^*) - (p + cs^*)v''(s^*))v(t^*)} \\ &= \frac{(q + ct^*)v'(t^*)}{v''(t^*)(p + cs^*)v(t^*)} \cdot \frac{cv(s^*)v'(s^*) - (p + cs^*)v'(s^*)^2 + (p + cs^*)v(s^*)v''(s^*)}{(cv'(s^*) - (p + cs^*)v''(s^*))} \\ &= \frac{2cv(s^*)}{v''(t^*)(p + cs^*)} \cdot \frac{(-cv'(s^*) + (p + cs^*)v''(s^*))}{(cv'(s^*) - (p + cs^*)v''(s^*))} = -\frac{v'(s^*)}{v''(t^*)}; \end{aligned}$$

$$\begin{aligned} \ell(s^*, t^*) &= \\ &= \frac{p + cs^*}{q + ct^*} (1 - e^{-s^*}) \frac{cv(t^*)(e^{t^*} - 1) - (q + ct^*)(v'(t^*)(e^{t^*} - 1) - v(t^*)e^{t^*})}{(cv'(t^*) - (q + ct^*)v''(t^*))v(s^*)} \\ &= \frac{(p + cs^*)v'(s^*)}{(q + ct^*)v(s^*)} \cdot \frac{cv(t^*)v'(t^*)e^{t^*} - (q + ct^*)v'(t^*)^2e^{t^*} + (q + ct^*)v(t^*)e^{t^*}}{(cv'(t^*) - (q + ct^*)v''(t^*))} \\ &= \frac{2cv(s^*)}{v''(t^*)(q + ct^*)} \cdot \frac{(-cv'(t^*) + (q + ct^*))}{(cv'(t^*) - (p + ct^*)v''(t^*))} \\ &= \frac{v'(t^*)}{v''(t^*)} \cdot \frac{q + cv(t^*)}{(cv'(t^*) - (q + ct^*)v''(t^*))}. \end{aligned}$$

Now we insert this in the objective function (2.16), and similarly to the fBm case we obtain

$$\begin{aligned} \frac{(p + cs)^2}{v(s)} &= \frac{(p + cs^*)^2}{v(s^*)} (1 + o(e^{-T})); \\ \frac{(q + ct)^2}{v(t)} &= \frac{(q + ct^*)^2}{v(t^*)} (1 + o(e^{-T})); \\ \frac{(p + cs)(q + ct)(1 - e^{-s})(e^t - 1)e^{-T}}{v(s)v(t)} &= \frac{(p + cs^*)(q + ct^*)(1 - e^{-s^*})(e^{t^*} - 1)e^{-T}}{v(s^*)v(t^*)} + o(e^{-T}). \end{aligned}$$

Thus we get for iOU input the desired result:

$$\begin{aligned}\kappa(T) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \kappa_n(T) \\ &= \frac{(p + cs^*)(q + ct^*)}{v(s^*)v(t^*)} \cdot \frac{1}{2}(1 - e^{-s^*})(e^{t^*} - 1)e^{-T} + o(e^{-T}),\end{aligned}$$

which simplifies to  $2c^2e^{-(T-t^*)} + o(e^{-T})$ .

## 2.5 Discussion and concluding remarks

*A. Generalizations.* Theorems 2.3.4 and 2.3.6 suggest that our results can be generalized considerably, in that it can be expected that their counterparts can be stated for a substantially broader class of Gaussian processes with stationary increments. First observe that Expression (2.16) can alternatively be written as

$$\frac{1}{2} \left( \frac{(p + cs)^2}{v(s)} + \frac{(q + ct)^2}{v(t)} - 2 \frac{(p + cs)(q + ct)\Gamma_T(s, t)}{v(s)v(t)} \right) + o(\Gamma_T(s, t)). \quad (2.28)$$

Now suppose we wish to evaluate  $\inf_t (f(t) + \varepsilon g(t)) - f(t^*)$ , where  $t^*$  is minimizer of  $f(\cdot)$ . A Taylor expansion of  $f'(t) + \varepsilon g'(t)$  in  $t_\varepsilon^* = t^* + \varepsilon \bar{t}$  reads

$$f'(t^*) + \varepsilon \bar{t} f''(t^*) + \varepsilon g'(t^*) + O(\varepsilon^2) = f'(t^*) + \varepsilon (\bar{t} f''(t^*) + g'(t^*)) + O(\varepsilon^2),$$

so that we obtain  $\bar{t} = -g'(t^*)/f''(t^*)$ . Hence, under appropriate regularity conditions,

$$\inf_t (f(t) + \varepsilon g(t)) = f(t^*) - \varepsilon \frac{g'(t^*)}{f''(t^*)} f'(t^*) + \varepsilon g(t^*) + O(\varepsilon^2).$$

Now using  $f'(t^*) = 0$  it follows that

$$\inf_t (f(t) + \varepsilon g(t)) - f(t^*) = \varepsilon g(t^*) + O(\varepsilon^2). \quad (2.29)$$

In the same way, a 2-dimensional counterpart of (2.29) can be stated. Now suppose that (for large  $T$ )  $\Gamma_T(s, t)$  decouples as  $\varphi(s, t) \cdot \varepsilon(T)$ ; here  $\varepsilon(T)$  does not depend on  $s$  and  $t$ , and converges to 0 as  $T \rightarrow \infty$ . Applying then the two-dimensional version of (2.29) to (2.28),

$$\begin{aligned}\kappa(T) &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \kappa_n(T) \\ &= \frac{(p + cs^*)(q + ct^*)}{v(s^*)v(t^*)} \varphi(s^*, t^*) \varepsilon(T) + o(\varepsilon(T)) \\ &= 4c^2 \frac{\varphi(s^*, t^*)}{v'(s^*)v'(t^*)} \varepsilon(T) + o(\varepsilon(T)).\end{aligned}$$

*B. Long-range dependence.* Based on Theorems 2.3.4 and 2.3.6, one may conjecture that long-range dependence of the input process carries over to the workload process. We now provide additional heuristic support for this claim.

First consider fBm. Heuristically reasoning, Theorem 2.3.4 entails that, for some constant  $\kappa_0$ ,

$$\kappa_n(T) \approx \exp(n\kappa_0 T^{2H-2}).$$

Hence, the correlation coefficient of the indicator functions  $1\{Q^n(0) > np\}$  and  $1\{Q^n(T) > nq\}$  roughly equals (using that  $x(1-x) \approx x$  for  $x$  small)

$$\begin{aligned} & \frac{\mathbb{P}(Q^n(0) > np, Q^n(T) > nq) - \mathbb{P}(Q^n(0) > np)\mathbb{P}(Q^n(T) > nq)}{\sqrt{\mathbb{P}(Q^n(0) > np)\mathbb{P}(Q^n(T) > nq)}} \\ & \approx (e^{n\kappa_0 T^{2H-2}} - 1) \cdot \sqrt{\mathbb{P}(Q^n(0) > np)\mathbb{P}(Q^n(0) > nq)}. \end{aligned}$$

Using  $e^x \approx 1+x$  for  $x$  small, we find that for  $T$  large, the above display is of the form  $\psi(n)T^{2H-2}$ , where the function  $\psi(\cdot)$  does not depend on  $T$ . Observe that the latter expression is non-summable (over  $T$ ) for  $H > \frac{1}{2}$ . This intuitive argument suggests that the long-range dependence of the input process propagates to the queueing process.

Likewise, for iOU we find that the correlation coefficient introduced above is roughly proportional to  $e^{-T}$ , and hence corresponds to a short-range dependent process.

Further research on this issue could make use of the concept of *Hurstiness*, as introduced in [112]. Hurstiness is a property of the queue's input process (closely related to long-range dependence), and it is shown that the Hurstiness is preserved by several fundamental operators; for instance the Hurstiness of the departure process equals that of the arrival process. It is not immediately clear, however, whether results as those presented in the present chapter can be found relying on the notion of Hurstiness. As there is a clear relation between the departure process and the workload dynamics, one would think so, but a technical issue is that Hurstiness relates to *cumulative* processes, such as arrival and departure processes, whereas our focus is on the dependence between 'instantaneous values' of the workload at time 0 and  $T$ . Also, Hurstiness relates to just the rate of decay, and in view of this it is not likely that it would help to (for instance) find the constant in front of  $T^{2H-2}$  in Theorem 2.3.4.

*C. Remarks on asymptotics for iOU.* It may be surprising, at first glance, that the asymptotics of  $\kappa(T)$  for iOU, that is  $2c^2 e^{-(T-t^*)}$ , depend on  $q$ , but *do not depend on  $p$* . This can be understood as follows.

First observe that for iOU input (unlike for fBm input) there is a notion of a *traffic rate* process  $S(\cdot)$ , where  $S(t) = X'(t)$ . It can be checked easily that

- (i)  $S(t)$  is Normally distributed with mean 0 and variance  $\frac{1}{2}$ ,
- (ii)  $\text{Cov}(S(0), S(T)) = \frac{1}{2}e^{-T}$ ,
- (iii) the conditional distribution of  $A(T-t, T)$  given  $S(0) = x$  is Normal with mean and variance, respectively,

$$\begin{aligned}\mu_T(t | x) &= \mathbb{E}(A(T-t, T) | S(0) = x) = x(e^t - 1)e^{-T}, \\ v_T(t | x) &= \text{Var}(A(T-t, T) | S(0) = x) = v(t) - e^{-2T}(e^t - 1)^2,\end{aligned}$$

as follows from standard formulae for conditional Normal distributions (cf. [81, Section 4.3]).

Also, rewrite  $\kappa_n(T)$  as

$$\frac{\mathbb{P}(Q^n(T) > nq | Q^n(0) > np)}{\mathbb{P}(Q^n(0) > np)}.$$

The decay rate of the probability in the denominator is given by (2.7). Now focus on the decay rate of the probability in the numerator. Realize that, as the condition  $Q^n(0) > np$  is binding, the most likely path (in the ‘Schilder sense’) must be such that the traffic rate at time 0 is  $c$  (which means that the aggregate input process is generating traffic at a rate  $nc$ ); otherwise the queue grows even beyond  $np$ . Also notice that the most likely path is such that the buffer has been empty between 0 and  $T$ . These observations, in conjunction with the Markovian nature of the rate process of iOU, entail that all the information about the system at time 0 that has impact on the system at time  $T$ , is contained in the fact that the rate is (most likely)  $nc$  at time 0. To find the decay rate of  $\mathbb{P}(Q^n(T) > nq | Q^n(0) > np)$ , we therefore have to solve

$$\inf_{t>0} \frac{(q + ct - \mu_T(t | c))^2}{2v_T(t | c)}.$$

The above formulae for the conditional mean and variance entail that this optimization problem reduces to

$$\inf_{t>0} \left( \frac{(q + ct)^2}{2v(t)} - \frac{(q + ct)c(e^t - 1)e^{-T}}{v(t)} + o(e^{-T}) \right).$$

Applying Equation (2.29) once again, inserting (2.10), and using that

$$v'(t) = 1 - e^{-t} = e^{-t}(e^t - 1),$$

we indeed obtain that  $\kappa(T)$  equals  $2c^2e^{-(T-t^*)} + o(e^{-T})$ , as expected.

The above reasoning explains why the decay rate does not depend on  $p$ ; as an aside we mention that also  $\ell(s^*, t^*)$  does not depend on  $p$ .

*D. Other regimes.* In this chapter we have focused on the metric  $R(T|p, q)$  under the many-sources scaling, and that was intended to express the level of correlation between the workloads at time 0 and  $T$ . Then we studied the asymptotics of  $R(T|p, q)$  for large  $T$ . Evidently, many other measures for correlation can be thought of. One could for instance consider similar measures, but then in the large-buffer regime.

In this respect, we could consider a queue fed by a single Gaussian input, emptied at a constant rate  $c > 0$ . Then an interesting measure could be, for fixed  $p, q$ , and  $T$ ,

$$R(TB|pB, qB) = \frac{\mathbb{P}(Q(0) > pB, Q(TB) > qB)}{\mathbb{P}(Q(0) > pB)\mathbb{P}(Q(TB) > qB)} = \frac{\mathbb{P}(Q(0) > pB, Q(TB) > qB)}{\mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB)},$$

and its asymptotics for large  $B$ . The analysis of  $R(TB|pB, qB)$  is radically different from that of  $\kappa_n(T)$ ; the reason for this is that in the many-sources regime the most likely timescales to overflow are more or less constant in the scaling parameter (i.e.,  $n$ ), whereas in the large-buffer one would expect that these timescales are roughly proportional to the scaling parameter (i.e.,  $B$ ).

In this case, when analyzing  $\mathbb{P}(Q(0) > pB, Q(TB) > qB)$ , we expect different regimes. More precisely: for  $B$  large it is not always true that, in the most likely scenario, both constraints are tightly met; for some values of  $p, q, T$  this will be the case, while for others just one constraint will be tightly met (and the other event ‘comes for free’). In case both constraints are tightly met, again two cases can be distinguished: a first in which the queue has not become empty between 0 and  $TB$  (which we expect is the case for  $T$  smaller than some critical timescale  $T^*$ ), and a second in which epochs 0 and  $TB$  lie in different busy periods (for  $T$  larger than  $T^*$ ), cf. [108, Section 11.2].

In Chapter 3, we will consider the large buffer scaling and we will make rigorous the heuristics stated above.

## Appendix

### 2.A Proof of an auxiliary result

In this appendix we prove a lemma that is needed to establish Proposition 2.4.3. We first determine the closure of the set  $\mathcal{S}_T$ . We define

$$\begin{aligned} \mathcal{S}(s) &:= \{f \in \Omega : -f(-s) > p + cs\}; \\ \mathcal{S}_T(t) &:= \{f \in \Omega : f(T) - f(T-t) > q + ct\}; \end{aligned}$$

also

$$\begin{aligned}\overline{\mathcal{S}(s)} &:= \{f \in \Omega : -f(-s) \geq p + cs\}; \\ \overline{\mathcal{S}_T(t)} &:= \{f \in \Omega : f(T) - f(T-t) \geq q + ct\}.\end{aligned}$$

Notice that evidently

$$\mathcal{S}_T = \bigcup_{s,t>0} (\mathcal{S}(s) \cap \mathcal{S}_T(t)).$$

**Lemma 2.A.1.** *For any  $T$ , we have that the closure  $\overline{\mathcal{S}_T}$  of  $\mathcal{S}_T$  is given by*

$$\bigcup_{s,t>0} (\overline{\mathcal{S}(s)} \cap \overline{\mathcal{S}_T(t)}).$$

*Proof.* The proof is similar to those in [86, 94]. We prove both inclusions separately.

- We first show the inclusion “ $\subseteq$ ”. For any  $f \in \overline{\mathcal{S}_T}$  there exists a sequence  $f_n \in \mathcal{S}_T$  such that  $\|f_n - f\|_\Omega \rightarrow 0$  as  $n \rightarrow \infty$ . Now since  $f_n \in \mathcal{S}_T$  there is an  $s_n > 0$  and a  $t_n > 0$  such that  $f_n \in \mathcal{S}(s_n) \cap \mathcal{S}_T(t_n)$ , so that we have  $-f_n(-s_n) > p + cs_n$  and  $f_n(T) - f_n(T - t_n) > q + ct_n$ . The sequence  $s_n$  is bounded, because, if not, we would have a subsequence satisfying

$$\begin{aligned}0 &= \lim_{n \rightarrow \infty} \|f - f_n\|_\Omega \geq \lim_{n \rightarrow \infty} \frac{f(-s_n) - f_n(-s_n)}{1 + s_n} \\ &\geq \lim_{n \rightarrow \infty} \left( \frac{f(-s_n)}{1 + s_n} + \frac{p + cs_n}{1 + s_n} \right) = c,\end{aligned}$$

(use that  $f \in \Omega!$ ), which gives a contradiction (recall that  $c > 0$ ). Along the same lines it can be shown that  $t_n$  is bounded. Hence there are subsequences  $s_{n_k} \rightarrow s_0$  and  $t_{n_k} \rightarrow t_0$ , for finite  $s_0$  and  $t_0$ . We conclude that for large enough  $k$

$$-f_{n_k}(-s_0) \geq p + cs_0 \quad \text{and} \quad f_{n_k}(T) - f_{n_k}(T - t_0) \geq q + ct_0.$$

We conclude that

$$f \in \left( \overline{\mathcal{S}(s_0)} \cap \overline{\mathcal{S}_T(t_0)} \right) \subseteq \left( \overline{\mathcal{S}(s_0)} \cap \overline{\mathcal{S}_T(t_0)} \right).$$

- For the other inclusion, “ $\supseteq$ ”, let

$$f \in \bigcup_{s,t>0} (\overline{\mathcal{S}(s)} \cap \overline{\mathcal{S}_T(t)}).$$

Then there exist  $s_0, t_0 > 0$  such that  $f \in \overline{\mathcal{S}(s_0)} \cap \overline{\mathcal{S}_T(t_0)}$ . Let  $\eta(\cdot)$  be a path in  $\mathbb{R}$  that is strictly increasing and taking negative values for  $u \in (-\infty, 0)$  and

positive values for  $u \in (0, \infty)$  (for instance  $\eta(u) := \operatorname{sgn}(u)\sqrt{|u|}$  or  $\arctan u$ ). Define

$$f_n(u) := f(u) + \frac{\eta(u)}{n}.$$

Then  $f_n \in \mathcal{S}(s_0) \cap \mathcal{S}_T(t_0)$  as

$$-f_n(-s_0) = -f(-s_0) - \frac{\eta(-s_0)}{n} \geq p + cs_0 - \frac{\eta(-s_0)}{n} > p + cs_0$$

and

$$\begin{aligned} f_n(T) - f_n(T - t_0) &= f(T) - f(T - t_0) + \frac{\eta(T) - \eta(T - t_0)}{n} \\ &\geq q + ct_0 + \frac{\eta(T) - \eta(T - t_0)}{n} > q + ct_0. \end{aligned}$$

Moreover, we have, for  $n \rightarrow \infty$ , that  $\|f_n - f\|_\Omega \rightarrow 0$  (use that  $\eta \in \mathcal{R} \subset \Omega$ ), and hence

$$f \in \left( \overline{\mathcal{S}(s_0) \cap \mathcal{S}_T(t_0)} \right) \subseteq \left( \overline{\mathcal{S}(s_0)} \cap \overline{\mathcal{S}_T(t_0)} \right) \subseteq \overline{\mathcal{S}_T}.$$

This proves the second inclusion. □

## Chapter 3

---

# Gaussian queue under large buffer scaling

While in Chapter 2 we considered a queue fed by Gaussian input under the many sources scaling, in the present chapter we study the same queue under a different scaling regime, *viz.*, the so-called large buffer scaling. The following section gives a short review of the results obtained in Chapter 2 and describes heuristically different regimes that are envisaged under the scaling considered in this chapter.

### 3.1 Introduction

Over the past decade a substantial research effort has been devoted to the analysis of queues with Gaussian input [80, 88, 93]. It is noted, however, that the vast majority of papers on these *Gaussian queues* address issues related to the corresponding *steady-state* distribution. These results are predominantly of an asymptotic nature, in that they identify the tail asymptotics [48, 63, 89, 92]. Importantly, however, so far hardly any attention has been paid to *transient* properties. A notable exception is the recent article [53], see also Chapter 2, where asymptotics of transient probabilities under a so-called many-sources scaling were found (for specific Gaussian inputs).

In more detail, in Chapter 2 the following model was considered. A queue is fed by  $n$  i.i.d. Gaussian processes with stationary increments, and emptied at a constant rate  $nc$  (with  $c$  large enough to ensure stability). With  $Q^n(t)$  denoting the buffer content at time  $t$ , the logarithmic asymptotics

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(Q^n(0) > np, Q^n(T) > nq)$$

were determined for  $T$  large (assuming the queue is in stationarity at time 0). A crucial element in the reasoning is that for  $T$  large enough, the time epochs 0 and  $T$  lie in separate busy periods, thus simplifying the analysis substantially. A conclusion drawn in Chapter 2 is that the correlation structure of the input process essentially carries over to the workload process.

In the present chapter we consider a different scaling, *viz.* the so-called *large-buffer scaling*. Then the queue is fed by just a single Gaussian process with stationary increments (with the associated variance curve denoted by  $v(\cdot)$ ), and emptied at a

constant rate  $c$ . With  $Q(t)$  denoting the buffer content at time  $t$ , the first goal of this chapter is to determine the decay rate

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > pB, Q(TB) > qB). \quad (3.1)$$

Interestingly, in view of earlier work, see e.g. [82] and [108, Section 11.7], multiple regimes are envisaged depending on the value of  $T$ . For small values of  $T$ , typically one of the events  $\{Q(0) > pB\}$  and  $\{Q(TB) > qB\}$  will essentially imply the other; in the sequel we call this regime (A). For instance if  $p$  is substantially larger than  $q$  (and  $T$  small), then it is likely that (3.1) equals the decay rate of just  $\mathbb{P}(Q(0) > pB)$  — we say that in this case the event  $\{Q(0) > pB\}$  is ‘tight’. Likewise, if  $q$  is substantially larger than  $p$ , then we expect that only  $\{Q(TB) > qB\}$  is tight. Then there is an intermediate range of values of  $T$ , regime (B), for which it is to be expected that both events  $\{Q(0) > pB\}$  and  $\{Q(TB) > qB\}$  are tight, but that the time epochs 0 and  $T$  lie in the same busy period with overwhelming probability. Finally, for large  $T$  still both events are tight, but now they occur in different busy periods with overwhelming probability; to this regime we refer as regime (C). A second goal of the chapter is to make the above statements rigorous.

The remainder of this chapter is organized as follows. In Section 3.2 we present the model and give a problem description. Then Section 3.3 introduces additional notation, and we establish a useful reduction property. Our first main result, namely an explicit representation of the decay rate (3.1), is given in Section 3.4. The cases of short-range dependent and long-range dependent input are dealt with in Section 3.5; in both cases the regimes (A), (B), and (C) are studied.

## 3.2 Model and problem description

Let  $\{X(t) : t \in \mathbb{R}\}$  be a Gaussian process with *stationary increments* and a.s. continuous sample paths, starting off at 0 (that is,  $X(0) = 0$ , a.s.). Without loss of generality we assume that the process is *centered*, i.e.,  $\mathbb{E}X(t) = 0$  for any  $t$ . Furthermore, the variance function is given through  $v(t) := \mathbb{V}\text{ar}X(t)$ .

Throughout the present chapter we impose the following assumption.

**Assumption 3.2.1.**  $v(\cdot)$  is continuous, and regularly varying (at  $\infty$ ) of index  $\alpha \in (0, 2)$ .

In this chapter we analyze a queue fed by input process  $X(\cdot)$ , emptied at a constant rate  $c > 0$ . More formally, we define the steady-state buffer content process  $\{Q(t) : t \geq 0\}$  by the following representation:

$$Q(t) = \sup_{s \geq 0} (A(t-s, t) - cs) \stackrel{d}{=} Q_e = \sup_{s \geq 0} (A(-s, 0) - cs), \quad (3.2)$$

where  $A(s, t)$  for  $s \leq t$ , is to be interpreted as the amount of traffic having entered the system between  $s$  and  $t$ .

As mentioned in Section 3.1, this chapter focuses on analyzing transient properties of the buffer content process, or more specifically, we wish to determine, under Assumption 3.2.1, the asymptotics of

$$\begin{aligned} N(B) &\equiv N_{p,q,T}(B) := \mathbb{P}(Q(0) > pB, Q(TB) > qB) \\ &= \mathbb{P}(\exists s \geq 0 : A(-s, 0) > pB + cs, \exists t \geq 0 : A(TB - t, TB) > qB + ct); \end{aligned} \quad (3.3)$$

for  $B$  large and  $p, q, T > 0$  given (the latter identity follows from a direct interpretation of the definition of the supremum in (3.2)).

For the univariate case these logarithmic asymptotics are known (and in fact even the *exact* asymptotics have been found); these are (roughly) Weibullian:

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_e > B) = -\frac{1}{2} \left( \frac{2}{2-\alpha} \right)^{2-\alpha} \left( \frac{2c}{\alpha} \right)^\alpha. \quad (3.4)$$

We refer to, e.g., [36]; studies on the accuracy of the resulting approximations are, e.g., [4, 85].

In Section 3.4 it will turn out that the nature of the decay rate (3.1) crucially depends on the values of  $p$ ,  $q$ , and  $T$ . Typically, we will have that for  $p$  and  $q$  given and  $T$  small the joint asymptotics (3.1) reduce to the one-dimensional asymptotics; in light of (3.4) this means that for  $p > q$  and  $T$  small, we have

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > pB, Q(TB) > qB) = -\frac{1}{2} \left( \frac{2p}{2-\alpha} \right)^{2-\alpha} \left( \frac{2c}{\alpha} \right)^\alpha,$$

while for  $q > p$ , we have the same result but with  $p$  replaced by  $q$ . We will, for any pair  $(p, q)$ , show in Section 3.5 that the joint asymptotics reduce to one-dimensional asymptotics if and only if  $T$  is smaller than some threshold (being the unique solution of an explicit equation). For  $T$  larger than this threshold, we may have two types of behavior: the queue can have been empty (with overwhelming probability) or not. Typically, when  $T$  is large it is more likely that the buffer content first reaches  $pB$  at time 0, then drops to 0, and only just before  $TB$  increases again, to reach level  $qB$  at time  $TB$ ; for smaller  $T$  (with overwhelming probability) the queue has not been empty between 0 and  $TB$ . In Section 3.5 we will explicitly give a threshold above which time 0 and time  $TB$  lie in separate busy periods (with overwhelming probability).

### 3.3 Notation and preliminaries

In this section we first derive a useful reduction property. We then introduce the notation that we use throughout the chapter.

#### 3.3.1 Reduction property

The following result appears to be useful later on. After the proof, we also give a more intuitive reasoning why it is valid. Let

$$\mathcal{E}_T := \{(s, t) : s \geq 0, t \in [0, T) \cup \{T + s\}\}.$$

**Lemma 3.3.1.** *For any  $p, q, T > 0$ ,*

$$\begin{aligned} & \mathbb{P}(\exists s \geq 0, t \geq 0 : A(-s, 0) - cs > p, A(T - t, T) - ct > q) \\ &= \mathbb{P}(\exists (s, t) \in \mathcal{E}_T : A(-s, 0) - cs > p, A(T - t, T) - ct > q). \end{aligned}$$

*Proof.* Let  $\check{s}$  be the optimizer in  $\sup_{s \geq 0} (A(-s, 0) - cs)$ . Also,

$$\begin{aligned} \mathcal{A}_T &:= \{\exists (s, t) \in \mathcal{E}_T : A(-s, 0) - cs > p, A(T - t, T) - ct > q\}, \\ \mathcal{A} &:= \{\exists (s, t) \in \mathbb{R}_+^2 : A(-s, 0) - cs > p, A(T - t, T) - ct > q\}. \end{aligned}$$

We prove the stated by showing  $\mathcal{A}_T = \mathcal{A}$ . As  $\mathcal{A}_T \subseteq \mathcal{A}$ , it is left to show  $\mathcal{A}_T \supseteq \mathcal{A}$ .

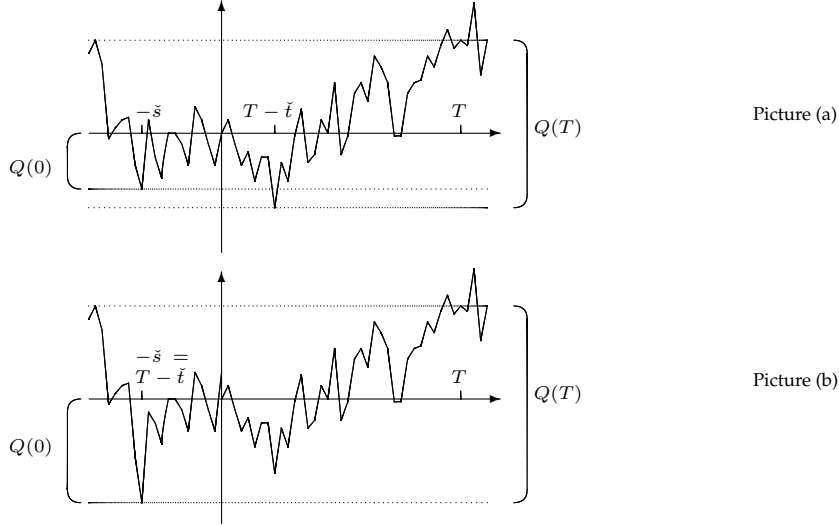
Take a realization from  $\mathcal{A}$  and suppose for  $t \in [T, T + \check{s}) \cup (T + \check{s}, \infty)$  we have that  $A(T - t, T) - ct > q$  (as for all other  $t$  the claimed is clear). Then also, by definition of  $\check{s}$ ,

$$\begin{aligned} A(-\check{s}, T) - c(T + \check{s}) &= (A(-\check{s}, 0) - c\check{s}) + (A(0, T) - cT) \\ &\geq (A(T - t, 0) - c(t - T)) + (A(0, T) - cT) = A(T - t, T) - ct > q. \end{aligned}$$

Hence the realization was also in  $\mathcal{A}_T$ , which proves the stated.  $\square$

**Remark 3.3.2.** An alternative, more intuitive but essentially equivalent, line of reasoning is the following. Let  $\check{t}$  be the optimizer in  $\sup_{t \geq 0} A(T - t, T) - ct$ . The optimizers  $\check{s}$  and  $\check{t}$  can be interpreted as the starting epochs of the busy periods in which 0 and  $T$ , respectively, are contained, see Figure 3.1.

- It is clear that  $\check{t}$  cannot lie in  $(T, T + \check{s})$ : it cannot be that a busy period starts in  $(-\check{s}, 0)$ , as the buffer has been non-empty in this interval all the time (since the busy period in which 0 is contained started at  $\check{s}$ ).
- Similarly,  $\check{t}$  cannot lie in  $(T + \check{s}, \infty)$ : it cannot be that a busy period starts before  $\check{s}$  and lasts till at least  $T$ , as the buffer was empty just before  $\check{s}$  (since a busy period started at  $\check{s}$ ).  $\spadesuit$



**Figure 3.1:** Proof of Lemma 3.3.1. In picture (a) the busy period in which time  $T$  is contained starts after time  $0$ ; in picture (b) the busy periods in which time  $0$  and time  $T$  are contained start at the same moment. Here  $Q(u) := \sup_{v \leq u} (A(v, u) - c(u - v))$ .

The following corollary is an immediate consequence of Lemma 3.3.1. It means that we can restrict ourselves to  $(s, t) \in \mathcal{D}_B$  rather than  $\mathbb{R}^2$  when analyzing  $N(B)$ .

**Corollary 3.3.3.** With  $\mathcal{D}_B := \mathcal{E}_{TB}$ ,

$$N(B) = \mathbb{P}(\exists (s, t) \in \mathcal{D}_B : A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB).$$

### 3.3.2 Notation

In the sequel we extensively use the following Gaussian processes:

$$Y_B(s) := \frac{A(-s, 0)}{pB + cs}; \quad Z_B(t) \equiv Z_{B,T}(t) := \frac{A(TB - t, TB)}{qB + ct};$$

observe that neither  $Y_B(\cdot)$  nor  $Z_B(\cdot)$  has stationary increments. Define the ‘standard deviation curve’ by  $\sigma(s) := \sqrt{v(s)}$ . Also

$$\sigma_Y(s) := \sqrt{\text{Var}Y_B(s)} = \frac{\sigma(s)}{pB + cs}; \quad \sigma_Z(t) := \sqrt{\text{Var}Z_B(t)} = \frac{\sigma(t)}{qB + ct}.$$

Notice that  $\sigma_Y(s), \sigma_Z(t)$  depend on  $p, q$  and  $B$ , but not on  $T$ . Furthermore, we define

$$\gamma(s, t) \equiv \gamma_{B,p,q}(s, t) = \min \left\{ \frac{\sigma_Y(s)}{\sigma_Z(t)}, \frac{\sigma_Z(t)}{\sigma_Y(s)} \right\}.$$

We also define the correlation between  $Y_B(s)$  and  $Z_B(t)$ , which does not depend on  $p$  and  $q$ :

$$r(s, t) \equiv r_{B,T}(s, t) = \text{Corr}(Y_B(s), Z_B(t)) = \frac{\text{Cov}(A(-s, 0), A(TB - t, TB))}{\sigma(s)\sigma(t)}.$$

Realizing that  $v(-s) = v(s)$ , it is readily checked that for  $t \in (0, TB) \cup \{TB + s\}$

$$r(s, t) = \frac{1}{2} \frac{v(TB + s) + v(TB - t) - v(TB) - v(TB - t + s)}{\sigma(s)\sigma(t)}.$$

A crucial role will be played by the function

$$\begin{aligned} \xi_{X;B}(s, t) &\equiv \xi_{X;B,p,q,T}(s, t) \\ &:= \frac{1}{2 \min\{\sigma_Y^2(s), \sigma_Z^2(t)\}} \left( 1 + \frac{(\gamma(s, t) - r(s, t))^2}{1 - r^2(s, t)} I(s, t) \right), \end{aligned}$$

with  $I(s, t) := 1_{\{r(s, t) < \gamma(s, t)\}}$ . As will appear later on, it turned out practical to add the subscript 'X' that indicates the underlying Gaussian process (that in turn defines the processes  $Y_B$  and  $Z_B$ ).

### 3.4 General results

The following general result can be deduced. It is a generalization of the one-dimensional logarithmic asymptotics of [36], and extension of [97], where the two-dimensional logarithmic asymptotics for the class of centered Gaussian processes was considered. The only assumption required is that the variance curve is regularly varying at  $\infty$ . Let  $\mathbb{B}_\alpha(\cdot)$  denote (standard) fBm with Hurst parameter  $H = \alpha/2$ , i.e., a Gaussian process with stationary increments and variance curve  $v(t) = t^{2H}$ .

**Theorem 3.4.1.** *Assume that  $\{X(t) : t \in \mathbb{R}\}$  satisfies Assumption 3.2.1 with  $\alpha \in (0, 2)$ . Then for each  $p, q, T > 0$ ,*

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) = - \inf_{s \geq 0} \inf_{t \in [0, T] \cup \{T+s\}} \xi_{\mathbb{B}_\alpha; 1}(s, t).$$

Notice that the above theorem entails that, under Assumption 3.2.1, the bivariate asymptotics of  $N(B)$  reduce to the bivariate asymptotics of a queue with fBm input. In the remainder of this section we present the complete proof of Theorem 3.4.1. We start by establishing a lemma that is also of independent interest.

**Lemma 3.4.2.** *For arbitrary  $0 < \underline{\varepsilon} < \bar{\varepsilon} < \infty$ ,*

(i) *Uniformly in*  $s \in [\underline{\varepsilon}, \bar{\varepsilon}]$ , *as*  $B \rightarrow \infty$ ,

$$\sigma_Y^2(sB) \frac{B^2}{v(B)} \rightarrow \frac{s^\alpha}{(cs + p)^2};$$

(ii) *Uniformly in*  $t \in [\underline{\varepsilon}, \bar{\varepsilon}]$ , *as*  $B \rightarrow \infty$ ,

$$\sigma_Z^2(tB) \frac{B^2}{v(B)} \rightarrow \frac{t^\alpha}{(ct + q)^2};$$

(iii) *Uniformly in*  $(s, t) \in [\underline{\varepsilon}, \bar{\varepsilon}]^2$ , *as*  $B \rightarrow \infty$ ,

$$\gamma(sB, tB) \rightarrow \min \left\{ \frac{s^{\alpha/2}/(p + cs)}{t^{\alpha/2}/(q + ct)}, \frac{t^{\alpha/2}/(q + ct)}{s^{\alpha/2}/(p + cs)} \right\};$$

(iv) *Uniformly in*  $(s, t) \in [\underline{\varepsilon}, \bar{\varepsilon}]^2$ , *as*  $B \rightarrow \infty$ ,

$$r(sB, tB) \rightarrow \frac{(T + s)^\alpha - T^\alpha + |T - t|^\alpha - |T - t + s|^\alpha}{2s^{\alpha/2}t^{\alpha/2}}.$$

*Proof.* The proof of Lemma 3.4.2 follows straightforwardly from Assumption 3.2.1, combined with standard properties of regularly varying functions.  $\square$

**Lemma 3.4.3.** *For each*  $0 < \underline{\varepsilon} < \bar{\varepsilon} < \infty$ ,

$$\xi_{X;B}(sB, tB) \cdot \frac{v(B)}{B^2} \rightarrow \xi_{\mathbb{B}_\alpha;1}(s, t)$$

*as*  $B \rightarrow \infty$  *uniformly in*  $(s, t) \in [\underline{\varepsilon}, \bar{\varepsilon}]^2$ .

*Proof.* The claim follows from applying Lemma 3.4.2 to the definition of  $\xi_{X;B}(s, t)$ .  $\square$

**Lemma 3.4.4.** *For each*  $0 < \underline{\varepsilon} < \bar{\varepsilon} < \infty$ ,

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P} \left( \begin{array}{l} A(-sB, 0) - csB > pB; \\ A(TB - tB, TB) - ctB > qB \end{array} \right) = -\xi_{\mathbb{B}_\alpha;1}(s, t)$$

*uniformly in*  $(s, t) \in [\underline{\varepsilon}, \bar{\varepsilon}]^2$ .

*Proof.* Follows from the combination of classical asymptotics of the bivariate Normal random variable, in conjunction with Lemma 3.4.3; see Equation (3) in [97], [75, Appx. A] and the references therein, and also Example 4.1.9 in [80].  $\square$

Corollary 3.3.3 indicated that we can restrict ourselves, when analyzing  $N(B)$ , to  $s \geq 0$  and  $t \in [0, TB) \cup \{TB + s\}$ . The following lemma is useful in that we can restrict ourselves, for  $B$  large, even further, viz. to finite  $s$  and  $t$  that are bounded away from zero. This property will appear to be useful later on when applying the standard inequalities for suprema of Gaussian processes. We first introduce some useful additional notation. For given  $0 < \underline{\varepsilon} < \bar{\varepsilon}$  (where  $\underline{\varepsilon} < T$ ), we let

$$\mathcal{C}_B := \{(s, t) : s \in [\underline{\varepsilon}B, \bar{\varepsilon}B], t \in [\underline{\varepsilon}B, TB) \cup \{TB + s\}\}.$$

**Lemma 3.4.5.** *There exist  $\bar{\varepsilon} > \underline{\varepsilon} > 0$  such that*

$$N(B) = \mathbb{P} \left( \exists (s, t) \in \mathcal{C}_B : A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB \right) \cdot (1 + o(1)), \text{ as } B \rightarrow \infty.$$

*Proof.* In view of Corollary 3.3.3 it suffices to establish an upper bound. An obvious inequality is

$$\mathbb{P}(\exists (s, t) \in \mathcal{D}_B : A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB) \leq \pi_1 + \pi_2,$$

where  $\pi_1 \equiv \pi_1(B)$  and  $\pi_2 \equiv \pi_2(B)$  are given through

$$\pi_1 := \mathbb{P}(\exists (s, t) \in \mathcal{C}_B : A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB);$$

$$\pi_2 := \mathbb{P}(\exists (s, t) \in \mathcal{D}_B \setminus \mathcal{C}_B : A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB).$$

Observe that it suffices to show that  $\pi_2 = o(\pi_1)$  as  $B \rightarrow \infty$ . We do so by bounding  $\pi_1$  from below and  $\pi_2$  from above, as follows.

Let  $\bar{\varepsilon} > \underline{\varepsilon} > 0$  be such that  $\bar{s} := \alpha p / ((2 - \alpha)c) \in [\underline{\varepsilon}, \bar{\varepsilon}]$ . Then, by virtue of Lemma 3.4.4, we have

$$\begin{aligned} \log \pi_1 &\geq \log \mathbb{P}(A(-\bar{s}B, 0) - c\bar{s} > pB; A(-\bar{s}B, TB) - c(\bar{s} + T)B > qB) \\ &= -\frac{B^2}{v(B)} \xi_{\mathbb{B}_{\alpha, 1}}(\bar{s}, \bar{s} + T)(1 + o(1)), \end{aligned} \quad (3.5)$$

as  $B \rightarrow \infty$ . Moreover, for each  $B > 0$ , it holds that  $\pi_2 \leq \pi_3 + \pi_4$ , with

$$\begin{aligned} \pi_3 \equiv \pi_3(B) &:= \mathbb{P} \left( \sup_{s \in [0, \underline{\varepsilon}B]} (A(-s, 0) - cs) > pB \right); \\ \pi_4 \equiv \pi_4(B) &:= \mathbb{P} \left( \sup_{s \in [\bar{\varepsilon}B, \infty)} (A(-s, 0) - cs) > pB \right). \end{aligned}$$

By applying Borell's inequality (see, e.g., Adler [5, Theorem 2.1], or, alternatively, see the remark on p. 147, combined with Theorem 1 of [78, Section 12]), we can bound

both probabilities from above. Let us first focus on  $\pi_3$ . For  $B \rightarrow \infty$ ,

$$\begin{aligned} \log \pi_3 &= \log \mathbb{P} \left( \sup_{s \in [0, \underline{\varepsilon}B]} \frac{A(-s, 0)}{cs + pB} > 1 \right) \\ &\leq -\frac{1}{2} \inf_{s \in [0, \underline{\varepsilon}B]} \frac{(cs + pB)^2}{v(s)} (1 + o(1)) \leq -\frac{p^2}{4\underline{\varepsilon}^\alpha} \frac{B^2}{v(B)} (1 + o(1)); \end{aligned}$$

this is due to the fact that  $v(\cdot)$  is regularly varying (and continuous, as we assumed that the sample paths of  $\{X(t) : t \in \mathbb{R}\}$  are continuous), so that  $v(s)$  for  $s \in [0, \underline{\varepsilon}B]$  can be bounded from above by  $2\underline{\varepsilon}^\alpha v(B)$ .

Analogously, for any  $\zeta \leq (2 - \alpha)/2$  and  $B$  sufficiently large,

$$\begin{aligned} \log \pi_4 &\leq -\frac{1}{2} \inf_{s \in [\bar{\varepsilon}B, \infty)} \frac{(cs + pB)^2}{v(s)} (1 + o(1)) \\ &= -\frac{1}{2} \inf_{s \in [\bar{\varepsilon}, \infty)} (cs + p)^2 \frac{v(B)}{v(sB)} \frac{B^2}{v(B)} (1 + o(1)) \\ &\leq -\frac{1}{2} \inf_{s \in [\bar{\varepsilon}, \infty)} (1 - \zeta) \frac{(cs + p)^2}{s^{\alpha + \zeta}} \frac{B^2}{v(B)} (1 + o(1)). \end{aligned}$$

We have now collected all the prerequisites to prove the claim  $\pi_2 = o(\pi_1)$  as  $B \rightarrow \infty$ . First realize that  $p^2/(4\underline{\varepsilon}^\alpha) \rightarrow \infty$ , as  $\underline{\varepsilon} \rightarrow 0$ , and (because  $s^{2-\alpha-\zeta} \rightarrow \infty$  as  $s \rightarrow \infty$ )

$$\inf_{s \in [\bar{\varepsilon}, \infty)} \frac{(cs + p)^2}{s^{\alpha + \zeta}} \rightarrow \infty$$

as  $\bar{\varepsilon} \rightarrow \infty$ . This means that, in order to have  $\pi_2 = o(\pi_1)$ , we can choose  $\bar{\varepsilon} > \underline{\varepsilon} > 0$  such that

$$\xi_{\mathbb{B}_\alpha; 1}(\bar{s}, \bar{s} + T) < \frac{p^2}{4\underline{\varepsilon}^\alpha} \quad \text{and} \quad \xi_{\mathbb{B}_\alpha; 1}(\bar{s}, \bar{s} + T) < \frac{1}{2} \inf_{s \in [\bar{\varepsilon}, \infty)} (1 - \zeta) \frac{(cs + p)^2}{s^{\alpha + \zeta}}.$$

This completes the proof.  $\square$

Before proving Theorem 3.4.1, we first prove a useful lemma.

**Lemma 3.4.6.** *With*

$$\theta(s, t) \equiv \theta_{Y, Z}(s, t) := 1 - r(s, t) \cdot \max\{r(s, t), \gamma(s, t)\}$$

$$\beta(s, t) \equiv \beta_{Y, Z}(s, t) := \max\{r(s, t), \gamma(s, t)\} - r(s, t),$$

it holds for any  $s, t$  that

$$\begin{aligned} &\frac{1}{2} \left( \frac{\theta(s, t) + \beta(s, t)\gamma(s, t)}{\min\{\sigma_Y(s), \sigma_Z(t)\}} \right)^2 \bigg/ \mathbb{E} \left( \left( \frac{\theta(s, t)Y_B(s)}{\sigma_Y(s)} + \frac{\beta(s, t)Z_B(t)}{\sigma_Z(t)} \right)^2 \right) \\ &= \frac{1}{2} \frac{\theta(s, t) + \beta(s, t)\gamma(s, t)}{(1 - r^2(s, t)) (\min\{\sigma_Y(s), \sigma_Z(t)\})^2}. \end{aligned}$$

*Proof.* As we keep  $s$  and  $t$  fixed throughout the proof, we can suppress the dependence on these arguments. Write  $m := \max\{r, \gamma\}$ . First observe

$$\begin{aligned}\Xi &:= \mathbb{E} \left( \frac{\theta Y_B(s)}{\sigma_Y(s)} + \frac{\beta Z_B(t)}{\sigma_Z(t)} \right)^2 \\ &= \frac{\theta^2 \mathbb{E}(Y_B(s))^2}{\sigma_Y(s)^2} + \frac{\beta^2 \mathbb{E}(Z_B(t))^2}{\sigma_Z(t)^2} + 2 \frac{\theta \beta \mathbb{E}(Y_B(s)Z_B(t))}{\sigma_Y(s)\sigma_Z(t)}.\end{aligned}$$

Then it follows that

$$\begin{aligned}\Xi &= \theta^2 + \beta^2 + 2\theta\beta r = (1 - rm)^2 + (m - r)^2 + 2r(1 - rm)(m - r) \\ &= 1 + r^2 m^2 - 2rm + m^2 + r^2 - 2rm + 2rm - 2r^2 - 2r^2 m^2 + 2r^3 m \\ &= 1 - r^2 + m^2 - r^2 m^2 + 2r^3 m - 2rm = (1 - r^2)(1 - 2rm + m^2) \\ &= (1 - r^2)(1 - rm) + (m - r)m = (1 - r^2)(\theta + \beta m).\end{aligned}$$

If  $r \geq \gamma$ , then  $\beta = 0$ , and consequently we have  $\theta + \beta m = \theta + \beta \gamma$ . If  $r < \gamma$ , then  $m = \gamma$ , and hence again  $\theta + \beta m = \theta + \beta \gamma$ . This proves the claim.  $\square$

*Proof of Theorem 3.4.1.* In this proof (and in the sequel), we choose  $\underline{\varepsilon}$  and  $\bar{\varepsilon}$  as indicated in Lemma 3.4.5. We subsequently prove the lower bound and upper bound.

*Lower bound.* We use the argumentation of [93]. An evident lower bound is

$$\begin{aligned}N(B) &\geq \mathbb{P}(\exists(s, t) \in \mathcal{C}_B : A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB) \\ &\geq \sup_{(s, t) \in \mathcal{C}_B} \mathbb{P}(A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB).\end{aligned}$$

Hence, due to Lemma 3.4.4, we have

$$\lim_{B \rightarrow \infty} \log N(B) \cdot \frac{v(B)}{B^2} \geq - \inf_{s \in [\underline{\varepsilon}, \bar{\varepsilon}]; t \in [\underline{\varepsilon}, T) \cup \{T+s\}} \xi_{\mathbb{B}_{\alpha;1}}(s, t).$$

Now it suffices to observe that, for appropriately chosen  $\underline{\varepsilon}, \bar{\varepsilon}$ ,

$$\inf_{s \in [\underline{\varepsilon}, \bar{\varepsilon}]; t \in [\underline{\varepsilon}, T) \cup \{T+s\}} \xi_{\mathbb{B}_{\alpha;1}}(s, t) = \inf_{s \in [0, \infty); t \in [0, T) \cup \{T+s\}} \xi_{\mathbb{B}_{\alpha;1}}(s, t),$$

which follows from the fact that  $\sigma_Y(s) \rightarrow 0$  as  $s \rightarrow 0$  or  $s \rightarrow \infty$ , and  $\sigma_Z(t) \rightarrow 0$  as  $t \rightarrow 0$ .

*Upper bound.* The upper bound is considerably more involved than the lower bound. Due to Lemma 3.4.5 we have

$$\begin{aligned}N(B) &\leq \mathbb{P}(\exists(s, t) \in \mathcal{C}_B : A(-s, 0) - cs > pB, A(TB - t, TB) - ct > qB) (1 + o(1)) \\ &= \mathbb{P}(\exists(s, t) \in \mathcal{C}_B : Y_B(s) > 1, Z_B(t) > 1) (1 + o(1)).\end{aligned}$$

In this proof we need the following notation:

$$\mathcal{D}_B^{(1)} := \{(s, t) \in \mathcal{D}_B : \sigma_Y(s) \leq \sigma_Z(t)\}; \quad \mathcal{D}_B^{(2)} := \{(s, t) \in \mathcal{D}_B : \sigma_Y(s) > \sigma_Z(t)\}.$$

The union bound trivially gives  $\mathbb{P}(\exists(s, t) \in \mathcal{D}_B : Y_B(s) > 1, Z_B(t) > 1) \leq \bar{\pi}_1 + \bar{\pi}_2$ , where

$$\begin{aligned} \bar{\pi}_1 &:= \mathbb{P}\left(\exists(s, t) \in \mathcal{D}_B^{(1)} : Y_B(s) > 1, Z_B(t) > 1\right); \\ \bar{\pi}_2 &:= \mathbb{P}\left(\exists(s, t) \in \mathcal{D}_B^{(2)} : Y_B(s) > 1, Z_B(t) > 1\right). \end{aligned}$$

We subsequently asymptotically analyze  $\bar{\pi}_1$  and  $\bar{\pi}_2$ . The following upper bound on  $\bar{\pi}_1$  is straightforward, as  $\sigma_Y(s) \leq \sigma_Z(t)$  on  $\mathcal{D}_B^{(1)}$ :

$$\begin{aligned} \bar{\pi}_1 &= \mathbb{P}\left(\exists(s, t) \in \mathcal{D}_B^{(1)} : \frac{Y_B(s)}{\sigma_Y(s)} > \frac{1}{\min\{\sigma_Y(s), \sigma_Z(t)\}}, \frac{Z_B(t)}{\sigma_Z(t)} > \frac{\gamma(s, t)}{\min\{\sigma_Y(s), \sigma_Z(t)\}}\right) \\ &= \mathbb{P}\left(\exists(s, t) \in \mathcal{D}_B^{(1)} : \frac{\theta(s, t)Y_B(s)}{\sigma_Y(s)} > \frac{\theta(s, t)}{\min\{\sigma_Y(s), \sigma_Z(t)\}}, \frac{\beta(s, t)Z_B(t)}{\sigma_Z(t)} > \frac{\beta(s, t)\gamma(s, t)}{\min\{\sigma_Y(s), \sigma_Z(t)\}}\right) \\ &\leq \mathbb{P}\left(\exists(s, t) \in \mathcal{D}_B^{(1)} : \frac{\theta(s, t)Y_B(s)}{\sigma_Y(s)} + \frac{\beta(s, t)Z_B(t)}{\sigma_Z(t)} > \frac{\theta(s, t)}{\min\{\sigma_Y(s), \sigma_Z(t)\}} + \frac{\beta(s, t)\gamma(s, t)}{\min\{\sigma_Y(s), \sigma_Z(t)\}}\right) \\ &= \mathbb{P}\left(\exists(s, t) \in \mathcal{D}_B^{(1)} : \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s, t) + \beta(s, t)\gamma(s, t)} \left(\frac{\theta(s, t)Y_B(s)}{\sigma_Y(s)} + \frac{\beta(s, t)Z_B(t)}{\sigma_Z(t)}\right) > 1\right). \end{aligned}$$

We now prove that

$$\mathbb{E}\left(\sup_{(s, t) \in \mathcal{D}_B^{(1)}} \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s, t) + \beta(s, t)\gamma(s, t)} \left(\frac{\theta(s, t)Y_B(s)}{\sigma_Y(s)} + \frac{\beta(s, t)Z_B(t)}{\sigma_Z(t)}\right)\right) \rightarrow 0 \quad (3.6)$$

as  $B \rightarrow \infty$ . This is done as follows. Trivially,

$$\mathbb{E}\left(\sup_{(s, t) \in \mathcal{D}_B^{(1)}} \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s, t) + \beta(s, t)\gamma(s, t)} \left(\frac{\theta(s, t)Y_B(s)}{\sigma_Y(s)} + \frac{\beta(s, t)Z_B(t)}{\sigma_Z(t)}\right)\right) \leq \psi_1 + \psi_2$$

where

$$\begin{aligned} \psi_1 &\equiv \psi_1(B) := \mathbb{E}\left(\sup_{(s, t) \in \mathcal{D}_B^{(1)}} \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s, t) + \beta(s, t)\gamma(s, t)} \frac{\theta(s, t)Y_B(s)}{\sigma_Y(s)}\right); \\ \psi_2 &\equiv \psi_2(B) := \mathbb{E}\left(\sup_{(s, t) \in \mathcal{D}_B^{(1)}} \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s, t) + \beta(s, t)\gamma(s, t)} \frac{\beta(s, t)Z_B(t)}{\sigma_Z(t)}\right). \end{aligned}$$

Then realize that

$$\psi_1 \leq \sup_{(s,t) \in \mathcal{D}_B^{(1)}} \left( \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s,t) + \beta(s,t)\gamma(s,t)} \frac{\theta(s,t)}{\sigma_Y(s)} \right) \mathbb{E} \left( \sup_{(s,t) \in \mathcal{D}_B^{(1)}} Y_B(s) \right),$$

where, due to Lemma 3.4.2,

$$\sup_{(s,t) \in \mathcal{D}_B^{(1)}} \left( \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s,t) + \beta(s,t)\gamma(s,t)} \frac{\theta(s,t)}{\sigma_Y(s)} \right)$$

is bounded from above as  $B \rightarrow \infty$ , and following Lemma 2.2 in [36],

$$\mathbb{E} \left( \sup_{(s,t) \in \mathcal{D}_B^{(1)}} Y_B(s) \right) \rightarrow 0,$$

as  $B \rightarrow \infty$ . Hence  $\psi_1 \rightarrow 0$  as  $B \rightarrow \infty$ . Analogously,  $\psi_2 \rightarrow 0$  as  $B \rightarrow \infty$ . Hence, we have proved (3.6).

The fact that (3.6) applies means that Borell's inequality [5, pages 43-44] yields ( $B$  large)

$$\begin{aligned} \log \bar{\pi}_1 &\leq - \inf_{(s,t) \in \mathcal{D}_B^{(1)}} \frac{1}{2} \frac{(\theta(s,t) + \beta(s,t)\gamma(s,t))^2}{(\min\{\sigma_Y(s), \sigma_Z(t)\})^2 \mathbb{E} \left( \left( \frac{\theta(s,t)Y_B(s)}{\sigma_Y(s)} + \frac{\beta(s,t)Z_B(t)}{\sigma_Z(t)} \right)^2 \right)} \\ &= - \inf_{(s,t) \in \mathcal{D}_B^{(1)}} \frac{1}{2} \frac{\theta(s,t) + \beta(s,t)\gamma(s,t)}{(1 - r^2(s,t)) (\min\{\sigma_Y(s), \sigma_Z(t)\})^2}; \end{aligned}$$

the last step is due to Lemma 3.4.6. The latter expression equals, by virtue of Lemma 3.4.3, as  $B \rightarrow \infty$ ,

$$\begin{aligned} &- \inf_{(s,t) \in \mathcal{D}_B^{(1)}} \frac{1}{2} \frac{\theta(s,t) + \beta(s,t)\gamma(s,t)}{(1 - r^2(s,t)) (\min\{\sigma_Y(s), \sigma_Z(t)\})^2} \\ &= - \inf_{(s,t) \in \mathcal{D}_1^{(1)}} \frac{1}{2} \frac{\theta(sB, tB) + \beta(sB, tB)\gamma(sB, tB)}{(1 - r^2(sB, tB)) (\min\{\sigma_Y(sB), \sigma_Z(tB)\})^2} \\ &= - \frac{B^2}{v(B)} \inf_{(s,t) \in \mathcal{D}_1^{(1)}} \xi_{\mathbb{B}_\alpha;1}(s,t)(1 + o(1)). \end{aligned}$$

Analogously, we have, as  $B \rightarrow \infty$ ,

$$\log \bar{\pi}_2 \leq - \frac{B^2}{v(B)} \inf_{(s,t) \in \mathcal{D}_1^{(2)}} \xi_{\mathbb{B}_\alpha;1}(s,t)(1 + o(1)).$$

We conclude, as  $B \rightarrow \infty$ ,

$$\begin{aligned} &\frac{v(B)}{B^2} \log \mathbb{P}(\exists(s,t) \in \mathcal{D}_B : Y_B(s) > 1, Z_B(t) > 1) \\ &\leq \frac{v(B)}{B^2} \log(\bar{\pi}_1 + \bar{\pi}_2) \leq \frac{v(B)}{B^2} \log(2 \max\{\bar{\pi}_1, \bar{\pi}_2\}) = - \inf_{(s,t) \in \mathcal{D}_1} \xi_{\mathbb{B}_\alpha;1}(s,t)(1 + o(1)). \end{aligned}$$

This completes the proof.  $\square$

**Remark 3.4.7.** Using a different approach, based on Schilder's theorem, we can give a different representation for the rate function

$$\inf_{s \geq 0} \inf_{t \in [0, T) \cup \{T+s\}} \xi_{\mathbb{B}_\alpha; 1}(s, t)$$

in Theorem 3.4.1.

Assume that  $X(t) = \mathbb{B}_\alpha(t)$  is a fractional Brownian motion with Hurst parameter  $\alpha/2$ . It appears that the self-similar structure of fBm enables, for this special case, a rather straightforward proof of Theorem 3.4.1. First observe that

$$\begin{aligned} N(B) &= \mathbb{P}(\exists s, t \geq 0 : A(-sB, 0) > (p + cs)B, : A(TB - tB, TB) > (q + ct)B) \\ &= \mathbb{P}\left(\exists s \geq 0 : \frac{A(-sB, 0)}{B} > p + cs, \exists t \geq 0 : \frac{A(TB - tB, TB)}{B} > q + ct\right) \\ &\stackrel{(i)}{=} \mathbb{P}\left(\exists s \geq 0 : \frac{A(-s, 0)}{B^{1-\alpha/2}} > p + cs, \exists t \geq 0 : \frac{A(T - t, T)}{B^{1-\alpha/2}} > q + ct\right) \\ &= \mathbb{P}\left(\exists s \geq 0 : \frac{A(-s, 0)}{p + cs} > B^{1-\alpha/2}, \exists t \geq 0 : \frac{A(T - t, T)}{q + ct} > B^{1-\alpha/2}\right), \end{aligned}$$

where in equality (i) the self-similarity has been used. We are now in a position to apply the Schilder-type sample-path large deviations [16, 80]. To this end, define the set of paths causing overflow over level  $p$  at time 0, and over level  $q$  at time  $T$ , as follows:

$$\mathcal{S}^0 := \bigcup_{s \geq 0} \mathcal{S}_s^0; \quad \mathcal{S}^T := \bigcup_{t \geq 0} \mathcal{S}_t^T,$$

where  $\mathcal{S}_s^0 := \{f \mid -f(-s) > p + cs\}$  and  $\mathcal{S}_t^T := \{f \mid f(T) - f(T - t) > q + ct\}$ . We also define the set of paths in the intersection of these events:

$$\begin{aligned} \mathcal{S}^{0,T} &:= \{f \mid \exists s \geq 0 : -f(-s) > p + cs; \exists t \geq 0 : f(T) - f(T - t) > q + ct\} \\ &= \bigcup_{s \geq 0} \bigcup_{t \geq 0} \mathcal{S}_{s,t}^{0,T} = \mathcal{S}^0 \cap \mathcal{S}^T. \end{aligned}$$

Now let  $X(t)$  satisfy Assumption 3.2.1 with  $\alpha \in (1, 2)$ . Schilder's theorem 2.2.5 combined with Theorem 3.4.1 entails the following result (as  $B \rightarrow \infty$ ):

$$\begin{aligned} -\frac{v(B)}{B^2} \log N(B) &\rightarrow \inf_{f \in \mathcal{S}^{0,T}} \mathbb{I}(f) = \inf_{s \geq 0, t \geq 0} \left( \inf_{f \in \mathcal{S}_{s,t}^{0,T}} \mathbb{I}(f) \right) \\ &= \inf_{s \geq 0} \inf_{t \in [0, T) \cup \{T+s\}} \left( \inf_{f \in \mathcal{S}_{s,t}^{0,T}} \mathbb{I}(f) \right). \end{aligned}$$

Here  $\mathbb{I}(f)$  is the rate function of a path  $f$ ; for a detailed introduction and a formal framework, see e.g. [4, 16, 88]. The last equality is due to Lemma 3.3.1. Now consider the evaluation of the inner infimum (for fixed  $s, t$ ). The key observation is that

$$\begin{aligned} \xi(s, t) &:= \inf_{f \in \mathcal{S}_{s,t}^{0,T}} \mathbb{I}(f) \\ &= - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{A(-s, 0)}{\sqrt{n}} \geq p + cs, \frac{A(T-t, T)}{\sqrt{n}} \geq q + ct \right). \end{aligned}$$

In other words:  $\xi(s, t)$ , for given  $s, t \geq 0$ , represents the large deviations rate function of a bivariate Normally distributed random variable. Now [80, Exercise 4.1.9] can be applied, and three cases are to be distinguished:

- If  $r(s, t) \geq \gamma(s, t)$  and  $\sigma_Y^2(s) \leq \sigma_Z^2(t)$ , then only the first requirement is ‘tight’ and  $\xi(s, t)$  is independent of  $t$ :

$$\xi(s, t) = \frac{1}{2} \frac{1}{\sigma_Y^2(s)} = \frac{1}{2} \frac{(p + cs)^2}{v(s)}. \quad (3.7)$$

- If  $r(s, t) \geq \gamma(s, t)$  and  $\sigma_Y^2(s) > \sigma_Z^2(t)$ , then only the first requirement is ‘tight’ and  $\xi(s, t)$  is independent of  $s$ :

$$\xi(s, t) = \frac{1}{2} \frac{1}{\sigma_Z^2(t)} = \frac{1}{2} \frac{(q + ct)^2}{v(t)}. \quad (3.8)$$

- If  $r(s, t) < \gamma(s, t)$ , then, with  $\Gamma_T(s, t) := \text{Cov}(A(-s, 0), A(T-t, T))$ , both requirements are ‘tight’:

$$\begin{aligned} \xi(s, t) &= \frac{1}{2} (p + cs, q + ct) \begin{pmatrix} v(s) & \Gamma_T(s, t) \\ \Gamma_T(s, t) & v(t) \end{pmatrix}^{-1} \begin{pmatrix} p + cs \\ q + ct \end{pmatrix} \\ &= \frac{1}{2} \frac{1}{1 - r^2(s, t)} \left( \frac{(p + cs)^2}{v(s)} - 2 \frac{\Gamma_T(s, t)(p + cs)(q + ct)}{v(t)v(s)} + \frac{(q + ct)^2}{v(t)} \right). \end{aligned} \quad (3.9)$$

Notice that the criterion  $r(s, t) < \gamma(s, t)$  can be rewritten as

$$\frac{\Gamma_T(s, t)}{(p + cs)(q + ct)} < \min\{\sigma_Y^2(s), \sigma_Z^2(t)\}.$$

We thus retrieve

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) = - \inf_{s \geq 0} \inf_{t \in [0, T] \cup \{T+s\}} \xi_{\mathbb{B}_{\alpha;1}}(s, t).$$

♠

**Remark 3.4.8.** It is noted that Theorem 3.4.1 can be extended to any dimension larger than 2, i.e., we can analyze in a similar fashion the decay rates of probabilities of the type

$$\mathbb{P}(Q(0) > p_0 B, Q(T_1) > p_1 B, \dots, Q(T_n) > p_n B),$$

for any  $n = 1, 2, \dots$ ,  $p_i > 0$  (for  $i = 0, \dots, n$ ) and  $T_n > T_{n-1} > \dots > T_1 > 0$ . The key observations are that an analogous reduction property applies, and that a Borell-based proof essentially goes through for  $n = 2, 3, \dots$  ♠

## 3.5 Special cases

In this section we apply Theorem 3.4.1 to two special cases, viz.

- Gaussian input processes which possess a *short-range dependent* (SRD) structure, by which we mean that  $v(\cdot)$  is regularly varying with parameter  $\alpha = 1$ ;
- Gaussian input processes which possess a *long-range dependent* (LRD) structure, by which we mean that  $v(\cdot)$  is regularly varying with parameter  $\alpha \in (1, 2)$ .

In particular, one could think of the following special cases which have been studied intensively in the literature.

- (i) *Integrated Gaussian processes.* In this case  $X(t) = \int_0^t S(s) ds$ , where  $S(\cdot)$  is a centered stationary Gaussian process with a continuous covariance function  $G(t) := \text{Cov}(S(s), S(s+t)) > 0$ . Note that if  $\int_0^\infty G(v) dv < \infty$ , then

$$\text{Var}X(t) = v(t) = 2 \left( \int_0^\infty G(v) dv \right) \cdot t(1 + o(1))$$

as  $t \rightarrow \infty$ , and hence  $X(\cdot)$  has an SRD structure. If  $R(t)$  is regularly varying at  $\infty$  with index  $\alpha - 2$ , for  $\alpha \in (1, 2)$ , then  $\text{Var}X(t)$  is regularly varying at  $\infty$  with index  $\alpha$ , which implies an LRD structure.

- (ii) *Fractional Brownian motions.* Then  $X(t) = B_{\alpha/2}(t)$ . Recall that for the case of  $\alpha = 1$  we are in the SRD scenario, while  $\alpha \in (1, 2)$  corresponds to the LRD case.

The relevance of integrated Gaussian input processes in the theory of fluid models is discussed in e.g. [41, 42]; see also [40, 90]. The use of fractional Brownian motions in modeling input processes has been advocated by e.g. [93, 110].

### 3.5.1 The SRD case

In this section we focus on the class of input processes with a short range dependence structure, i.e., we assume that  $\text{Var}X(t) = v(t)$  is regularly varying at infinity with index  $\alpha = 1$ .

**Proposition 3.5.1.** *Assume that  $\{X(t) : t \in \mathbb{R}\}$  satisfies Assumption 3.2.1 with  $\alpha = 1$ .*

(i) *If  $p > q > 0$ , then*

$$\begin{aligned} & \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) \\ &= - \begin{cases} 2pc & \text{if } T \leq \frac{p-q}{c}; \\ 2pc + \frac{(cT + q - p)^2}{2T} & \text{if } \frac{p-q}{c} < T \leq \frac{(\sqrt{p} + \sqrt{q})^2}{c}; \\ 2pc + 2qc & \text{if } T > \frac{(\sqrt{p} + \sqrt{q})^2}{c}. \end{cases} \end{aligned} \quad (3.10)$$

(ii) *If  $p = q > 0$ , then*

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) = - \begin{cases} 2pc + \frac{c^2 T}{2} & \text{if } T \leq \frac{4p}{c}; \\ 4pc & \text{if } T > \frac{4p}{c}. \end{cases} \quad (3.11)$$

(iii) *If  $q > p > 0$ , then*

$$\begin{aligned} & \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) \\ &= - \begin{cases} 2qc & \text{if } T \leq \frac{q-p}{c}; \\ 2pc + \frac{(cT + q - p)^2}{2T} & \text{if } \frac{q-p}{c} < T \leq \frac{(\sqrt{p} + \sqrt{q})^2}{c}; \\ 2pc + 2qc & \text{if } T > \frac{(\sqrt{p} + \sqrt{q})^2}{c}. \end{cases} \end{aligned} \quad (3.12)$$

*Proof.* By virtue of Theorem 3.4.1, we analyze

$$\inf_{s \geq 0} \inf_{t \in [0, T] \cup \{T+s\}} \xi_{\mathbb{B}_1;1}(s, t) = \min \left\{ \inf_{s \geq 0} \inf_{t \in [0, T]} \xi_{\mathbb{B}_1;1}(s, t), \inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s+T) \right\}.$$

Note that  $r(s, t) \equiv 0$  for all  $s \geq 0, t \in [0, T]$ , and hence

$$\begin{aligned} \inf_{s \geq 0} \inf_{t \in [0, T]} \xi_{\mathbb{B}_1;1}(s, t) &= \inf_{s \geq 0} \inf_{t \in [0, T]} \frac{1}{2} \left( \frac{(p+cs)^2}{s} + \frac{(q+ct)^2}{t} \right) \\ &= 2pc + \frac{1}{2} \frac{(q + c \min\{T, q/c\})^2}{\min\{T, q/c\}}. \end{aligned} \quad (3.13)$$

Case (i):  $p > q > 0$ . It is convenient to split this scenario into two subcases:  $T \leq (p - q)/c$  and  $T > (p - q)/c$ . Let us first consider  $T \leq (p - q)/c$ . This case follows from combining the fact that for each  $s, t$ ,

$$\xi_{\mathbb{B}_1;1}(s, t) \geq \frac{1}{2 \min\{\sigma_Y^2(s), \sigma_Z^2(t)\}} \geq \frac{1}{2\sigma_Y^2(s^*)} = 2pc,$$

with  $\xi_{\mathbb{B}_1;1}(s^*, s^* + T) = 2pc$  for  $s^* = p/c$ . Then consider  $T > (p - q)/c$ . Let

$$\mathcal{S}_1 := \{s \geq 0 : \sigma_Y(s) \leq \sigma_Z(s + T)\}, \quad \mathcal{S}_2 := \{s \geq 0 : \sigma_Y(s) > \sigma_Z(s + T)\}.$$

Note that  $\{s \geq 0\} = \mathcal{S}_1 \cup \mathcal{S}_2$ . Let us first analyze  $\inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s + T)$ . Note that for each  $s \geq 0$

$$r(s, s + T) = r_{1,T}(s, t) < \gamma_{1,p,q}(s, s + T) = \gamma(s, s + T).$$

Indeed, for  $s \in \mathcal{S}_1$  (using that  $T > (p - q)/c$ ) we have

$$\gamma(s, s + T) - r(s, s + T) = \sqrt{\frac{s}{s + T}} \frac{cT + q - p}{p + cs} > 0$$

while, for  $s \in \mathcal{S}_2$ , we have

$$\begin{aligned} \gamma(s, s + T) - r(s, s + T) &= \sqrt{\frac{s}{s + T}} \left( \frac{(T + s)(p + cs)}{s(q + c(s + T))} - 1 \right) \\ &= \sqrt{\frac{s}{s + T}} \frac{Tp + s(p - q)}{s(q + c(s + T))} > 0. \end{aligned}$$

Hence:

- if  $s \in \mathcal{S}_1$ , then

$$\xi_{\mathbb{B}_1;1}(s, s + T) = \frac{1}{2} \frac{(p + cs)^2}{s} + \frac{1}{2} \frac{(cT + q - p)^2}{T};$$

- if  $s \in \mathcal{S}_2$ , then

$$\begin{aligned} \xi_{\mathbb{B}_1;1}(s, s + T) &= \frac{1}{2} \frac{(q + c(T + s))^2}{T + s} + \frac{1}{2} \frac{(pT + s(p - q))^2}{sT(s + T)} \\ &= \frac{1}{2} \frac{(p + cs)^2}{s} + \frac{1}{2} \frac{(cT + q - p)^2}{T}. \end{aligned}$$

The above implies that

$$\begin{aligned} \inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s + T) &= \inf_{s \geq 0} \frac{1}{2} \frac{(p + cs)^2}{s} + \frac{1}{2} \frac{(cT + q - p)^2}{T} \\ &= 2pc + \frac{1}{2} \frac{(cT + q - p)^2}{T}. \end{aligned} \tag{3.14}$$

Finally, in order to complete the proof of (i), it suffices to check that combination of (3.13) with (3.14) leads to

$$\inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s+T) \leq \inf_{s \geq 0} \inf_{t \in [0, T)} \xi_{\mathbb{B}_1;1}(s, t) \quad \text{for } \frac{p-q}{c} < T \leq \frac{(\sqrt{p} + \sqrt{q})^2}{c},$$

$$\inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s+T) \geq \inf_{s \geq 0} \inf_{t \in [0, T)} \xi_{\mathbb{B}_1;1}(s, t) \quad \text{for } T > \frac{(\sqrt{p} + \sqrt{q})^2}{c}.$$

*Case (ii):*  $p = q > 0$ . This case follows from the same arguments as used in case (i). We omit the details.

*Case (iii):*  $q > p > 0$ . Analogously to case (i) we separately analyze the scenarios  $T \leq (q-p)/c$  and  $T > (q-p)/c$ . First consider  $T \leq (q-p)/c$ . The result directly follows from

$$\xi_{\mathbb{B}_1;1}(s, t) \geq \frac{1}{2 \min\{\sigma_Y^2(s), \sigma_Z^2(t)\}} \geq \frac{1}{2\sigma_Z^2(t^*)} = 2qc,$$

for each  $s, t$ , in conjunction with  $\xi_{\mathbb{B}_1;1}(t^* - T, t^*) = 2qc$  for  $t^* = q/c$ . Then focus on  $T > (q-p)/c$ . Let

$$\mathcal{S}_{21} := \{s \geq 0 : \sigma_Y(s) > \sigma_Z(s+T), r(s, s+T) < \gamma(s, s+T)\},$$

$$\mathcal{S}_{22} := \{s \geq 0 : \sigma_Y(s) > \sigma_Z(s+T), r(s, s+T) \geq \gamma(s, s+T)\}.$$

We analyze  $\inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s+T)$ .

- If  $s \in \mathcal{S}_1$ , then

$$r(s, s+T) = \sqrt{\frac{s}{s+T}} < \sqrt{\frac{s}{s+T}} \frac{c(s+T) + q}{p + cs} = \gamma(s, s+T),$$

and therefore

$$\xi_{\mathbb{B}_1;1}(s, s+T) = \frac{1}{2} \frac{(p+cs)^2}{s} + \frac{1}{2} \frac{(cT+q-p)^2}{T}. \quad (3.15)$$

- If  $s \in \mathcal{S}_{21}$ , then standard calculation leads to the same formula as in (3.15), i.e.,

$$\begin{aligned} \xi_{\mathbb{B}_1;1}(s, s+T) &= \frac{1}{2} \frac{(q+c(T+s))^2}{T+s} + \frac{1}{2} \frac{(pT+s(p-q))^2}{sT(s+T)} \\ &= \frac{1}{2} \frac{(p+cs)^2}{s} + \frac{1}{2} \frac{(cT+q-p)^2}{T}. \end{aligned} \quad (3.16)$$

Hence, using that  $p/c \in \mathcal{S}_{21}$ , we have

$$\inf_{s \in \mathcal{S}_1 \cup \mathcal{S}_{21}} \xi_{\mathbb{B}_1;1}(s, s+T) = 2pc + \frac{1}{2} \frac{(cT+q-p)^2}{T}. \quad (3.17)$$

- If  $s \in \mathcal{S}_{22}$ , then

$$\xi_{\mathbb{B}_1;1}(s, s+T) = \frac{1}{2 \min\{\sigma_Y^2(s), \sigma_Z^2(s+T)\}} = \frac{(q + c(s+T))^2}{2(s+T)}. \quad (3.18)$$

Moreover, the fact that  $s \in \mathcal{S}_{22}$  implies

$$r(s, s+T) \geq \gamma(s, s+T) \Leftrightarrow s \geq \frac{pT}{q-p}.$$

We conclude that

$$\begin{aligned} \inf_{s \in \mathcal{S}_{22}} \xi_{\mathbb{B}_1;1}(s, s+T) &= \xi_{\mathbb{B}_1;1}\left(\frac{pT}{q-p}, \frac{pT}{q-p} + T\right) \\ &= \frac{1}{2} \frac{q(cT + q - p)^2}{(q-p)T}. \end{aligned} \quad (3.19)$$

The comparison of (3.17) with (3.19) now implies that

$$\inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s+T) = 2pc + \frac{1}{2} \frac{(cT + q - p)^2}{T}. \quad (3.20)$$

Analogously to the proof of (i), the combination of (3.13) with (3.20) completes the proof.  $\square$

**Remark 3.5.2.** Related results for queues fed by Brownian motion have recently been obtained in [77]. There also emphasis was put on the nature of the decay rates, and the shape of the *most likely path* towards the rare event [4, 80]. In accordance with Proposition 3.5.1, it was found that for  $T$  up to some threshold, the decay rate of the joint probability equals the decay rate of  $\mathbb{P}(Q_e > \max\{p, q\}B)$ , with  $Q_e$  denoting the steady-state workload: if  $p > q$  then  $\{Q(0) > pB\}$  essentially implies  $\{Q(TB) > qB\}$  for  $T$  small, and if  $p < q$  then  $\{Q(TB) > qB\}$  essentially implies  $\{Q(0) > pB\}$  for  $T$  small — this is regime (A), as it was mentioned in the introduction. Then there is an intermediate range of values of  $T$ , regime (B), in which the event of interest is roughly equal to

$$\{Q(0) > pB, A(0, TB) \geq qB + cT - pB\};$$

in this range the buffer does not become empty between 0 and  $TB$ . For large  $T$  (regime (C)) the most likely scenario is that the queue reaches level  $pB$  at time 0, drains, and starts building up just before  $TB$ , to reach value  $qB$  at  $TB$ . In the Brownian case the most likely path of this scenario consists of two independent busy periods.  $\spadesuit$

### 3.5.2 The LRD case

In this subsection we focus on the scenario  $\alpha \in (1, 2)$ . Whereas for the case of  $\alpha = 1$  we could rely on explicit computations, for  $\alpha \in (1, 2)$  the analysis of the rate function

$$\inf_{s \geq 0} \inf_{t \in [0, T) \cup \{T+s\}} \xi_{\mathbb{B}_\alpha; 1}(s, t)$$

turns out to be substantially harder. Before presenting the main results of this section, we introduce some additional notation. Define, for a given  $\alpha \in (1, 2)$ , and  $p, q, c > 0$ ,

$$s^* := \arg \max_{s \geq 0} \left\{ \frac{s^{\alpha/2}}{p + cs} \right\} = \frac{p}{c} \frac{\alpha}{2 - \alpha},$$

$$t^* := \arg \max_{t \geq 0} \left\{ \frac{t^{\alpha/2}}{q + ct} \right\} = \frac{q}{c} \frac{\alpha}{2 - \alpha},$$

and

$$D(x) := \frac{1}{2} \left( \frac{2x}{2 - \alpha} \right)^{2 - \alpha} \left( \frac{2c}{\alpha} \right)^\alpha.$$

Note that for  $X(t) \equiv \mathbb{B}_\alpha(t)$  we have that

$$\max_{s \geq 0} \text{Var} Y_1(s) = \text{Var} Y_1(s^*) = \frac{1}{2D(p)}, \quad \max_{t \geq 0} \text{Var} Z_1(t) = \text{Var} Z_1(t^*) = \frac{1}{2D(q)}.$$

The following general bounds hold. The upper bound in (3.21) essentially says that the decay rate of the joint probability is smaller than the decay rate of the least likely event; the lower bound in (3.21) says that the joint probability is larger than the product of the individual probabilities (which makes sense in view of the positive correlation).

**Proposition 3.5.3.** *Assume that  $\{X(t) : t \in \mathbb{R}\}$  satisfies Assumption 3.2.1 with  $\alpha \in (1, 2)$ . Then*

$$-\max\{D(p), D(q)\} \geq \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) > -(D(p) + D(q)). \quad (3.21)$$

*Proof.* The upper bound follows immediately from

$$\begin{aligned} & \inf_{s \geq 0} \inf_{t \in [0, T) \cup \{T+s\}} \xi_{\mathbb{B}_\alpha; 1}(s, t) \geq \inf_{s \geq 0} \inf_{t \geq 0} \xi_{\mathbb{B}_\alpha; 1}(s, t) \\ & = \inf_{s \geq 0, t \geq 0} \frac{1}{2} \frac{1}{\min\{\sigma_Y^2(s), \sigma_Z^2(t)\}} \left( 1 + \frac{(\gamma(s, t) - r(s, t))^2}{1 - r^2(s, t)} I(s, t) \right) \\ & \geq \max \left\{ \inf_{s \geq 0} \frac{(p + cs)^2}{2v(s)}, \inf_{t \geq 0} \frac{(q + ct)^2}{2v(t)} \right\} = \max\{D(p), D(q)\}. \end{aligned}$$

The lower bound is due to the fact that, due to Lemma 3.4.5, for some  $\bar{\varepsilon} > \underline{\varepsilon} > 0$ ,

$$\begin{aligned} \inf_{s \geq 0} \inf_{t \in [0, T) \cup \{T+s\}} \xi_{\mathbb{B}; \alpha; 1}(s, t) &= \min_{s \in [\underline{\varepsilon}, \bar{\varepsilon}]} \min_{t \in [\underline{\varepsilon}, T) \cup \{T+s\}} \xi_{\mathbb{B}; \alpha; 1}(s, t) \\ &= \min_{s \in [\underline{\varepsilon}, \bar{\varepsilon}]} \min_{t \in [\underline{\varepsilon}, T) \cup \{T+s\}} \frac{1}{2} \frac{1}{\min\{\sigma_Y^2(s), \sigma_Z^2(t)\}} \left( 1 + \frac{(\gamma(s, t) - r(s, t))^2}{1 - r^2(s, t)} I(s, t) \right). \end{aligned} \quad (3.22)$$

Moreover the assumption that  $\alpha > 1$  straightforwardly implies  $r(s, t) > 0$  (positive correlation of the input traffic!). Realize that  $(\gamma^2 + 1)r < 2\gamma$  holds for all  $r \in (0, 1)$  and  $\gamma \in [0, 1]$ ; after elementary calculus this yields

$$\frac{(\gamma(s, t) - r(s, t))^2}{1 - r^2(s, t)} < \gamma^2(s, t)$$

for each  $s, t > 0$ , and therefore (3.22) is majorized by

$$\begin{aligned} \min_{s \in [\underline{\varepsilon}, \bar{\varepsilon}]} \min_{t \in [\underline{\varepsilon}, T) \cup \{T+s\}} \frac{1}{2} \frac{1}{\min\{\sigma_Y^2(s), \sigma_Z^2(t)\}} (1 + \gamma^2(s, t)) \\ = \min_{s \in [\underline{\varepsilon}, \bar{\varepsilon}]} \min_{t \in [\underline{\varepsilon}, T) \cup \{T+s\}} \frac{1}{2} \left( \frac{1}{\sigma_Y^2(s)} + \frac{1}{\sigma_Z^2(t)} \right) = D(p) + D(q). \end{aligned}$$

This completes the proof.  $\square$

In the following we determine the values of  $T$  for which the lower bound in (3.21) is tight.

**Proposition 3.5.4.** *Assume that  $\{X(t) : t \in \mathbb{R}\}$  satisfies Assumption 3.2.1 with  $\alpha \in (1, 2)$ . (i) If  $p > q > 0$ , then there exists a unique  $T^*$  solving the equation*

$$\gamma(s^*, s^* + T^*) = r(s^*, s^* + T^*) \quad (3.23)$$

such that

$$\begin{aligned} \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) &= -D(p) \text{ for } T \leq T^*; \\ \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) &< -D(p) \text{ for } T > T^*. \end{aligned}$$

(ii) If  $q > p > 0$ , then there exists a unique  $T_*$  solving the equation

$$\gamma(t^* - T_*, t^*) = r(t^* - T_*, t^*) \quad (3.24)$$

such that

$$\begin{aligned} \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) &= -D(q) \text{ for } T \leq T_*; \\ \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) &< -D(q) \text{ for } T > T_*. \end{aligned}$$

*Proof.* First consider the case  $p > q > 0$ . Note that in order to have

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) = -D(p)$$

we need the following two conditions to be satisfied:

$$\gamma(s^*, s^* + T) \leq r(s^*, s^* + T) \quad (3.25)$$

$$\sigma_Y(s^*) \leq \sigma_Z(s^* + T). \quad (3.26)$$

Under (3.26) we have

$$\begin{aligned} r(s^*, s^* + T) &= \frac{(T + s^*)^\alpha - T^\alpha + (s^*)^\alpha}{2(s^*(s^* + T))^{\alpha/2}} \\ &= \frac{1}{2} \left( \frac{s^*}{s^* + T} \right)^{\alpha/2} \left( \left( \frac{T + s^*}{s^*} \right)^\alpha - \left( \frac{T}{s^*} \right)^\alpha + 1 \right); \\ \gamma(s^*, s^* + T) &= \left( \frac{s^*}{s^* + T} \right)^{\alpha/2} \frac{q + cs^* + cT}{p + cs^*} \\ &= \left( \frac{s^*}{s^* + T} \right)^{\alpha/2} \left( 1 + \frac{q-p}{p + cs^*} + \frac{cs^*}{p + cs^*} \frac{T}{s^*} \right). \end{aligned}$$

Noticing that

$$\frac{cs^*}{p + cs^*} = \alpha/2; \quad -1 < \frac{q-p}{p + cs^*} = \frac{q-p}{p} \left( 1 - \frac{\alpha}{2} \right) < 0,$$

Inequalities (3.25) and (3.26) are equivalent to respectively

$$1 + 2 \frac{q-p}{p} (1 - \alpha/2) + \alpha \frac{T}{s^*} \leq \left( 1 + \frac{T}{s^*} \right)^\alpha - \left( \frac{T}{s^*} \right)^\alpha, \quad (3.27)$$

$$1 + \frac{q-p}{p} (1 - \alpha/2) + \frac{\alpha T}{2 s^*} \leq \left( 1 + \frac{T}{s^*} \right)^{\alpha/2}. \quad (3.28)$$

Interestingly, however, we have that Inequality (3.27) implies Inequality (3.28). This can be shown as follows. First rewrite Inequality (3.27) to

$$1 + \frac{q-p}{p} (1 - \alpha/2) + \frac{\alpha T}{2 s^*} \leq \frac{1}{2} \left( 1 + \left( 1 + \frac{T}{s^*} \right)^\alpha - \left( \frac{T}{s^*} \right)^\alpha \right). \quad (3.29)$$

Let  $\tilde{X}(t)$  correspond to fBm with variance curve  $v(t) = t^\alpha$ , and let  $\check{A}(s, t) := \tilde{X}(t) - \tilde{X}(s)$ . Then

$$\begin{aligned} \frac{\text{Cov}(\check{A}(0, s^*), \check{A}(0, s^* + T))}{\text{Var}\check{A}(0, s^*)} &= \frac{1}{2} \left( 1 + \left( 1 + \frac{T}{s^*} \right)^\alpha - \left( \frac{T}{s^*} \right)^\alpha \right); \\ \sqrt{\frac{\text{Var}\check{A}(0, s^* + T)}{\text{Var}\check{A}(0, s^*)}} &= \left( 1 + \frac{T}{s^*} \right)^{\alpha/2}. \end{aligned}$$

Consequently, using the fact that the correlation coefficient is smaller than 1,

$$\begin{aligned} 0 &< \frac{1}{2} \left( 1 + \left( 1 + \frac{T}{s^*} \right)^\alpha - \left( \frac{T}{s^*} \right)^\alpha \right) / \left( 1 + \frac{T}{s^*} \right)^{\alpha/2} \\ &= \frac{\text{Cov}(\check{A}(0, s^*), \check{A}(0, s^* + T))}{\sqrt{\text{Var}\check{A}(0, s^* + T)\text{Var}\check{A}(0, s^*)}} = \text{Corr}(\check{A}(0, s^*), \check{A}(0, s^* + T)) < 1. \end{aligned}$$

Hence the right-hand side of Inequality (3.29) is smaller than the right-hand side of Inequality (3.28), and we indeed have that Inequality (3.27) implies Inequality (3.28).

Now it suffices to show that the functions

$$f(x) := (1+x)^\alpha - x^\alpha \quad \text{and} \quad g(x) := 1 + 2 \left( 1 - \frac{\alpha}{2} \right) \frac{q-p}{p} + \alpha x$$

intersect in a unique point  $x^* > 0$ . Indeed the function  $g(\cdot)$  is increasing and

$$g(0) = 1 + (2 - \alpha) \frac{q-p}{p} < 1 = f(0).$$

Now notice that  $f(\cdot)$  is increasing and concave, since  $f'(x) = \alpha((1+x)^\alpha - 1 - x^{\alpha-1}) > 0$  and  $f''(x) = \alpha(\alpha-1)((1+x)^{\alpha-2} - x^{\alpha-2}) < 0$ . Then the graphs of the two functions must intersect in a unique point  $x^* > 0$ . We have thus found that there exists a unique  $T^* \geq 0$  such that for all  $T \leq T^*$  we have that Inequality (3.25) is satisfied.

Since the idea of the proof for the case  $q > p > 0$  is analogous to the proof for the case  $p > q > 0$ , we omit the details.  $\square$

In the next proposition we give a lower bound on  $T^*$  and  $T_*$ .

**Proposition 3.5.5.** (i) If  $p > q > 0$ , then  $T^* \geq (p-q)/c$ . (ii) If  $q > p > 0$ , then  $T_* \geq (q-p)/c$ .

*Proof.* Since the proofs of (i) and (ii) are analogous, we focus on the argument that shows (i). We need to check whether  $T = (p-q)/c$  satisfies (3.25).

First notice that (under the notation used in the proof of Proposition 3.5.4)

$$g\left(\frac{p-q}{cs^*}\right) = 1 + 2 \frac{q-p}{p+cs^*} + \alpha \frac{p-q}{cs^*} = 1$$

and we have that  $f(x)$  and  $g(x)$  are increasing and  $f(0) = 1$ . Hence we have

$$f\left(\frac{p-q}{cs^*}\right) \geq f(0) = g\left(\frac{p-q}{cs^*}\right).$$

This proves the claim in part (i).  $\square$

**Remark 3.5.6.** Conditions  $T < T^*$  and  $T < T_*$  have interesting interpretations. Consider for instance  $T < T^*$ . Elementary computations with the conditional distribution of Normal random variables yield that  $T < T^*$  is equivalent to

$$\mathbb{E}(A(0, T) \mid A(-s^*, 0) = p + cs^*) \geq q - p + cT.$$

The interpretation is that, given the queue exceeds  $pB$  at 0, exceeding  $qB$  at time  $TB$  is not a rare event anymore. A similar interpretation can be given to condition  $T < T_*$ .  $\spadesuit$

Proposition 3.5.4 says that, just as in the SRD case, if and only if  $T$  is smaller than some threshold, then the decay rate of the joint probability equals the decay rate of  $\mathbb{P}(Q > \max\{p, q\}B)$ , with  $Q$  denoting the steady-state workload. In other words:  $T^*$  (in case  $p > q$ ) or  $T_*$  (in case  $p < q$ ) separates regime (A) from regime (B). In the SRD case, we found a second threshold, separating regime (B) from regime (C): below this threshold the buffer does not become empty (most likely) before time  $TB$ , and above it *does* (for large values of  $T$ ). In the LRD case we believe that this structure still applies, but we have been able to prove just a partial result, which is stated in Proposition 3.5.8. It says that for  $T$  large enough, we are in Regime (C).

**Lemma 3.5.7.**

$$\inf_{s \geq 0} \xi_{\mathbb{B}_\alpha}(s, T + s) \geq \frac{1}{2}c^2T^{2-\alpha}.$$

*Proof.* Uniformly in  $s \geq 0$ ,

$$\xi_{\mathbb{B}_\alpha}(s, T + s) \geq \frac{1}{2} \frac{(q + c(T + s))^2}{(T + s)^\alpha} \geq \frac{1}{2}c^2T^{2-\alpha}.$$

This proves the stated.  $\square$

Due to Proposition 3.5.3, for  $\alpha \in (1, 2)$ , we have

$$\inf_{s \geq 0} \inf_{t \in [0, T] \cup \{T+s\}} \xi_{\mathbb{B}_\alpha}(s, t) \leq \xi^* := D(p) + D(q).$$

Upon combining the above, we obtain the following result. On an intuitive level, it says that for  $T$  larger than some explicitly given threshold, with overwhelming probability the most likely path is such that the busy period in which 0 is contained does not coincide with the busy period in which  $T$  is contained.

**Proposition 3.5.8.** *For*

$$T > T^\# := \left( \frac{2\xi^*}{c^2} \right)^{1/(2-\alpha)}$$

*we have that*

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) = \inf_{s \geq 0} \inf_{t \in [0, T]} \xi_{\mathbb{B}_\alpha}(s, t).$$

**Remark 3.5.9.** We finish this section with a few observations on the (practically less relevant) case  $\alpha \in (0, 1)$  (i.e., the input stream has negative correlation).

- It is anticipated that for  $T$  small, still the most demanding event will determine the asymptotics. In other words: up to some threshold the decay rate will be  $-\max\{D(p), D(q)\}$ ; the value of this threshold can be determined as in Remark 3.5.6.
- Now consider large  $T$ . Then time epochs 0 and  $TB$  will be in different busy periods. For the  $s, t$  of interest we have  $r \equiv r(s, t) < 0$ , which implies

$$\frac{(\gamma(s, t) - r(s, t))^2}{1 - r^2(s, t)} > \gamma^2(s, t)$$

(to see this, realize that  $\gamma \equiv \gamma(s, t) \in [0, 1]$ , and verify that the above relation reduces to  $(\gamma^2 + 1)r < 2\gamma$ ; it is immediate that this holds for all  $r < 0$  and  $\gamma \in [0, 1]$ ). We therefore obtain:

$$\lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log N(B) < -(D(p) + D(q));$$

in other words: in order to achieve a high buffer content at time  $TB$  it is for large  $T$  *disadvantageous* to have a large buffer content at time 0. ♠

## 3.6 Discussion and concluding remarks

*Exact asymptotics.* In this chapter we analyzed the logarithmic asymptotics of the joint probability  $\mathbb{P}(Q(0) > pB, Q(TB) > qB)$ . We have identified the corresponding decay rate. An open issue concerns the *exact* asymptotics, i.e., can we find an explicit function  $\varphi(\cdot)$  such that

$$\mathbb{P}(Q(0) > pB, Q(TB) > qB) \cdot \varphi(B) \rightarrow 1$$

as  $B \rightarrow \infty$ ? It is noted that for the single-dimensional case this was already a highly non-trivial task [63, 89, 92], and the answer involves the so-called *Pickands constant*.

*Regimes.* Then we considered the decay rate of the probability of interest in more detail, and identified three regimes for  $T$ . The SRD case could be dealt with explicitly, in that we presented closed-form expressions for the decay rate, as well as for the critical values of  $T$  that separate regime (A) from regime (B), and regime (B) from regime (C). In the LRD case we found an explicit expression for the decay rate in regime (A), and we showed that the critical value of  $T$ , which we called  $T^*$  for  $p > q$  and  $T_*$  for  $p < q$ , that separates regime (A) from regime (B) is the solution to some

algebraic equation. In addition we showed that for  $T$  larger than some explicitly given number  $T^\sharp$ , we are in regime (C). This in principle still allows oscillations between regimes (B) and (C) in the region between  $T^*$  ( $T_*$ , respectively) and  $T^\sharp$ . We conjecture that such oscillations do not occur.

*Scaling of time and space.* In our analysis we scaled space and time in the same way, i.e., both the buffer level and the length of the time interval are multiples of  $B$ . As is immediately visible from Remark 3.4.7, essentially due to the self-similarity, this scaling leads for fBm to well-defined decay rates; in the non-fBm case some sort of approximate self-similarity is enforced by imposing Assumption 3.2.1.

In view of the results in [37], it is anticipated that if  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ ,

$$\begin{aligned} & \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > pB, Q(T_B) > qB) \\ &= \min \left\{ \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > pB), \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > qB) \right\} \\ &= -\max\{D(p), D(q)\}; \end{aligned}$$

if  $T_B/B \rightarrow \infty$  as  $B \rightarrow \infty$ ,

$$\begin{aligned} & \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > pB, Q(T_B) > qB) \\ &= \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > pB) + \lim_{B \rightarrow \infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q(0) > qB) \\ &= -(D(p) + D(q)), \end{aligned}$$

using the function  $D(\cdot)$  as introduced in Section 3.5.2.

## Chapter 4

---

# Correlation structure of Lévy-driven queues

In Chapters 2 and 3 we have studied a single queue fed by Gaussian sources. The present chapter and Chapter 5 deal with another class of input processes, namely, Lévy processes. In this chapter we specialize to the class of spectrally-positive Lévy processes, for which we prove a number of structural properties of the correlation function  $r(t)$  of the stationary workload process. Furthermore, we study the asymptotics of  $r(t)$  for  $t \rightarrow \infty$  for light and heavy-tailed spectrally-positive Lévy inputs.

### 4.1 Introduction

Consider a queueing system, and, more particularly, its workload process  $\{Q(t) : t \geq 0\}$ . Where one usually focuses on the characterization of the (transient or steady-state) workload, another interesting problem relates to the identification of the *correlation function*  $r(t)$ , cf. (1.10). For several queueing systems this correlation function has been explicitly computed; Morse [91], for instance, analyzes the number of customers in the M/M/1 queue. Often explicit formulae were hard to obtain, but the analysis simplified greatly when looking at the Laplace transform

$$\rho(\vartheta) := \int_0^\infty r(t)e^{-\vartheta t} dt.$$

Beneš [18] managed to compute  $\rho(\cdot)$  for the workload in the M/G/1 queue; relying on the concept of complete monotonicity, Ott [95] elegantly proved that, in this case,  $r(\cdot)$  is positive, decreasing and convex. We further mention the survey by Reynolds [102], and interesting results by Abate and Whitt [2].

The primary aim of this chapter is to extend the results mentioned above to the class of single-server queues fed by *Lévy processes*. Notice that the M/G/1 queue is contained in this class (then the Lévy process under consideration is a compound Poisson process with drift). One could expect that such an extension is possible, as the classical Pollaczek-Khintchine result for the M/G/1 queue, carries over to queues with general Lévy input, see Zolotarev [113] for an early reference; we refer also to Bingham [22], and references therein, and the book by Kyprianou [74], for an extensive account of fluctuation theory for Lévy processes. The only condition one

usually needs to impose in order to obtain explicit results, is that no negative jumps are allowed.

In more detail, the setting we consider is the following. We define a ‘net input process’  $\{X(t) : t \geq 0\}$ , which is assumed to be a Lévy process with no negative jumps. Then the workload process  $\{Q(t) : t \geq 0\}$  is defined as the reflected process of  $\{X(t) : t \geq 0\}$  at 0. Because of the lack of explicit formulae for the probability distributions of the processes considered, we will work most of the time with their Laplace transforms; in our analysis the Laplace exponent  $\varphi(\cdot)$  of the process  $\{X(t) : t \geq 0\}$ , as well as its inverse  $\psi(\cdot)$ , play an important role.

We first obtain an explicit expression, in terms of  $\varphi(\cdot)$  and  $\psi(\cdot)$ , of the transform  $\rho(\cdot)$  of the correlation function. Using the concept of complete monotonicity, we use this transform to establish a series of structural properties of the correlation function, viz. we prove that  $r(\cdot)$  is positive, decreasing, and convex. These results indeed generalize those obtained by Ott [95] and Abate and Whitt [2] for the M/G/1 queue. We then consider the asymptotic behavior of  $r(t)$  for  $t$  large. For light-tailed Lévy input these asymptotics are essentially exponential; for the M/G/1 case they resemble those of the busy period. For heavy-tailed input we can use results for regularly varying functions, e.g. Karamata’s Tauberian theorem, to obtain the asymptotics of  $r(\cdot)$ .

The remainder of this chapter is organized as follows. In Section 4.2 we obtain the Laplace transform of the correlation function, where in Section 4.3 its structural properties are studied. The cases of light-tailed and heavy-tailed input are treated in Sections 4.4 and 4.5, respectively. Concluding remarks are found in Section 4.6.

## 4.2 Laplace transform of the correlation function

In this section we find an expression for the transform  $\rho(\cdot)$  of the correlation function. We start this section, however, with a formal introduction of our queueing system.

*Lévy processes.* Let  $\{X(t) : t \geq 0\}$  be a Lévy process without negative jumps, with drift  $\mathbb{E}X(1) < 0$ . Its Laplace exponent is given by the function  $\varphi(\cdot) : [0, \infty) \mapsto [0, \infty)$ , i.e.,  $\varphi(\alpha) := \log \mathbb{E}e^{-\alpha X(1)}$ . It is known that  $\varphi(\cdot)$  is increasing and convex on  $[0, \infty)$ , with slope  $\varphi'(0) = -\mathbb{E}X(1)$  in the origin. Therefore the inverse  $\psi(\cdot)$  of  $\varphi(\cdot)$  is well-defined on  $[0, \infty)$ . In the sequel we also require that  $X(t)$  is not a *subordinator*, i.e., a monotone process; thus  $X(1)$  has probability mass on the positive half-line, which implies that  $\lim_{\alpha \rightarrow -\infty} \varphi(\alpha) = \infty$ .

Important examples of such Lévy processes are the following.

- (1) *Brownian motion with drift.* We write  $X \in \mathbb{Bm}(-\delta, \sigma^2)$  when  $\varphi(\alpha) = \alpha\delta + \frac{1}{2}\alpha^2\sigma^2$ .

- (2) *Compound Poisson with drift.* Jobs arrive according to a Poisson process of rate  $\lambda$ ; the jobs  $J_1, J_2, \dots$  are i.i.d. samples from a distribution with Laplace transform  $\beta(\alpha) := \mathbb{E}e^{-\alpha J}$ ; the storage system is continuously depleted at a rate  $-1$ . We write  $X \in \mathbb{CP}(\lambda, \beta(\cdot))$ ; it can be verified that  $\varphi(\alpha) = \alpha - \lambda + \lambda\beta(\alpha)$ .

*Reflected Lévy processes; queues.* We consider the reflection of  $\{X(t) : t \geq 0\}$  at 0, which we denote by  $\{Q(t) : t \geq 0\}$ . It is formally introduced as follows, see for instance [11, Ch. IX]. Define the increasing process  $\{L(t) : t \geq 0\}$  by

$$L(t) = - \inf_{0 \leq s \leq t} X(s).$$

Then the reflected process (or: workload process, queueing process)  $\{Q(t) : t \geq 0\}$  is given through

$$Q(t) := X(t) + \max\{L(t), Q(0)\};$$

observe that  $Q(t) \geq 0$  for all  $t \geq 0$ . Then the steady-state distribution of  $Q_e := \lim_{t \rightarrow \infty} Q(t)$  is characterized by [113]:

$$\kappa(\alpha) := \mathbb{E}e^{-\alpha Q_e} = \frac{\alpha \varphi'(0)}{\varphi(\alpha)}; \quad (4.1)$$

for the special case of CP input this is the celebrated Pollaczek-Khintchine formula. This relation reveals all moments of the steady-state queue  $Q_e$ , and in particular its mean and variance:

$$\mu := \mathbb{E}Q_e = - \frac{d}{d\alpha} \frac{\alpha \varphi'(0)}{\varphi(\alpha)} \Big|_{\alpha \downarrow 0} = \frac{\varphi''(0)}{2\varphi'(0)}, \quad (4.2)$$

and similarly

$$v := \text{Var}Q_e = \frac{1}{4} \left( \frac{\varphi''(0)}{\varphi'(0)} \right)^2 - \frac{1}{3} \frac{\varphi'''(0)}{\varphi'(0)}, \quad (4.3)$$

which from now on are assumed to be finite.

*Correlation structure of the queue.* In this chapter we are interested in the correlation structure of the queue process  $\{Q(t) : t \geq 0\}$ . Our analysis relies on the following useful relation, see e.g. [11, Section IX.3] and [65]:

$$\mathbb{E} \left( e^{-\alpha Q(\tau)} \mid Q(0) = q \right) = \frac{\vartheta}{\vartheta - \varphi(\alpha)} \left( e^{-\alpha q} - \alpha \cdot \frac{e^{-\psi(\vartheta)q}}{\psi(\vartheta)} \right), \quad (4.4)$$

where  $\tau$  is exponentially distributed with mean  $\vartheta^{-1}$ , independently of the Lévy process. (As an aside we mention that (4.4) implies Pollaczek-Khintchine in at least two ways: (a) let  $\vartheta \downarrow 0$ , so that  $\tau$  corresponds with some epoch infinitely far away, and

use elementary calculus ('L'Hôpital'); (b) find  $\mathbb{E}e^{-\alpha Q(\tau)}$  by deconditioning, use that in stationarity  $\mathbb{E}e^{-\alpha Q(\tau)}$  should coincide with  $\mathbb{E}e^{-\alpha Q(0)}$ , and then solve  $\mathbb{E}e^{-\alpha Q(0)}$ .

Formula (4.4) enables us to find explicitly the Laplace transform  $\rho(\cdot)$  of

$$r(t) := \text{Corr}(Q(0), Q(t)) = \frac{\text{Cov}(Q(0), Q(t))}{\sqrt{\text{Var}Q(0) \cdot \text{Var}Q(t)}} = \frac{\mathbb{E}Q(0)Q(t) - (\mathbb{E}Q_e)^2}{\text{Var}Q_e},$$

as we show now. Here it is assumed that the system is in steady-state at time 0, that is,  $Q_e$  obeys the 'generalized Pollaczek-Khintchine' formula (4.1). First realize that

$$\mathbb{E}\left(e^{-\alpha Q(\tau)} \mid Q(0) = q\right) = \int_0^\infty \vartheta e^{-\vartheta t} \mathbb{E}\left(e^{-\alpha Q(t)} \mid Q(0) = q\right) dt.$$

By differentiation with respect to  $\alpha$  and subsequently letting  $\alpha \downarrow 0$ , we obtain

$$\int_0^\infty \vartheta e^{-\vartheta t} \mathbb{E}(Q(t) \mid Q(0) = q) dt = -\frac{\varphi'(0)}{\vartheta} + q + \frac{e^{-\psi(\vartheta)q}}{\psi(\vartheta)}. \quad (4.5)$$

Concentrate on the Laplace transform  $\gamma(\vartheta)$  of  $R(t)$ . Straightforward calculus reveals that

$$\begin{aligned} \gamma(\vartheta) &:= \int_0^\infty \text{Cov}(Q(0), Q(t)) e^{-\vartheta t} dt = \int_0^\infty (\mathbb{E}Q(0)Q(t) - \mu^2) e^{-\vartheta t} dt \\ &= \int_0^\infty \int_0^\infty q \cdot \mathbb{E}(Q(t) \mid Q(0) = q) \cdot e^{-\vartheta t} d\mathbb{P}(Q(0) \leq q) dt - \frac{\mu^2}{\vartheta}; \end{aligned}$$

it is assumed that the queue is in stationarity at time 0 (and hence it is in stationarity at time  $t$  as well). By invoking (4.5) we find that the expression in the previous display equals

$$\begin{aligned} &\int_0^\infty \frac{q}{\vartheta} \left( -\frac{\varphi'(0)}{\vartheta} + q + \frac{e^{-\psi(\vartheta)q}}{\psi(\vartheta)} \right) d\mathbb{P}(Q(0) \leq q) - \frac{\mu^2}{\vartheta} \\ &= -\frac{\mu\varphi'(0)}{\vartheta^2} + \frac{v}{\vartheta} + \frac{1}{\vartheta\psi(\vartheta)} \mathbb{E}\left(Q(0)e^{-\psi(\vartheta)Q(0)}\right). \end{aligned} \quad (4.6)$$

From the generalized Pollaczek-Khintchine formula (4.1) we obtain by differentiating

$$\mathbb{E}\left(Q(0)e^{-\alpha Q(0)}\right) = \varphi'(0) \left( -\frac{1}{\varphi(\alpha)} + \alpha \frac{\varphi'(\alpha)}{(\varphi(\alpha))^2} \right).$$

Inserting this relation, in addition to (4.2), into (4.6) we obtain the Laplace transform of  $\text{Cov}(Q(0), Q(t))$ :

$$\gamma(\vartheta) = -\frac{\varphi''(0)}{2\vartheta^2} + \frac{v}{\vartheta} + \frac{\varphi'(0)}{\vartheta^2} \left( \frac{1}{\vartheta\psi'(\vartheta)} - \frac{1}{\psi(\vartheta)} \right).$$

This trivially also provides us with the Laplace transform of the correlation function  $r(t) = \text{Corr}(Q(0), Q(t))$ , as stated in the following theorem. When specializing to CP input, we retrieve Equation (6.2) of Beneš [18].

**Theorem 4.2.1.** For any  $\vartheta \geq 0$ , and  $v$  as in (4.3),

$$\begin{aligned} \rho(\vartheta) &:= \int_0^\infty r(t) e^{-\vartheta t} dt = \frac{\gamma(\vartheta)}{v} \\ &= \frac{1}{\vartheta} - \frac{\varphi''(0)}{2v\vartheta^2} + \frac{\varphi'(0)}{v\vartheta^2} \left[ \frac{1}{\vartheta\psi'(\vartheta)} - \frac{1}{\psi(\vartheta)} \right]. \end{aligned} \quad (4.7)$$

**Remark 4.2.2.** Using the generalized Pollaczek-Khintchine formula (4.1), it is readily verified that the result in Theorem 4.2.1 can be simplified to

$$\rho(\vartheta) = \frac{1}{\vartheta} - \frac{1}{v} \left( \frac{\varphi''(0)}{2\vartheta^2} + \frac{\kappa'(\psi(\vartheta))}{\vartheta\psi(\vartheta)} \right).$$



**Example 4.2.3.** Consider the situation that  $\{X(t) : t \geq 0\}$  corresponds to standard Brownian motion decreased by a linear drift (say of rate 1, so  $X \in \mathbb{Bm}(-1, 1)$ ). In other words: the Laplace exponent of the Lévy process is given by  $\varphi(\alpha) = \alpha + \frac{1}{2}\alpha^2$ , and its inverse is  $\psi(\vartheta) = -1 + \sqrt{1 + 2\vartheta}$ . Now consider the workload process  $\{Q(t) : t \geq 0\}$  and its correlation function. The above theory yields that the Laplace transform of  $r(\cdot)$  is given by

$$\rho(\vartheta) = \frac{1}{\vartheta} - \frac{2}{\vartheta^2} + \frac{2}{\vartheta^3} \left( \sqrt{1 + 2\vartheta} - 1 \right).$$

It turns out to be possible to explicitly invert  $\rho(\cdot)$ :

$$r(t) = 2(1 - 2t - t^2) \left( 1 - \Phi_{\mathbb{N}}(\sqrt{t}) \right) + 2\sqrt{t}(1 + t)\phi_{\mathbb{N}}(\sqrt{t}), \quad (4.8)$$

with  $\Phi_{\mathbb{N}}(\cdot)$  (resp.  $\phi_{\mathbb{N}}(\cdot)$ ) the standard Normal distribution (resp. density). Equation (4.8) is in agreement with the results in [1] and [80, Section 12.1].  $\diamond$

### 4.3 Structural properties of the correlation function

This section concentrates on the derivation of a number of key structural properties of the correlation function  $r(\cdot)$ . More specifically, relying on the concept of completely monotone functions [19, 95], we prove in Theorem 4.3.6 that  $r(\cdot)$  is a positive, decreasing, and convex function. To this end, we first establish a number of auxiliary results; a key result is Proposition 4.3.1.

**Proposition 4.3.1.** Define  $\xi(\vartheta)$  by

$$\xi(\vartheta) := \frac{1}{\mu} \left( \frac{1}{\vartheta\psi'(\vartheta)} - \frac{1}{\psi(\vartheta)} \right); \quad (4.9)$$

then  $\xi(\vartheta)$  is the Laplace transform of a (non-negative) random variable  $Z$ .

**Remark 4.3.2.** The Laplace transform of the stationary-excess distribution  $Z_e$  associated with  $Z$  is given by [2]

$$\xi_e(\vartheta) = \frac{\xi(\vartheta) - 1}{\vartheta \xi'(\vartheta)} = \frac{\varphi''(0)}{2v\vartheta} (1 - \xi(\vartheta)). \quad (4.10)$$

Hence, the first moment of  $Z$  is  $2v/\varphi''(0)$ . ♠

To prove Proposition 4.3.1, we need a number of lemmas. These are stated and proved now. They extensively use the concept of *complete monotonicity* [19, 54]. The class  $\mathcal{C}$  of completely monotone functions is defined in the Appendix, where also a series of standard properties is given.

**Lemma 4.3.3.**  $\psi'(\vartheta) \in \mathcal{C}$ .

*Proof.* Consider for  $x \geq 0$ ,

$$H(x) := \inf\{t \geq 0 : X(t) = -x\};$$

then  $H(x)$  is a Lévy process with Laplace exponent  $-\psi(\vartheta)$ , see e.g. [106, Theorem 46.3]. More specifically,  $H(x)$  is a subordinator. Now apply Lemma 4.A.4. □

**Lemma 4.3.4.** If  $f(\alpha) \in \mathcal{C}$ , then so does

$$\frac{f(0) - f(\alpha) + \alpha f'(\alpha)}{\alpha^2}.$$

*Proof.* The result is a consequence of subsequently applying Lemma 4.A.3.(4) and 4.A.3.(5). □

**Lemma 4.3.5.** For  $\sigma^2 > 0$  and measure  $\Pi_\varphi(\cdot)$  such that

$$\begin{aligned} \int_{(0,\infty)} \min\{1, x^2\} \Pi_\varphi(dx) &< \infty, \\ \frac{\alpha\varphi'(\alpha) - \varphi(\alpha)}{\alpha^2} &= \frac{1}{2}\sigma^2 + \frac{1}{\alpha^2} \int_{(0,\infty)} (1 - e^{-\alpha x} - \alpha x e^{-\alpha x}) \Pi_\varphi(dx) \in \mathcal{C}. \end{aligned} \quad (4.11)$$

*Proof.* The Laplace exponent  $\varphi(\alpha)$  can be written as, with  $\sigma^2 > 0$  and measure  $\Pi_\varphi(\cdot)$  such that

$$\begin{aligned} \int_{(0,\infty)} \min\{1, x^2\} \Pi_\varphi(dx) &< \infty, \\ \varphi(\alpha) &= \alpha\delta + \frac{1}{2}\alpha^2\sigma^2 + \int_{(0,\infty)} (e^{-\alpha x} - 1 + \alpha x 1_{(0,1)}) \Pi_\varphi(dx), \end{aligned}$$

which immediately yields the equality in (4.11). The claim that this function is in  $\mathcal{C}$  follows from the fact that any positive constant is in  $\mathcal{C}$ , Lemma 4.3.4, and Lemma 4.A.3.(1). □

*Proof of Proposition 4.3.1.* We first decompose

$$\frac{1}{\vartheta\psi'(\vartheta)} - \frac{1}{\psi(\vartheta)} = \eta_1(\vartheta)\eta_2(\vartheta),$$

with

$$\eta_1(\vartheta) := \frac{\psi(\vartheta)}{\vartheta}, \quad \eta_2(\vartheta) := \frac{1}{\psi(\vartheta)\psi'(\vartheta)} - \frac{\vartheta}{(\psi(\vartheta))^2}.$$

Because of (4.1), we have that  $\alpha/\varphi(\alpha) \in \mathcal{C}$ ; now applying Lemma 4.A.3.(3), in conjunction with Lemma 4.3.3, we obtain that  $\eta_1(\vartheta) \in \mathcal{C}$ .

To prove that also  $\eta_2(\vartheta) \in \mathcal{C}$ , we first recall from Lemma 4.3.5 that

$$\frac{\alpha\varphi'(\alpha) - \varphi(\alpha)}{\alpha^2} \in \mathcal{C}.$$

Again applying Lemma 4.A.3.(3), in conjunction with Lemma 4.3.3, it follows that  $\eta_2(\vartheta) \in \mathcal{C}$ .

As both  $\eta_1(\vartheta)$  and  $\eta_2(\vartheta)$  are in  $\mathcal{C}$ , Lemma 4.A.3.(2) yields that  $\xi(\vartheta) \in \mathcal{C}$ . Applying 'L'Hôpital' twice, and using that  $\psi''(0)(\varphi'(0))^3 = -\varphi''(0)$ , it is readily verified that

$$\xi(0) = \lim_{\vartheta \downarrow 0} \xi(\vartheta) = 1,$$

Now Theorem 4.A.2 yields the stated.  $\square$

Let  $\rho^{(1)}(\vartheta)$  and  $\rho^{(2)}(\vartheta)$  be the Laplace transforms of, respectively,  $r'(t)$  and  $r''(t)$ . Their expressions are given respectively as follows

$$\rho^{(1)}(\vartheta) := \int_0^\infty r'(t) e^{-\vartheta t} dt = -\frac{\varphi''(0)}{2v\vartheta} (1 - \xi(\vartheta)) = -\xi_e(\vartheta); \quad (4.12)$$

$$\rho^{(2)}(\vartheta) := \int_0^\infty r''(t) e^{-\vartheta t} dt = \frac{\varphi''(0)}{2v} \xi(\vartheta), \quad (4.13)$$

for  $\vartheta \geq 0$ . Here the properties that  $r(0) = 1$  and

$$r'(0) = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E}Q(0)Q(\varepsilon) - \mu^2}{\varepsilon v} = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E}Q(0)X(\varepsilon)}{\varepsilon v} = -\frac{\varphi''(0)}{2v},$$

in conjunction with integration by parts, are used.

**Theorem 4.3.6.**  *$r(t)$  is positive, decreasing and convex. Moreover,  $r(t)$  can be written as the tail of the stationary-excess distribution function associated with  $Z$ . More specifically,  $r(t) = \mathbb{P}(Z_e > t)$ . Furthermore, if  $Z$  has a finite second moment, then  $r(t)$  is integrable and*

$$\int_0^\infty r(t) dt = \frac{1}{8v} \frac{\varphi^{(4)}(0)}{\varphi'(0)^2} - \frac{5}{12v} \frac{\varphi''(0)\varphi^{(3)}(0)}{\varphi'(0)^3} + \frac{1}{4v} \frac{\varphi''(0)^3}{\varphi'(0)^4} \quad (4.14)$$

*Proof.* Convexity follows from the expression for  $\rho^{(2)}(\vartheta)$  in (4.13); it is concluded from Proposition 4.3.1 that  $\rho^{(2)}(\vartheta) \in \mathcal{C}$ , thus  $r''(t)$  is non-negative (for  $t \geq 0$ ). The monotonicity follows from the expression for  $\rho^{(1)}(\vartheta)$  in Equation (4.12), by applying Lemma 4.A.3.(4) to  $\rho^{(2)}(\vartheta) \in \mathcal{C}$ ; we find that  $-\rho^{(1)}(\vartheta)$  is in  $\mathcal{C}$ , implying that  $r'(t) \leq 0$  (for  $t \geq 0$ ). Then it is easily verified that applying Lemma 4.A.3.(4) to  $-\rho^{(1)}(\vartheta) \in \mathcal{C}$ , in conjunction with Equation (4.7), implies  $\rho(\vartheta) \in \mathcal{C}$ , and hence  $r(t) \geq 0$  (for  $t \geq 0$ ).

Observe that combining Equations (4.7) and (4.10) yields

$$\rho(\vartheta) = \frac{1 - \xi_e(\vartheta)}{\vartheta}. \quad (4.15)$$

It is straightforward to verify that the right-hand side of the previous display is just the Laplace transform of  $\mathbb{P}(Z_e > t)$ . It is concluded that  $r(t) = \mathbb{P}(Z_e > t)$  by the uniqueness of the Laplace transform. Equation (4.14) is found, after considerable calculus (i.e., application of ‘L’Hôpital’ several times, and various series expansions), by evaluating

$$\int_0^\infty r(t) dt = \rho(0) = \lim_{\vartheta \downarrow 0} \rho(\vartheta);$$

it is noted that  $\varphi^{(4)}(0)$  exists if the second moment of  $Z$  is finite.  $\square$

**Remark 4.3.7.** For the GI/G/1 queue, Borovkov *et al.* [27, Theorem 7.3] obtained the following expression for  $\int_0^\infty R(t) dt$ :

$$\int_0^\infty R(t) dt = \frac{1}{2} \left( \left( \frac{a_{\mathcal{A}}}{a_l} \sigma_l \right)^2 + \sigma_{\mathcal{A}}^2 - 2 \frac{a_{\mathcal{A}}}{a_l} \sigma_l \sigma_{\mathcal{A}} c(\mathcal{A}, l) \right)$$

where  $\mathcal{A}$  is the area swept under the workload process  $Q(t)$  during the busy period and  $l$  is the length of the busy cycle. Furthermore,  $a_{\mathcal{A}}$ ,  $\sigma_{\mathcal{A}}^2$  and  $a_l$ ,  $\sigma_l^2$  denote the mean and the variance of  $\mathcal{A}$  and  $l$ , respectively, whereas  $c(\mathcal{A}, l)$  is the correlation between  $\mathcal{A}$  and  $l$ . It can be checked that this formula coincides with ours in the M/G/1 queue case.  $\spadesuit$

## 4.4 Correlation asymptotics for light-tailed input

When  $\varphi(\cdot)$  has an analytic continuation for  $\alpha < 0$ , we are in the regime of light tails, as *a fortiori* then all moments  $(-1)^n \varphi^{(n)}(0)$  of  $X(1)$  exist. When  $\{X(t) : t \geq 0\}$  does not correspond to a decreasing subordinator, we also have that  $\lim_{\alpha \rightarrow -\infty} \varphi(\alpha) = \infty$ . Bearing in mind the fact that  $\varphi(\cdot)$  has a positive slope at the origin, and that convexity of  $\varphi(\cdot)$  implies continuity, there is a unique minimizer  $\zeta < 0$  such that  $\varphi(\zeta) < 0$ ,  $\varphi'(\zeta) = 0$  and  $\varphi''(\zeta) > 0$ .

In this situation, also  $\psi(\cdot)$  is well-defined for negative arguments; more precisely: for all  $\vartheta \geq \varphi(\zeta)$  the inverse  $\psi(\vartheta)$  has a meaningful interpretation. In fact,  $\vartheta^* := \varphi(\zeta)$  can be regarded as *branching point*. We thus see that Theorem 4.2.1 does not only apply for  $\vartheta \geq 0$ , but also for  $\vartheta \in [\vartheta^*, 0)$ . Around  $\zeta$ , we can write  $\varphi(\cdot)$  as

$$\varphi(\alpha) = \varphi(\zeta) + \frac{1}{2}(\alpha - \zeta)^2 \varphi''(\zeta) + O((\alpha - \zeta)^3),$$

and hence for  $\theta \downarrow \vartheta^*$

$$\psi(\vartheta) - \zeta \sim \sqrt{\frac{2}{\varphi''(\zeta)}} \cdot \sqrt{\vartheta - \varphi(\zeta)} = \sqrt{\frac{2}{\varphi''(\zeta)}} \cdot \sqrt{\vartheta - \vartheta^*}$$

(as before ‘ $\sim$ ’ indicates that the ratio of the left-hand side and right-hand side tends to 1). Routine calculations reveal that, for  $\theta \downarrow \vartheta^*$ , we have that  $\rho(\vartheta)$  looks like

$$\frac{1}{v} \left( -\frac{\varphi''(0)}{2(\vartheta^*)^2} + \frac{1}{4\vartheta^*} \left( \frac{\varphi''(0)}{\varphi'(0)} \right)^2 - \frac{1}{3\vartheta^*} \frac{\varphi'''(0)}{\varphi'(0)} - \frac{1}{(\vartheta^*)^2} \frac{\varphi'(0)}{\psi(\vartheta)} + \frac{1}{(\vartheta^*)^3} \frac{\varphi'(0)}{\psi'(\vartheta)} \right),$$

or, more precisely,

$$\begin{aligned} \rho(\vartheta) &- \frac{1}{v} \left( -\frac{\varphi''(0)}{2(\vartheta^*)^2} + \frac{1}{4\vartheta^*} \left( \frac{\varphi''(0)}{\varphi'(0)} \right)^2 - \frac{1}{3\vartheta^*} \frac{\varphi'''(0)}{\varphi'(0)} - \frac{1}{(\vartheta^*)^2} \frac{\varphi'(0)}{\zeta} \right) \\ &\sim \frac{\sqrt{2}\varphi'(0)}{\sqrt{\varphi''(\zeta)v(\vartheta^*)^2} \left( \frac{1}{\zeta^2} + \frac{\varphi''(\zeta)}{\vartheta^*} \right)} \sqrt{\vartheta - \vartheta^*}. \end{aligned}$$

We now relate the behavior of a transform  $\int_0^\infty e^{-\vartheta t} f(t) dt$  (around a branching point  $\vartheta^* < 0$ ) to the behavior of the ‘transformed’ function  $f(t)$  (for  $t$  large). We heuristically obtain the following result, cf. for instance the ‘Heaviside approach’ of [3, Equations (3.21)–(3.23)]; see also [33, pp. 153–154]: Suppose  $\varphi(\alpha) < \infty$  for some  $\alpha < 0$ . Then

$$r(t) \sim \ell \cdot \frac{e^{\vartheta^* t}}{t\sqrt{t}} \quad \text{as } t \rightarrow \infty, \quad (4.16)$$

where

$$\ell := -\frac{\varphi'(0)}{\sqrt{2\pi\varphi''(\zeta)v(\vartheta^*)^2} \left( \frac{1}{\zeta^2} + \frac{\varphi''(\zeta)}{\vartheta^*} \right)}. \quad (4.17)$$

**Remark 4.4.1.**  $\ell$ , as given in (4.17), is positive, as is seen as follows. From (4.11) we know that

$$f(\alpha) := \frac{2}{\varphi''(0)} \frac{\alpha\varphi'(\alpha) - \varphi(\alpha)}{\alpha^2}$$

is a Laplace transform, and hence also  $-f'(\alpha)/-f'(0)$ , so that for all  $\alpha$  holds that  $f'(\alpha) < 0$ , or

$$\alpha^3 \varphi''(\alpha) - 2\alpha^2 \varphi'(\alpha) + 2\alpha \varphi(\alpha) < 0.$$

Now insert  $\alpha := \zeta < 0$ ; using  $\varphi'(\zeta) = 0$  and  $\zeta < 0$ , we obtain  $\zeta^2 \varphi''(\zeta) + 2\varphi(\zeta) > 0$ , which implies

$$\frac{\varphi''(\zeta)}{\varphi(\zeta)} + \frac{2}{\zeta^2} < 0,$$

(use  $\varphi(\zeta) < 0$ ), and hence also

$$-\frac{\varphi''(\zeta)}{\vartheta^*} > \frac{2}{\zeta^2} > \frac{1}{\zeta^2},$$

thus implying  $\ell > 0$ . ♠

**Example 4.4.2.** It can be checked that for Brownian motion with drift, i.e.,  $X \in \mathbb{Bm}(-1, 1)$  as in the setting of Example 4.2.3,

$$r(t) \sim 8 \sqrt{\frac{2}{\pi}} \frac{e^{-t/2}}{t\sqrt{t}};$$

this could be found directly from (4.8) as well, cf. again [1] and [80, Section 12.1]. ◇

**Example 4.4.3.** For the compound Poisson model with exponential jobs (i.e., M/M/1 queue), it can be checked that

$$\psi(\vartheta) = \frac{1}{2} \left( \lambda - \mu + \vartheta + \sqrt{(\lambda - \mu + \vartheta)^2 + 4\vartheta\mu} \right),$$

so that the branching point is  $\vartheta^* = -(\sqrt{\mu} - \sqrt{\lambda})^2$ . Also,  $\zeta = -\mu + \sqrt{\lambda\mu}$ . Equation (4.16) now yields an explicit expression for the correlation asymptotics:

$$r(t) \sim \frac{1}{2\rho\sqrt{\pi}} \left( \frac{1 - \sqrt{\rho}}{1 + \sqrt{\rho}} \right)^3 \frac{\exp(-(1 - \sqrt{\rho})\sqrt{\mu t})}{(\sqrt{\mu t})^{3/2}} \text{ as } t \rightarrow \infty. \quad \diamond$$

**Remark 4.4.4.** For compound Poisson input, that is,  $X \in \mathbb{CP}(\lambda, \beta(\cdot))$ , the tail asymptotics of the correlation function are proportional to those of the busy period, at least in this light-tailed regime (where light-tailedness here means that we should require that  $\beta(\alpha) < \infty$  for some  $\alpha < 0$ ). This can be seen as follows.

First recall that the Laplace exponent is  $\varphi(\alpha) = \alpha - \lambda + \lambda\beta(\alpha)$ . With  $\pi(\cdot)$  the Laplace transform of the busy period, it is known that it satisfies  $\pi(\vartheta) = \beta(\vartheta + \lambda - \lambda\pi(\vartheta))$ . Therefore

$$0 = \beta(\vartheta + \lambda - \lambda\pi(\vartheta)) - \pi(\vartheta) = \frac{1}{\lambda} \varphi(\vartheta + \lambda - \lambda\pi(\vartheta)) - \frac{\vartheta}{\lambda},$$

and hence  $\varphi(\vartheta + \lambda - \lambda\pi(\vartheta)) = \vartheta$ . Apply  $\psi(\cdot)$  to both sides, and we obtain

$$\pi(\vartheta) = \frac{\lambda + \vartheta}{\lambda} - \frac{1}{\lambda}\psi(\vartheta).$$

Considering the tail asymptotics of the busy period, first observe that  $\pi(\cdot)$  also has a branching point at  $\vartheta^* < 0$  (i.e., it has the same branching point as  $\rho(\vartheta)$ ), such that, for  $\vartheta \downarrow \vartheta^*$ ,

$$\pi(\vartheta) \sim \frac{\lambda - \vartheta}{\lambda} - \frac{1}{\lambda} \cdot \left( \zeta + \sqrt{\frac{2}{\varphi''(\zeta)}} \cdot \sqrt{\vartheta - \vartheta^*} \right).$$

Applying ‘Heaviside’ now yields, with  $P$  the busy period,

$$\frac{d}{dt}\mathbb{P}(P \leq t) \sim \frac{1}{\lambda} \sqrt{\frac{2}{\varphi''(\zeta)}} \cdot \frac{1}{2\sqrt{\pi}} \frac{e^{\vartheta^* t}}{t\sqrt{t}} = \frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{1}{\beta''(\zeta)}} \frac{e^{\vartheta^* t}}{\lambda t \sqrt{\lambda t}},$$

in line with the results of Cox and Smith [33, Section 5.6]. These asymptotics are indeed proportional to those of Equation (4.16). Similarly, applying the relation

$$\mathbb{E}e^{-\vartheta P} = 1 - \vartheta \int_0^\infty \mathbb{P}(P > t) dt, \quad (4.18)$$

we obtain

$$\mathbb{P}(P > t) \sim -\sqrt{\frac{2}{\varphi''(\zeta)}} \cdot \frac{1}{2\sqrt{\pi}} \frac{e^{\vartheta^* t}}{\vartheta^* \lambda t \sqrt{t}} = -\frac{1}{\sqrt{2\pi}} \cdot \sqrt{\frac{1}{\beta''(\zeta)}} \frac{e^{\vartheta^* t}}{\vartheta^* \lambda t \sqrt{\lambda t}}.$$

♠

## 4.5 Correlation asymptotics for heavy-tailed input

Where the previous section focused on light-tailed Lévy input, we now consider the heavy-tailed case. We extensively use the concept of slowly (and regularly) varying functions. Proposition 4.5.4 is the main result of this section; in Corollary 4.5.5 it is applied to the situation of a queue with CP input with regularly varying jobs.

The following class of functions plays a crucial role in our analysis.

**Definition 4.5.1.** We say that  $f(x) \in \mathcal{R}_\delta(n, \sigma)$ , with  $n \in \mathbb{N}$ ,  $\sigma \in \mathbb{R}$ , and  $\delta \in (n, n+1)$ , for  $x \downarrow 0$  if

$$f(x) = \sum_{i=0}^n \frac{f^{(i)}(0)}{i!} x^i + \sigma x^\delta l(1/x) \quad (x \downarrow 0),$$

for a slowly varying function  $l(\cdot)$  (i.e.,  $l(x)/l(tx) \rightarrow 1$  for  $x \rightarrow \infty$ , for any  $t > 0$ ).

**Lemma 4.5.2.** *Suppose  $\varphi'(\alpha) \in \mathcal{R}_{\delta-1}(n-1, \sigma)$ . Then*

$$\varphi(\alpha) \in \mathcal{R}_{\delta}(n, \sigma/\delta);$$

$$\psi(\vartheta) \in \mathcal{R}_{\delta}(n, \tau), \quad \text{with } \tau := -\frac{\sigma}{\delta(\varphi'(0))^{\delta+1}};$$

$$\psi'(\vartheta) \in \mathcal{R}_{\delta-1}(n-1, \tau\delta),$$

for  $\alpha \downarrow 0$ , resp.  $\vartheta \downarrow 0$ .

*Proof.* The first statement is an immediate consequence of Karamata's theorem; the second statement follows from  $\psi(\varphi(\alpha)) = \alpha$ ; the third statement follows in an elementary way by using  $\psi'(\vartheta) = 1/\varphi'(\psi(\vartheta))$ .  $\square$

The following lemma presents the behavior of  $\xi_e(\vartheta)$  as  $\vartheta \downarrow 0$ . We need this type of results, as Karamata's Tauberian theorem then enables us to translate the behavior of transforms around 0 into the behavior of  $r(t)$  for  $t$  large.

**Lemma 4.5.3.** *If  $\varphi'(\alpha) \in \mathcal{R}_{\delta-1}(n-1, \sigma)$ , with  $n \in \{3, 4, \dots\}$  and  $\delta \in (n, n+1)$ , then*

$$\xi_e(\vartheta) = 1 - \vartheta\rho(\vartheta) \in \mathcal{R}_{\delta-3}(n-3, \omega), \quad \text{with}$$

$$\omega := \frac{(\delta-1)}{v\delta(\varphi'(0))^{\delta-2}}\sigma.$$

*Proof.* Recall Equations (4.9) and (4.10). The crucial step is to verify that

$$\frac{\vartheta}{\psi(\vartheta)} \in \mathcal{R}_{\delta-1}(n-1, -\tau(\varphi'(0))^2) \quad \text{and} \quad \frac{1}{\psi'(\vartheta)} \in \mathcal{R}_{\delta-1}(n-1, -\tau\delta(\varphi'(0))^2);$$

use Lemma 4.5.2. Verification of the claim is now straightforward (though tedious).  $\square$

The Tauberian theorem in Bingham, Goldie, and Teugels [24, Theorem 8.1.6] now yields the following result; see also [23].

**Proposition 4.5.4.** *If  $\varphi'(\alpha) \in \mathcal{R}_{\delta-1}(n-1, \sigma)$ , with  $n \in \{3, 4, \dots\}$  and  $\delta \in (n, n+1)$ , then*

$$r(t) \sim \frac{\omega \cdot (-1)^{n+1}}{\Gamma(4-\delta)} t^{3-\delta} l(t) \quad \text{as } t \rightarrow \infty.$$

*Proof.* First recall that  $r(t) = \mathbb{P}(Z_e > t)$ , and that  $Z_e$  has transform  $\xi_e(\cdot)$ . Lemma 4.5.3 and Theorem 8.1.6 of [24] yield the stated.  $\square$

**Remark 4.5.5.** Interestingly, we can now also find a criterion for long-range dependence, cf. the remarks in the introduction of [95].

Suppose  $\varphi'(\alpha) \in \mathcal{R}_{\delta-1}(n-1, \sigma)$ , with  $n \in \{3, 4, \dots\}$  and  $\delta \in (n, n+1)$ . Then the queueing process is long-range dependent if  $n = 3$ , as in this case  $\int_0^\infty r(t)dt = \infty$ . Consider for instance the case that  $X \in \mathbb{CP}(\lambda, \beta(\cdot))$ , with  $\mathbb{P}(B > t) \sim t^{-\nu}$ , for  $\nu \in (3, 4)$ . Then the first three moments of  $B$  exist, and hence also the first two moments of the steady-state queue length, as well as the covariance  $R(t)$ . The tail of  $J$ , however, is so heavy that  $R(t)$  decays roughly as  $t^{3-\nu}$ , giving rise to a long-range dependent queueing process.

Likewise, it follows that the queueing process is short-range dependent if  $n \in \{4, 5, \dots\}$ , for instance when considering  $\mathbb{CP}$  input with  $\mathbb{P}(B > t) \sim t^{-\nu}$ , for  $\nu \in (4, \infty)$ . ♠

## 4.6 Concluding remarks

In this chapter we studied the correlation function of the workload process of a single-queue fed by a Lévy process (that is, a Lévy process reflected at 0). Relying on the concept of complete monotonicity we have been able to derive a set of structural properties of the correlation function, viz. that it is a positive, decreasing, and convex function. Importantly, we have shown how to represent the correlation function  $r(\cdot)$  as the complementary distribution function of a specific random variable. This representation, as well as an explicit characterization of the Laplace transform of  $r(\cdot)$ , enabled the analysis of the asymptotic behavior of  $r(t)$  for  $t$  large; both the light-tailed and heavy-tailed cases were studied.

An alternative way to conclude that the correlation function is positive, decreasing, and convex, may be the following. The Laplace exponent of any Levy process can be approximated arbitrarily closely by that of an appropriately chosen  $\mathbb{CP}$  process, see e.g. [54, Theorem XVII.1]. As the claim has been proved for  $\mathbb{CP}$  input [95], a limit argument may lead to an alternative proof of Theorem 4.3.6. Exploration of this approach is a subject for further research.

Restricting ourselves to the case of  $\mathbb{CP}$  input, one could say that Section 4 covers the case in which the jobs have a finite moment generating function in a neighborhood of the origin:  $\beta(\alpha) < \infty$  for some  $\alpha < 0$ , and hence all moments are finite. On the other hand, Section 5 addresses the situation in which just a finite set of moments are finite. In between, however, there is a third class of distributions: those for which all moments are finite, but without an analytic continuation for  $\alpha < 0$  (that is  $\beta(\alpha) = \infty$  for all  $\alpha < 0$ ). Examples of distributions in this class are the Weibull and LogNormal distributions. A subject for further research would be the analysis of the correlation asymptotics for this class of distributions.

As we lack, in most cases, an explicit formula for  $r(t)$ , one may attempt to estimate it through simulation. This is particularly challenging, as  $r(t)$  can be extremely small for large  $t$ , and is the difference of two (potentially large) numbers. A way to circumvent this problem is to use *importance sampling* [12, Section V.1], that is, sampling under an alternative measure and correcting the simulation output by likelihood ratios (that keep track of the relative likelihood of the realization under the actual measure, relative to the alternative measure). The resemblance with the busy period asymptotics suggests that, for light-tailed input, the (exponentially-twisted) change of measure proposed in [87] may work well; it is noted that the analysis of [87] indicates that the twisting of the work present at time 0 should be handled with care. An other option could be to rely on the representation of the correlation function  $r(\cdot)$  as the complementary distribution function of the random variable  $Z_e$ , see Theorem 4.3.6.

A potential application area of our results is the following. Suppose that no measurements of the queue's input process can be made, and hence estimation of the probabilistic law of the input process has to be performed in an alternative manner. One approach could be to measure the queue's workload (for instance periodically), and to infer the input characteristics from the resulting measurements. Insight into the correlation between subsequent measurements, as obtained in the present chapter, may be useful when devising such a procedure. Work along these lines for queues with Gaussian input was done in [84] (in a somewhat more experimental context), and for M/G/ $\infty$  systems in [25] (building on the results presented in [102]); see also [60].

## Appendix

### 4.A Complete monotonicity

The concept of complete monotonicity is summarized in the following definition.

**Definition 4.A.1.** A function  $f(\alpha)$  on  $[0, \infty)$  is completely monotone if for all  $n \in \mathbb{N}$ ,  $\alpha \geq 0$ ,

$$(-1)^n \frac{d^n}{d\alpha^n} f(\alpha) \geq 0.$$

We write  $f(\alpha) \in \mathcal{C}$ .

The following deep and powerful result is due to Bernstein [19]. It says that there is equivalence between  $f(\alpha)$  being completely monotone, and the possibility of

writing  $f(\alpha)$  as a Laplace transform. For more background on completely monotone functions, see [54, pp. 439-442].

**Theorem 4.A.2.** *A function  $f(\alpha)$  on  $[0, \infty)$  is the Laplace transform of a non-negative random variable if and only if (i)  $f(\alpha) \in \mathcal{C}$ , and (ii)  $f(0) = 1$ .*

The concept of complete monotonicity is easy to work with, as one can use a set of practical properties.

**Lemma 4.A.3.** *The following properties apply:*

- (1)  $\mathcal{C}$  is closed under addition: if  $f(\alpha) \in \mathcal{C}$  and  $g(\alpha) \in \mathcal{C}$ , then  $f(\alpha) + g(\alpha) \in \mathcal{C}$ . This extends to: if  $f_x(\alpha) \in \mathcal{C}$  for  $x \in \Xi$ , then  $\int_{x \in \Xi} f_x(\alpha) \mu(dx) \in \mathcal{C}$  for any measure  $\mu(\cdot)$ .
- (2)  $\mathcal{C}$  is closed under multiplication: if  $f(\alpha) \in \mathcal{C}$  and  $g(\alpha) \in \mathcal{C}$ , then so does  $f(\alpha)g(\alpha) \in \mathcal{C}$ .
- (3) Properties of composite  $\mathcal{C}$  functions: if  $f(\alpha) \in \mathcal{C}$  and  $g(\alpha) \geq 0$  with  $g'(\alpha) \in \mathcal{C}$ , then  $f(g(\alpha)) \in \mathcal{C}$ .
- (4) Let  $U(\alpha)$  non-decreasing on  $[0, \infty)$ , and  $U(0) = 0$ ,  $u := \lim_{\alpha \rightarrow \infty} U(\alpha) < \infty$ , and

$$f(\alpha) := \int_{[0, \infty)} e^{-\alpha x} dU(x);$$

clearly  $f(\alpha) \in \mathcal{C}$  and  $u = f(0)$ . Then also

$$g(\alpha) := \frac{f(0) - f(\alpha)}{\alpha} \in \mathcal{C}.$$

- (5)  $\mathcal{C}$  closed under differentiation: if  $f(\alpha) \in \mathcal{C}$ , then  $-f'(\alpha) \in \mathcal{C}$ .

*Proof.* (1) is trivial from the definition. (2) follows from [54, Criterion 1], and (3) from [54, Criterion 2]. Property (4) can be found in for instance [95, Equation (4.2)]. (5) is trivial.  $\square$

**Lemma 4.A.4.** *Let  $\{Y(t) : t \geq 0\}$  be an increasing subordinator Lévy process, with Laplace exponent  $\xi(\alpha)$ , then  $-\xi'(\alpha) \in \mathcal{C}$ .*

*Proof.* According to Bertoin [20, Ch. III, Equation (3)], we can write

$$\xi(\alpha) = -d\alpha + \int_{(0, \infty)} (e^{-\alpha x} - 1) \Pi_{\xi}(dx),$$

with  $d \geq 0$ , and measure  $\Pi_{\xi}(\cdot)$  such that  $\int_{(0, \infty)} \min\{1, x^2\} \Pi_{\xi}(dx) < \infty$ . This implies that

$$-\xi'(\alpha) = d + \int_{(0, \infty)} x e^{-\alpha x} \Pi_{\xi}(dx),$$

so that  $-\xi'(\alpha) \in \mathcal{C}$ ; use Lemma 4.A.3.(1).  $\square$



## Chapter 5

---

# Transient asymptotics of Lévy-driven queues

We have seen in Chapter 4 that for spectrally-positive Lévy input the correlation function of the workload process is captured via the Laplace transform even though the correlation function is not explicitly known. Unfortunately this is not the case for general Lévy input. Therefore, in the present chapter we consider the alternative metric  $\mathbb{R}(T|p, q)$  introduced in (1.13) and we study its asymptotics under the large buffer scaling introduced in Chapter 3.

### 5.1 Introduction

Lévy processes are widely used to model various real-life phenomena, for instance in finance and networking, see e.g. [74, 90]. In the literature special attention is paid to two intimately related subjects: fluctuation theory for Lévy processes (predominantly focusing on the analysis of the distribution of the maximal value attained by a Lévy process with negative drift), and to queues fed by Lévy input (studying the probabilistic properties of the workload).

Assuming that the Lévy process does not make negative jumps (i.e., the Lévy process is *spectrally-positive*), the Laplace transform of the steady-state workload  $Q_e$  has been known for over four decades, and is referred to as the (generalized) Pollaczek-Khintchine formula [113]; see also [22] for more background. In addition, the *asymptotics* of  $\mathbb{P}(Q_e > B)$  ( $B$  large) have been identified, in various regimes. Asymptotically exact results for the light-tailed case (or: Cramér case) are presented in [21], cf. also [59], whereas the heavy-tailed case was covered by e.g. [10]; it is furthermore noted that there is also an intermediate case, cf. e.g. [68].

Substantially less attention has been paid to the analysis of transient characteristics of Lévy-driven queues. Again for the case of spectrally-positive Lévy input, in principle the full transient distribution is known, as we have an explicit expression for the double transform

$$F(q, \alpha) := \int_0^\infty e^{-qt} \mathbb{E} \left( e^{-\alpha Q(t)} \mid Q(0) = q \right) dt,$$

with  $Q(t)$  denoting the workload at time  $t > 0$ , and  $q \geq 0$ ; see e.g. [65]. In order to get a handle on the transient distribution, one may use inversion techniques.

Note however that essentially two inversions then need to be performed: one to obtain  $\mathbb{E}(e^{-\alpha Q(t)} \mid Q(0) = q)$  from  $F(q, \alpha)$ , and one to obtain the transient distribution  $\mathbb{P}(Q(t) \leq \cdot \mid Q(0) = q)$  from  $\mathbb{E}(e^{-\alpha Q(t)} \mid Q(0) = q)$ . We remark that [51], see also Chapter 4 in this monograph, uses the double transform mentioned above to analyze the covariance function  $R(t) := \text{Cov}(Q(0), Q(t))$ ; more specifically, it is proved that  $R(\cdot)$  is positive, decreasing, and concave, and in addition its asymptotics (for large  $t$ ) are determined.

In this chapter we choose an alternative approach to analyze transient workload probabilities. Our goal is to assess to what extent the workload at time 0 has impact on the workload at time  $T_B$ , by concentrating on probabilities of the type

$$\Pi_B := \mathbb{P}(Q(0) > pB, Q(T_B) > qB), \quad (5.1)$$

where  $p$  and  $q$  are two positive constants, and  $T_B$  is a given positive function of  $B$ . More specifically, one of our aims is to identify conditions under which  $\Pi(B)$  essentially factorizes (when  $B$  grows large) into  $\mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB)$ , so that it is justified to approximate  $\mathbb{P}(Q(T_B) > qB \mid Q(0) > pB)$  by  $\mathbb{P}(Q_e > qB)$ . It is stressed that we do not impose the assumption that the Lévy input process, say  $\{X(t) : t \in \mathbb{R}\}$ , be spectrally-positive.

Interestingly, the shape of the function  $T_B$  essentially dictates the asymptotics of  $\Pi_B$ . More specifically, this chapter makes the following contributions.

- (i) Our first contribution is the identification of conditions under which

$$\Pi_B \sim \mathbb{P}(Q_e > \max\{p, q\}B), \quad (5.2)$$

or, in other words, the most demanding requirement determines the asymptotics (here ' $\sim$ ' means that the ratio of the left-hand side and right-hand side converges to 1). These conditions essentially boil down to requiring that  $T_B$  is sublinear, that is  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ . The idea behind this property is that the most demanding requirement essentially implies the other requirement with overwhelming probability, as  $B \rightarrow \infty$ .

- (ii) A second contribution is the identification of a condition on  $T_B$  such that

$$\Pi_B \sim \mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB). \quad (5.3)$$

If  $\mathbb{P}(Q_e > B)$  decays (roughly) like  $e^{-B}$  (exponential decay),  $\exp(-B^\alpha)$  with  $0 < \alpha < 1$  (Weibullian decay), then this 'decoupling condition' reduces to  $T_B/B \rightarrow \infty$ . If  $\mathbb{P}(Q_e > B)$  roughly looks like  $B^{-\alpha}$  (polynomial decay), however, then the condition reads  $T_B/B^2 \rightarrow \infty$ ; this class of queues includes two relevant 'heavy-tailed' cases, viz. the situations in which the Lévy input process corresponds to an  $\alpha$ -stable process, and to a compound Poisson process with regularly varying job sizes.

- (iii) For the two ‘heavy-tailed’ cases mentioned above, we determine the asymptotics of  $\Pi_B$  for  $T_B$  increasing superlinearly but subquadratically; in this case the rare event under consideration is essentially due to a single big jump. In the superquadratic case two big jumps are needed, leading to asymptotics (5.3).
- (iv) We pay special attention to the linear case, that is,  $T_B = RB$  for some  $R > 0$ . For light-tailed input we derive intuitively appealing logarithmic asymptotics. If  $R$  is small (that is, fulfilling an explicit criterium in terms of  $p, q$ , and the characteristics of the Lévy process  $\{X(t) : t \in \mathbb{R}\}$ , then we have asymptotics as in (5.2). If this condition does not apply, two cases are possible: for large  $R$  the most likely scenario is that the buffer drains, remains empty for a while, and starts building up relatively short before  $R$  (in this case the asymptotics look like the decoupled asymptotics (5.3)), and for moderate  $R$  the buffer remains (most likely) non-empty between 0 and  $R$ . These three regimes are in line with those identified in e.g. [38] for Gaussian input, see also Chapter 3, [82] for exponential on-off input, as well as [108, Section 11.7] in the setting of an M/M/1 queue. The proofs of our ‘trichotomy’ rely intensively on large deviations techniques, e.g., sample-path large deviations results [43].

The remainder of the chapter is as follows. In Section 2 we introduce the model, and present a number of preliminaries, such as a useful lemma taken from [38], see also Lemma 3.3.1. In Section 3 we address contributions (i) and (ii). Section 4 is devoted to the situation in which  $\mathbb{P}(Q_e > B)$  decays polynomially, that is, contribution (iii). Finally, contribution (iv) is covered by Section 5. Section 6 contains a short summary, discussion, and directions for future research.

## 5.2 Notation and preliminaries

In this chapter we consider a queue fed by a Lévy process  $\{X(t) : t \in \mathbb{R}\}$ , emptied at a constant rate  $c > 0$ ; recall that Lévy processes are stochastic processes with stationary independent increments [74]. Assume that  $\mathbb{E}A(0, 1) = \varpi < c$ , to ensure that the stationary workload exists.

More formally, the steady-state buffer-content process  $\{Q(t) : t \geq 0\}$  is given through

$$Q(t) = \sup_{s \geq 0} (A(t-s, t) - cs) \stackrel{d}{=} Q_e = \sup_{s \geq 0} (A(-s, 0) - cs), \quad (5.4)$$

where  $A(s, t) := X(t) - X(s)$  for  $s \leq t$ .

As mentioned in the introduction, this chapter focuses on analyzing transient

characteristics of the buffer-content process. We define

$$\Pi_B := \mathbb{P}(Q(0) > pB, Q(T_B) > qB).$$

In this chapter the primary focus lies on the asymptotics of  $\Pi_B$  as  $B \rightarrow \infty$ , for given  $p, q > 0$  and some function  $T_B$  that tends to  $\infty$  as  $B \rightarrow \infty$ .

We finish this section with two general lemmas that are used later in the chapter. Directly from (5.4) it can be found that

$$\Pi_B = \mathbb{P}(\exists s \geq 0, t \geq 0 : A(-s, 0) - cs > pB, A(T_B - t, T_B) - ct > qB). \quad (5.5)$$

The following lemma, featuring a reduction property proven in Chapter 3, see also [38], formalizes the evident property that the start of the busy period in which  $T_B$  is contained (corresponding to time, say,  $T_B - t^*$ ), cannot take place before the start  $-s^*$  of the busy period in which 0 is contained, but also not in the interval  $(-s^*, 0]$ . In other words: the only two options are that both busy periods start at the same epoch (then  $t^* = T_B + s^*$ ), and that the busy period in which 0 is contained ends before  $T_B$  (then  $t^* \in [0, T_B)$ ). It means that in (5.5) we can restrict ourselves to a subset of  $s, t \geq 0$ .

**Lemma 5.2.1.** *Let*

$$\mathcal{E} := \{(s, t) : s \geq 0, t \in [0, T_B) \cup \{T_B + s\}\}.$$

*Then*

$$\Pi_B = \mathbb{P}(\exists (s, t) \in \mathcal{E} : A(-s, 0) - cs > pB, A(T_B - t, T_B) - ct > qB).$$

We finally state a weak law of large numbers, which holds due to the fact that  $X(t)$  is integrable.

**Lemma 5.2.2.** *For any  $\delta > 0$ ,*

$$\lim_{t \rightarrow \infty} \mathbb{P}\left(\frac{X(t)}{t} < \varpi - \delta\right) = \lim_{t \rightarrow \infty} \mathbb{P}\left(\frac{X(t)}{t} > \varpi + \delta\right) = 0.$$

### 5.3 General results

In this section we prove two general results. The first says that (5.2) holds under the plausible condition that  $T_B/B \rightarrow 0$ ; in the sequel we call this the *short time-scale regime*. The second identifies a condition under which the asymptotic decoupling (5.3) holds; notably, as mentioned in the introduction, this condition does not necessarily reduce to  $T_B/B \rightarrow \infty$ . We refer to the latter regime as the *long time-scale regime*.

### 5.3.1 Short time-scale regime

In this subsection we prove our result for the short time-scale regime; as before  $Q_e$  denotes the stationary workload. It consists of two cases: the case  $p > q$  which holds under the condition  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ , and the case  $q > p$  which holds under Assumption 5.3.1. We stress that later in this chapter we will show that both in heavy-tailed scenarios and in light-tailed scenarios Assumption 5.3.1 is fulfilled as long as  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ .

**Assumption 5.3.1.** *One of the following two properties holds:*

(i) *The sequence  $T_B$  is such that, for all  $\eta > 0$ ,*

$$\limsup_{B \rightarrow \infty} \frac{\mathbb{P}(\exists t \in (0, T_B) : X(t) - ct > \eta B)}{\mathbb{P}(Q_e > qB)} = 0.$$

(ii) *The sequence  $T_B$  is such that, for all  $\eta > 0$ ,*

$$\mathbb{P}(Q_e > qB + \eta T_B) \sim \mathbb{P}(Q_e > qB) \text{ as } B \rightarrow \infty.$$

**Theorem 5.3.2.** *Case  $p > q$ : If  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ , then*

$$\Pi_B \equiv \mathbb{P}(Q(0) > pB, Q(T_B) > qB) \sim \mathbb{P}(Q_e > pB).$$

*Case  $q > p$ : Under Assumption 5.3.1,*

$$\Pi_B \equiv \mathbb{P}(Q(0) > pB, Q(T_B) > qB) \sim \mathbb{P}(Q_e > qB).$$

*Proof.* First consider the case  $p > q$ . We are left to prove that

$$\liminf_{B \rightarrow \infty} \frac{\mathbb{P}(Q(0) > pB, Q(T_B) > qB)}{\mathbb{P}(Q_e > pB)} \geq 1.$$

This is proven as follows. Fix  $\varepsilon > 0$ . Let  $B$  be sufficiently large such that

$$(p - q)B > (c - \varpi + \varepsilon)T_B$$

(which is possible due to  $T_B/B \rightarrow 0$  and  $p > q$ ). Then

$$\mathbb{P}(Q(0) > pB, Q(T_B) > qB) \geq \mathbb{P}(Q_e > pB) \cdot \mathbb{P}(X(T_B) > (\varpi - \varepsilon)T_B).$$

This is evidently true if  $T_B$  is bounded, and if it is not, then due to Lemma 5.2.2 we have that for any  $\delta > 0$  and for  $B$  large enough  $\mathbb{P}(X(T_B) > (\varpi - \varepsilon)T_B) > 1 - \delta$ . The stated follows by letting  $\delta \downarrow 0$ .

Now focus on  $q > p$ , first under Assumption 5.3.1.(i). Now it suffices to prove that, as  $B \rightarrow \infty$ , we have that  $\mathbb{P}(Q(0) < pB, Q(T_B) > qB) = o(\mathbb{P}(Q_e > qB))$ . Let  $\mathcal{T}_B$  be the event that  $Q(t) > 0$  for all  $t \in (0, T_B)$ . First observe that, with  $\eta := q - p > 0$ ,

$$\begin{aligned} \mathbb{P}(Q(0) < pB, Q(T_B) > qB, \mathcal{T}_B) &\leq \mathbb{P}(X(T_B) > \eta B + cT_B) \\ &\leq \mathbb{P}(\exists t \in (0, T_B) : X(t) - ct > \eta B), \end{aligned}$$

which is  $o(\mathbb{P}(Q_e > qB))$  due to Assumption 5.3.1.(i). Also,

$$\begin{aligned} \mathbb{P}(Q(0) < pB, Q(T_B) > qB, \mathcal{T}_B^c) &\leq \mathbb{P}(Q(T_B) > qB, \mathcal{T}_B^c) \\ &\leq \mathbb{P}(\exists t \in (0, T_B) : A(T_B - t, T_B) - ct > qB), \end{aligned}$$

which is also  $o(\mathbb{P}(Q_e > qB))$ , again by Assumption 5.3.1.(i).

Again consider the case  $q > p$ , but now under Assumption 5.3.1.(ii). It is clear that it suffices to show that

$$\liminf_{B \rightarrow \infty} \frac{\Pi_B}{\mathbb{P}(Q > qB)} \geq 1.$$

For each positive  $N$ ,

$$\begin{aligned} \Pi_B &\geq \mathbb{P}(Q(0) > qB + NT_B, Q(T_B) > qB) \\ &\geq \mathbb{P}(Q_e > qB + NT_B) \cdot \mathbb{P}(X(T_B) > (c - N)T_B). \end{aligned}$$

Now observe that, by assumption,  $\mathbb{P}(Q_e > qB + NT_B) \sim \mathbb{P}(Q_e > qB)$  as  $B \rightarrow \infty$ . Moreover, for each  $\epsilon > 0$  there exists an  $N_0$  such that, for each  $N \geq N_0$ , it holds that  $\mathbb{P}(X(T_B) > (c - N)T_B) \geq 1 - \epsilon$  for sufficiently large  $B$ . Thus, as  $B \rightarrow \infty$ ,

$$\mathbb{P}(Q_e > qB + NT_B) \cdot \mathbb{P}(X(T_B) > (c - N)T_B) \sim \mathbb{P}(Q_e > qB) \cdot \mathbb{P}(X(T_B) > (c - N)T_B),$$

which is larger than  $(1 - \epsilon)\mathbb{P}(Q_e > qB)$ . The stated follows by letting  $\epsilon \downarrow 0$ . This completes the proof.  $\square$

**Remark 5.3.3.** *The case  $p = q$ .* The case  $p = q$  should be handled with care; it is readily checked from the proof of Theorem 5.3.2 that the argumentation for  $q > p$  works for  $q \geq p$  under Assumption 5.3.1.(ii), but not under Assumption 5.3.1.(i).

Let us now check how Assumption 5.3.1.(ii) relates to the condition  $T_B/B \rightarrow 0$ . In case that  $\mathbb{P}(Q_e > B)$  decays (roughly) polynomially (i.e.,  $\mathbb{P}(Q_e > B) \sim KB^{-\zeta}$ ), then Assumption 5.3.1.(ii) indeed reduces to  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ . It is noted, however, that if  $\mathbb{P}(Q_e > B)$  decays (roughly) exponentially, then Assumption 5.3.1.(ii) reads  $T_B \rightarrow 0$ .

We now argue that Assumption 5.3.1.(ii) is, in the case  $p = q$ , ‘minimal’ if the probability  $\mathbb{P}(Q_e > B)$  decays exponentially, in the sense that

$$\liminf_{B \rightarrow \infty} T_B = M > 0 \text{ leads to } \limsup_{B \rightarrow \infty} \frac{\Pi_B}{\mathbb{P}(Q_e > pB)} < 1,$$

as follows. Consider for instance the case that  $\{X(t) : t \in \mathbb{R}\}$  corresponds to (standard) Brownian motion. Decompose  $\Pi_B$  into  $\Pi_B^{(1)} + \Pi_B^{(2)}$ , where  $\mathcal{T}_B$  is defined in the proof of Theorem 5.3.2 and

$$\begin{aligned}\Pi_B^{(1)} &:= \mathbb{P}(Q(0) > pB, Q(T_B) > pB, \mathcal{T}_B), \\ \Pi_B^{(2)} &:= \mathbb{P}(Q(0) > pB, Q(T_B) > pB, \mathcal{T}_B^c).\end{aligned}$$

First observe that

$$\begin{aligned}\Pi_B^{(2)} &\leq \mathbb{P}(Q(0) > pB, \exists t \in [0, T_B] : A(t, T_B - t) > pB + ct) \\ &= \mathbb{P}(Q(0) > pB) \cdot \mathbb{P}(\exists t \in [0, T_B] : A(t, T_B - t) > pB + ct) \\ &\leq (\mathbb{P}(Q_e > pB))^2 = o(\mathbb{P}(Q_e > pB)).\end{aligned}$$

Regarding  $\Pi_B^{(1)}$ , first recall that  $\mathbb{P}(Q_e > B) = e^{-2cB}$ . We find

$$\begin{aligned}\Pi_B^{(1)} &\leq \mathbb{P}(Q(0) > pB, Q(0) + X(T_B) > pB + cT_B) \\ &= \int_{pB}^{\infty} \mathbb{P}(X(T_B) > pB + cT_B - x) \cdot 2ce^{-2cx} dx \\ &= \int_0^{\infty} \mathbb{P}(X(T_B) > cT_B - y) \cdot 2ce^{-2c(y+pB)} dy \\ &= \mathbb{P}(Q(0) + X(T_B) > cT_B) \cdot \mathbb{P}(Q_e > pB).\end{aligned}$$

Since  $\liminf_{B \rightarrow \infty} T_B = M > 0$ , we have that

$$\limsup_{B \rightarrow \infty} \mathbb{P}(Q(0) + X(T_B) > cT_B) < 1,$$

and as a consequence that

$$\limsup_{B \rightarrow \infty} \frac{\Pi_B^{(1)}}{\mathbb{P}(Q_e > pB)} < 1,$$

and therefore also

$$\limsup_{B \rightarrow \infty} \frac{\Pi_B}{\mathbb{P}(Q_e > pB)} < 1.$$

This shows that Assumption 5.3.1.(ii) is ‘minimal’ for the case  $p = q$ . ♠

### 5.3.2 Long time-scale regime

The main goal of this section is to prove our result for the long time-scale regime. A crucial role is played by the following assumption. Recall that  $Q_e$  denotes the stationary workload; we also define (for  $d > \varpi$ )  $Q_e^d$  as the stationary workload if the queue were emptied at rate  $d$  rather than  $c$ .

**Assumption 5.3.4.** *The sequence  $T_B$  is such that, for all  $\eta > 0$ ,  $d > \varpi$ ,*

$$\limsup_{B \rightarrow \infty} \frac{\mathbb{P}(Q_e^d > \eta T_B)}{\mathbb{P}(Q_e > pB) \cdot \mathbb{P}(Q_e > qB)} = 0.$$

In the next sections we relate this assumption to the behavior of  $T_B$  as  $B \rightarrow \infty$ . It turns out that depending on the driving Lévy process being heavy-tailed or light-tailed, various regimes need to be distinguished.

**Theorem 5.3.5.** *Under Assumption 5.3.4, it holds that*

$$\Pi_B \equiv \mathbb{P}(Q(0) > pB, Q(T_B) > qB) \sim \mathbb{P}(Q_e > pB) \cdot \mathbb{P}(Q_e > qB).$$

*Proof.* Let us start by establishing the lower bound. By definition,

$$\begin{aligned} & \mathbb{P}(Q(0) > pB, Q(T_B) > qB) \\ &= \mathbb{P}(\exists s \geq 0 : A(-s, 0) > pB + cs, \exists t \geq 0 : A(T_B - t, T_B) > qB + ct). \end{aligned}$$

The probability in the right-hand side of the previous display majorizes

$$\begin{aligned} & \mathbb{P}(\exists s \geq 0 : A(-s, 0) > pB + cs, \exists t \in (0, T_B) : A(T_B - t, T_B) > qB + ct) \\ &= \mathbb{P}(\exists s \geq 0 : A(-s, 0) > pB + cs) \cdot \mathbb{P}(\exists t \in (0, T_B) : A(T_B - t, T_B) > qB + ct) \\ &= \mathbb{P}(Q_e > pB) \cdot \mathbb{P}(\exists t \in (0, T_B) : A(-t, 0) > qB + ct). \end{aligned}$$

We observe that it is left to prove that

$$\frac{\mathbb{P}(\exists t > T_B : A(-t, 0) > qB + ct)}{\mathbb{P}(Q_e > qB)} \rightarrow 0 \tag{5.6}$$

as  $B \rightarrow \infty$ . Let us consider the numerator of (5.6). It is trivial that

$$\begin{aligned} & \mathbb{P}(\exists t > T_B : A(-t, 0) > qB + ct) \\ &= \mathbb{P}\left(\left(\sup_{t > T_B} A(-t, -T_B) - c(t - T_B)\right) + A(-T_B, 0) > qB + cT_B\right) \\ &= \mathbb{P}(Q(-T_B) + A(-T_B, 0) > qB + cT_B). \end{aligned}$$

We now distinguish between  $Q(-T_B)$  being either smaller or larger than  $\delta cT_B$ , so that the previous expression is not larger than

$$\mathbb{P}(Q(-T_B) + A(-T_B, 0) > qB + cT_B, Q(-T_B) < \delta cT_B) + \mathbb{P}(Q(-T_B) \geq \delta cT_B).$$

First consider the second probability, which evidently equals  $\mathbb{P}(Q_e \geq \delta cT_B)$ . Due to Assumption 5.3.4,  $\mathbb{P}(Q_e \geq \delta cT_B)$  is  $o(\mathbb{P}(Q_e > qB))$  — in fact, Assumption 5.3.4 implies that it is even  $o(\mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB))$ , as we will need below.

To deal with the first probability, pick  $\varepsilon > 0$  such that  $c^* := \varpi + \varepsilon < (1 - \delta)c$ ; then

$$\begin{aligned} & \mathbb{P}(Q(-T_B) + A(-T_B, 0) > qB + cT_B, Q(-T_B) < \delta cT_B) \\ & \leq \mathbb{P}(A(-T_B, 0) > (1 - \delta)cT_B) = \mathbb{P}(A(-T_B, 0) - c^*T_B > ((1 - \delta)c - c^*)T_B) \\ & \leq \mathbb{P}(\exists t \geq 0 : A(-t, 0) - c^*t > ((1 - \delta)c - c^*)T_B) = \mathbb{P}(Q_e^{c^*} > ((1 - \delta)c - c^*)T_B), \end{aligned}$$

which is  $o(\mathbb{P}(Q_e > qB))$  due to Assumption 5.3.4 — again, it is even

$$o(\mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB)).$$

We now proceed by establishing the upper bound. In view of Lemma 5.2.1, we can split the probability of interest on the basis of the queue having been empty in  $(0, T_B)$  or not, thus obtaining

$$\mathbb{P}(Q(0) > pB, Q(T_B) > qB, \mathcal{T}_B^c) + \mathbb{P}(Q(0) > pB, Q(T_B) > qB, \mathcal{T}_B). \quad (5.7)$$

The first of the probabilities in (5.7) equals

$$\begin{aligned} & \mathbb{P}(\exists s \geq 0 : A(-s, 0) > pB + cs, \exists t \in (0, T_B) : A(T_B - t, T_B) > qB + ct) \\ & = \mathbb{P}(\exists s \geq 0 : A(-s, 0) > pB + cs) \cdot \mathbb{P}(\exists t \in (0, T_B) : A(T_B - t, T_B) > qB + ct) \\ & \leq \mathbb{P}(\exists s \geq 0 : A(-s, 0) > pB + cs) \cdot \mathbb{P}(\exists t \geq 0 : A(-t, 0) > qB + ct) \\ & = \mathbb{P}(Q_e > pB) \cdot \mathbb{P}(Q_e > qB). \end{aligned}$$

The second of the probabilities in (5.7) equals

$$\begin{aligned} & \mathbb{P}(\exists s > 0 : A(-s, 0) > pB + cs, A(-s, T_B) > qB + c(T_B + s)) \\ & \leq \mathbb{P}(\exists s > 0 : A(-s, T_B) > qB + c(T_B + s)) \\ & = \mathbb{P}(\exists s > T_B : A(-s, 0) > qB + cs). \end{aligned}$$

Above we saw that  $\mathbb{P}(\exists s > T_B : A(-s, 0) > qB + cs)$  is

$$o(\mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB))$$

as  $B$  grows large. This observation completes the proof.  $\square$

## 5.4 Heavy-tailed Lévy input

In this section we focus on the situation that the tail distribution of  $Q_e$  decays essentially polynomially.

**Assumption 5.4.1.** For a  $\zeta$ , all  $d > \varpi$ , and  $K(\cdot) > 0$ ,

$$\mathbb{P}(Q_e^d > qB) \cdot B^\zeta \rightarrow K(d),$$

as  $B \rightarrow \infty$ .

Let us first check what Assumptions 5.3.1 and 5.3.4 look like in this situation.

- Consider Assumption 5.3.1(ii). As has been noticed in Remark 5.3.3, this assumption is valid under  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ .
- Now consider Assumption 5.3.4. It is readily checked that under Assumption 5.4.1 this does *not* reduce to  $T_B/B \rightarrow \infty$ , but to  $T_B/B^2 \rightarrow \infty$ .

We here mention that, interestingly, Assumption 5.3.4 *does* reduce to requiring that  $T_B/B \rightarrow \infty$  for  $B \rightarrow \infty$  in a number of specific situations in which the tail distribution of  $Q_e$  decays subexponentially (but faster than polynomially); this is for instance the case when  $\log \mathbb{P}(Q_e^d > B)/B^\alpha \rightarrow -\kappa(d)$  as  $B \rightarrow \infty$  for  $\alpha \in (0, 1)$  and some  $\kappa(\cdot) > 0$  (Weibullian decay). Interestingly, in the situation that  $\log \mathbb{P}(Q_e^d > B)/(\log B)^2 \rightarrow -\kappa(d)$  (which is a tail that resembles that of the lognormal distribution), Assumption 5.3.4 holds if

$$(\log(\eta T_B))^2 - (\log(pB))^2 - (\log(qB))^2 \rightarrow \infty;$$

with  $T_B$  of the type  $B^\beta$ , this simplifies to requiring that  $\beta > \sqrt{2}$ .

The above observations indicate that, for  $\mathbb{P}(Q_e > B)$  behaving as  $B^{-\zeta}$ , the situations that are left to investigate are those in which  $T_B$  is between linear and quadratic. In this section we analyze this case.

As a first observation, we notice that Lemma 5.2.1 entails that we can decompose  $\Pi_B$  into

$$\begin{aligned} & \mathbb{P} \left( \begin{array}{l} \exists s \geq 0, \exists t \in [0, T_B] : A(-s, 0) - cs > pB, A(T_B - t, T_B) - ct > qB \\ \vee \exists s \geq 0 : A(-s, 0) - cs > pB, A(-s, T_B) - c(s + T_B) > qB \end{array} \right) \\ &= \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2), \end{aligned}$$

where

$$\begin{aligned} E_1 &:= \{ \exists s \geq 0, \exists t \in [0, T_B] : A(-s, 0) - cs > pB, A(T_B - t, T_B) - ct > qB \}, \\ E_2 &:= \{ \exists s \geq 0 : A(-s, 0) - cs > pB, A(-s, T_B) - c(s + T_B) > qB \}. \end{aligned}$$

The following two lemmas are useful in our proofs.

**Lemma 5.4.2.** *The following three statements hold:*

(i) for any  $B > 0$ ,

$$\mathbb{P}(E_1) = \mathbb{P}(Q_e > pB) \cdot \mathbb{P} \left( \sup_{t \in [0, T_B]} (X(t) - ct) > qB \right);$$

(ii) for  $\varepsilon \in (0, 1)$ , if  $T_B \rightarrow \infty$  as  $B \rightarrow \infty$ , then  $\mathbb{P}(E_2) \sim p_1(B) + p_2(B)$ , where

$$\begin{aligned} p_1(B) &:= \mathbb{P}(Q_e > pB) \cdot \mathbb{P}(X(T_B) > cT_B + (q-p)B), \\ p_2(B) &:= \mathbb{P}(Q_e > qB + (c-\varpi)T_B + \varepsilon(T_B + B)) \\ &\quad \cdot \mathbb{P}(X(T_B) \in [\varpi T_B - \varepsilon(T_B + B), cT_B + (q-p)B]); \end{aligned}$$

(iii) if  $T_B = RB^2$  for some  $R > 0$ , then under Assumption 5.4.1 we have

$$\mathbb{P}(E_1 \cap E_2) = o(\mathbb{P}(E_1)) \quad \text{as } B \rightarrow \infty.$$

*Proof.* Claim (i) follows directly from the independence of the increments of the process  $\{X(t) : t \in \mathbb{R}\}$ . Now concentrate on Claim (ii). To make the notation a bit lighter, we write  $T$  instead of  $T_B$  throughout this proof. Observe that

$$\begin{aligned} \mathbb{P}(E_2) &= \mathbb{P}(Q(0) > \max\{pB, qB + cT - X(T)\}) \\ &= \mathbb{P}(Q(0) > \max\{pB, qB + cT - X(T)\}, X(T) > cT + (q-p)B) \\ &\quad + \mathbb{P}(Q(0) > \max\{pB, qB + cT - X(T)\}, X(T) \leq cT + (q-p)B) \\ &= \mathbb{P}(Q_e > pB)\mathbb{P}(X(T) > cT + (q-p)B) + \mathbb{P}(E_{21}), \end{aligned}$$

where  $E_{21} := \{Q(0) > qB + cT - X(T), X(T) \leq cT + (q-p)B\}$ . We first consider  $\mathbb{P}(E_{21})$ . Let  $\varepsilon \in (0, 1)$ . Then

$$\mathbb{P}(E_{21}) = \mathbb{P}(E_{211}) + \mathbb{P}(E_{212}), \quad (5.8)$$

with

$$\begin{aligned} E_{211} &:= \{Q(0) > qB + cT - X(T), X(T) \in [\varpi T - \varepsilon(T+B), cT + (q-p)B]\} \\ E_{212} &:= \{Q(0) > qB + cT - X(T), X(T) < \varpi T - \varepsilon(T+B)\}. \end{aligned}$$

First consider  $\mathbb{P}(E_{211})$  which equals

$$\begin{aligned} &\mathbb{P}\left(\begin{array}{l} Q(0) > qB + cT + \max\{\varepsilon(T+B) - \varpi T, -X(T)\}, \\ X(T) \in [\varpi T - \varepsilon(T+B), cT + (q-p)B] \end{array}\right) \\ &\quad + \mathbb{P}\left(\begin{array}{l} qB + cT - X(T) < Q(0) \leq qB + (c-\varpi)T + \varepsilon(T+B), \\ X(T) \in [\varpi T - \varepsilon(T+B), cT + (q-p)B] \end{array}\right) \\ &= \mathbb{P}\left(\begin{array}{l} Q(0) > qB + cT + \max\{\varepsilon(T+B) - \varpi T, -X(T)\}, \\ X(T) \in [\varpi T - \varepsilon(T+B), cT + (q-p)B] \end{array}\right) \quad (5.9) \\ &= \mathbb{P}\left(\begin{array}{l} Q(0) > qB + (c-\varpi)T + \varepsilon(T+B), \\ X(T) \in [\varpi T - \varepsilon(T+B), cT + (q-p)B] \end{array}\right) \\ &= \mathbb{P}(Q(0) > qB + (c-\varpi)T + \varepsilon(T+B)) \\ &\quad \cdot \mathbb{P}(X(T) \in [\varpi T - \varepsilon(T+B), cT + (q-p)B]). \end{aligned}$$

Now consider  $\mathbb{P}(E_{212})$ . Observe that

$$\begin{aligned}\mathbb{P}(E_{212}) &\leq \mathbb{P}(Q(0) > qB + (c - \varpi)T + \varepsilon(T + B), X(T) < \varpi T - \varepsilon(T + B)) \\ &= \mathbb{P}(Q(0) > qB + (c - \varpi)T + \varepsilon(T + B)) \mathbb{P}(X(T) < \varpi T - \varepsilon(T + B)) \\ &= o(\mathbb{P}(Q_e > qB + (c - \varpi)T))\end{aligned}\quad (5.10)$$

as  $B \rightarrow \infty$ , because

$$\begin{aligned}\mathbb{P}(X(T) < \varpi T - \varepsilon(T + B)) &= \mathbb{P}\left(\frac{X(T)}{T} < \varpi - \frac{\varepsilon(T + B)}{T}\right) \\ &\leq \mathbb{P}\left(\frac{X(T)}{T} < \varpi - \varepsilon\right) \rightarrow 0\end{aligned}$$

due to Lemma 5.2.2. Upon combining (5.8) with (5.9) and (5.10), we have established Claim (ii).

Finally consider Claim (iii). Let  $\delta \in (0, \frac{1}{2})$  and  $\varepsilon > 0$ . We have

$$\begin{aligned}\mathbb{P}(E_1 \cap E_2) &= \mathbb{P}(E_1 \cap E_2, X(T) \geq (\varpi + \varepsilon)T + T^{1-\delta}) \\ &\quad + \mathbb{P}(E_1 \cap E_2, X(T) \leq (\varpi + \varepsilon)T + T^{1-\delta}) \\ &\leq \mathbb{P}(Q(0) > pB, X(T) \geq (\varpi + \varepsilon)T + T^{1-\delta}) \\ &\quad + \mathbb{P}\left(Q(0) > qB + (c - \varpi - \varepsilon)T - T^{1-\delta}, \sup_{t \in [0, T]} (A(T - t, T) - ct) > qB\right) \quad (5.11) \\ &= \mathbb{P}(Q(0) > pB) \mathbb{P}(X(T) \geq (\varpi + \varepsilon)T + T^{1-\delta}) \\ &\quad + \mathbb{P}(Q(0) > qB + (c - \varpi - \varepsilon)T - T^{1-\delta}) \mathbb{P}\left(\sup_{t \in [0, T]} (A(0, t) - ct) > qB\right).\end{aligned}$$

Since  $T = RB^2$  and  $\delta \in (0, \frac{1}{2})$ , for some constant  $\bar{K} > 0$ ,

$$\begin{aligned}\mathbb{P}(X(T) \geq (\varpi + \varepsilon)T + T^{1-\delta}) &\leq \mathbb{P}\left(\sup_{t \geq 0} (X(t) - (\varpi + \varepsilon)t) \geq T^{1-\delta}\right) \\ &\sim \bar{K}(T^{1-\delta})^{-\zeta} = o(B^{-\zeta});\end{aligned}$$

use Assumption 5.4.1. We thus conclude that

$$\mathbb{P}(Q_e > pB) \mathbb{P}(X(T) \geq (\varpi + \varepsilon)T + T^{1-\delta}) = o(\mathbb{P}(E_1)).$$

We also have

$$\begin{aligned}\mathbb{P}(Q_e > qB + (c - \varpi - \varepsilon)T - T^{1-\delta}) &\sim \mathbb{P}(Q_e > (c - \varpi - \varepsilon)T) \\ &= O(B^{-2\zeta}) = O(\mathbb{P}(E_1));\end{aligned}$$

$$\begin{aligned} \mathbb{P}\left(\sup_{t \in [0, T]} (X(t) - ct) > qB\right) &\leq \mathbb{P}\left(\sup_{t \geq 0} (X(t) - ct) > qB\right) \\ &= \mathbb{P}(Q_e > qB) = O(B^{-\zeta}) \end{aligned}$$

as  $B \rightarrow \infty$ , which in view of (5.11) implies that  $\mathbb{P}(E_1 \cap E_2) = o(\mathbb{P}(E_1))$ . This completes the proof of (iii).  $\square$

**Lemma 5.4.3.** *Under Assumption 5.4.1, for each  $R > 0$ , as  $B \rightarrow \infty$ ,*

$$\mathbb{P}\left(\sup_{t \in [0, RB^2]} (X(t) - ct) > B\right) \sim \mathbb{P}(Q_e > B).$$

*Proof.* It clearly suffices to establish the lower bound. We have

$$\mathbb{P}\left(\sup_{t \in [0, RB^2]} (X(t) - ct) > B\right) \geq \mathbb{P}(Q_e > B) - \mathbb{P}\left(\sup_{t > RB^2} (X(t) - ct) > B\right). \quad (5.12)$$

Also, with  $Q^*$  denoting a random variable that is distributed as the stationary workload  $Q_e$ , and which is independent of  $X(RB^2)$ ,

$$\begin{aligned} &\mathbb{P}\left(\sup_{t > RB^2} (X(t) - ct) > B\right) \\ &= \mathbb{P}\left(X(RB^2) + \sup_{t > RB^2} (X(t) - X(RB^2) - c(t - RB^2)) > B + cRB^2\right) \\ &= \mathbb{P}(X(RB^2) + Q^* > B + cRB^2) \sim \mathbb{P}(X(RB^2) + Q^* > cRB^2). \end{aligned}$$

Now realize that by Assumption 5.4.1  $\mathbb{P}(Q_e > cRB^2)$  is asymptotically proportional to  $B^{-2\zeta}$  as  $B \rightarrow \infty$ , and

$$\begin{aligned} \mathbb{P}(X(RB^2) > cRB^2) &= \mathbb{P}(X(RB^2) - (\varpi + \varepsilon)RB^2 > (c - \varpi - \varepsilon)RB^2) \\ &\leq \mathbb{P}\left(\sup_{t > 0} (X(t) - (\varpi + \varepsilon)t) > (c - \varpi - \varepsilon)RB^2\right) \\ &\sim \check{K}B^{-2\zeta}, \end{aligned}$$

for some positive constant  $\check{K}$  (again by Assumption 5.4.1), so that the probability  $\mathbb{P}(X(RB^2) + Q^* > cRB^2)$  is roughly proportional to  $B^{-2\zeta}$  as well, as follows from basic properties of regularly varying distributions. The stated is now a direct consequence of (5.12) and the fact that  $\mathbb{P}(Q_e > B)$  is asymptotically proportional to  $B^{-\zeta}$ .  $\square$

We now present two propositions that, for the case that  $T_B$  is at least linear but slower than quadratic, express the asymptotics of  $\Pi_B$  in terms of the asymptotics of  $\mathbb{P}(Q_e > B)$ , viz. Proposition 5.4.4 for the case  $q \geq p$  and Proposition 5.4.6 for the case  $p > q$ . Corollaries 5.4.5 and 5.4.7 summarize the findings so far.

**Proposition 5.4.4.** *Let  $q \geq p$ .*

(i) *If  $\liminf_{B \rightarrow \infty} T_B/B \geq R$  for some  $R > 0$  and  $T_B/B^2 \rightarrow 0$  as  $B \rightarrow \infty$ , then*

$$\Pi_B \sim \mathbb{P}(Q_e > qB + (c - \varpi)T_B); \quad (5.13)$$

(ii) *If  $T_B = RB^2$  for some  $R > 0$ , then*

$$\Pi_B \sim \mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB) + \mathbb{P}(Q_e > (c - \varpi)T_B). \quad (5.14)$$

*Proof.* To prove Claim (i), it suffices to show  $\mathbb{P}(E_1) = o(\mathbb{P}(E_2))$ . From Lemma 5.4.2.(i) it immediately follows that  $\mathbb{P}(E_1) \leq \mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB)$ . As, for  $q \geq p$ , we have that  $p_1(B) = o(p_2(B))$ , we also have, by letting  $\varepsilon \downarrow 0$  in Lemma 5.4.2.(ii), that  $\mathbb{P}(E_2) \sim \mathbb{P}(Q_e > qB + (c - \varpi)T_B)$ . It also holds that

$$\mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB) = o(\mathbb{P}(Q_e > qB + (c - \varpi)T_B))$$

as  $B \rightarrow \infty$ . This completes the proof of Claim (i).

Now consider Claim (ii). If  $T_B = RB^2$ , then, following Lemmas 5.4.2 and 5.4.3,

$$\mathbb{P}(E_1) = \mathbb{P}(Q_e > pB)\mathbb{P}\left(\sup_{t \in [0, T_B]} (X(t) - ct) > qB\right) \sim \mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB)$$

and

$$\mathbb{P}(E_2) \sim \mathbb{P}(Q_e > pB + (c - \varpi)T_B) \sim \mathbb{P}(Q_e > (c - \varpi)T_B),$$

as  $B \rightarrow \infty$ . Since  $\mathbb{P}(E_1) = O(\mathbb{P}(E_2))$ , it now suffices to recall that due to Lemma 5.4.2.(iii) it holds that  $\mathbb{P}(E_1 \cap E_2) = o(\mathbb{P}(E_1))$ . We thus establish Claim (ii).  $\square$

The following corollary is an immediate consequence of Theorems 5.3.2, 5.3.5, and 5.4.4, Remark 5.3.3 and Lemma 5.4.3.

**Corollary 5.4.5.** *Let  $q \geq p$ .*

(i) *If  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ , then  $\Pi_B \sim Kq^{-\zeta}B^{-\zeta}$ ;*

(ii) *If  $T = RB$  for some  $R > 0$ , then  $\Pi_B \sim K(q + cR)^{-\zeta}B^{-\zeta}$ ;*

(iii) *If  $T_B/B \rightarrow \infty$  and  $T_B/B^2 \rightarrow 0$  as  $B \rightarrow \infty$ , then  $\Pi_B \sim K(cT_B)^{-\zeta}$ ;*

(iv) *If  $T_B = RB^2$  for some  $R > 0$ , then*

$$\Pi_B \sim (K^2(pq)^{-\zeta} + (cR)^{-\zeta})B^{-2\zeta};$$

(v) *If  $T_B/B^2 \rightarrow \infty$  as  $B \rightarrow \infty$ , then  $\Pi_B \sim K^2(pq)^{-\zeta}B^{-2\zeta}$ .*

We now switch to the case  $q < p$ .

**Proposition 5.4.6.** *Let  $q < p$ .*

(i) *If  $T_B = RB$  with  $R \leq (p - q)/B$ , then  $\Pi_B \sim \mathbb{P}(Q_e > pB)$ ;*

(ii) *if  $\liminf_{B \rightarrow \infty} T_B/B > (p - q)/c$  and  $T_B/B^2 \rightarrow 0$  as  $B \rightarrow \infty$ , then*

$$\Pi_B \sim \mathbb{P}(Q_e > qB + (c - \varpi)T_B);$$

(iii) *if  $T = RB^2$  as  $B \rightarrow \infty$  for some  $R > 0$ , then*

$$\Pi_B \sim \mathbb{P}(Q_e > pB)\mathbb{P}(Q_e > qB) + \mathbb{P}(Q_e > (c - \varpi)T_B).$$

*Proof.* We only consider Claim (i); the other claims can be proven as the corresponding statements in Proposition 5.4.4. Notice that

$$\mathbb{P}(Q_e > pB) \cdot \mathbb{P}(X(T_B) > \varpi T_B + (c - \varpi + (q - p)/R)T_B) \sim \mathbb{P}(Q_e > pB),$$

due to the weak law of large numbers (Lemma 5.2.2); the probability  $p_2(B)$  corresponds to two rare events (use  $c - \varpi + (q - p)/R < 0$ ), such that  $p_2(B) = o(p_1(B))$ . As a consequence, Lemma 5.4.2.(ii) entails that  $\mathbb{P}(E_2) \sim \mathbb{P}(Q_e > pB)$ . Combining this with Lemma 5.4.2.(i), we conclude that  $\mathbb{P}(E_1) = o(\mathbb{P}(E_2))$ . This implies  $\Pi_B \sim \mathbb{P}(Q_e > pB)$ , which completes the proof of (i).  $\square$

**Corollary 5.4.7.** *Let  $q < p$ .*

(i) *If  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ , or  $T = RB$  with  $R \leq (p - q)/c$ , then  $\Pi_B \sim Kp^{-\zeta}B^{-\zeta}$ ;*

(ii) *If  $T_B = RB$  for  $R > (p - q)/c$ , then  $\Pi_B \sim K(q + cR)^{-\zeta}B^{1-\zeta}$ ;*

(iii) *If  $T_B/B \rightarrow \infty$  and  $T_B/B^2 \rightarrow 0$  as  $B \rightarrow \infty$ , then  $\Pi_B \sim K(cT_B)^{-\zeta}$ ;*

(iv) *If  $T_B = RB^2$  as  $B \rightarrow \infty$  for some  $R > 0$ , then*

$$\Pi_B \sim (K^2(pq)^{-\zeta} + (cR)^{-\zeta})B^{-2\zeta};$$

(v) *If  $T_B/B^2 \rightarrow \infty$  as  $B \rightarrow \infty$ , then  $\Pi_B \sim K^2(pq)^{-\zeta}B^{-2\zeta}$ .*

In the remainder of this section we consider two special cases: (A)  $\alpha$ -stable input, and (B) compound Poisson input with polynomially decaying job size distribution.

(A)  *$\alpha$ -stable input.* Let  $X(t)$  be an  $\alpha$ -stable Lévy process [105] with  $\alpha \in (1, 2)$  and  $\beta \in (-1, 1]$ . We use the notation

$$\mathbb{B}(\alpha, \beta) := \frac{\Gamma(1 + \alpha)}{\pi} \sqrt{1 + \beta^2 \tan^2(\pi\alpha/2)} \sin\left(\frac{\pi\alpha}{2} + \arctan(\beta \tan(\pi\alpha/2))\right).$$

Then, due to [98], Assumption 5.4.1 is valid with

$$K = \frac{\mathbb{B}(\alpha, \beta)}{c\alpha(\alpha - 1)},$$

and  $\zeta = \alpha - 1$ . Hence the theory developed earlier in this section can be applied.

(B) *Compound Poisson input with polynomially decaying job sizes.* Consider a Poissonian arrival stream (with rate  $\lambda$ ) of i.i.d. jobs. Let the distribution of the jobs obey  $\mathbb{P}(J^r > x) \sim \kappa x^{-\zeta}$ , for positive  $\zeta, \kappa$ , where  $J^r$  denotes the *residual* job length:

$$\mathbb{P}(J^r > x) = \frac{1}{\mathbb{E}J} \int_x^\infty \mathbb{P}(J > y) dy.$$

Note that  $\varpi = \lambda \cdot \mathbb{E}J$ . Then [26, 32]

$$\mathbb{P}(Q_e > x) \sim \frac{\varpi}{c - \varpi} \kappa x^{-\zeta}.$$

Conclude that again Assumption 5.4.1 (and hence the theory of this section) applies, with an obvious value for  $K$ .

## 5.5 Light-tailed input

In this section we derive the logarithmic asymptotics of  $\Pi_B$  as  $B \rightarrow \infty$ , for the case of light-tailed input. We impose the following assumption.

**Assumption 5.5.1.** *With*

$$\beta^* := \inf\{\beta : \mathbb{E}e^{-\beta X(1)} < \infty\},$$

assume that  $\beta^* < 0$ . Let  $\varphi(\vartheta) := \log \mathbb{E} \exp(-\vartheta X(1))$ , and assume that there exists  $\vartheta^* \in (\beta^*, 0)$ , such that  $\varphi(\vartheta^*) + c\vartheta^* = 0$ .

We first recall in Proposition 5.5.2 a result that is a special case of [59, Theorem 4], which states that the tail probabilities of the steady-state workload decay essentially exponentially. Bearing in mind Assumption 5.3.4, this means that Theorem 5.3.5 holds when  $T_B/B \rightarrow \infty$ . In Lemma 5.5.5 we will show that Assumption 5.3.1 applies if  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ , and hence the case  $T_B/B \rightarrow 0$  is covered by Theorem 5.3.2.

The above means that the only case left to analyze is the linear case, and therefore the rest of this section concentrates on  $T_B = RB$  for some  $R > 0$ . It turns out that three intuitively appealing regimes can be distinguished (small  $R$ , moderate  $R$ , large  $R$ ); at the end of this section we provide more insight in these regimes.

In the following proposition, we let  $Q_e$  denote the stationary workload of a Lévy-driven queue, i.e., let  $Q_e$  be distributed as  $\sup_{t>0}(X(t) - ct)$ .

**Proposition 5.5.2.** *Under Assumption 5.5.1 it holds that*

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \mathbb{P}(Q_e > B) = \vartheta^*. \quad (5.15)$$

**Remark 5.5.3.** We give here an alternative proof of the upper bound associated with the above result, as it provides interesting additional insight, and the proof technique will be used again in the proof of Lemma 5.5.5. Importantly, we obtain the uniform upper bound  $\mathbb{P}(Q_e > B) \leq e^{\vartheta^* B}$ .

Under the assumption  $\varpi < c$ , evidently the queueing system is stable under the measure  $\mathbb{P}$ . We will now perform a change of measure, with which we associate  $\mathbb{Q}$ , under which overflow occurs with probability 1, by application of an exponential twist  $\vartheta^*$ . Under the light-tailed assumption, we have that the Laplace exponent  $\varphi(\vartheta)$  of  $X(t)$  is well defined and characterized through, with  $\sigma^2 > 0$  and a measure  $\Pi_\varphi(\cdot)$  such that  $\int_{(0,\infty)} \min\{1, x^2\} \Pi_\varphi(dx) < \infty$ ,

$$\varphi(\vartheta) = \vartheta\delta + \frac{1}{2}\vartheta^2\sigma^2 + \int_{(0,\infty)} (e^{-\vartheta x} - 1 + \vartheta x 1_{(0,1)}) \Pi_\varphi(dx).$$

It is now a matter of straightforward calculations to show that

$$\bar{\varphi}(\vartheta) := \varphi(\vartheta + \vartheta^*) - \varphi(\vartheta^*)$$

is a Laplace exponent as well. Under  $\mathbb{Q}$ , the Lévy process has Laplace exponent  $\bar{\varphi}(\vartheta)$ ; from the convexity of  $\varphi(\cdot)$  it is concluded that (in self-evident notation)  $\mathbb{E}_{\mathbb{Q}} X(1) = -\bar{\varphi}'(\vartheta^*) > \varpi$ , so that the system under the new measure is indeed unstable. (One can check that under  $\mathbb{Q}$  the drift has increased to  $\delta - \vartheta^* \sigma^2$ , the Brownian term remains unchanged, whereas the measure  $\Pi_{\bar{\varphi}}(dx)$  is given through the exponentially twisted version  $e^{-\vartheta^* x} \Pi_\varphi(dx)$ ).

Suppose one would compute  $\mathbb{P}(\sup_{t>0} X(t) - ct > B)$  by simulating under  $\mathbb{Q}$ . There is the fundamental equality, with  $\chi$  denoting the indicator function of the event  $\{\sup_{t>0} X(t) - ct > B\}$

$$\mathbb{P}\left(\sup_{t>0} X(t) - ct > B\right) = \mathbb{E}_{\mathbb{Q}}(L\chi),$$

cf. [11, Theorem XIII.3.2], where  $L$  denotes the likelihood ratio (to be understood as a Radon-Nikodým derivative) of the value of the Lévy process under  $\mathbb{P}$  with respect to  $\mathbb{Q}$ ; it is a standard result that at time  $t$  this likelihood ratio equals  $e^{\vartheta^* X(t)} \exp(\varphi(\vartheta^*)t)$ . Let  $\sigma_B$  be defined as the first epoch at which  $X(t)$  exceeds  $B + ct$  (which is a stopping time); as  $\chi = 1$  with  $\mathbb{Q}$ -probability 1, we thus obtain

$$\mathbb{P}\left(\sup_{t>0} X(t) - ct > B\right) = \mathbb{E}_{\mathbb{Q}} e^{\vartheta^* X(\sigma_B)} e^{\varphi(\vartheta^*)\sigma_B} = \mathbb{E}_{\mathbb{Q}} e^{\vartheta^* X(\sigma_B)} e^{-c\vartheta^* \sigma_B}.$$

As by definition  $X(\sigma_B) \geq B + c\sigma_B$ , we thus find that  $\mathbb{P}(Q_e > B) \leq e^{\vartheta^* B}$ .  $\spadesuit$

In the next lemma we relate the decay rate  $\vartheta^*$  to the large deviations rate function, defined through  $I(r) := \sup_{\vartheta \geq 0} (\vartheta r - \varphi(-\vartheta))$ , and an associated variational problem.

**Lemma 5.5.4.** *It holds that*

$$-\vartheta^* = \inf_{r > c} \frac{I(r)}{r - c}. \quad (5.16)$$

*Proof.* Let the minimizer in the right-hand side of (5.16) be  $r^*$ , satisfying

$$(r^* - c)I'(r^*) = I(r^*).$$

Define in addition

$$\vartheta(r) := \arg \sup_{\vartheta \geq 0} (\vartheta r - \varphi(-\vartheta)),$$

so that  $I(r) = \vartheta(r)r - \varphi(-\vartheta(r))$ . Noticing that  $\vartheta(r)$  satisfies  $r + \varphi'(-\vartheta) = 0$ , we find that

$$I'(r) = \vartheta'(r)r + \vartheta(r) + \vartheta'(r)\varphi(-\vartheta(r)) = \vartheta(r).$$

From the facts that  $\vartheta^*$  solves  $\varphi(\vartheta^*) + c\vartheta^* = 0$  and

$$\vartheta(r^*)r^* - \varphi(-\vartheta(r^*)) = I(r^*) = (r^* - c)I'(r^*) = (r^* - c)\vartheta(r^*),$$

we conclude that  $-\vartheta(r^*) = \vartheta^*$ , which proves the claim.  $\square$

As indicated in the beginning of this section, like in the heavy-tailed case, in this light-tailed case we again have that Assumption 5.3.1 is valid if  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ . This is proven in the following lemma. We recall that it entails that the only case left to analyze is the linear case, that is,  $T_B = RB$  for some  $R > 0$ .

**Lemma 5.5.5.** *Under Assumption 5.5.1, Assumption 5.3.1.(i) applies if  $T_B/B \rightarrow 0$  as  $B \rightarrow \infty$ .*

*Proof.* Let  $\mathbb{Q}(\vartheta)$  be the probability measure obtained after exponentially twisting the original probability measure  $\mathbb{P}$  with twist  $\vartheta < 0$ , as in Remark 5.5.3. In a similar fashion, it follows that

$$\mathbb{P}(\exists t \in (0, T_B) : X(t) - ct > qB) \leq \mathbb{E}_{\mathbb{Q}(\vartheta)} e^{\vartheta(B+c\sigma_B)} e^{\varphi(\vartheta)\sigma_B},$$

where  $\sigma_B$  is the minimum of  $T_B$  and the first epoch at which  $X(t) - ct$  exceeds  $B$  (which is a stopping time). It then follows that for all  $\vartheta < 0$ , bearing in mind that  $\sigma_B \leq T_B = o(B)$ ,

$$\begin{aligned} & \limsup_{B \rightarrow \infty} \frac{1}{B} \log \mathbb{P}(\exists t \in (0, T_B) : X(t) - ct > qB) \\ & \leq \limsup_{B \rightarrow \infty} \left( \vartheta + \vartheta c \frac{\sigma_B}{B} + \varphi(\vartheta) \frac{\sigma_B}{B} \right) = \vartheta. \end{aligned}$$

This entails that  $\mathbb{P}(\exists t \in (0, T_B) : X(t) - ct > qB)$  decays superexponentially:

$$\limsup_{B \rightarrow \infty} \frac{1}{B} \log \mathbb{P}(\exists t \in (0, T_B) : X(t) - ct > qB) \leq \inf_{\vartheta < 0} \vartheta = -\infty.$$

Combining this with Proposition 5.5.2, the stated follows.  $\square$

From now on we just consider the case that  $T_B = RB$ . The next proposition shows that for small  $R$  the decay rate of interest equals the decay rate of the ‘most binding event’, cf. Theorem 5.3.2. We denote

$$\bar{R} := \max \left\{ \frac{p-q}{c-\varpi}, \frac{q-p}{r^*-c} \right\}.$$

**Proposition 5.5.6.** *If  $R < \bar{R}$ , then*

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B = \max\{p, q\} \vartheta^*.$$

*Proof.* First suppose  $p > q > 0$ . The upper bound follows immediately from Proposition 5.5.2:

$$\limsup_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B \leq \limsup_{B \rightarrow \infty} \frac{1}{B} \log \mathbb{P}(Q_e > pB) = p\vartheta^*.$$

Now consider the lower bound, which we establish by applying the lower bound of a sample-path large deviations result. We here rely on de Acosta [43, Theorem 5.1], which can be applied to obtain

$$\liminf_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B \geq -\mathbb{I}(f), \quad (5.17)$$

for any  $f \in \mathcal{A}$ , where

$$\mathbb{I}(f) := \int_{-\infty}^{\infty} I(f'(\tau)) d\tau,$$

and the set of paths  $\mathcal{A}$  is given by

$$\mathcal{A} := \{f : \exists(\sigma, \tau) \in \mathcal{E} : -f(-\sigma) \geq c\sigma + p, f(R) - f(R - \tau) \geq c\tau + q\}.$$

Now consider the continuous path  $f^*$  through the origin that has slope  $r^*$  between  $-p/(r^* - c)$  and 0, and slope  $\varpi$  elsewhere; clearly

$$\mathbb{I}(f) := \int_{-p/(r^*-c)}^0 I(f'(\tau)) d\tau = \frac{p}{r^*-c} \cdot I(r^*) = -p\vartheta^*.$$

Claim 1 now follows from the observation that  $f^* \in \mathcal{A}$ , as

$$-f\left(-\frac{p}{r^*-c}\right) = \frac{pr^*}{r^*-c} = \frac{pc}{r^*-c} + p,$$

and, by virtue of  $R < (p - q)/(c - \varpi)$ ,

$$f(R) - f\left(-\frac{p}{r^* - c}\right) = \varpi R + \frac{pr^*}{r^* - c} > c\left(R + \frac{p}{r^* - c}\right) + q.$$

Claim (2) can be proven along the same lines. The upper bound is identical, and in the lower bound we again use Theorem 5.1 of [43], but now with a path  $f^*$  that has slope  $r^*$  between  $R - q/(r^* - c)$  and  $R$ , and  $\varpi$  elsewhere. The stated follows after checking that this path is in  $\mathcal{A}$  if  $R < (q - p)/(r^* - c)$ .  $\square$

In the sequel we use the following lemma extensively, see [44, Lemma 1.2.15].

**Lemma 5.5.7.** *For any finite integer  $M$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left( \sum_{i=1}^M \alpha_{i,n} \right) = \max_{i=1, \dots, M} \left( \lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_{i,n} \right).$$

**Proposition 5.5.8.** *If  $R > \bar{R}$ , then*

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B = p\vartheta^* + \max\{q\vartheta^*, -\psi(R)\},$$

where

$$\psi(R) := R \cdot I\left(c + \frac{q - p}{R}\right).$$

*Proof.* First we establish the upper bound, which consists of five steps.

STEP I. The probability of interest  $\Pi_B$  can be decomposed as  $\Pi_B^{(1)} + \Pi_B^{(2)}$ , with

$$\begin{aligned} \Pi_B^{(1)} &:= \mathbb{P}(Q(0) > pB, Q(RB) > qB, \forall t \in (0, RB) : Q(t) > 0), \\ \Pi_B^{(2)} &:= \mathbb{P}(Q(0) > pB, Q(RB) > qB, \exists t \in (0, RB) : Q(t) = 0). \end{aligned}$$

STEP II. We first observe that we can bound  $\Pi_B^{(2)}$  as follows:

$$\begin{aligned} \Pi_B^{(2)} &= \mathbb{P}(Q(0) > pB, \exists t \in (0, RB) : A(RB - t, RB) - ct \geq qB) \\ &= \mathbb{P}(Q(0) > pB) \mathbb{P}(\exists t \in (0, RB) : A(RB - t, RB) - ct \geq qB) \\ &\leq \mathbb{P}(Q_e > pB) \mathbb{P}(\exists t \geq 0 : A(RB - t, RB) - ct \geq qB) \\ &= \mathbb{P}(Q_e > pB) \mathbb{P}(Q_e > qB), \end{aligned}$$

and hence

$$\begin{aligned} \lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B^{(2)} &\leq \lim_{B \rightarrow \infty} \frac{1}{B} \log \mathbb{P}(Q_e > pB) + \lim_{B \rightarrow \infty} \frac{1}{B} \log \mathbb{P}(Q_e > qB) \\ &= (p + q)\vartheta^*. \end{aligned} \tag{5.18}$$

STEP III. Now let us focus on  $\Pi_B^{(1)}$ ; in this scenario the busy period in which  $R$  is contained starts at the same epoch as the busy period in which  $0$  is contained. Hence

$$\Pi_B^{(1)} = \mathbb{P}(\exists s \geq 0 : A(-s, 0) - cs > pB, A(-s, RB) - c(RB + s) > qB).$$

Let  $\varepsilon > 0$  be picked arbitrary; let  $M$  be some natural number, whose value we specify later. Then  $\Pi_B^{(1)}$  is majorized by

$$\begin{aligned} \sum_{k=0}^{M-1} \mathbb{P} \left( \exists s \geq 0 : \begin{array}{l} A(-s, 0) - cs \in ((p + k\varepsilon)B, (p + (k+1)\varepsilon)B]; \\ A(-s, RB) - c(RB + s) > qB \end{array} \right) \\ + \mathbb{P}(\exists s \geq 0 : A(-s, 0) - cs > (p + M\varepsilon)B). \end{aligned} \quad (5.19)$$

Now the  $k$ -th term in the summation of the previous display is bounded from above by

$$\mathbb{P}(\exists s \geq 0 : A(-s, 0) - cs > (p + k\varepsilon)B) \times \mathbb{P}(A(0, RB) - cRB > (q - (p + (k+1)\varepsilon))B),$$

which we call  $\zeta_B^{(k)}$ . Due to Proposition 5.5.2 and Cramér's theorem,

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \zeta_B^{(k)} = (p + k\varepsilon)\vartheta^* - R \cdot I \left( c + \frac{q - p - (k+1)\varepsilon}{R} \right).$$

We have now found that (5.19) is not larger than

$$\sum_{k=0}^{M-1} \zeta_B^{(k)} + \mathbb{P}(Q_e > (p + M\varepsilon)B),$$

and therefore, due to Lemma 5.5.7,

$$\begin{aligned} & \lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B^{(1)} \\ & \leq \max \left\{ \begin{array}{l} \max_{0 \leq k \leq M-1} \left\{ (p + k\varepsilon)\vartheta^* - R \cdot I \left( c + \frac{q - p - (k+1)\varepsilon}{R} \right) \right\}, \\ (p + M\varepsilon)\vartheta^* \end{array} \right\}. \end{aligned}$$

STEP IV. We now study how  $g_k := (p + k\varepsilon)\vartheta^* - R \cdot I(\Delta_k/R)$  behaves when varying  $k$ , with  $\Delta_k := cR + q - p - (k+1)\varepsilon$ . Because of the convexity of  $I(\cdot)$ , we see that  $g_k$  is concave in  $k$ . This means that proving  $g_1 \leq g_0$  also yields that  $\max_{k=0, \dots, M-1} g_k = g_0$ . To this end, first observe that, owing to the convexity of  $I(\cdot)$  and using that  $\Delta_1 < \Delta_0$ ,

$$\begin{aligned} g_0 - g_1 &= -\varepsilon\vartheta^* + R \left( I \left( \frac{\Delta_1}{R} \right) - I \left( \frac{\Delta_0}{R} \right) \right) \\ &\geq -\varepsilon\vartheta^* + (\Delta_1 - \Delta_0) I' \left( \frac{\Delta_1}{R} \right) \\ &= -\varepsilon \left( I' \left( \frac{\Delta_1}{R} \right) + \vartheta^* \right). \end{aligned}$$

Now recall that  $\vartheta^* = -I'(r^*)$ , and that  $I'(\cdot)$  is increasing. It follows that  $g_1 \leq g_0$  if  $\Delta_1 < r^*R$ , which is true under  $R > (q-p)/(r^* - c)$  and  $\varepsilon$  sufficiently small. We conclude, noting that we can take  $M$  arbitrarily large,

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B^{(1)} \leq g_0 = p\vartheta^* - R \cdot I\left(c + \frac{q-p-\varepsilon}{R}\right). \quad (5.20)$$

STEP V. By letting  $\varepsilon \downarrow 0$  in (5.20), applying the upper bound on the decay rates of both  $\Pi_B^{(1)}$  and  $\Pi_B^{(2)}$ , and Lemma 5.5.7 once more, we have

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B \leq p\vartheta^* + \max\left\{q\vartheta^*, -R \cdot I\left(c + \frac{q-p}{R}\right)\right\}.$$

This completes the proof of the upper bound.

The lower bound follows again from sample-path large deviations arguments [43].

- Let us first consider the case that

$$q\vartheta^* > -R \cdot I\left(c + \frac{q-p}{R}\right). \quad (5.21)$$

Condition (5.21) implies that  $R \geq q/(r^* - c)$ , as can be seen as follows. Supposing  $R < q/(r^* - c)$ , and recalling that we have  $R > (q-p)/(r^* - c)$ , it would follow that

$$R \cdot I\left(c + \frac{q-p}{R}\right) < \frac{q}{r^* - c} I(r) = -q\vartheta^*,$$

which is a contradiction; note that we also used that  $c + (q-p)/R > \varpi$ .

Using that we know that (5.21) implies  $R \geq q/(r^* - c)$ , it can be seen that the path  $f^*$  through the origin that has slope  $r^*$  between  $-p/(r^* - c)$  and 0, and also between  $R - q/(r^* - c) > 0$  and  $R$ , and slope  $\varpi$  elsewhere, is indeed feasible (that is, lies in  $\mathcal{A}$ ). It is also readily verified that  $\mathbb{I}(f^*) = -(p+q)\vartheta^*$ , as required.

- Now suppose that (5.21) does not hold. Define  $f^*$  as the path through the origin with slope  $r^*$  between  $-p/(r^* - c)$  and 0, slope  $c + (q-p)/R$  between 0 and  $R$ , and slope  $\varpi$  elsewhere. It is easily seen that this path is feasible and, by applying the definition of  $\mathbb{I}(\cdot)$ ,

$$\mathbb{I}(f^*) = -p\vartheta^* + R \cdot I\left(c + \frac{q-p}{R}\right),$$

as desired.

This concludes the proof of the lower bound.  $\square$

**Lemma 5.5.9.** *For all  $R > \bar{R}$ , we have that  $\psi(R)$  is increasing. In addition we have that  $\psi(\bar{R}) \leq -q\vartheta^*$ .*

*Proof.* Observe, recalling that  $I'(r) = \vartheta(r)$ , that

$$\psi'(R) = -\frac{q-p}{R} \cdot \vartheta\left(c + \frac{q-p}{R}\right) + I\left(c + \frac{q-p}{R}\right).$$

First consider the case  $p > q$ , such that  $\bar{R} = (p-q)/(c-\varpi)$ . It then holds that  $c + (q-p)/\bar{R} = \varpi$ , so that

$$\psi'(\bar{R}) = -\frac{q-p}{\bar{R}} \cdot I'(\varpi) + I(\varpi) = 0,$$

due to  $I(\varpi) = I'(\varpi) = 0$ . We are done if we can prove that  $\psi'(R)$  increases for  $R \geq \bar{R}$ . To this end, we compute  $\psi''(R)$ ; it is easily verified that  $I'(r) = \vartheta(r)$  entails that

$$\psi''(R) = \frac{(q-p)^2}{R^3} I''\left(c + \frac{q-p}{R}\right),$$

which is indeed non-negative because of the convexity of  $I(\cdot)$ .

We now consider the case  $q \geq p$ , i.e.,  $\bar{R} = (q-p)/(r^* - c)$ . It then holds that  $c + (q-p)/\bar{R} = r^*$ , so that

$$\psi'(\bar{R}) = (c - r^*) \cdot I'(r^*) + I(r^*) = 0,$$

see the proof of Lemma 5.5.4. Again, we are done if we can prove that  $\psi'(R)$  increases for  $R \geq \bar{R}$ , which follows in the same fashion as above.

We finally consider  $\psi(\bar{R})$ . In case  $p > q$ , this equals 0, which is evidently below  $-q\vartheta^*$ . In case  $q \geq p$ , we have

$$\psi(\bar{R}) = \frac{q-p}{r^* - c} I(r^*) = -(q-p)\vartheta^* \leq -q\vartheta^*.$$

This completes the proof. □

The following claim is an immediate consequence of the previous lemma.

**Corollary 5.5.10.** *There is a unique solution (larger than  $\bar{R}$ ) to  $\psi(R) = -q\vartheta^*$ , say  $\check{R}$ . For all  $R \in (\bar{R}, \check{R})$  we have  $\psi(\bar{R}) \leq -q\vartheta^*$ , for all  $R > \check{R}$  we have  $\psi(\bar{R}) > -q\vartheta^*$ .*

Application of Props. 5.5.6, 5.5.8 and this corollary immediately lead to the following theorem.

**Theorem 5.5.11.** (i) *For  $R \leq \bar{R}$  we have*

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B = \max\{p, q\}\vartheta^*.$$

(ii) For  $R \in (\bar{R}, \check{R})$  we have

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B = p\vartheta^* - \psi(R).$$

(iii) For  $R \geq \check{R}$  we have

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B = (p + q)\vartheta^*.$$

Summarizing, we have identified the decay rate of  $\Pi(B)$ , and we found three regimes for  $R$ . This could be dealt with explicitly, in that we presented closed-form expressions for the decay rate, as well as for the critical values of  $R$  that separate three regimes, which could be anticipated in view of earlier work, see e.g. [38, 82] and [108, Section 11.7]. The three regimes have an appealing intuitive explanation.

- For small values of  $R$ , that is the case of 5.5.11(i), the ‘tightest’ of the events  $\{Q(0) > pB\}$  and  $\{Q(RB) > qB\}$  will essentially imply the other event, which is leading to the decay rate  $\max\{p, q\}\vartheta^*$ .
- Then there is an intermediate range of values of  $R$ , case 5.5.11(ii), for which both the events  $\{Q(0) > pB\}$  and  $\{Q(RB) > qB\}$  are tight, but that the time epochs 0 and  $RB$  lie in the same busy period with overwhelming probability. The decay rate  $p\vartheta^*$  represents the requirement that  $pB$  has to be exceeded at time 0, and then  $cRB + (q - p)B$  traffic has to be generated in the next  $RB$  time units, leading to the contribution  $-\psi(R)$ .
- Finally, for large  $R$ , case 5.5.11(iii), still both events are tight, but now they occur in different busy periods with overwhelming probability, so that the joint probability effectively decouples (thus leading to the decay rate  $(p + q)\vartheta^*$ ).

Theorem 5.5.11 has made this heuristic rigorous. We finish this section with an example.

**Example 5.5.12.** Consider the Brownian case, that is,  $\varphi(\vartheta) = -\varpi\vartheta + \frac{1}{2}\vartheta^2$ . It is easy to derive that  $I(a) = \frac{1}{2}(a - \varpi)^2$  and  $\vartheta^* = -2(c - \varpi)$ . The solution  $\check{R}$  (larger than  $\bar{R}$ ) of  $q\vartheta^* = -\psi(R)$  is

$$\check{R} = \frac{(\sqrt{p} + \sqrt{q})^2}{(c - \varpi)},$$

in line with Proposition 5.1 of [38].

◇

## 5.6 Discussion and concluding remarks

In this chapter we analyzed the asymptotics of  $\Pi_B$  for  $B$  large. We showed that for  $T_B$  increasing sublinearly, the asymptotics reduce to those of the most demanding event, cf. (5.2). We also identified a criterion under which the events become asymptotically independent ('decoupling'), cf. (5.3). The latter criterion reduces to  $T_B/B \rightarrow \infty$  in many situations, a notable exception being the case that  $\mathbb{P}(Q_e > B)$  decays polynomially (in which case the condition is  $T_B/B^2 \rightarrow \infty$ ).

While this chapter gives a fairly complete picture of all possible regimes, a number of special cases are still open. For instance when  $\mathbb{P}(Q_e > B)$  looks like  $\exp(-B^\alpha)$  for some  $\alpha \in (0, 1)$ , or like  $B^{-\log B}$ , the above mentioned criterion for decoupling is  $T_B/B \rightarrow \infty$ , but it remains unclear what happens when  $T_B = RB$  for some  $R > 0$ . It is expected that delicate analysis is needed to obtain the asymptotics in these situations. Another topic for future research concerns the exact asymptotics for the light-tailed case and  $T_B = RB$ .



## Chapter 6

---

# Markov fluid-driven queues

The present chapter considers another important class of input processes used in modeling storage systems, viz., Markov-fluid input processes. For this class of processes we will mainly focus on two transient characteristics: the busy period of the system and the covariance  $R(t)$  of the workload process at time 0 and time  $t$ .

### 6.1 Introduction

Markov fluid models have been widely studied in a variety of application domains, with significant contributions made in the areas of queueing theory, storage processes, communication networking, insurance risk, etc., see for instance [8, 14, 70, 71, 99, 107]. A *Markov-fluid-driven queue* is a storage system which is fed by a source whose transmission rate modulates between multiple values in a Markovian manner, and which is emptied at constant speed. Traditionally in the literature emphasis was laid on computing *steady-state* characteristics of this class of queueing systems — in particular the distribution of the stationary workload; see for a nice (recent) literature overview for instance [72] and the introduction of [35] — whereas considerably less attention has been paid to *transient analysis*. The motivation behind studying the transient characteristics is that knowledge of the dependence structure of the queueing process is in many practical situations quite relevant. Indeed, this knowledge would give a handle on the time scale after which it is justified to approximate transient probabilities by their steady-state counterpart. The main goal of the present chapter is the analysis of two such transient characteristics: (i) the distribution of the busy period, and (ii) the correlation function of the workload process.

Let us first give a brief (non-exhaustive) account of the literature on transient analysis of Markov-fluid-driven queues. Restricting themselves to the special case in which the Markov fluid source is actually a superposition of on-off sources, Ren and Kobayashi [101] were able to convert the partial differential equations [8, 70, 71] that govern the queue's transient behavior, into a matrix equation in the Laplace domain. Roughly simultaneously, a paper by Asmussen [9] appeared, where the Laplace transform of the busy period was computed, mainly relying on martingale techniques; the resulting transform is in terms of a matrix of probabilities, which

are not given explicitly, but are fixed points of a specific integral equation (which are proven to be unique). Narayanan and Kulkarni [73] showed that the probability distribution of the first time the buffer becomes empty (starting at an arbitrary positive level  $x$ ), satisfies a system of partial differential equations; furthermore, they show that its Laplace transform is a solution of a specific differential equation. Barbot *et al.* [17] mainly focused on numerical issues: using the fact that the probability distribution of the busy period obeys a certain backward differential equation, they proposed an efficient numerical procedure. Finally, recent work by Ahn and Ramaswami [7] provides an efficient (quadratically convergent) procedure for computing the Laplace transform of the busy period, exploiting relations with so-called quasi-birth-death processes. To the best of our knowledge, no results on the correlation function have been reported so far.

As said above, this chapter focuses on the distribution of the busy period, as well as the workload's correlation function. More specifically, the main contributions are the following.

- *Busy period.* In the first place we adopt a new approach for computing the Laplace transform of the busy period. This approach, some steps of which resonate elements of [7], first uses elementary calculations to express the Laplace transform in terms of a number of auxiliary transforms (as many as there are states with net buffer increase). Then an important role is played by a lemma that provides us with a sufficient number of additional constraints in order to uniquely determine these auxiliary transforms. The proof of this lemma is based on a powerful result by Sonneveld [109], in conjunction with Geršgorin's circle theorem, see [83].

It is stressed that this new methodology has a number of attractive properties. Most of the analysis is based on first principles; the above-mentioned lemma is the only technical element. In addition, the analysis essentially carries over to the correlation function, see below. Also, in special cases, such as the case of a two-state modulating Markov chain, the analysis can be done explicitly, and the resulting transform can be inverted.

Then we focus on the logarithmic asymptotics of the tail distribution of the busy period. It is shown that these can be written in terms of the minimum of the so-called *cumulant function* (i.e., the asymptotic log-moment generating function) of the input process. The upper bound is elementary, viz. a direct application of the Gärtner-Ellis theorem. The lower bound, on the contrary, is considerably more technical and relies on sample-path large deviations [31].

- *Correlation function.* With  $Q(t)$  denoting the buffer content at time  $t$ , the covariance function  $R(t) = \text{Cov}(Q(0), Q(t))$  is a measure of dependence between the workload at time 0, and the workload at time  $t$  (and the correlation function is

defined as the covariance function divided by  $\sqrt{\text{Var}Q(0)\text{Var}Q(t)}$ . Using the methodology that we developed for analyzing the busy period, we uniquely characterize the covariance function by its Laplace transform. Assuming that the workload was in stationarity at time 0, this has been done before in the cases of compound Poisson input [95] and spectrally-positive Lévy input [51], but the case of Markov-fluid input was not addressed yet. We do not restrict ourselves to the case that  $Q(0)$  is distributed according to the workload's equilibrium distribution; in fact, for any initial distribution of phase-type the Laplace transform of  $R(t)$  has a relatively manageable form.

Again for the case of on-off Markov-fluid input, the correlation function can be determined explicitly. Interestingly, its asymptotics are equal (up to a multiplicative constant) to those of the tail distribution of the busy period — a property that was observed before in the case of queues with spectrally-positive Lévy input, see Chapter 4.

The remainder of this chapter is organized as follows. In Section 6.2 we describe our model and recapitulate a number of known results on the steady-state workload distribution. In Section 6.3 we identify the Laplace transform of the busy period. Its logarithmic asymptotics are derived in Section 6.4, invoking sample-path large deviations. In Section 6.5 we concentrate on the covariance function of the workload process. Section 6.6 presents an example that illustrates the results obtained in this chapter. We draw conclusions and identify a number of open problems in Section 6.7.

## 6.2 Model and preliminaries

Let  $\{J(t), t \geq 0\}$  be an irreducible continuous-time Markov process with finite state space  $\mathcal{E} = \{1, 2, \dots, N\}$ . This modulating Markov process drives a buffer in the following way: if it is in state  $i$ , the buffer content changes at rate  $r_i$  (which can be both positive and negative); there is *reflection at zero*, meaning that if the buffer is empty, and the Markov process is in a state  $i$  with  $r_i < 0$ , then the buffer remains empty. We denote by  $\{Q(t), t \geq 0\}$  the buffer content process (or: workload process). The buffer size is assumed to be infinite, and hence  $Q(t)$  can attain any value in  $[0, \infty)$ .

In order to avoid confusions in the notation, in the sequel the bold small letters will denote vectors, and bold large letters will denote matrices. Let  $\mathbf{\Lambda} = (\lambda_{ij})_{1 \leq i, j \leq N}$  be the intensity matrix (or: rate matrix) of the Markov process  $J(t)$ , with  $\lambda_i = -\lambda_{ii}$ . Also, denote by  $\boldsymbol{\pi} \equiv (\pi_1, \pi_2, \dots, \pi_N)^T$  its invariant distribution (where the superscript T denotes the transpose); then  $\pi_i$  is the stationary probability that  $J(t)$  is in state  $i$ . Because of the above assumptions, this distribution exists and is unique. Furthermore, let  $\mathcal{E}^+$  be the states  $i$  in  $\mathcal{E}$  such that  $r_i > 0$  ('up-states'), and  $\mathcal{E}^-$  the states

such that  $r_i < 0$  ('down-states'); we assume for ease that  $r_i \neq 0$  for all  $i \in \mathcal{E}$ . Define the traffic rate matrix  $\mathbf{R} = \text{diag}\{r_1, \dots, r_N\}$ . Let  $N^+$  be the number of up-states, and  $N^-$  the number of down-states. For ease, we let the state-space of the modulating process  $J(t)$  be labeled such that the first  $N^+$  states correspond to the up-states, whereas states  $N^+ + 1$  up to  $N$  correspond to the down-states. In self-evident notation, we write

$$\mathbf{r} \equiv \begin{pmatrix} \mathbf{r}^+ \\ \mathbf{r}^- \end{pmatrix}.$$

This straightforward partition is used frequently in this chapter.

The transient probabilities  $q_i(t, x) := \mathbb{P}(Q(t) \leq x, J(t) = i)$  of the bivariate Markov process  $(Q(t), J(t))$ , defined on  $[0, \infty) \times \mathcal{E}$ , are known to satisfy a system of partial differential equations [70, 71], namely

$$\frac{\partial}{\partial t} q_i(t, x) + r_i \frac{\partial}{\partial x} q_i(t, x) = \sum_{j \in \mathcal{E}} \lambda_{ji} q_j(t, x), \quad \forall i \in \mathcal{E}, x > 0.$$

It is also well-known that under the condition

$$\sum_{i=1}^N r_i \pi_i < 0, \tag{6.1}$$

the workload process is stable, that is, ergodic. Moreover, there exists a stochastic vector  $(Q_e, J_e)$  to which the process  $(Q(t), J(t))$  converges in distribution as  $t \rightarrow \infty$ , and the stationary distribution of  $(Q_e, J_e)$ , say  $\mathbf{q}(x) \equiv (q_1(x), \dots, q_N(x))^T$ , exists and satisfies

$$\mathbf{R} \frac{d}{dx} \mathbf{q}(x) = \mathbf{\Lambda}^T \mathbf{q}(x).$$

As a solution to the above system one could try  $\mathbf{q}(x) = e^{\xi x} \mathbf{v}$ , where  $\xi \in \mathbb{C}$  and  $\mathbf{v}$  is a  $N$ -dimensional vector. Inserting this into the differential equation yields  $(\xi \mathbf{R} - \mathbf{\Lambda}^T) \mathbf{v} = 0$ . A non-trivial solution  $\mathbf{v}$  exists if

$$\det(\xi \mathbf{R} - \mathbf{\Lambda}^T) = 0. \tag{6.2}$$

Sonneveld [109] showed that there are  $N$  eigenvalues  $\xi_j$  (counting multiplicities) satisfying Equation (6.2), of which  $N^+$  have negative real parts, one is zero, and  $N^- - 1$  have positive real parts.

If the eigenvalues  $\xi_j$  are simple then

$$\mathbf{q}(x) = \boldsymbol{\pi} + \sum_{j=1}^{N^+} c_j e^{\xi_j x} \mathbf{v}^j; \tag{6.3}$$

where  $(\xi_j, \mathbf{v}^j)$  satisfy  $(\xi_j \mathbf{R} - \mathbf{\Lambda}^T) \mathbf{v}^j = 0$ , and the constants  $c_j$ ,  $j = 1, \dots, N^+$  are determined by the boundary conditions  $\pi_i + \sum_{j=1}^{N^+} c_j v_i^j = 0$  if  $r_i > 0$ ; if the eigenvalues are non-simple, elementary results from the theory of linear differential equations entail that there are terms in the above spectral expansion of the form  $c_j x^\ell e^{\xi_j x} \mathbf{v}^j$ , with  $\ell = 0, \dots, k-1$ , in case of an eigenvalue of multiplicity of order  $k$ .

Finally, we mention that we let  $p_i(x)$  denote the density corresponding to the distribution function  $q_i(x)$  and  $\mathbf{p}(x) \equiv (p_1(x), \dots, p_N(x))^T$ .

### 6.3 Analysis of the busy period

In this section we determine the Laplace transform of the busy period. Denoting by  $P$  the remaining time till the buffer becomes empty,  $P := \inf\{t > 0 : Q(t) = 0\}$ , the Laplace transform of the busy period is defined by

$$f_i(s) := \mathbb{E}(e^{-sP} \mid Q(0) = 0, J(0) = i). \quad (6.4)$$

Realize that any busy period starts in a state  $i \in \mathcal{E}^+$ .

In our analysis, the following transforms play an important role: for  $i \in \mathcal{E}$ ,  $x \geq 0$ ,  $s \geq 0$ , and  $t > 0$ , define

$$\begin{aligned} \zeta(s \mid x, i) &:= \mathbb{E}(e^{-sP} \mid Q(0) = x, J(0) = i); \\ f_i(s, t) &:= \int_0^\infty e^{-tx} \zeta(s \mid x, i) dx. \end{aligned}$$

Furthermore, let  $t_i \equiv t_i(s) := (\lambda_i + s)/r_i$  for  $i \in \mathcal{E}^+$ .

For now we will focus on the double transforms  $f_i(s, t)$ ; later we explain how these relate to the Laplace transforms of the busy period, i.e., the  $f_i(s)$ . Notice that when analyzing  $f_i(s, t)$ , we have to consider both  $i \in \mathcal{E}^+$  and  $i \in \mathcal{E}^-$ ; we deal with these cases separately.

Let us first consider  $i \in \mathcal{E}^+$ . It is evident that the busy period cannot end before a transition of the modulating Markov process  $J(\cdot)$ . By virtue of the memoryless property of the exponential distribution, one immediately obtains

$$f_i(s, t) = \sum_{k \neq i} \frac{\lambda_{ik}}{\lambda_i} \int_0^\infty e^{-tx} \int_0^\infty \lambda_i e^{-\lambda_i u} e^{-su} \zeta(s \mid x + r_i u, k) du dx.$$

After some algebra (change-of-variables and interchanging order of integrals) this reduces to

$$f_i(s, t) = \sum_{k \neq i} \frac{\lambda_{ik}}{\lambda_i + s - tr_i} (f_k(s, t) - f_k(s, t_i)).$$

Now proceed with  $i \in \mathcal{E}^-$ . Then the busy period may end before the first transition of the modulating process. We obtain

$$\begin{aligned} f_i(s, t) &= \sum_{k \neq i} \frac{\lambda_{ik}}{\lambda_i} \int_0^\infty e^{-tx} \int_0^{-x/r_i} \lambda_i e^{-\lambda_i u} e^{-su} \zeta(s | x + r_i u, k) du dx \\ &\quad + \sum_{k \neq i} \frac{\lambda_{ik}}{\lambda_i} \int_0^\infty e^{-tx} \int_{-x/r_i}^\infty \lambda_i e^{-\lambda_i u} e^{sx/r_i} du dx. \end{aligned}$$

It is readily verified (again by a change-of-variable and in addition interchanging the order of the integrals) that this reduces to

$$f_i(s, t) = \sum_{k \neq i} \frac{\lambda_{ik}}{\lambda_i + s - tr_i} f_k(s, t) - \frac{r_i}{\lambda_i + s - tr_i}.$$

Summarizing, we have found that  $f_i(s, t)$ ,  $i \in \mathcal{E}$  is a solution of the following system

$$(-\lambda_i - s + tr_i) f_i(s, t) + \sum_{k \neq i} \lambda_{ik} f_k(s, t) = \sum_{k \neq i} \lambda_{ik} f_k(s, t_i), \quad i \in \mathcal{E}^+; \quad (6.5)$$

$$(-\lambda_i - s + tr_i) f_i(s, t) + \sum_{k \neq i} \lambda_{ik} f_k(s, t) = r_i, \quad i \in \mathcal{E}^-. \quad (6.6)$$

Now consider Equations (6.5) and (6.6). It is clear that if the auxiliary transforms  $f_k(s, t_i)$  would be known, for  $(k, i) \in \mathcal{E} \times \mathcal{E}^+$ , then  $\mathbf{f}(s, t)$  follow from Cramer's rule:

$$f_i(s, t) = \frac{\det(\mathbf{\Lambda} + t\mathbf{R} - s\mathbf{I} | \mathbf{g}(s), i)}{\det(\mathbf{\Lambda} + t\mathbf{R} - s\mathbf{I})}, \quad i \in \mathcal{E}; \quad (6.7)$$

where for  $i \in \mathcal{E}$ ,  $(\mathbf{\Lambda} + t\mathbf{R} - s\mathbf{I} | \mathbf{g}(s), i)$  is equal to the matrix  $(\mathbf{\Lambda} + t\mathbf{R} - s\mathbf{I})$  with its  $i^{\text{th}}$  column replaced by a vector  $\mathbf{g}(s)$  defined by

$$g_i(s) := \begin{cases} \sum_{k \neq i} \lambda_{ik} f_k(s, t_i), & i \in \mathcal{E}^+ \\ r_i, & i \in \mathcal{E}^-. \end{cases} \quad (6.8)$$

Therefore, it remains to identify the  $N \cdot N^+$  auxiliary transforms  $f_k(s, t_i)$ , for  $(k, i) \in \mathcal{E} \times \mathcal{E}^+$ , and  $s > 0$  given.

Now first observe that inserting  $t = t_j$  (for  $j \in \mathcal{E}^+$ ) into the system (6.5)–(6.6), yields  $N^+ \cdot (N - 1)$  linear equations for the  $f_k(s, t_i)$ ; realize that in equation (6.5) for  $t = t_i$  we obtain a meaningless relation (that is,  $0 = 0$ ). In other words, we can express all  $f_k(s, t_i)$ , for  $i \in \mathcal{E}^+$  and  $k \in \mathcal{E} \setminus \{i\}$ , in terms of the  $N^+$  unknowns  $f_j(s, t_j)$ , where  $j \in \mathcal{E}^+$ . Put differently, we have identified functions  $\gamma_{kij}(s)$  and  $\sigma_{ki}(s)$  such that

$$f_k(s, t_i) = \sum_{j \in \mathcal{E}^+} \gamma_{kij}(s) f_j(s, t_j) + \sigma_{ki}(s), \quad i \in \mathcal{E}^+, \quad k \in \mathcal{E} \setminus \{j\}. \quad (6.9)$$

Hence it remains to identify the  $f_j(s, t_j)$ , for  $j \in \mathcal{E}^+$ ; if we would know them we could rewrite the vector  $\mathbf{g}^+(s)$  as

$$g_i(s) = \sum_{j \in \mathcal{E}^+} \sum_{k \neq i} \lambda_{ik} \gamma_{kij}(s) f_j(s, t_j) + \sum_{k \neq i} \lambda_{ik} \sigma_{ki}(s). \quad (6.10)$$

With  $\mathbf{f}^+(s, t_+) \equiv (f_1(s, t_1), \dots, f_{N^+}(s, t_{N^+}))^T$ , the above means that we have, in self-evident notation, constructed a matrix  $\mathbf{M}(s)$  (square; of dimension  $N^+$ ) and a vector  $\boldsymbol{\omega}^+(s)$  such that

$$\mathbf{g}^+(s) = \mathbf{M}(s) \mathbf{f}^+(s, t_+) + \boldsymbol{\omega}^+(s); \quad (6.11)$$

where the matrix  $\mathbf{M}$  and the vector  $\boldsymbol{\omega}^+$  are given by

$$m_{ij}(s) := \sum_{k \neq i} \lambda_{ik} \gamma_{kij}(s), \quad i, j \in \mathcal{E}^+; \quad (6.12)$$

$$\omega_i(s) := \sum_{k \neq i} \lambda_{ik} \sigma_{ki}(s), \quad i \in \mathcal{E}^+. \quad (6.13)$$

As mentioned above, it remains to determine  $f_j(s, t_j)$ ,  $j \in \mathcal{E}^+$  for a given  $s > 0$ . This can be done by using the following powerful lemma.

**Lemma 6.3.1.** *For fixed  $s > 0$ , consider the equation  $\det(\boldsymbol{\Lambda} - s\mathbf{I} + t\mathbf{R}) = 0$  for  $t \in \mathbb{C}$ . There are  $N^+$  values of  $t$  with  $\operatorname{Re} t > 0$  satisfying this equation.*

For a special choice of  $\boldsymbol{\Lambda}$  and  $\mathbf{R}$  the above lemma was proven in [83], but it is readily checked that the result carries over to general  $\boldsymbol{\Lambda}$  and  $\mathbf{R}$  (as long as  $\boldsymbol{\Lambda}$  corresponds to an irreducible Markov chain, and as long as the stability condition (6.1) is fulfilled). The proof is identical to the one given in [83], and is based on [109], and intensively uses Geršgorin's circle theorem [58].

Fixing  $s > 0$ , realize that the transform (6.7) should have a finite norm for any  $t$  in the right half-plane. Hence, for any  $t$  in the right half plane for which the denominator in (6.7) equals 0 (that is,  $\det(\boldsymbol{\Lambda} + t\mathbf{R} - s\mathbf{I}) = 0$ ), also the numerator should equal 0. From Lemma 6.3.1 we know that there are, for any  $s > 0$ , exactly  $N^+$  such zeros in the right half-plane. Inserting these zeros into the numerator of (6.7) and equating it to 0, we obtain exactly  $N^+$  linear equations that determine  $\mathbf{f}^+(s, t_+)$ . Using (6.11) we can now determine  $\mathbf{g}(s)$ , and using (6.7) also  $\mathbf{f}(s, t)$ .

We conclude this section by arguing that, knowing the  $g_i(s)$ , we have also identified the Laplace transform of the busy period  $f_i(s)$ . This is seen as follows. Consid-

ering  $f_i(s)$ , with  $i \in \mathcal{E}^+$ , straightforward arguments yield

$$\begin{aligned} f_i(s) &= \sum_{k \neq i} \lambda_{ik} \int_0^\infty e^{-\lambda_i u} e^{-s u} \mathbb{E}(e^{-sP} \mid Q(0) = r_i u, J(0) = k) du \\ &= \sum_{k \neq i} \frac{\lambda_{ik}}{r_i} \int_0^\infty \exp\left(-\frac{\lambda_i + s}{r_i} v\right) \mathbb{E}(e^{-sP} \mid Q(0) = v, J(0) = k) dv \\ &= \sum_{k \neq i} \frac{\lambda_{ik}}{r_i} f_k(s, t_i) = \frac{g_i(s)}{r_i}. \end{aligned}$$

The above findings are summarized in the following theorem.

**Theorem 6.3.2.** For  $s, t > 0$ ,

$$\mathbf{f}(s, t) = (\mathbf{\Lambda} - s\mathbf{I} + t\mathbf{R})^{-1} \mathbf{g}(s),$$

with  $\mathbf{g}^-(s) \equiv \mathbf{r}^-$  and  $\mathbf{g}^+(s)$  given by (6.11) and obtained by solving

$$\det(\mathbf{\Lambda} - s\mathbf{I} + \tau_i(s)\mathbf{R} \mid \mathbf{g}(s), k) = 0,$$

for  $i = 1, \dots, N^+$ ; here  $\tau_i(s)$ ,  $i = 1, \dots, N^+$ , are, for  $s > 0$  given, the  $N^+$  values of  $\tau$  in the right half-plane that satisfy  $\det(\mathbf{\Lambda} - s\mathbf{I} + \tau\mathbf{R}) = 0$ .

Furthermore, the Laplace transform of the busy period, starting in state  $i \in \mathcal{E}^+$ , is given by

$$f_i(s) = \frac{g_i(s)}{r_i}, \quad i = 1, \dots, N^+. \quad (6.14)$$

## 6.4 Busy period asymptotics

In this section we derive the logarithmic asymptotics of the probability that, starting in an up-state  $i$ , the busy period lasts longer than  $t$ , for  $t \rightarrow \infty$ . To this end, we define the *cumulant function*

$$\Gamma(\vartheta) := \lim_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{E} \exp(\vartheta A(0, t)),$$

here  $A(0, t) := \int_0^t r_{X(s)} ds$ ; notice that the cumulant function can be regarded as an asymptotic log-moment generating function, and is, as a consequence, convex.  $\Gamma'(0)$  equals the drift  $\sum_{i=1}^N r_i \pi_i$ , which we assumed to be negative. Define by  $\vartheta^*$  the minimizer of  $\Gamma(\vartheta)$ ; observe that necessarily  $\vartheta^* > 0$  and  $\Gamma(\vartheta^*) < 0$ .

We introduce the short notations

$$\mathbb{P}_i(\cdot) := \mathbb{P}(\cdot \mid J(0) = i), \quad \text{and} \quad \varrho_i(t) := \mathbb{P}_i(P > t \mid Q(0) = 0).$$

We can now state the main result of this section.

**Theorem 6.4.1.** For  $i \in \mathcal{E}^+$

$$\lim_{t \rightarrow \infty} \frac{1}{t} \log \varrho_i(t) = \Gamma(\vartheta^*);$$

here  $\vartheta^*$  is the minimizing point of  $\Gamma(\vartheta)$ .

*Proof.* Let  $i \in \mathcal{E}^+$ . First we prove the upper bound. Evidently, we have

$$\varrho_i(t) \leq \mathbb{P}(A(0, t) > 0 \mid J(0) = i).$$

Now the Gärtner-Ellis theorem [44] immediately yields, with  $A[k] := A(0, k+1) - A(0, k)$ , that

$$\begin{aligned} \limsup_{t \rightarrow \infty} \frac{1}{t} \log \varrho_i(t) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}_i \left( \sum_{k=0}^{n-1} A[k] > 0 \right) \\ &= -\sup_{\vartheta \geq 0} (\vartheta \cdot 0 - \Gamma(\vartheta)) = \inf_{\vartheta \geq 0} \Gamma(\vartheta). \end{aligned}$$

The infimum of  $\Gamma(\vartheta)$  is attained at  $\vartheta^*$  as introduced above, which proves the upper bound.

We now proceed by proving the lower bound. For any  $\delta \in (0, 1)$ ,

$$\varrho_i(t) \geq \mathbb{P}_i(\forall u \in [0, \delta t] : J(u) = i; \forall s \in (\delta t, t] : A(0, s) > 0).$$

But using the conditional independence, the expression in the right-hand side of the previous display equals

$$\mathbb{P}_i(\forall u \in [0, \delta t] : J(u) = i) \mathbb{P}_i(A(0, s) > -r_i \delta t, \forall s \in (0, (1-\delta)t).$$

The first factor of the above display is equal to  $e^{-\lambda_i \delta t}$ . Now concentrate on the second factor, with  $r_{\max} := \max_{i \in \mathcal{E}} r_i$ :

$$\begin{aligned} &\frac{1}{t} \log \mathbb{P}_i(A(0, s) > -r_i \delta t, \forall s \in (0, (1-\delta)t]) \\ &= \frac{1}{t} \log \mathbb{P}_i(A(0, \gamma(1-\delta)t) > -r_i \delta t, \forall \gamma \in [0, 1]) \\ &= \frac{1}{t} \log \mathbb{P}_i \left( \frac{A(0, \gamma(1-\delta)t)}{(1-\delta)t} > -r_i \frac{\delta}{1-\delta}, \forall \gamma \in [0, 1] \right) \\ &\geq \frac{1}{t} \log \mathbb{P}_i \left( \frac{A(0, \lceil \gamma(1-\delta)t \rceil)}{\lceil (1-\delta)t \rceil} > -r_i \frac{\delta}{1-\delta} + r_{\max} \frac{1}{(1-\delta)t}, \forall \gamma \in [0, 1] \right) \\ &\geq \frac{1}{t} \log \mathbb{P}_i \left( \frac{A(0, \lceil \gamma(1-\delta)t \rceil)}{\lceil (1-\delta)t \rceil} > -\frac{r_i}{2} \frac{\delta}{(1-\delta)}, \forall \gamma \in [0, 1] \right); \end{aligned}$$

in the first inequality we use the fact that

$$\begin{aligned} A(0, \gamma(1-\delta)t) &= A(0, \lceil \gamma(1-\delta)t \rceil) - A(\gamma(1-\delta)t, \lceil \gamma(1-\delta)t \rceil) \\ &\geq A(0, \lceil \gamma(1-\delta)t \rceil) - r_{\max}, \end{aligned}$$

and the last inequality holds for all  $t > t^* := 2r_{\max}/(r_i\delta)$ .

The process  $A_n(0, \gamma) := n^{-1} \cdot A(0, \lceil n\gamma \rceil)$ , with  $\gamma \in [0, 1]$  and  $n := \lceil (1 - \delta)t \rceil$  fits in the framework of the sample-path large deviations principle in Example 2.5 of Chang [31]. As a consequence,

$$\begin{aligned} & (1 - \delta) \liminf_{t \rightarrow \infty} \frac{1}{(1 - \delta)t} \log \mathbb{P}_i \left( \frac{A(0, \lceil \gamma(1 - \delta)t \rceil)}{\lceil (1 - \delta)t \rceil} > -\frac{r_i}{2} \frac{\delta}{(1 - \delta)}, \forall \gamma \in [0, 1] \right) \\ & \geq -(1 - \delta) \inf_{f \in \mathcal{A}^\circ} \mathbb{I}(f), \end{aligned}$$

where  $\mathbb{I}(\cdot)$  is the ‘rate functional’:

$$\mathbb{I}(f) := \int_0^1 \sup_{\vartheta} (\vartheta f'(t) - \Gamma(\vartheta)) dt,$$

and  $\mathcal{A}^\circ$  is the interior of  $\mathcal{A}$ , which is the set of paths of interest:

$$\mathcal{A} := \left\{ f \in \text{AC}([0, 1], (\mathbb{R}, \|\cdot\|_\infty)) : f(\gamma) > -\frac{r_i}{2} \frac{\delta}{(1 - \delta)}, \forall \gamma \in [0, 1] \right\}.$$

Here  $\text{AC}([0, 1], (\mathbb{R}, \|\cdot\|_\infty))$  is the space of absolutely continuous functions  $f$  such that  $f(0) = 0$ , equipped with the supremum norm topology, i.e.,

$$\|f\|_\infty = \sup_{t \in [0, 1]} |f(t)|.$$

It is seen that the set  $\mathcal{A}$  is open (and consequently  $\mathcal{A} = \mathcal{A}^\circ$ ), as follows. Since the functions considered are absolutely continuous, thus continuous, over the *closed* interval  $[0, 1]$ , any function  $f$  attains a minimum at some point  $\gamma_f \in [0, 1]$ ; as  $f \in \mathcal{A}$ , we have that

$$f(\gamma_f) > -\frac{r_i}{2} \frac{\delta}{(1 - \delta)}.$$

Then consider the ball  $B(f, \epsilon)$  around  $f$  with radius

$$\epsilon := \frac{1}{2} \left( f(\gamma_f) + \frac{r_i}{2} \frac{\delta}{(1 - \delta)} \right) > 0;$$

this ball is evidently contained in  $\mathcal{A}$ , and hence  $\mathcal{A}$  is open.

Then observe that the path  $f_0 \equiv 0$  is in  $\mathcal{A}^\circ = \mathcal{A}$ . Hence

$$-\inf_{f \in \mathcal{A}} \mathbb{I}(f) \geq -\mathbb{I}(f_0) = -\sup_{\vartheta} (0 - \Gamma(\vartheta)) = \Gamma(\vartheta^*).$$

Summarizing, we have

$$\begin{aligned} \liminf_{t \rightarrow \infty} \frac{1}{t} \log \varrho_i(t) & \geq \liminf_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}_i(J(\delta t) = i) \\ & \quad + \liminf_{t \rightarrow \infty} \frac{1}{t} \log \mathbb{P}_i(A(0, s) > -r_i\delta t, \forall s \in (0, (1 - \delta)t]) \\ & \geq -\lambda_i\delta + (1 - \delta)\Gamma(\vartheta^*). \end{aligned}$$

Now letting  $\delta \downarrow 0$  yields the lower bound.  $\square$

## 6.5 The covariance function of the workload process

In this section we analyze the Laplace transform of the covariance function of the workload process  $Q(u)$ . Let  $R(u) := \text{Cov}(Q(0), Q(u))$  and  $\gamma(\vartheta)$  be its Laplace transform, i.e.,

$$\gamma(\vartheta) := \int_0^\infty e^{-\vartheta u} R(u) du = \int_0^\infty e^{-\vartheta u} [\mathbb{E}Q(0)Q(u) - \mathbb{E}Q(0)\mathbb{E}Q(u)] du.$$

As an important special case, we later consider the situation that  $Q(0)$  is distributed according to the stationary distribution (6.3). Then the correlation coefficient  $r(u)$  between  $Q(0)$  and  $Q(u)$  reads

$$r(u) := \frac{\text{Cov}(Q(0), Q(u))}{\sqrt{\text{Var}Q(0) \cdot \text{Var}Q(u)}} = \frac{\mathbb{E}Q(0)Q(u) - (\mathbb{E}Q_e)^2}{\text{Var}Q_e},$$

using that the queue is still in equilibrium at time  $u$ . We denote by  $\rho(\vartheta)$  the Laplace transform of the correlation  $r(u)$ .

In our derivation of the Laplace transform  $\gamma(\vartheta)$ , we condition on the state of the system at time zero. More specifically, we assume throughout that the probability distribution of  $(Q(0), J(0))$  is given, and denote its density by

$$\mathbf{p}^0(x) \equiv (p_1^0(x), \dots, p_N^0(x))^T, \text{ for } x \geq 0;$$

as indicated above we will later specialize to the special case of  $\mathbf{p}^0(x) = \mathbf{p}(x)$ , with  $\mathbf{p}(x)$  distributed according to (6.3). In the sequel let  $\tau$  be an exponentially distributed random variable with parameter  $\vartheta$ , independently of the modulating process  $J(t)$ . For  $s \geq 0$ ,  $t > 0$ , we also introduce

$$\begin{aligned} \eta_i(\vartheta, s | x) &:= \mathbb{E}\left(e^{-sQ(\tau)} \mid Q(0) = x, J(0) = i\right); \\ \ell_i(\vartheta, s, t) &:= \int_0^\infty e^{-tx} \eta_i(\vartheta, s | x) dx. \end{aligned} \quad (6.15)$$

For later use, define  $\vartheta_i := \vartheta + \lambda_i$ , and in addition, the ‘derivatives’ of  $\eta(\vartheta, s | x)$  and  $\ell(\vartheta, s, t)$ , for  $i \in \mathcal{E}$ :

$$\begin{aligned} \eta_i^{(s)}(\vartheta, s | x) &:= \frac{\partial}{\partial s} \eta_i(\vartheta, s | x), \\ \ell_i^{(s)}(\vartheta, s, t) &:= \frac{\partial}{\partial s} \ell_i(\vartheta, s, t), \quad \ell_i^{(s,t)}(\vartheta, s, t) := \frac{\partial^2}{\partial s \partial t} \ell_i(\vartheta, s, t). \end{aligned}$$

These functions will turn out to play a pivotal role in determining the Laplace transform of the covariance function  $R(u)$ . The following lemma relates  $\eta(\vartheta, s | 0)$  and  $\ell(\vartheta, s, t)$ .

**Lemma 6.5.1.** *The vectors  $\eta(\vartheta, s | 0)$  and  $\ell(\vartheta, s, t)$  satisfy the following relation:*

$$(tr_i - \vartheta)\ell_i(\vartheta, s, t) + \sum_{k \in \mathcal{E}} \lambda_{ik} \ell_k(\vartheta, s, t) = r_i \eta_i(\vartheta, s | 0) - \frac{\vartheta}{s+t}, \quad (6.16)$$

for  $i \in \mathcal{E}$ . In addition,

$$\eta_i(\vartheta, s | 0) = \sum_{k \neq i} \frac{\lambda_{ik}}{r_i} \cdot \ell_k \left( \vartheta, s, \frac{\vartheta_i}{r_i} \right) + \frac{\vartheta}{sr_i + \vartheta_i}, \quad i \in \mathcal{E}^+; \quad (6.17)$$

$$\eta_i(\vartheta, s | 0) = \sum_{k \neq i} \frac{\lambda_{ik}}{\vartheta_i} \cdot \eta_k(\vartheta, s | 0) + \frac{\vartheta}{\vartheta_i}, \quad i \in \mathcal{E}^-. \quad (6.18)$$

*Proof.* First we stress that there are strong similarities between this proof and the steps used in Section 6.3, when we determined the Laplace transform of the busy period.

Notice that for  $i \in \mathcal{E}^+$  the buffer cannot become empty before the first jump of the modulating Markov process, whereas for  $i \in \mathcal{E}^-$  this is possible; we deal with the two cases differently. First consider the case  $i \in \mathcal{E}^+$ . Then, conditioning on the jump epoch of the modulating Markov process,

$$\begin{aligned} \ell_i(\vartheta, s, t) &= \int_0^\infty e^{-tx} \int_0^\infty \vartheta_i e^{-\vartheta_i u} \left\{ \sum_{k \neq i} \frac{\lambda_{ik}}{\vartheta_i} \eta_k(\vartheta, s | x + r_i u) \right. \\ &\quad \left. + \frac{\vartheta}{\vartheta_i} e^{-s(x+r_i u)} \right\} dudx \\ &= \sum_{k \neq i} \frac{\lambda_{ik}}{tr_i - \vartheta_i} \left( \ell_k \left( \vartheta, s, \frac{\vartheta_i}{r_i} \right) - \ell_k(\vartheta, s, t) \right) + \frac{\vartheta}{(s+t)(sr_i + \vartheta_i)}; \end{aligned}$$

the last equality followed after elementary calculus. The above relation can then be rewritten to

$$(tr_i - \vartheta_i)\ell_i(\vartheta, s, t) + \sum_{k \neq i} \lambda_{ik} \ell_k(\vartheta, s, t) = \sum_{k \neq i} \lambda_{ik} \ell_k \left( \vartheta, s, \frac{\vartheta_i}{r_i} \right) + \frac{\vartheta(tr_i - \vartheta_i)}{(s+t)(sr_i + \vartheta_i)}. \quad (6.19)$$

Now consider the case  $i \in \mathcal{E}^-$ . Taking into account that the buffer can become empty before the first jump of the modulating Markov process,

$$\begin{aligned} \ell_i(\vartheta, s, t) &= \int_0^\infty e^{-tx} \int_0^\infty \vartheta_i e^{-\vartheta_i u} \left\{ \sum_{k \neq i} \frac{\lambda_{ik}}{\vartheta_i} \eta_k(\vartheta, s | (x + r_i u)^+) \right. \\ &\quad \left. + \frac{\vartheta}{\vartheta_i} e^{-s(x+r_i u)^+} \right\} dudx \\ &= - \sum_{k \neq i} \frac{\lambda_{ik}}{tr_i - \vartheta_i} \left\{ \ell_k(\vartheta, s, t) - \frac{r_i}{\vartheta_i} \eta_k(\vartheta, s | 0) \right\} + \frac{\vartheta((s+t)r_i - \vartheta_i)}{(s+t)\vartheta_i(tr_i - \vartheta_i)}, \end{aligned}$$

which can be written as

$$(tr_i - \vartheta_i)\ell_i(\vartheta, s, t) + \sum_{k \neq i} \lambda_{ik} \ell_k(\vartheta, s, t) = \sum_{k \neq i} \frac{\lambda_{ik} r_i}{\vartheta_i} \eta_k(\vartheta, s | 0) + \frac{\vartheta((s+t)r_i - \vartheta_i)}{\vartheta_i(s+t)}. \quad (6.20)$$

Now let us evaluate  $\eta_i(\vartheta, s | 0)$  further. First consider the case  $i \in \mathcal{E}^+$ ; then the buffer immediately becomes non-empty, and hence

$$\begin{aligned} \eta_i(\vartheta, s | 0) &= \int_0^\infty \vartheta_i e^{-\vartheta_i u} \left\{ \sum_{k \neq i} \frac{\lambda_{ik}}{\vartheta_i} \eta(\vartheta, s | r_i u, k) + \frac{\vartheta}{\vartheta_i} e^{-sr_i u} \right\} du \\ &= \sum_{k \neq i} \frac{\lambda_{ik}}{r_i} \cdot \ell_k \left( \vartheta, s, \frac{\vartheta_i}{r_i} \right) + \frac{\vartheta}{sr_i + \vartheta_i}, \end{aligned}$$

which proves (6.17). For  $i \in \mathcal{E}^-$  the buffer remains empty until the first jump, and hence

$$\begin{aligned} \eta_i(\vartheta, s | 0) &= \int_0^\infty \vartheta_i e^{-\vartheta_i u} \left\{ \sum_{k \neq i} \frac{\lambda_{ik}}{\vartheta_i} \cdot \eta_k(\vartheta, s | 0) + \frac{\vartheta}{\vartheta_i} \right\} du \\ &= \sum_{k \neq i} \frac{\lambda_{ik}}{\vartheta_i} \eta_k(\vartheta, s | 0) + \frac{\vartheta}{\vartheta_i}, \end{aligned}$$

which proves (6.18). Equation (6.16) is obtained by inserting (6.17)–(6.18) into (6.19)–(6.20).  $\square$

We now explain how the transform  $\ell(\vartheta, s, t)$  can be identified. Equation (6.16) can be rewritten in matrix form:

$$(\mathbf{\Lambda} - \vartheta \mathbf{I} + t \mathbf{R}) \ell(\vartheta, s, t) = \mathbf{w}(\vartheta, s, t); \quad (6.21)$$

here  $w_i(\vartheta, s, t) := r_i \eta_i(\vartheta, s | 0) - \vartheta/(s+t)$ , for  $i \in \mathcal{E}$ . In other words, assuming for the moment that  $\eta(\vartheta, s | 0)$  is known, application of Cramer's rule leads to

$$\ell_i(\vartheta, s, t) = \frac{\det(\mathbf{\Lambda} - \vartheta \mathbf{I} + t \mathbf{R} | \mathbf{w}(\vartheta, s, t), i)}{\det(\mathbf{\Lambda} - \vartheta \mathbf{I} + t \mathbf{R})}, \quad i \in \mathcal{E}. \quad (6.22)$$

It is observed that, if we are able to determine  $\eta(\vartheta, s | 0)$ , then we also have found  $\ell(\vartheta, s, t)$ . We now identify  $N$  linear equations that enable us to compute  $\eta(\vartheta, s | 0)$ . Equation (6.18) already gives  $N^-$  equations, so that it remains to determine the other  $N^+$  linear equations. Since  $\vartheta$  is fixed, using Lemma 6.3.1 with  $\vartheta$  instead of  $s$ , there are  $N^+$  values  $\tau_i \equiv \tau_i(\vartheta)$  ( $i = 1, \dots, N^+$ ) in the right half-plane satisfying

$$\det(\mathbf{\Lambda} - \vartheta \mathbf{I} + \tau_i \mathbf{R}) = 0.$$

Since Equation (6.22) should give a finite norm for any  $\vartheta > 0$ , these  $N^+$  values  $\tau_i \equiv \tau_i(\vartheta)$  should also satisfy

$$\det(\mathbf{A} - \vartheta \mathbf{I} + \tau_i \mathbf{R} | \mathbf{w}(\vartheta, s, \tau_i), k) = 0, \quad i \in \mathcal{E}^+.$$

In other words: we have now obtained  $N$  linear equations; solving these yields  $\boldsymbol{\eta}(\vartheta, s | 0)$ .

Above we developed a procedure for determining  $\ell(\vartheta, s, t)$ . Clearly  $\ell(\vartheta, s, t)$  uniquely defines  $\boldsymbol{\eta}(\vartheta, s | x)$ , see (6.15). We now focus on how this procedure can be used to obtain an expression for the Laplace transform  $\gamma(\cdot)$  of the covariance function.

**Theorem 6.5.2.** For  $\vartheta > 0$ ,

$$\gamma(\vartheta) = \frac{1}{\vartheta} \cdot \sum_{i \in \mathcal{E}} \int_0^\infty [\mathbb{E}Q(0) - x] \eta_i^{(s)}(\vartheta, 0 | x) p_i^0(x) dx, \quad (6.23)$$

where

$$\mathbb{E}Q(0) = \sum_{i \in \mathcal{E}} \int_0^\infty x p_i^0(x) dx.$$

*Proof.* With  $\mathbb{E}_{x,i}(\cdot) := \mathbb{E}(\cdot | Q(0) = x, J(0) = i)$ , conditioning on the state of the system at time 0 yields

$$\begin{aligned} \vartheta \gamma(\vartheta) &= \int_0^\infty \vartheta e^{-\vartheta u} R(u) du \\ &= \int_0^\infty \vartheta e^{-\vartheta u} \left( \sum_{i \in \mathcal{E}} \int_0^\infty [x \mathbb{E}_{x,i}Q(u) - \mathbb{E}Q(0) \mathbb{E}_{x,i}Q(u)] p_i^0(x) dx \right) du \\ &= \sum_{i \in \mathcal{E}} \int_0^\infty [x - \mathbb{E}Q(0)] \left( \int_0^\infty \vartheta e^{-\vartheta u} \mathbb{E}_{x,i}Q(u) du \right) p_i^0(x) dx \\ &= \sum_{i \in \mathcal{E}} \int_0^\infty [\mathbb{E}Q(0) - x] \eta_i^{(s)}(\vartheta, 0 | x) p_i^0(x) dx, \end{aligned}$$

where in the last equality we used the fact that

$$\eta_i^{(s)}(\vartheta, 0 | x) = - \int_0^\infty \vartheta e^{-\vartheta u} \mathbb{E}_{x,i}Q(u) du.$$

□

Now we consider a number of special cases. If the density  $p^0(x)$  is given by

$$p_i^0(x) = \sum_{j=1}^k \sigma_{ij} e^{-\zeta_j x}, \quad x \geq 0, \quad i \in \mathcal{E}; \quad (6.24)$$

for constants  $\sigma_{ij}$  and  $\zeta_j > 0$ , then the Laplace transform  $\gamma(\vartheta)$  is given by

$$\gamma(\vartheta) = \frac{1}{\vartheta} \cdot \sum_{i \in \mathcal{E}} \sum_{j=1}^k \sigma_{ij} \left[ \ell_i^{(s,t)}(\vartheta, 0, \zeta_j) + \mathbb{E}Q(0) \ell_i^{(s)}(\vartheta, 0, \zeta_j) \right]. \quad (6.25)$$

Formula (6.25) extends in a straightforward way to the case in which among the  $\zeta_j$  there are pairs of complex conjugates (with necessarily positive real parts). Importantly, this observation entails that we have now identified the Laplace transform of the covariance function in case  $Q(0)$  obeys the stationary workload distribution (6.3). Also the case of eigenvalues with multiplicity  $k$  larger than 1 can be solved; then the density of the stationary workload has terms proportional to  $x^j e^{-\zeta_j x}$ , with  $j = 0, \dots, k-1$ , which is reflected in the appearance of higher order derivatives of  $\ell_i(\vartheta, s, t)$  (where  $s$  and  $t$  should be replaced by 0 and  $\zeta_j$ , respectively) in the expression for  $\gamma(\vartheta)$ .

## 6.6 Example

The following example illustrates the results of this chapter. We concentrate on the two-state case, and compute the busy period distribution, as well as the covariance function. For  $\alpha, \beta, \lambda$  and  $\mu$  positive, we denote

$$\mathbf{\Lambda} := \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}, \quad \mathbf{r} := \begin{pmatrix} \alpha \\ -\beta \end{pmatrix}, \quad \boldsymbol{\pi} := \psi \begin{pmatrix} \mu \\ \lambda \end{pmatrix}, \quad \text{with } \psi := \frac{1}{\lambda + \mu};$$

it is easily verified that  $\boldsymbol{\pi}$  is the invariant distribution of  $\mathbf{\Lambda}$ . We call the state space  $\mathcal{E} = \{+, -\}$ . The stability condition is satisfied if  $\beta\lambda > \alpha\mu$ , which in the sequel is assumed to hold true. The pairs of eigenvalues-vectors of  $\mathbf{R}^{-1} \mathbf{\Lambda}^T$  are given by

$$\xi_0 = 0, \quad \text{with } v^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \quad \xi_+ = \frac{\alpha\mu - \beta\lambda}{\alpha\beta}, \quad \text{with } v^{(+)} = \begin{pmatrix} \lambda/\alpha \\ \mu/\beta \end{pmatrix};$$

note that  $\xi_+ < 0$ . The stationary distribution of  $(Q_e \leq x, J_e = \pm)$  and its density are

$$\begin{aligned} \mathbb{P}(Q_e \leq x, J_e = +) &= \mu\psi (1 - \exp(\xi_+ x)), & p_+(x) &= -\mu\xi_+ \psi \exp(\xi_+ x); \\ \mathbb{P}(Q_e \leq x, J_e = -) &= \lambda\psi \left( 1 - \frac{\alpha\mu}{\beta\lambda} \exp(\xi_+ x) \right), & p_-(x) &= -\frac{\alpha}{\beta} \mu\xi_+ \psi \exp(\xi_+ x). \end{aligned} \quad (6.26)$$

The mean and variance of  $Q$  are finite and given by

$$\mathbb{E}Q_e = \frac{\alpha\mu(\alpha + \beta)}{(\lambda + \mu)(\beta\lambda - \alpha\mu)}, \quad \mathbb{V}\text{ar } Q_e = \alpha^2 \left( \frac{\beta^2}{(\beta\lambda - \alpha\mu)^2} - \frac{1}{(\lambda + \mu)^2} \right).$$

*Busy period.* We first determine the distribution of the busy period  $P$ , as well as its tail asymptotics. The system (6.5)–(6.6) can be rewritten as, with  $f_{ij}(s) := f_i(s, t_j(s))$ , for  $i, j \in \{+, -\}$ ,

$$\begin{cases} -(\lambda + s - \alpha t)f_+(s, t) + \lambda f_-(s, t) &= \lambda f_{-+}(s); \\ \mu f_+(s, t) - (\mu + s + \beta t)f_-(s, t) &= -\beta. \end{cases}$$

From the second equation we have, by inserting  $t = (\lambda + s)/\alpha$ ,

$$f_{-+}(s) = \frac{\alpha}{\alpha(\mu + s) + \beta(\lambda + s)} (\mu f_{++}(s) + \beta),$$

and hence the vector  $\mathbf{g}(s)$  is given by

$$\mathbf{g}(s) = \left( \frac{\lambda\alpha(\mu f_{++}(s) + \beta)}{\alpha(\mu + s) + \beta(\lambda + s)}, -\beta \right)^T.$$

To determine  $f_{++}(s)$ , we first compute the zeros of the determinant of  $(\mathbf{\Lambda} + t\mathbf{R} - s\mathbf{I})$  for given  $s > 0$ :

$$\tau_{\pm}(s) = \frac{1}{2\alpha\beta} \cdot \left( \beta(\lambda + s) - \alpha(\mu + s) \pm \sqrt{[\beta(\lambda + s) + \alpha(\mu + s)]^2 - 4\alpha\beta\lambda\mu} \right).$$

Notice that  $\tau_-(s)$  is negative and  $\tau_+(s)$  is positive. We focus on the positive root  $\tau_+(s)$ . It is clear that  $\tau_+(s)$  must also be a zero of the determinant of  $(\mathbf{\Lambda} + t\mathbf{R} - s\mathbf{I} | \mathbf{g}(s), +)$ . It can now be verified that

$$f_{++}(s) = \frac{\beta^2}{\mu\alpha} \cdot \frac{\lambda + s - \alpha\tau_+(s)}{\mu + s + \beta\tau_+(s)}, \quad f_{-+}(s) = \frac{\beta}{\mu + s + \beta\tau_+(s)}.$$

The Laplace transform of  $P$ , starting off at buffer level 0 (so that the busy period necessarily starts in +), is then given by

$$f_+(s) = \frac{1}{2\alpha\mu} \left( (\alpha + \beta)s + (\beta\lambda + \alpha\mu) - \sqrt{[(\alpha + \beta)s + (\beta\lambda + \alpha\mu)]^2 - 4\alpha\beta\lambda\mu} \right).$$

This transform can be explicitly inverted, yielding the density of the busy period; with  $F_i(t) := \mathbb{P}_i(P \leq t | Q(0) = 0)$  denoting the distribution function of the busy period, we have that the density of the busy period equals

$$\frac{d}{dt} F_+(t) = \sqrt{\frac{\beta\lambda}{\alpha\mu}} \cdot \mathbf{I}_1 \left( 2 \frac{\sqrt{\alpha\beta\lambda\mu}}{\alpha + \beta} t \right) \cdot \frac{1}{t} \exp \left( - \frac{\alpha\mu + \beta\lambda}{\alpha + \beta} t \right);$$

here  $\mathbf{I}_1(x)$  is the modified Bessel function of the first kind. By differentiating  $f_+(s)$  and inserting  $s = 0$  we can now find all moments of the busy-period  $P$ . The first moment is

$$\mathbb{E}_+(P | Q(0) = 0) = -f'_+(0) = \frac{\alpha + \beta}{\beta\lambda - \alpha\mu}.$$

The asymptotics of the density and the tail distribution of  $F_+(t)$  (which was already introduced as  $\varrho_+(t)$  in Section 6.4) are given by, as  $t \rightarrow \infty$ ,

$$\begin{aligned} \frac{d}{dt}F_+(t) &\sim \frac{(\beta\lambda)^{1/4}}{(\alpha\mu)^{3/4}} \frac{\sqrt{\alpha+\beta}}{2\sqrt{\pi}} \cdot \frac{1}{t\sqrt{t}} \exp\left(-\frac{(\sqrt{\beta\lambda}-\sqrt{\alpha\mu})^2}{\alpha+\beta}t\right), \\ \varrho_+(t) &\sim \frac{(\beta\lambda)^{1/4}}{(\alpha\mu)^{3/4}} \frac{(\alpha+\beta)^{3/2}}{2\sqrt{\pi}(\sqrt{\beta\lambda}-\sqrt{\alpha\mu})^2} \cdot \frac{1}{t\sqrt{t}} \exp\left(-\frac{(\sqrt{\beta\lambda}-\sqrt{\alpha\mu})^2}{\alpha+\beta}t\right); \end{aligned} \quad (6.27)$$

here ' $\sim$ ' means that the ratio of both sides tends to 1 as  $t \rightarrow \infty$ .

Now consider a more specific example. Taking  $\alpha = \lambda = \mu = 1$  and  $\beta = 2$ , we are in the setting of [9, Section 9]. Then

$$f_+(s) = \frac{3(1+s) - \sqrt{9(1+s)^2 - 8}}{2}; \quad \mathbb{E}_+(P) = 3,$$

in agreement with the findings of [9]. We find, however, a number of new results:

$$\begin{aligned} \frac{d}{dt}F_+(t) &= \sqrt{2} \cdot \frac{1}{t} e^{-t} I_1\left(\frac{2\sqrt{2}}{3}t\right), \quad t > 0; \\ \varrho_+(t) &\sim \frac{\sqrt{3}(3+2\sqrt{2})}{\sqrt{2}\sqrt{2} \cdot \pi} \cdot \frac{1}{t\sqrt{t}} e^{-\frac{3-2\sqrt{2}}{3}t}, \quad t \rightarrow \infty. \end{aligned}$$

Now consider the logarithmic asymptotics of  $\mathbb{P}_+(P > t)$ . The cumulant function is given by  $\Gamma(\vartheta) = \log \mathbf{sp}(\mathbf{A} + \vartheta \mathbf{B})$ , where  $\mathbf{sp}(\mathbf{M})$  is the largest eigenvalue of the matrix  $\mathbf{M}$ , see [66]. In our example,

$$\Gamma(\vartheta) = -\frac{(\lambda + \mu + (\beta - \alpha)\vartheta)}{2} + \frac{\sqrt{((\beta + \alpha)\vartheta + (\mu - \lambda))^2 + 4\mu\lambda}}{2}.$$

Furthermore we have under our stability condition that  $\Gamma'(0) < 0$ , and  $\Gamma(\cdot)$  attains its minimum at

$$\vartheta^* = \frac{\lambda - \mu}{\beta + \alpha} + \sqrt{\frac{\lambda\mu}{\alpha\beta} \frac{\beta - \alpha}{\alpha + \beta}} = \frac{(\sqrt{\lambda\alpha} + \sqrt{\mu\beta})(\sqrt{\beta\lambda} - \sqrt{\alpha\mu})}{\sqrt{\alpha\beta}(\alpha + \beta)} > 0,$$

so that

$$\Gamma(\vartheta^*) = -\frac{(\sqrt{\beta\lambda} - \sqrt{\alpha\mu})^2}{\alpha + \beta} < 0.$$

Hence, by virtue of Theorem 6.4.1, the decay rate of  $\mathbb{P}_+(P > t)$  is  $\Gamma(\vartheta^*)$ , which agrees with the asymptotics given in (6.27).

*Covariance function.* Equations (6.16) are written as

$$\begin{cases} -(\lambda + \vartheta - \alpha t)l_+(\vartheta, s, t) + \lambda l_-(\vartheta, s, t) = \alpha \eta_+(\vartheta, s | 0) - \frac{\vartheta}{s + t}; \\ \mu l_+(\vartheta, s, t) - (\mu + \vartheta + \beta t)l_-(\vartheta, s, t) = -\beta \eta_-(\vartheta, s | 0) - \frac{\vartheta}{s + t}, \end{cases}$$

whereas (6.18) reads

$$\eta_-(\vartheta, s | 0) = \frac{\mu}{\mu + \vartheta} \eta_+(\vartheta, s | 0) + \frac{\vartheta}{\mu + \vartheta}.$$

The vector  $\mathbf{w}(s, t)$  is given by

$$\mathbf{w}(s, t) = \left( \alpha \eta_+(\vartheta, s | 0) - \frac{\vartheta}{s+t}, -\beta \eta_-(\vartheta, s | 0) - \frac{\vartheta}{s+t} \right)^T.$$

Let us first compute the zeros of the determinant of  $(\mathbf{\Lambda} + t\mathbf{R} - \vartheta\mathbf{I})$ , for given  $\vartheta$ . We find

$$\tau_{\pm}(\vartheta) = \frac{1}{2\alpha\beta} \left( \beta(\lambda + \vartheta) - \alpha(\mu + \vartheta) \pm \sqrt{[\beta(\lambda + \vartheta) + \alpha(\mu + \vartheta)]^2 - 4\alpha\beta\lambda\mu} \right);$$

where  $\tau_-(\vartheta)$  is negative and  $\tau_+(\vartheta)$  is positive. We focus on the positive root  $\tau_+(\vartheta)$ . The procedure described in Section 6.5 now yields

$$\eta_+(\vartheta, s | 0) = \frac{\vartheta}{s + \tau_+(\vartheta)} \frac{\lambda\beta(s + \tau_+(\vartheta)) + (\vartheta + \mu)(\beta\tau_+(\vartheta) + \lambda + \mu + \vartheta)}{(\alpha(\vartheta + \mu)(\beta\tau_+(\vartheta) + \mu + \vartheta) - \lambda\mu\beta)}.$$

Then, due to Equation (6.17),

$$\eta_-(\vartheta, s | 0) = \frac{\vartheta}{\mu + \vartheta} \frac{(\beta\tau_+(\vartheta) + \mu + \vartheta)(\alpha(s + \tau_+(\vartheta)) + \mu) + \lambda\mu}{(s + \tau_+(\vartheta))(\alpha(\vartheta + \mu)(\beta\tau_+(\vartheta) + \mu + \vartheta) - \lambda\mu\beta)}.$$

It now follows that

$$\begin{aligned} \ell_+(\vartheta, s, t) &= \frac{\left( \frac{\lambda\beta\mu}{\mu + \vartheta} - \alpha(\beta t + \mu + \vartheta) \right) \eta_+(\vartheta, s | 0) + \frac{\vartheta}{s+t} (\beta t + \lambda + \mu + \vartheta) + \frac{\lambda\beta\vartheta}{\mu + \vartheta}}{-\alpha\beta t^2 + [\beta\lambda - \alpha\mu + \vartheta(\beta - \alpha)]t + \vartheta(\vartheta + \lambda + \mu)}, \\ \ell_-(\vartheta, s, t) &= \frac{(\beta(\lambda + \vartheta - \alpha t) - \alpha(\mu + \vartheta)) \eta_-(\vartheta, s | 0) + \frac{\vartheta}{s+t} (\alpha s + \lambda + \mu + \vartheta)}{-\alpha\beta t^2 + [\beta\lambda - \alpha\mu + \vartheta(\beta - \alpha)]t + \vartheta(\vartheta + \lambda + \mu)}. \end{aligned}$$

If the distribution of  $(Q(0), J(0))$  is given we can then compute the Laplace transform of the covariance function, by relying on Theorem 6.5.2.

In the remainder of this example we consider the situation that the system is in stationarity at time 0, and hence  $\mathbf{p}^0(x)$  is given by (6.26).

Since the formulae of the above functions are long and cumbersome, we prefer to treat a more specific example: as in [9], we choose  $\alpha = \lambda = \mu = 1$  and  $\beta = 2$ . Then we have

$$\begin{aligned} \ell_+^{(s,t)}(\vartheta, 0, \xi_+) &= 2 \frac{-8 - 77\vartheta + 392\vartheta^2 + 549\vartheta^3 + 232\vartheta^4 + 32\vartheta^5 + (8 + 5\vartheta + \vartheta^2)\sqrt{9(1 + \vartheta)^2 - 8}}{\vartheta^2(\vartheta + 2)(2\vartheta + 5)^2}, \\ \ell_-^{(s,t)}(\vartheta, 0, \xi_+) &= 2 \frac{-4 - 39\vartheta + 267\vartheta^2 + 432\vartheta^3 + 208\vartheta^4 + 32\vartheta^5 + (4 + 3\vartheta)\sqrt{9(1 + \vartheta)^2 - 8}}{\vartheta^2(\vartheta + 2)(2\vartheta + 5)^2}. \end{aligned}$$

The Laplace transform  $\rho(\vartheta)$  of the correlation  $r(t)$  is given by

$$\rho(\vartheta) = \frac{-4 - 37\vartheta + 15\vartheta^2(\vartheta + 3)(2\vartheta + 3) + (\vartheta + 4)\sqrt{9(1 + \vartheta)^2 - 8}}{15\vartheta^3(\vartheta + 2)(2\vartheta + 5)}.$$

Clearly  $\rho(\vartheta)$  is a well-defined function for  $\vartheta \geq 0$ , but from its expression we observe that  $\rho(\vartheta)$  can be continued analytically to the left up to the point

$$\bar{\vartheta} = -\frac{(\sqrt{\beta\lambda} - \sqrt{\alpha\mu})^2}{(\alpha + \beta)} = -\frac{1}{3}(\sqrt{2} - 1)^2.$$

Around this branching point,

$$\rho(\vartheta) \sim \frac{\sqrt[4]{2}(4\sqrt{2} + 18)}{15\sqrt{3} \left(\frac{2\sqrt{2}}{3} - 1\right)^3 \left(\frac{2\sqrt{2}}{3} + 1\right)^3 \left(\frac{4\sqrt{2}}{3} + 3\right)} \sqrt{\vartheta + \frac{1}{3}(\sqrt{2} - 1)^2}, \text{ as } \vartheta \downarrow \bar{\vartheta}.$$

Relying on standard techniques, we have, for  $t \rightarrow \infty$ ,

$$r(t) \sim \frac{\sqrt[4]{2}(2\sqrt{2} + 9)}{15\sqrt{3\pi} \left(1 - \frac{2\sqrt{2}}{3}\right)^3 \left(\frac{2\sqrt{2}}{3} + 1\right) \left(\frac{4\sqrt{2}}{3} + 3\right)} \cdot \frac{1}{t\sqrt{t}} \exp\left(-\frac{3 - 2\sqrt{2}}{3}t\right).$$

Notice that in this example the asymptotics of the busy-period distribution and the correlation function coincide up to a constant factor; we have encountered the same proportionality property for queues with spectrally-positive Lévy input in Chapter 4.

## 6.7 Concluding remarks

In this chapter we have considered transient characteristics of a Markov-fluid-driven queue, viz., the busy period and the covariance function of the workload process, by studying their Laplace transforms. In the case of the busy period we used sample-path large deviations to obtain its logarithmic asymptotics.

We conclude by listing a number of open issues. In Theorem 6.4.1 we found the logarithmic asymptotics of tail distribution of the busy period. The results for the two-state case in Section 6.6, however, lead to the conjecture that, for  $i \in \mathcal{E}^+$ , there is a constant  $\omega$  such that

$$\mathbb{P}_i(P > t \mid Q(0) = 0) \sim \frac{\omega}{t\sqrt{t}} e^{t\Gamma(\vartheta^*)}.$$

A similar relation can be conjectured for  $R(u) = \mathbb{Cov}(Q(0), Q(u))$ , as  $u \rightarrow \infty$ , in view of the findings of Section 6.6. In this case, however, not even the logarithmic

asymptotics are known. It is not clear, for instance if (and, if yes, how) sample-path large deviations [31] are of any help here. In fact, it is not even clear *a priori* that  $R(\cdot)$  is positive and decreasing; in case of spectrally-positive Lévy input these properties were shown relying on the concept of completely monotone functions, cf. Chapter 4 and [51, 95].

## Chapter 7

---

# Exact multivariate workload asymptotics

In this chapter we consider a discrete-time queue fed by a general process assuming only stationarity of the increments. The main contribution of this chapter concerns the derivation of the exact asymptotics of the joint probability  $\mathbb{P}(Q(0) > p, Q(T) > q)$  under the many sources scaling.

### 7.1 Introduction

As already established in the preceding chapters, a way to measure the degree of dependence between the workloads  $Q(0)$  at time 0 and  $Q(T)$  at time  $T$ , is to consider the measure  $R(T|p, q)$  defined in (1.13) for (given) positive  $p$  and  $q$ . While for various input models, considerable insight has been gained into the steady-state distribution  $Q_e$ , less is known about  $\mathbb{P}(Q(0) > p, Q(T) > q)$ , the joint probability of the workload exceeding the threshold  $p$  at time 0 and exceeding the threshold  $q$  at time  $T$ . This joint probability gives an insight in the way the events  $\{Q(0) > p\}$  and  $\{Q(T) > q\}$  are dependent. Clearly, the analysis of  $\mathbb{P}(Q(0) > p, Q(T) > q)$  is of practical and theoretical interest.

The objective of this chapter is to analyze the joint probability given above, in case a large number  $n$  of i.i.d. sources feed into the queue, with the queueing resources (buffer and service speed) scaled by  $n$  as well. In this many-sources framework, considered already in Chapter 2, we find exact asymptotics (as  $n \rightarrow \infty$ ) of the probability of interest. Our approach relies on ideas developed by Likhanov and Mazumdar [79] to obtain the exact asymptotics of the steady-state distribution of the workload under the many-sources scaling, in conjunction with results by Chaganty and Sethuraman [30] for the large deviations of sample means of multivariate random variables. As in [79], we consider a slotted-time model, i.e., a discrete-time model.

The remainder of this chapter is organized as follows. In Section 7.2 we introduce the model, and determine the tail probabilities of bivariate sample means. These are used in Section 7.3 to determine the exact asymptotics of the probability of interest. We also include a number of remarks, and indicate how to extend the results to the setting where one would consider more than two time epochs.

## 7.2 Model, objective, and preliminaries

*Traffic model.* In this chapter we consider a queueing resource fed by  $n$  i.i.d. sources. Let  $A_i\{s, t\}$ , with  $s, t \in \mathbb{Z}$  such that  $s < t$ , be the amount of traffic generated by the  $i$ -th source in timeslots  $s + 1, \dots, t$ . The  $A_i\{s, t\}$ ,  $i = 1, \dots, n$ , are distributed as the (generic) stochastic process  $A\{s, t\}$ . It is assumed that this process has stationary increments:  $A\{s, t\}$  has the same distribution as  $A\{s + u, t + u\}$  for all  $u \in \mathbb{Z}$ .

We define the cumulant function of  $A\{0, s\}$  by  $\Lambda(\vartheta|s) := \log \mathbb{E} \exp(\vartheta A\{0, s\})$ , which we assume to exist for some positive  $\vartheta$ ; the corresponding Legendre-Fenchel transform is given by  $I(x|s) := \sup_{\vartheta} (\vartheta x - \Lambda(\vartheta|s))$ . We also need the two-dimensional counterparts of these objects:

$$\begin{aligned} \check{\Lambda}_T(\vartheta, \eta|s, t) &:= \log \mathbb{E} \exp(\vartheta A\{-s, 0\} + \eta A\{T - t, T\}), \\ \check{I}_T(x, y|s, t) &:= \sup_{\vartheta, \eta} (\vartheta x + \eta y - \check{\Lambda}_T(\vartheta, \eta|s, t)). \end{aligned}$$

Let  $\vartheta(x|s)$  and  $(\vartheta(x, y|s, t), \eta(x, y|s, t))$  be the optimizing arguments in the definitions of  $I(x|s)$  and  $\check{I}_T(x, y|s, t)$ ; they may be found from the obvious first order conditions. We finally define, suppressing the arguments of  $\Lambda$  and  $\check{\Lambda}_T$ ,

$$\begin{aligned} \sigma^2(x|s) &:= \left. \frac{d^2 \Lambda}{d\vartheta^2} \right|_{\vartheta := \vartheta(x|s)}, \\ \check{\sigma}_T^2(x, y|s, t) &:= \det \left( \begin{array}{cc} \frac{\partial^2 \check{\Lambda}_T}{\partial \vartheta^2} & \frac{\partial^2 \check{\Lambda}_T}{\partial \vartheta \partial \eta} \\ \frac{\partial^2 \check{\Lambda}_T}{\partial \vartheta \partial \eta} & \frac{\partial^2 \check{\Lambda}_T}{\partial \eta^2} \end{array} \right) \bigg|_{\substack{\vartheta := \vartheta(x, y|s, t) \\ \eta := \eta(x, y|s, t)}}. \end{aligned}$$

*Queueing model, objective, reduction property.* The  $n$  sources fed into a buffered resource that is drained at a constant rate  $nc$ . To ensure stability, it is assumed that  $\mathbb{E}A\{0, 1\} < c$ . Let  $Q^n(t)$  denote the workload in the system at time  $t \in \mathbb{Z}$ . It is well-known that the stationary queue obeys

$$Q^n(t) = \sup_{s \in \mathbb{N}} \left( \sum_{i=1}^n A_i\{t - s, t\} - ncs \right), \quad (7.1)$$

where  $A\{t, t\}$  is to be understood as 0. The goal of this chapter is to find the exact asymptotics, as  $n \rightarrow \infty$ , of the probability  $\pi_T^n(p, q) := \mathbb{P}(Q^n(0) \geq np, Q^n(T) \geq nq)$ , for given positive  $p$  and  $q$ .

We will now rewrite  $\pi_T^n(p, q)$  as follows. With  $p_s := p + cs$  and  $q_t := q + ct$ , let  $E_T^n(p, q|s, t)$  denote the event

$$E_T^n(p, q|s, t) := \left\{ \sum_{i=1}^n A_i\{-s, 0\} > np_s, \sum_{i=1}^n A_i\{T - t, T\} > nq_t \right\}.$$

Then

$$\pi_T^n(p, q) = \mathbb{P}(\exists(s, t) \in \mathbb{N}^2 : E_T^n(p, q|s, t)) = \mathbb{P}(\exists(s, t) \in \mathcal{E} : E_T^n(p, q|s, t)), \quad (7.2)$$

where  $\mathcal{E}$  is the set of elements  $(s, t)$  such that  $s \in \mathbb{N}$  and  $t \in \{0, \dots, T\} \cup \{T + s\}$ . The first equality in (7.2) is directly from (7.1), whereas the second follows from a reduction property, established in [38], see also Chapter 3: we can reduce  $\mathbb{N}^2$  in (7.2) to  $\mathcal{E}$  (this is essentially due to the fact that the busy period in which  $T$  is contained can start (i) either after time 0, (ii) or at the same time as the start of the busy period in which 0 is contained).

The logarithmic asymptotics of  $\mathbb{P}(Q_e^n > np)$ , with  $Q_e^n$  denoting the stationary workload under the many-sources scaling, were found before [28]:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(Q_e^n > np) = - \inf_{s \in \mathbb{N}} I(p_s|s); \quad (7.3)$$

for exact asymptotics, see [79]. We denote by  $\bar{s}$  an optimizing argument in the right hand side of (7.3), which is not necessarily unique; also, let  $\bar{t}$  be a  $t$  for which  $I(q_t|t)$  is minimal.

*Exact two-dimensional sample-mean asymptotics.* Now consider

$$\pi_T^n(p, q|s, t) := \mathbb{P}(E_T^n(p, q|s, t)).$$

From (7.2) we find

$$\pi_T^n(p, q|s, t) \leq \pi_T^n(p, q) \leq \sum_{(s, t) \in \mathcal{E}} \pi_T^n(p, q|s, t). \quad (7.4)$$

As our goal is to derive the exact asymptotics of  $\pi_T^n(p, q)$ , we first consider in this section those of  $\pi_T^n(p, q|s, t)$ .

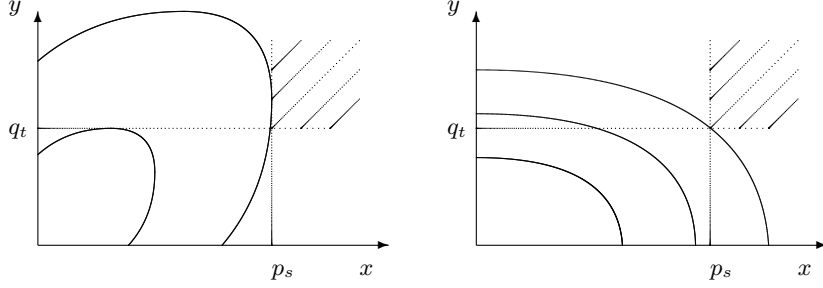
As a first remark, we note that  $\check{I}(x, y|s, t) \geq \max\{I(x|s), I(y|t)\}$ , as follows from

$$\sup_{(\vartheta, \eta) \in \mathbb{R}^2} (\vartheta x + \eta y - \check{\Lambda}_T(\vartheta, \eta|s, t)) \geq \sup_{(\vartheta, \eta) \in \mathbb{R} \times \{0\}} (\vartheta x + \eta y - \check{\Lambda}_T(\vartheta, \eta|s, t)) = I(p_s|s).$$

The bivariate version of ‘Cramér’ states

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \pi_T^n(p, q|s, t) = - \inf_{x \geq p_s, y \geq q_t} \check{I}(x, y|s, t). \quad (7.5)$$

Realizing (A) that  $\mathbb{E}A\{-s, 0\} < p_s$  and  $\mathbb{E}A\{T - t, T\} < q_t$ , and (B) that the contour lines of  $\check{I}(\cdot, \cdot|s, t)$  are convex, there are three possibilities for the optimizer  $p^*, q^*$  in the right-hand side of (7.5): (i)  $p^* = p_s$  and  $q^* > q_t$ ; (ii)  $p^* > p_s$  and  $q^* = q_t$ ; (iii)  $p^* = p_s$  and  $q^* = q_t$ . We refer to Figure 7.1 for a pictorial illustration; the left panel depicts Case (i), the right panel Case (iii).



**Figure 7.1:** Contour lines of the (two-dimensional) rate function  $\tilde{I}(x, y|s, t)$ ; the objective function is to be minimized over the shaded region.

Let us first consider Case (i). Write  $\pi_T^n(p, q|s, t)$  as

$$\begin{aligned} \pi_T^n(p, q|s, t) = & \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i\{-s, 0\} > p_s \right) \\ & - \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i\{-s, 0\} > p_s, \frac{1}{n} \sum_{i=1}^n A_i\{T-t, T\} \leq q_t \right). \end{aligned} \quad (7.6)$$

Using the Bahadur-Rao estimate [15], we have for the first probability in (7.6) that, as  $n \rightarrow \infty$ ,

$$\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i\{-s, 0\} > p_s \right) \sim \frac{e^{-nI(p_s|s)}}{\vartheta(p_s|s) \sqrt{2\pi n \sigma^2(p_s|s)}},$$

where  $g(n) \sim f(n)$  as  $n \rightarrow \infty$  denotes  $f(n)/g(n) \rightarrow 1$ . On the other hand, the second probability in (7.6) decays faster than the first probability, and is therefore asymptotically negligible: bearing in mind the left panel of Figure 7.1, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i\{-s, 0\} > p_s, \frac{1}{n} \sum_{i=1}^n A_i\{T-t, T\} \leq q_t \right) \\ & = -\tilde{I}(p_s, q_t|s, t) \\ & < -I(p_s|s). \end{aligned}$$

As Case (ii) can be dealt with similarly, we have arrived at the following result.

**Proposition 7.2.1.** *If  $p^* = p_s$  and  $q^* > q_t$ , then as  $n \rightarrow \infty$*

$$\pi_T^n(p, q|s, t) \sim \frac{e^{-nI(p_s|s)}}{\vartheta(p_s|s) \sqrt{2\pi n \sigma^2(p_s|s)}}.$$

*If  $p^* > p_s$  and  $q^* = q_t$ , then as  $n \rightarrow \infty$*

$$\pi_T^n(p, q|s, t) \sim \frac{e^{-nI(q_t|t)}}{\vartheta(q_t|t) \sqrt{2\pi n \sigma^2(q_t|t)}}.$$

Now we consider Case (iii). The decay rate of the probability of interest can alternatively be found through the Lagrangian  $\check{I}_T(x, y|s, t) - \lambda(x - p_s) - \mu(y - q_t)$ . As we know that the optimum is attained at  $p^* = p_s$  and  $q^* = q_t$ , we know that at the stationary point  $\lambda^* > 0$  and  $\mu^* > 0$  (complementary slackness). It is standard from convex analysis that

$$\frac{\partial}{\partial x} \check{I}_T(x, y|s, t) = \vartheta(x, y|s, t), \quad \frac{\partial}{\partial y} \check{I}_T(x, y|s, t) = \eta(x, y|s, t),$$

but at the same time these partial derivatives are, at  $(p^*, q^*)$ , equal to  $\lambda^*$  and  $\mu^*$ , respectively, and hence they are strictly positive. We conclude that Condition (3.4) of Chaganty and Sethuraman [30] is fulfilled, so that we can use their Theorem 3.4. We have the following result.

**Proposition 7.2.2.** *If  $p^* = p_s$  and  $q^* = q_t$ , then*

$$\pi_T^n(p, q|s, t) \sim \frac{e^{-n\check{I}_T(p_s, q_t|s, t)}}{\vartheta(p_s, q_t|s, t) \cdot \eta(p_s, q_t|s, t) \cdot 2\pi n \sqrt{\check{\sigma}_T^2(p_s, q_t|s, t)}}.$$

### 7.3 Exact workload asymptotics

In this section we use the estimates for exact bivariate sample-mean large deviations, as derived in the previous section, to determine the exact asymptotics of  $\pi_T^n(p, q)$ . As will become clear, the main idea is that these asymptotics are, under mild assumptions, fully determined by the contribution of busy periods starting at a *single* time epoch  $(s^*, t^*)$ , cf. [79].

Let  $\kappa^{(i)}(s, t)$  be 1 if Case (i) applies for  $s$  and  $t$  (that is  $p^* = p_s$  and  $q^* > q_t$ ) and 0 otherwise;  $\kappa^{(ii)}(s, t)$  and  $\kappa^{(iii)}(s, t)$  are defined likewise. Then we introduce, in self-evident notation,

$$K_T(s, t) := I(p_s|s) \cdot \kappa^{(i)}(s, t) + I(q_t|t) \cdot \kappa^{(ii)}(s, t) + \check{I}_T(p_s, q_t|s, t) \cdot \kappa^{(iii)}(s, t),$$

so that Props. 7.2.1-7.2.2 entail that  $n^{-1} \log \pi_T^n(p, q|s, t) \rightarrow -K_T(s, t)$  as  $n \rightarrow \infty$ .

We now impose the following two assumptions, in line with those needed to find the exact asymptotics of the stationary workload  $Q_e^n$ , see [79].

**Assumption 7.3.1.**  $(s^*, t^*) := \arg \min_{(s, t) \in \mathcal{E}} K_T(s, t)$  is unique.

**Assumption 7.3.2.**  $\liminf_{s \rightarrow \infty} I(p_s|s)/\log s > 0$ .

It will turn out that Assumption 7.3.1 entails that the event of overflow over level  $np$  at time 0 and over level  $nq$  at time  $T$  is essentially exclusively caused by the event  $E_T^n(p_{s^*}, q_{t^*}|s^*, t^*)$ . Assumption 7.3.2 will be needed to make sure that contributions

of  $E_T^n(p_s, q_t | s, t)$  for large  $s$  and  $t$  do not contribute significantly; below we will comment on what happens if the uniqueness assumption is not fulfilled. We are now ready to prove our main result.

**Theorem 7.3.3.** *As  $n \rightarrow \infty$ ,*

$$\frac{\pi_T^n(p, q)}{\pi_T^n(p, q | s^*, t^*)} \rightarrow 1.$$

*Proof.* The lower bound is evident due to (7.4), so let us focus on the upper bound. First observe that by applying (7.4), for any finite  $M$ ,

$$\begin{aligned} \pi_T^n(p, q) &\leq \sum_{(s,t) \in \mathcal{E}} \pi_T^n(p, q | s, t) \\ &\leq \sum_{s=0}^M \left( \sum_{t=0}^{T+s} \pi_T^n(p, q | s, t) \right) + \sum_{s=M+1}^{\infty} \left( \sum_{t=0}^{T+s} \pi_T^n(p, q | s, t) \right). \end{aligned} \quad (7.7)$$

For  $(s, t) \neq (s^*, t^*)$ , because of Assumption 7.3.1,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \pi_T^n(p, q | s, t) = -K_T(s, t) < -K_T(s^*, t^*) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \pi_T^n(p, q | s^*, t^*),$$

so that for these  $(s, t)$  it holds that  $\pi_T^n(p, q | s, t) = o(\pi_T^n(p, q | s^*, t^*))$ . Choosing  $M$  large enough such that  $s^* \in \{0, \dots, M\}$  and  $t^* \in \{0, \dots, T + s^*\}$ , it follows that

$$\sum_{s=0}^M \left( \sum_{t=0}^{T+s} \pi_T^n(p, q | s, t) \right) \sim \pi_T^n(p, q | s^*, t^*).$$

Now consider the second sum in the right-hand side of (7.7). Trivially,

$$\begin{aligned} \sum_{s=M+1}^{\infty} \left( \sum_{t=0}^{T+s} \pi_T^n(p, q | s, t) \right) &\leq \sum_{s=M+1}^{\infty} \left( \sum_{t=0}^{T+s} \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i \{-s, 0\} > p_s \right) \right) \\ &= \sum_{s=M+1}^{\infty} (T + s + 1) \mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i \{-s, 0\} > p_s \right). \end{aligned} \quad (7.8)$$

Now applying the Chernoff bound in conjunction with the fact that there is an  $\alpha > 0$  such that  $I(p_s | s) > \alpha \log s$  for  $s$  sufficiently large (due to Assumption 7.3.2), we have

$$\mathbb{P} \left( \frac{1}{n} \sum_{i=1}^n A_i \{-s, 0\} > p_s \right) \leq e^{-nI(p_s | s)} \leq s^{-n\alpha}.$$

Consequently, (7.8) is further bounded by, taking  $n > 2/\alpha$ ,

$$\begin{aligned} \sum_{s=M+1}^{\infty} (T + s + 1) s^{-n\alpha} &\leq \int_M^{\infty} (T + s + 1) s^{-n\alpha} ds \\ &= (T + 1) \frac{M^{-n\alpha+1}}{n\alpha - 1} + \frac{M^{-n\alpha+2}}{n\alpha - 2}, \end{aligned}$$

which is  $o(\pi_T^n(p, q|s^*, t^*))$  as  $n \rightarrow \infty$ , by picking  $M$  sufficiently large (that is,  $M$  should be chosen such that  $\alpha \log M > K_T(s^*, t^*)$ ).  $\square$

The following corollary is an immediate consequence of Props. 7.2.1-7.2.2 and Theorem 7.3.3.

**Corollary 7.3.4.** *If  $\kappa^{(i)}(s^*, t^*) = 1$ , then*

$$\pi_T^n(p, q) \sim \frac{e^{-nI(p_{s^*}|s^*)}}{\vartheta(p_{s^*}|s^*)\sqrt{2\pi n\sigma^2(p_{s^*}|s^*)}}.$$

*If  $\kappa^{(ii)}(s^*, t^*) = 1$ , then*

$$\pi_T^n(p, q) \sim \frac{e^{-nI(q_{t^*}|t^*)}}{\vartheta(q_{t^*}|t^*)\sqrt{2\pi n\sigma^2(q_{t^*}|t^*)}}.$$

*If  $\kappa^{(iii)}(s^*, t^*) = 1$ , then*

$$\pi_T^n(p, q) \sim \frac{e^{-n\tilde{I}_T(p_{s^*}, q_{t^*}|s^*, t^*)}}{\vartheta(p_{s^*}, q_{t^*}|s^*, t^*) \cdot \eta(p_{s^*}, q_{t^*}|s^*, t^*) \cdot 2\pi n \sqrt{\tilde{\sigma}_T^2(p_{s^*}, q_{t^*}|s^*, t^*)}}.$$

**Remark 7.3.5.** *Non-unique optimizers  $s^*$  and  $t^*$ .* Suppose  $K(s, t)$  is minimal at two  $(s, t)$ -pairs, viz.  $(s_1^*, t_1^*)$  and  $(s_2^*, t_2^*)$ . If  $\kappa^{(i)}(s_1^*, t_1^*) = 1$  and  $\kappa^{(iii)}(s_2^*, t_2^*) = 1$ , then we are essentially in case (i) of the above corollary (as the  $1/n$  factor is negligible compared to the  $1/\sqrt{n}$  factor). The same line of reasoning applies if  $\kappa^{(ii)}(s_1^*, t_1^*) = 1$  and  $\kappa^{(iii)}(s_2^*, t_2^*) = 1$ .

The other cases are harder to deal with. If  $\kappa^{(iii)}(s_1^*, t_1^*) = 1$  and  $\kappa^{(iii)}(s_2^*, t_2^*) = 1$ , then the asymptotics look like  $\gamma \exp(-nK(s_1^*, t_1^*))/n$ , but now the constant  $\gamma > 0$  cannot be determined explicitly. A similar property applies if

$$\sum_{k=1}^2 \kappa^{(i)}(s_k^*, t_k^*) + \kappa^{(ii)}(s_k^*, t_k^*) = 2;$$

then the asymptotics look like  $\delta \exp(-nK(s_1^*, t_1^*))/\sqrt{n}$ , with a constant  $\delta > 0$  that cannot be determined explicitly.  $\spadesuit$

**Remark 7.3.6.** The optimizing  $s^*$  and  $t^*$  can be interpreted as follows [38]. Given overflow over level  $np$  at time 0 and over level  $nq$  at time  $T$ , the busy period in which 0 is contained started with overwhelming probability at time  $-s^*$ , whereas the busy period in which  $T$  is contained started at time  $T - t^*$ . This means that if  $t^* = T + s^*$ , epochs 0 and  $T$  lie in the same busy period. It is expected that for large  $T$  this is typically not the case: then it is more likely that 0 and  $T$  are contained in separate busy periods. We now determine a  $T^-$  such that for  $T > T^-$  we have that  $t^* \in \{0, \dots, T\}$ .

We first observe that, due to Assumption 7.3.2, for some  $\alpha > 0$ ,

$$\begin{aligned} \inf_{s \in \mathbb{N}} \inf_{x \geq p_s, y \geq q_{T+s}} \check{I}(x, y|s, T+s) &\geq \inf_{s \in \mathbb{N}} \inf_{y \geq q_{T+s}} I(y|T+s) \\ &= \inf_{s \in \mathbb{N}} I(q_{T+s}|T+s) \geq \inf_{s \in \mathbb{N}} \alpha \log(T+s) \geq \alpha \log T. \end{aligned}$$

We impose the condition of *positive input correlation*:

$$\check{I}(x, y|s, t) \leq I(x|s) + I(y|t).$$

With  $\mathcal{E}^-$  denoting  $\mathbb{N} \times \{1, \dots, T\}$ ,

$$\inf_{(s,t) \in \mathcal{E}^-} \inf_{x \geq p_s, y \geq q_t} \check{I}(x, y|s, t) \leq \inf_{s \in \mathbb{N}} \inf_{x \geq p_s} I(x|s) + \inf_{t \in \{1, \dots, T\}} \inf_{y \geq q_t} I(y|t),$$

which equals  $I(p_{\bar{s}}|\bar{s}) + I(q_{\bar{t}}|\bar{t})$  for  $T \geq \bar{t}$  (recall that  $\bar{s}$  and  $\bar{t}$  were defined in Section 2); note that  $I(p_{\bar{s}}|\bar{s}) + I(q_{\bar{t}}|\bar{t})$  would be the decay rate if  $Q^n(0)$  and  $Q^n(T)$  would be independent. We conclude that if

$$T > T^- := \max \left\{ \exp \left( \frac{I(p_{\bar{s}}|\bar{s}) + I(q_{\bar{t}}|\bar{t})}{\alpha} \right), \bar{t} \right\},$$

we can restrict ourselves to  $(s, t) \in \mathcal{E}^-$ : the decay rate of interest equals

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \pi_T^n(p, q) = - \inf_{(s,t) \in \mathcal{E}^-} \inf_{x \geq p_s, y \geq q_t} \check{I}(x, y|s, t).$$

Intuitively, for  $T$  larger than  $T^-$  the time epochs 0 and  $T$  lie in separate busy periods with overwhelming probability. ♠

**Remark 7.3.7.** The bivariate results presented above can be easily extended to dimensions  $d \in \{3, 4, \dots\}$ . Then the probability  $\mathbb{P}(Q^n(T_i) \geq np_i, i = 1, \dots, d)$  is studied, for time epochs  $0 = T_1 \leq \dots \leq T_d$  and positive numbers  $p_1, \dots, p_d$ . Again, under mild assumptions, the corresponding asymptotics are fully determined by the contribution of busy periods starting at a single time epoch  $(s_1^*, \dots, s_d^*)$ . The result from [30] can be used again, to obtain that the asymptotics look like  $\gamma_d n^{-d^*/2} \exp(-nI_d)$ , for some positive  $\gamma_d$  and  $I_d$ ; here  $d^* \in \{1, \dots, d\}$  denotes the number of constraints that are tightly met in the  $d$ -dimensional counterpart of (7.5) evaluated at the point  $(s_1^*, \dots, s_d^*)$ . ♠

---

## Bibliography

- [1] J. Abate and W. Whitt. The correlation function of RBM and M/M/1. *Stoch. Mod.*, 4:315–359, 1988.
- [2] J. Abate and W. Whitt. Transient behavior of the M/G/1 workload process. *Oper. Res.*, 42:750–764, 1994.
- [3] J. Abate and W. Whitt. Asymptotics for M/G/1 low-priority waiting-time tail probabilities. *Queueing Syst.*, 25:173–233, 1997.
- [4] R. Addie, P. Mannersalo, and I. Norros. Most probable paths and performance formulae for buffers with Gaussian input traffic. *European Trans. Telecommun.*, 13:183–196, 2002.
- [5] R. Adler. *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Lecture Notes-Monograph Series, Vol. 12.* Institute of Mathematical Statistics, Hayward, CA, USA, 1990.
- [6] R. Adler and J. Taylor. *Random Fields and Geometry.* Springer, 2007.
- [7] S. Ahn and V. Ramaswami. Efficient algorithms for transient analysis of stochastic fluid flow models. *J. Appl. Probab.*, 42:531–549, 2005.
- [8] D. Anick, D. Mitra, and M. Sondhi. Stochastic theory of data-handling system with multiple sources. *Bell System Tech. J.*, 61:1871–1894, 1982.
- [9] S. Asmussen. Busy period analysis, rare events and transient behavior in fluid models. *J. Appl. Math. Stoch. Anal.*, 7:269–299, 1994.
- [10] S. Asmussen. Extreme value theory for queues via cycle maxima. *Extremes*, 1:137–168, 1998.
- [11] S. Asmussen. *Applied Probability and Queues, 2nd ed.* Springer, New York, NY, USA, 2003.

- 
- [12] S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, NY, USA, 2007.
- [13] S. Asmussen and C. Klüppelberg. Large deviations results for subexponential tails, with applications to insurance risk. *Stoch. Proc. Appl.*, 64:103–125, 1996.
- [14] S. Asmussen and T. Rolski. Risk theory in a periodic environment: the Cramer-Lundberg approximation and Lundberg’s inequality. *Math. Oper. Res.*, 19:410–433, 1994.
- [15] R. Bahadur and R. Rao. On deviations of the sample mean. *Ann. Math. Statist.*, 31:1015–1027, 1960.
- [16] R. Bahadur and S. Zabell. Large deviations of the sample mean in general vector spaces. *Ann. Probab.*, 7:587–621, 1979.
- [17] N. Barbot, B. Sericola, and M. Telek. Distribution of the busy period in stochastic fluid models. *Stoch. Mod.*, 17:407–427, 2001.
- [18] V. Beneš. On queues with Poisson arrivals. *Ann. Math. Statist.*, 28:670–677, 1957.
- [19] S. Bernstein. Sur les fonctions absolument monotones. *Acta Math.*, 52:1–66, 1929.
- [20] J. Bertoin. *Lévy Processes*. Cambridge University Press, Cambridge, UK, 1996.
- [21] J. Bertoin and R. Doney. Cramér’s estimate for Lévy processes. *Statist. Probab. Lett.*, 21:363–365, 1994.
- [22] N. Bingham. Fluctuation theory in continuous time. *Ann. Appl. Probab.*, 14:1766–1801, 1975.
- [23] N. Bingham and R. Doney. Asymptotic properties of subcritical branching processes I: the Galton-Watson process. *Adv. Appl. Probab.*, 6:711–731, 1974.
- [24] N. Bingham, C. Goldie, and J. Teugels. *Regular Variation*. Cambridge University Press, Cambridge, UK, 1987.
- [25] N. Bingham and S. Pitts. Non-parametric estimation for the  $M/G/\infty$  queue. *Ann. Inst. Statist. Math.*, 51:71–97, 1999.
- [26] A. Borovkov. *Stochastic Processes in Queueing Theory*. Springer, New York, NY, USA, 1976.
- [27] A. Borovkov, O. Boxma, and Z. Palmowski. On the integral of the workload process of the single server queue. *J. Appl. Probab.*, 40:200–225, 2003.

- [28] D. Botvich and N. Duffield. Large deviations, the shape of the loss curve, and economies of large scale multiplexers. *Queueing Syst.*, 20:293–320, 1995.
- [29] R. Bradley. Basic properties of strong mixing conditions – a survey and some open questions. *Probab. Surveys*, 2:107–144, 2005.
- [30] N. Chaganty and J. Sethuraman. Multidimensional strong large deviation theorems. *J. Statist. Plann. Inference*, 55:265–280, 1996.
- [31] C.-S. Chang. Sample path large deviations and intree networks. *Queueing Syst.*, 20:7–36, 1995.
- [32] J. Cohen. Some results on regular variation for distributions in queueing and fluctuation theory. *J. Appl. Probab.*, 10:343–353, 1973.
- [33] D. Cox and W. Smith. *Queues*. Methuen, London, UK, 1961.
- [34] H. Cramér and M. Leadbetter. *Stationary and Related Stochastic Processes: Sample Function Properties and their Applications*. Wiley, New York, USA, 1967.
- [35] A. da Silva Soares and G. Latouche. Matrix-analytic methods for fluid queues with finite buffers. *Perf. Eval.*, 63:295–314, 2006.
- [36] K. Dębicki. A note on LDP for supremum of Gaussian processes over infinite horizon. *Statist. Probab. Lett.*, 44:211–220, 1999.
- [37] K. Dębicki, A. Es-Saghouani, and M. Mandjes. Transient asymptotics of Lévy-driven queues. *Submitted*, 2009.
- [38] K. Dębicki, A. Es-Saghouani, and M. Mandjes. Transient characteristics of Gaussian queues. *Queueing Syst.*, to appear, 2009.
- [39] K. Dębicki and M. Mandjes. Exact overflow asymptotics for queues with many Gaussian inputs. *J. Appl. Probab.*, 40:702–720, 2003.
- [40] K. Dębicki and M. Mandjes. Traffic with an FBM limit: convergence of the workload process. *Queueing Syst.*, 46:113–127, 2004.
- [41] K. Dębicki and Z. Palmowski. Heavy traffic Gaussian asymptotics of on-off fluid model. *Queueing Syst.*, 33:327–338, 1999.
- [42] K. Dębicki and T. Rolski. A Gaussian fluid model. *Queueing Syst.*, 20:433–452, 1995.
- [43] A. de Acosta. Large deviations for vector-valued Lévy processes. *Stoch. Proc. Appl.*, 51:75–115, 1994.

- [44] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications, 2nd edition*. Springer, New York, NY, USA, 1998.
- [45] J.-D. Deuschel and D. Stroock. *Large Deviations*. Academic Press, Boston, MA, USA, 1989.
- [46] A. Dieker. Extremes of Gaussian processes over an infinite horizon. *Stoch. Proc. Appl.*, 115:207–248, 2005.
- [47] A. Dieker. Applications of factorization embeddings for Lévy processes. *Adv. Appl. Probab.*, 38:768–791, 2006.
- [48] N. Duffield and N. O’Connell. Large deviations and overflow probabilities for general single-server queue, with applications. *Math. Proc. Camb. Phil. Soc.*, 118:363–374, 1995.
- [49] A. Erlang. The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik B*, 20:33–39, 1909.
- [50] A. Erlang. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *The Post Office Electrical Engineers’ Journal (Translated from 1917 article in Danish in Elektroteknikeren, Vol. 13)*, 10:189–197, 1918.
- [51] A. Es-Saghouani and M. Mandjes. On the correlation structure of Lévy-driven queues. *J. Appl. Probab.*, 45:940–952, 2008.
- [52] A. Es-Saghouani and M. Mandjes. On the dependence structure of Gaussian queues. *Stoch. Mod.*, 25:221–247, 2009.
- [53] A. Es-Saghouani and M. Mandjes. Transient analysis of Markov fluid-driven queues. *TOP, Journal of the Spanish Society of Statistics and Operations Research*, to appear, 2009.
- [54] W. Feller. *An Introduction to Probability Theory and its Applications, 2nd ed.* Wiley, New York, NY, USA, 1971.
- [55] B. Fristedt. Sample functions of stochastic processes with stationary independent increments. In P. Ney and S. Port, editors, *Advances in Probability and Related Topics*, volume 3, pages 241–396. Marcel Dekker Inc., New York, NY, USA, 1974.
- [56] R. Gaigalas and I. Kaj. Convergence of scaled renewal processes and a packet arrival model. *Bernoulli*, 9:671–703, 2003.
- [57] A. Ganesh, N. O’Connell, and D. Wischik. *Big Queues. Lecture Notes in Mathematics, Vol. 1838*. Springer, Berlin, Germany, 2004.

- [58] S. Geršgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk. SSSR Ser. Mat.*, 1:749–754, 1931.
- [59] P. Glynn and W. Whitt. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. In J. Galambos and J. Gani, editors, *Studies in Applied Probability, Papers in Honour of Lajos Takács*, pages 131–156. Applied Probability Trust, 1994.
- [60] P. Hall and J. Park. Non-parametric inference about service-time distributions from indirect measurements. *J. Roy. Statist. Soc.*, B 66:861–875, 2004.
- [61] J. Harrison. *Brownian Motion and Stochastic Flow Systems*. Wiley, New York, NY, USA, 1985.
- [62] F. Hernandez-Campos, K. Jeffay, C. Park, J. Marron, and S. Resnick. Extremal dependence: Internet traffic applications. *Stoch. Mod.*, 21:1–35, 2005.
- [63] J. Hüsler and V. Piterbarg. Extremes of a certain class of Gaussian processes. *Stoch. Proc. Appl.*, 83:257–271, 1999.
- [64] I. Ibragimov and Y. Rozanov. *Gaussian Random Processes*. Springer-Verlag, New York, USA, 1978.
- [65] O. Kella, O. Boxma, and M. Mandjes. A Lévy process reflected at a Poisson age process. *J. Appl. Probab.*, 43:221–230, 2006.
- [66] G. Kesidis, J. Walrand, and C.-S. Chang. Effective bandwidths for multiclass Markov fluids and other ATM sources. *IEEE/ACM Trans. Netw.*, 1:424–428, 1993.
- [67] J. Kilpi and I. Norros. Testing the Gaussian approximation of aggregate traffic. In *Proceedings of the 2nd Internet Measurement Workshop*, pages 49–61, 2002.
- [68] C. Klüppelberg, A. Kyprianou, and R. Maller. Ruin probabilities and overshoots for general Lévy insurance risk processes. *Adv. Appl. Probab.*, 7:705–766, 1975.
- [69] T. Konstantopoulos and G. Last. On the dynamics and performance of stochastic fluid systems. *J. Appl. Probab.*, 37:652–667, 2000.
- [70] L. Kosten. Stochastic theory of a multi-entry buffer (I). *Delft Progress Report, Series F*, 1:10–18, 1974.
- [71] L. Kosten. Stochastic theory of data-handling systems with groups of multiple sources. In H. Rudin and W. Bux, editors, *Performance of Computer-Communication Systems*, pages 321–331. Elsevier, 1984.

- [72] V. Kulkarni. *Frontiers in Queueing*, chapter Fluid models for single buffer systems, pages 321–338. CRC Press, Boca Raton, FL, USA, 1997.
- [73] V. Kulkarni and A. Narayanan. First passage times in fluid models with an application to two priority fluid systems. In *Proceedings of the 2nd IPDS '96*, pages 166–175, 1996.
- [74] A. Kyprianou. *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin, Germany, 2006.
- [75] A. Ledford and J. Tawn. Modeling dependence within joint tail regions. *J. R. Statist. Soc. B*, 59:575–599, 1997.
- [76] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. Netw.*, 2:1–15, 1994.
- [77] P. Lieshout and M. Mandjes. Transient analysis of Brownian queues. *Preprint*, 2007.
- [78] M. Lifshits. *Gaussian Random Functions*. Kluwer, Dordrecht, The Netherlands, 1995.
- [79] N. Likhanov and R. Mazumdar. Cell loss asymptotics in buffers fed with a large number of independent stationary sources. *J. Appl. Probab.*, 36:86–96, 1999.
- [80] M. Mandjes. *Large Deviations for Gaussian Queues*. Wiley, Chichester, UK, 2007.
- [81] M. Mandjes, P. Mannersalo, I. Norros, and M. van Uitert. Large deviations of infinite intersections of events in Gaussian processes. *Stoch. Proc. Appl.*, 116:1269–1293, 2006.
- [82] M. Mandjes and A. Ridder. A large deviations analysis of the transient of a queue with many Markov fluid inputs: approximations and fast simulation. *ACM Trans. Mod. Comp. Sim.*, 12:1–26, 2002.
- [83] M. Mandjes and W. Scheinhardt. A fluid model for a relay node in an ad hoc network: evaluation of resource sharing policies. *J. Appl. Math. Stoch. Anal.*, 2008.
- [84] M. Mandjes and R. van de Meent. Inferring traffic burstiness by sampling the buffer occupancy. In R. Boutaba, K. Almeroth, R. Puigjaner, S. Shen, and J. Black, editors, *Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communication Systems. Fourth International IFIP-TC6 Networking Conference (Networking 2005), Waterloo, Canada. Lecture Notes in Computer Science (LNCS) Series, 3462*, pages 303–315. Springer, 2005.

- [85] M. Mandjes and R. van de Meent. Resource provisioning through buffer sampling. *IEEE/ACM Trans. Netw.*, to appear.
- [86] M. Mandjes and M. van Uitert. Sample-path large deviations for tandem and priority queues with Gaussian inputs. *Ann. Appl. Probab.*, 15:1193–1226, 2005.
- [87] M. Mandjes and B. Zwart. Large deviations for sojourn times in processor sharing queues. *Queueing Syst.*, 52:237–250, 2006.
- [88] P. Mannersalo and I. Norros. A most probable path approach to queueing systems with general Gaussian input. *Comp. Netw.*, 40:399–412, 2002.
- [89] L. Massoulié and A. Simonian. Large buffer asymptotics for the queue with FBM input. *J. Appl. Probab.*, 36:894–906, 1999.
- [90] T. Mikosch, S. Resnick, H. Rootzén, and A. Stegeman. Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Probab.*, 12:23–68, 2002.
- [91] P. Morse. Stochastic properties of waiting lines. *Oper. Res.*, 3:255–262, 1955.
- [92] O. Narayan. Exact asymptotic queue length distribution for fractional Brownian traffic. *Adv. Perf. Anal.*, 1:39–63, 1998.
- [93] I. Norros. A storage model with self-similar input. *Queueing Syst.*, 16:387–396, 1994.
- [94] I. Norros. Busy periods of fractional Brownian storage: a large deviations approach. *Adv. Perf. Anal.*, 2:1–20, 1999.
- [95] T. Ott. The covariance function of the virtual waiting-time process in an M/G/1 queue. *Adv. Appl. Probab.*, 9:158–168, 1977.
- [96] A. Pakes. On the tails of waiting-time distributions. *J. Appl. Probab.*, 12:555–564, 1975.
- [97] V. Piterbarg and B. Stamatović. Crude asymptotics of the probability of simultaneous high extrema of two Gaussian processes: the dual action functional. *Russ. Math. Surv.*, 60:167–168, 2005.
- [98] S. Port. Stable processes with drift on the line. *Trans. Am. Math. Soc.*, 313:805–841, 1989.
- [99] N. Prabhu. *Stochastic Storage Processes: Queues, Insurance Risk, Dams and Data Communication, 2nd edition*. Springer, New York, NY, USA, 1998.
- [100] E. Reich. On the integrodifferential equation of Takács I. *Ann. Math. Statist.*, 29:563–570, 1958.

- [101] Q. Ren and H. Kobayashi. Transient solution for the buffer behavior in statistical multiplexing. *Perf. Eval.*, 23:65–87, 1995.
- [102] J. Reynolds. The covariance structure of queues and related processes – a survey of recent work. *Adv. Appl. Probab.*, 7:383–415, 1975.
- [103] L. Rogers. Fluid models in queueing theory and Wiener-Hopf factorization of Markov chains. *Ann. Appl. Probab.*, 4:390–413, 1994.
- [104] G. Samorodnitsky. Long memory and self-similar processes. *Ann. de la Faculté des Sciences de Toulouse Série 6*, 15:107–123, 2006.
- [105] G. Samorodnitsky and M. Taqqu. *Stable non-Gaussian Random Processes*. Chapman & Hall, London, UK, 1990.
- [106] K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, Cambridge, UK, 1999.
- [107] W. Scheinhardt. *Markov-Modulated and Feedback Fluid Queues*. PhD thesis, University of Twente, The Netherlands, 1998.
- [108] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis. Queues, Communications, and Computing*. Chapman & Hall, London, UK, 1995.
- [109] P. Sonneveld. Some properties of the generalized eigenvalue problem  $Mx = \lambda(\Gamma - cI)x$ , where  $M$  is the infinitesimal generator of a Markov process, and  $\Gamma$  is a real diagonal matrix. *Delft University of Technology Report 04-02*, 2004.
- [110] M. Taqqu, W. Willinger, and R. Sherman. Proof for a fundamental result in self-similar traffic modeling. *Comp. Comm. Rev.*, 27:5–23, 1997.
- [111] A. Weiss. A new technique for analyzing large traffic systems. *Adv. Appl. Probab.*, 18:506–532, 1986.
- [112] D. Wischik and A. Ganesh. The calculus of Hurstiness. Available from <http://www.cs.ucl.ac.uk/staff/ucacdjw/research/hurstiness.pdf>. 2006.
- [113] V. Zolotarev. The first passage time of a level and the behaviour at infinity for a class of processes with independent increments. *Th. Probab. Appl.*, 9:653–661, 1964.
- [114] B. Zwart, S. Borst, and M. Mandjes. Exact asymptotics for fluid queues fed by multiple heavy-tailed on-off sources. *Ann. Appl. Probab.*, 14:903–957, 2004.

---

## Samenvatting (Summary)

Dit proefschrift richt zich op de evolutie van het bufferinhoudproces van een wachtrij. Voor specifieke wachtrijmodellen is het transiënte gedrag expliciet bepaald, maar voor wachtrijen met enigszins algemenere input is relatief weinig bekend. Dit is uiteraard het geval voor modellen waarvoor zelfs niet eens de stationaire verdeling van de bufferinhoud bepaald kan worden (laat staan de transiënte), maar er zijn ook tal van voorbeelden waarbij er wel uitdrukkingen zijn (al dan niet in termen van Laplace-getransformeerden) voor de stationaire bufferinhoud, maar niet voor de bijbehorende transiënte verdeling.

In iets specifiekere zin gaat dit proefschrift in op het volgende onderwerp: het analyseren van verschillende metrieken die ons inzicht verschaffen in de afhankelijkheidsstructuur van het stationaire bufferinhoudproces. De metrieken die we in detail bestuderen, zijn de volgende.

1. *Covariantie- en correlatiefunctie van het bufferinhoudproces.* Deze twee functies kunnen gezien worden als een maat voor de afhankelijkheid van de stationaire bufferinhoud, op twee verschillende momenten in de tijd (waarbij zij opgemerkt dat 'ongecorreleerdheid' niet equivalent is aan onafhankelijkheid).

Het is duidelijk dat voor het bepalen van deze beide metrieken het in elk geval noodzakelijk is dat men een uitdrukking heeft voor de gezamenlijke verdeling van de bufferinhoud op twee verschillende momenten in de tijd. Zoals hierboven betoogd, is dit niet altijd haalbaar als we algemeen inputverkeer beschouwen. Daarom hebben wij in ons onderzoek ook een alternatieve metriek geïntroduceerd.

2. *Alternatieve metriek.* Deze metriek is gedefinieerd als het quotiënt van de kans dat de bufferinhoud op twee verschillende tijdstippen bepaalde (gegeven) drempels overschrijdt, en het product van de corresponderende marginale kansen. Indien dit quotiënt dicht bij 1 ligt, is dit een indicatie van onafhankelijkheid.

De Hoofdstukken 2 en 3 beschouwen wachtrijmodellen met Gaussisch inputverkeer. Gebruikmakend van de theorie van de grote afwijkingen bepalen we het ge-

drag van de alternatieve metriek in twee verschillende asymptotische regimes. In Hoofdstuk 2 richten we ons op het regime dat bekend staat als het zgn. *many-sources regime* (waarbij de wachtrij wordt gevoed door een groot aantal bronnen); expliciete resultaten worden bepaald voor specifieke Gaussische inputprocessen, namelijk *fractionele Brownse beweging* en het *geïntegreerde Ornstein-Uhlenbeck* proces. Een ander regime, dat bekend staat als het *large-buffer regime*, wordt beschouwd in Hoofdstuk 3.

Hoofdstukken 4 en 5 behandelen een wachtrij met Lévy input. In Hoofdstuk 4 beschouwen we een speciale klasse van Lévy processen, de zogenaamde spectraal-positieve Lévy processen (wat wil zeggen dat alleen positieve sprongen zijn toegestaan). Voor deze klasse analyseren we de covariantie- en correlatiefunctie van het bufferinhoudproces. Daarna bepalen we ook enkele belangrijke structureigenschappen van deze functies; in het bijzonder tonen we aan dat ze positieve, dalend en convex zijn. Het algemene Lévy geval wordt geanalyseerd in Hoofdstuk 5; de resultaten zijn in termen van de alternatieve metriek in een large-buffer regime.

In Hoofdstuk 6 wordt een wachtrijmodel met Markov-gemoduleerde vloeistof input geanalyseerd. Om inzicht te krijgen in het transiënt gedrag van deze wachtrij, richten we ons eerst op de zgn. 'busy period', die opgevat kan worden als de tijd die het duurt voor de buffer voor het eerst leeg wordt. We bepalen de verdeling van die busy period in termen van haar Laplace-getransformeerde. Daarna beschouwen we de covariantie- en correlatiefunctie. Gebruikmakend van eerdere resultaten, kunnen we ook de getransformeerden hiervan bepalen, in termen van de oplossingen van een gerelateerd eigensysteem.

In Hoofdstuk 7 leggen we, in tegenstelling tot de eerdere hoofdstukken, geen eisen op aan het inputverkeer. Als we het corresponderende wachtrijsysteem bekijken in discrete tijd, blijkt het mogelijk te zijn de exacte asymptotiek te bepalen van de alternatieve metriek in het many-sources regime.

---

## About the author

Abdelghafour Es-Saghouani werd geboren op 24 september 1975 te Tazourakht, Marokko. In Al Hoceima maakte hij zijn middelbare onderwijs af met het behalen van zijn Baccalauréat in juni 1993. Hij genoot zijn universitaire opleiding aan de Université Mohammed I te Oujda, Marokko, waar hij is afgestudeerd in Wiskunde in juni 1997. In 2000 is hij begonnen aan een master Actuarial Sciences aan de Université Libre de Bruxelles, België. Na een korte verblijf in België is hij naar Nederland verhuisd om in 2001 een master aan de Universiteit van Amsterdam te volgen. In januari 2006 heeft hij zijn Master of Science in Financial Stochastic Mathematics behaald. Vervolgens is hij in februari 2006 begonnen als promovendus onder begeleiding van Michel Mandjes aan het Korteweg-de Vries Instituut voor Wiskunde der Universiteit van Amsterdam, waar hij zijn proefschrift zal verdedigen op 17 november 2009.

إِنْ شَاءَ اللَّهُ