

TRANSIENT ASYMPTOTICS OF LÉVY-DRIVEN QUEUES

KRZYSZTOF DĘBICKI,* *University of Wrocław*

ABDELGHAFOR ES-SAGHOUANI,** *University of Amsterdam*

MICHEL MANDJES,*** *University of Amsterdam, CWI, and EURANDOM*

Abstract

With $(Q_t)_t$ denoting the stationary workload process in a queue fed by a Lévy input process $(X_t)_t$, this paper focuses on the asymptotics of rare event probabilities of the type $P(Q_0 > pB, Q_{T_B} > qB)$ for given positive numbers p and q , and a positive deterministic function T_B . We first identify conditions under which the probability of interest is dominated by the ‘most demanding event’, in the sense that it is asymptotically equivalent to $P(Q > \max\{p, q\}B)$ for large B , where Q denotes the steady-state workload. These conditions essentially reduce to T_B being sublinear (i.e. $T_B/B \rightarrow 0$ as $B \rightarrow \infty$). A second condition is derived under which the probability of interest essentially ‘decouples’, in that it is asymptotically equivalent to $P(Q > pB)P(Q > qB)$ for large B . 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 α -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. We pay special attention to the case $T_B = RB$ for some $R > 0$; for light-tailed input, we derive intuitively appealing asymptotics, intensively relying on sample path large deviations results. The regimes obtained can be interpreted in terms of the most likely paths to overflow.

Keywords: Queues; transient analysis; Lévy process; large deviations

2000 Mathematics Subject Classification: Primary 60K25

1. Introduction

Lévy processes are widely used to model various real-life phenomena in, for instance, finance and networking; see, e.g. [14] and [16]. 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 queues fed by Lévy input (studying the probabilistic properties of the workload).

Received 11 December 2008; revision received 22 October 2009.

* Postal address: Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland. Email address: krzysztof.debicki@math.uni.wroc.pl

Supported by MNiSW grant number N N2014079 33 (2007–2009) and by a Marie Curie Transfer of Knowledge Fellowship of the European Community’s Sixth Framework Programme under contract MTKD-CT-2004-013389.

** Postal address: Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands. Email address: a.es-saghouani@uva.nl

*** Postal address: CWI, PO Box 94079, 1090 GB Amsterdam, The Netherlands.

emailm.r.h.mandjes@uva.nl

Part of this work was done at Stanford University.

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 has been known for over four decades, and is referred to as the (generalized) Pollaczek–Khinchine formula [20]; see also [4] for more background. In addition, the *asymptotics* of $P(Q > B)$ (B large) have been identified, in various regimes. Asymptotically, exact results for the light-tailed case (or Cramér case) are presented in [3] (cf. also [11]), whereas the heavy-tailed case was covered by, e.g. [1]; it is furthermore noted that there is also an intermediate case (cf., e.g. [13]).

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

$$F(q, \alpha) := \int_0^\infty e^{-qt} E(e^{-\alpha Q_t} \mid Q_0 = x) dt,$$

with Q_s denoting the workload at time $s > 0$ and $x \geq 0$; see, e.g. [12]. In order to get a handle on the transient distribution, we may use inversion techniques. Note, however, that essentially two inversions then need to be performed: one to obtain $E(e^{-\alpha Q_t} \mid Q_0 = x)$ from $F(q, \alpha)$ and another to obtain the transient distribution $P(Q_t \leq \cdot \mid Q_0 = x)$ from $E(e^{-\alpha Q_t} \mid Q_0 = x)$. We remark that Es-Saghouani and Mandjes [10] used the double transform mentioned above to analyze the covariance function $r(t) := \text{cov}(Q_0, Q_t)$; more specifically, it was proved that $r(\cdot)$ is positive, decreasing, and concave, and, in addition, its asymptotics (for large t) were determined.

In this paper we choose an alternative approach to analyze transient workload probabilities. Our goal is to assess to what extent the workload at time 0 has an impact on the workload at time T_B , by concentrating on probabilities of the type

$$\Pi_B := P(Q_0 > pB, Q_{T_B} > qB),$$

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 Π_B essentially factorizes (when B grows large) into $P(Q > pB)P(Q > qB)$, so that it is justified to approximate $P(Q_{T_B} > qB \mid Q_0 > pB)$ by $P(Q > qB)$. It is stressed that we do not impose the assumption that the Lévy input process, say $(X_t)_t$, be spectrally positive.

Interestingly, the shape of the function T_B essentially dictates the asymptotics of Π_B . More specifically, our paper makes the following contributions.

- (i) Our first contribution is the identification of conditions under which

$$\Pi_B \sim P(Q > \max\{p, q\}B) \tag{1.1}$$

(here ‘ \sim ’ means that the ratio of the left-hand side to the right-hand side converges to 1), or, in other words, the most demanding requirement determines the asymptotics. 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

$$\Pi_B \sim P(Q > pB)P(Q > qB). \tag{1.2}$$

If $P(Q > B)$ decays (roughly) like e^{-B} (exponential decay), or $\exp(-B^\alpha)$ with $0 < \alpha < 1$ (Weibullian decay), then this ‘decoupling condition’ reduces to $T_B/B \rightarrow \infty$. If $P(Q > B)$ 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 α -stable process, and to a compound Poisson process with regularly varying job sizes.

- (iii) For the two ‘heavy-tailed’ scenarios mentioned above, we determine the asymptotics of Π_B for T_B increasing superlinearly but subquadratically; in this case the rare event under consideration is essentially due to a single big jump (whereas in the superquadratic case two big jumps are needed, leading to the asymptotics in (1.2)).
- (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 criterion in terms of p, q , and the characteristics of the Lévy process $(X_t)_t$) then we have asymptotics as in (1.1). 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 shortly before R (in this case the asymptotics look like the decoupled asymptotics in (1.2)), and, for moderate R , the buffer remains (most likely) nonempty between 0 and R . These three regimes are in line with those identified in, e.g. [8] for Gaussian input, [15] for exponential on–off input, as well as [19, 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 [7].

The structure of the paper is as follows. In Section 2 we introduce the model, and present a number of preliminaries, such as a useful lemma taken from [8]. In Section 3 we address contributions (i) and (ii). Section 4 is devoted to the situation in which $P(Q > B)$ 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.

2. Notation and preliminaries

In this paper we consider a queue fed by a Lévy process $(X_t)_t$, emptied at a constant rate $C > 0$; recall that Lévy processes are stochastic processes with stationary independent increments [14]. Assume that $E(X_1) = \varrho < C$, to ensure that the stationary workload exists.

More formally, the steady-state buffer-content process $(Q_t)_t$ is given through

$$Q_t = \sup_{s \geq 0} (A(t - s, t) - Cs), \tag{2.1}$$

where $A(s, t) := X_t - X_s$ for $s \leq t$. Let the random variable Q denote the stationary buffer content; the law of Q coincides with that of $\sup_{t \geq 0} (X_t - Ct)$.

As mentioned in the introduction, in this paper we analyze transient characteristics of the buffer-content process. We define

$$\Pi_B := P(Q_0 > pB, Q_{T_B} > qB).$$

In this paper the primary focus is on the asymptotics of Π_B as $B \rightarrow \infty$, for given $p, q > 0$ and some function T_B that tends to ∞ as $B \rightarrow \infty$.

We finish this section with two general lemmas that are used later in the paper. Directly from (2.1), it can be found that

$$\mathbb{P}_B = \mathbb{P}(\text{there exist } s \geq 0, t \geq 0: A(-s, 0) - Cs > pB, A(T_B - t, T_B) - Ct > qB). \tag{2.2}$$

The following lemma, featuring a reduction property proven in [8], formalizes the evident property that the start of the busy period in which T_B is contained (corresponding to time $T_B - t^*$, say) 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 $t^* \in [0, T_B)$). This means that in (2.2) we can restrict ourselves to a subset of $s, t \geq 0$.

Lemma 2.1. *Let*

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

Then

$$\mathbb{P}_B = \mathbb{P}(\text{there exist } (s, t) \in \mathcal{E}: A(-s, 0) - Cs > pB, A(T_B - t, T_B) - Ct > qB).$$

We finally state a weak law of large numbers, which holds due to the fact that X_t is integrable.

Lemma 2.2. *For any $\delta > 0$,*

$$\lim_{t \rightarrow \infty} \mathbb{P}\left(\frac{X_t}{t} < \varrho - \delta\right) = \lim_{t \rightarrow \infty} \mathbb{P}\left(\frac{X_t}{t} > \varrho + \delta\right) = 0.$$

3. General results

In this section we prove two general results. The first says that (1.1) 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 (1.2) 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*.

3.1. Short time-scale regime

In this subsection we prove our result for the short time-scale regime; as before, Q denotes the stationary workload. It consists of two cases: the $p > q$ case which holds under the condition that $T_B/B \rightarrow 0$ as $B \rightarrow \infty$ and the $q > p$ case which holds under Assumption 3.1, below. We stress that later in this paper we will show that in both heavy-tailed and light-tailed scenarios Assumption 3.1 is fulfilled as long as $T_B/B \rightarrow 0$ as $B \rightarrow \infty$.

Assumption 3.1. *One of the following two properties holds.*

- (i) *The sequence T_B is such that, for all $\eta > 0$,*

$$\limsup_{B \rightarrow \infty} \frac{\mathbb{P}(\text{there exists } t \in (0, T_B): X_t - Ct > \eta B)}{\mathbb{P}(Q > qB)} = 0.$$

- (ii) *The sequence T_B is such that, for all $\eta > 0$, $\mathbb{P}(Q > qB + \eta T_B) \sim \mathbb{P}(Q > qB)$ as $B \rightarrow \infty$.*

Theorem 3.1. For the case in which $p > q$, if $T_B/B \rightarrow 0$ as $B \rightarrow \infty$ then

$$\Pi_B \equiv P(Q_0 > pB, Q_{T_B} > qB) \sim P(Q > pB).$$

For the case in which $q > p$, under Assumption 3.1,

$$\Pi_B \equiv P(Q_0 > pB, Q_{T_B} > qB) \sim P(Q > qB).$$

Proof. First consider the case in which $p > q$. We are left to prove that

$$\liminf_{B \rightarrow \infty} \frac{P(Q_0 > pB, Q_{T_B} > qB)}{P(Q > pB)} \geq 1.$$

This is proven as follows. Fix $\varepsilon > 0$. Let B be sufficiently large such that $(p - q)B > (C - \varrho + \varepsilon)T_B$ (which is possible due to $T_B/B \rightarrow 0$ and $p > q$). Then

$$P(Q_0 > pB, Q_{T_B} > qB) \geq P(Q_0 > pB)P(X_{T_B} > (\varrho - \varepsilon)T_B).$$

Recalling that we assumed that $T_B \rightarrow \infty$, owing to Lemma 2.2, we have, for any $\delta > 0$ and large enough B , $P(X_{T_B} > (\varrho - \varepsilon)T_B) > 1 - \delta$. The statement then follows by letting $\delta \downarrow 0$.

Now focus on $q > p$, first under Assumption 3.1(i). It suffices to prove that, as $B \rightarrow \infty$, we have $P(Q_0 < pB, Q_{T_B} > qB) = o(P(Q > qB))$. Let \mathcal{T}_B be the event that $Q_t > 0$ for all $t \in (0, T_B)$. First observe that, with $\eta := q - p > 0$,

$$\begin{aligned} P(Q_0 < pB, Q_{T_B} > qB, \mathcal{T}_B) &\leq P(X_{T_B} > \eta B + CT_B) \\ &\leq P(\text{there exists } t \in (0, T_B) : X_t - Ct > \eta B), \end{aligned}$$

which is $o(P(Q > qB))$ due to Assumption 3.1(i). Also,

$$\begin{aligned} P(Q_0 < pB, Q_{T_B} > qB, \mathcal{T}_B^c) &\leq P(Q_{T_B} > qB, \mathcal{T}_B^c) \\ &\leq P(\text{there exists } t \in (0, T_B) : A(T_B - t, T_B) - Ct > qB), \end{aligned}$$

which is also $o(P(Q > qB))$, again by Assumption 3.1(i).

Again, consider the case in which $q > p$, but now under Assumption 3.1(ii). It is clear that it suffices to show that $\liminf_{B \rightarrow \infty} \Pi_B/P(Q > qB) \geq 1$. For each positive N , we have $qB + NT_B > pB$, and so

$$\Pi_B \geq P(Q_0 > qB + NT_B, Q_{T_B} > qB) \geq P(Q > qB + NT_B)P(X_{T_B} > (C - N)T_B).$$

Now observe that, by assumption, $P(Q > qB + NT_B) \sim P(Q > qB)$ as $B \rightarrow \infty$. Moreover, for each $\varepsilon > 0$, there exists an N_0 such that, for each $N \geq N_0$, it holds that $P(X_{T_B} > (C - N)T_B) \geq 1 - \varepsilon$ for sufficiently large B . Thus, as $B \rightarrow \infty$,

$$P(Q > qB + NT_B)P(X_{T_B} > (C - N)T_B) \sim P(Q > qB)P(X_{T_B} > (C - N)T_B),$$

which is larger than $(1 - \varepsilon)P(Q > qB)$. The statement then follows by letting $\varepsilon \downarrow 0$. This completes the proof.

Remark 3.1. The case in which $p = q$ should be handled with care; it is readily checked from the proof of Theorem 3.1 that the argumentation for $q > p$ works for $q \geq p$ under Assumption 3.1(ii), but not under Assumption 3.1(i).

Let us now check how Assumption 3.1(ii) relates to the condition $T_B/B \rightarrow 0$. In the case that $P(Q > B)$ decays (roughly) polynomially (i.e. $P(Q > B) \sim KB^{-\zeta}$), Assumption 3.1(ii) indeed reduces to $T_B/B \rightarrow 0$ as $B \rightarrow \infty$. It is noted, however, that if $P(Q > B)$ decays (roughly) exponentially then Assumption 3.1(ii) reads $T_B \rightarrow 0$.

We now argue that Assumption 3.1(ii) is, in the case in which $p = q$, ‘minimal’ if $P(Q > B)$ decays exponentially, in the sense that $\liminf_{B \rightarrow \infty} T_B = M > 0$ leads to $\limsup_{B \rightarrow \infty} \Pi_B/P(Q > pB) < 1$, as follows. Consider, for instance, the case that $(X_t)_t$ corresponds to (standard) Brownian motion. Decompose Π_B into $\Pi_B^{(1)} + \Pi_B^{(2)}$, where

$$\Pi_B^{(1)} := P(Q_0 > pB, Q_{T_B} > pB, \mathcal{T}_B), \quad \Pi_B^{(2)} := P(Q_0 > pB, Q_{T_B} > pB, \mathcal{T}_B^c).$$

and \mathcal{T}_B is defined in the proof of Theorem 3.1. First observe that

$$\begin{aligned} \Pi_B^{(2)} &\leq P(Q_0 > pB, \text{there exists } t \in [0, T_B]: A(t, T_B - t) > pB + Ct) \\ &= P(Q_0 > pB) P(\text{there exists } t \in [0, T_B]: A(t, T_B - t) > pB + Ct) \\ &\leq (P(Q > pB))^2 \\ &= o(P(Q > pB)). \end{aligned}$$

Regarding $\Pi_B^{(1)}$, first recall that $P(Q > B) = e^{-2CB}$. We find that

$$\begin{aligned} \Pi_B^{(1)} &\leq P(Q_0 > pB, Q_0 + X_{T_B} > pB + CT_B) \\ &= \int_{pB}^{\infty} P(X_{T_B} > pB + CT_B - x) 2Ce^{-2Cx} dx \\ &= \int_0^{\infty} P(X_{T_B} > CT_B - y) 2Ce^{-2C(y+pB)} dy \\ &= P(Q + X_{T_B} > CT_B) P(Q > pB). \end{aligned}$$

Since $\liminf_{B \rightarrow \infty} T_B = M > 0$, we have $\limsup_{B \rightarrow \infty} P(Q + X_{T_B} > CT_B) < 1$, and, as a consequence, $\limsup_{B \rightarrow \infty} \Pi_B^{(1)}/P(Q > pB) < 1$; therefore, $\limsup_{B \rightarrow \infty} \Pi_B/P(Q > pB) < 1$ also. This shows that Assumption 3.1(ii) is ‘minimal’ for the case in which $p = q$.

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 denotes the stationary workload; we also define (for $D > \varrho$) Q^D as the stationary workload if the queue were emptied at rate D rather than C .

Assumption 3.2. *The sequence T_B is such that, for all $\eta > 0$ and $D > \varrho$,*

$$\lim_{B \rightarrow \infty} \frac{P(Q^D > \eta T_B)}{P(Q > pB) P(Q > qB)} = 0.$$

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 3.2. *Under Assumption 3.2, it holds that*

$$\Pi_B \equiv P(Q_0 > pB, Q_{T_B} > qB) \sim P(Q > pB) P(Q > qB).$$

Proof. Let us start by establishing the lower bound. By definition,

$$P(Q_0 > pB, Q_{T_B} > qB) = P(\text{there exists } s \geq 0: A(-s, 0) > pB + Cs, \\ \text{there exists } t \geq 0: A(T_B - t, T_B) > qB + Ct).$$

The probability on the right-hand side of this equation majorizes:

$$P(\text{there exists } s \geq 0: A(-s, 0) > pB + Cs, \\ \text{there exists } t \in (0, T_B): A(T_B - t, T_B) > qB + Ct) \\ = P(\text{there exists } s \geq 0: A(-s, 0) > pB + Cs) \\ \times P(\text{there exists } t \in (0, T_B): A(T_B - t, T_B) > qB + Ct) \\ = P(Q > pB) P(\text{there exists } t \in (0, T_B): A(-t, 0) > qB + Ct).$$

We observe that it remains to prove that

$$\frac{P(\text{there exists } t > T_B: A(-t, 0) > qB + Ct)}{P(Q > qB)} \rightarrow 0 \tag{3.1}$$

as $B \rightarrow \infty$. Let us consider the numerator of (3.1). It is trivial to see that

$$P(\text{there exists } t > T_B: A(-t, 0) > qB + Ct) \\ = P\left(\left(\sup_{t > T_B} A(-t, -T_B) - C(t - T_B)\right) + A(-T_B, 0) > qB + CT_B\right) \\ = P(Q_{-T_B} + A(-T_B, 0) > qB + CT_B).$$

We now distinguish between Q_{-T_B} being either smaller or larger than δCT_B , so that the previous expression is not larger than

$$P(Q_{-T_B} + A(-T_B, 0) > qB + CT_B, Q_{-T_B} < \delta CT_B) + P(Q_{-T_B} \geq \delta CT_B).$$

The second of these probabilities, which evidently equals $P(Q \geq \delta CT_B)$, is $o(P(Q > qB))$ due to Assumption 3.2—in fact, this assumption implies that it is even $o(P(Q > pB) P(Q > qB))$, as we will need below. To deal with the first probability, choose $\varepsilon > 0$ such that $C^* := q + \varepsilon < (1 - \delta)C$; then

$$P(Q_{-T_B} + A(-T_B, 0) > qB + CT_B, Q_{-T_B} < \delta CT_B) \\ \leq P(A(-T_B, 0) > (1 - \delta)CT_B) \\ = P(A(-T_B, 0) - C^*T_B > ((1 - \delta)C - C^*)T_B) \\ \leq P(\text{there exists } t \geq 0: A(-t, 0) - C^*t > ((1 - \delta)C - C^*)T_B) \\ = P(Q^{C^*} > ((1 - \delta)C - C^*)T_B),$$

which is $o(P(Q > qB))$ due to Assumption 3.2—again, it is even $o(P(Q > pB) P(Q > qB))$.

We now proceed by establishing the upper bound. In view of Lemma 2.1 we can split the probability of interest according to whether the queue was empty in $(0, T_B)$ or not, thus obtaining

$$P_B = P(Q_0 > pB, Q_{T_B} > qB, \mathcal{T}_B^c) + P(Q_0 > pB, Q_{T_B} > qB, \mathcal{T}_B). \tag{3.2}$$

The first of the probabilities in (3.2) equals

$$\begin{aligned} &P(\text{there exists } s \geq 0: A(-s, 0) > pB + Cs, \\ &\quad \text{there exists } t \in (0, T_B): A(T_B - t, T_B) > qB + Ct) \\ &= P(\text{there exists } s \geq 0: A(-s, 0) > pB + Cs) \\ &\quad \times P(\text{there exists } t \in (0, T_B): A(T_B - t, T_B) > qB + Ct) \\ &\leq P(\text{there exists } s \geq 0: A(-s, 0) > pB + Cs) \\ &\quad \times P(\text{there exists } t \geq 0: A(-t, 0) > qB + Ct) \\ &= P(Q > pB)P(Q > qB). \end{aligned}$$

The second of the probabilities in (3.2) equals

$$\begin{aligned} &P(\text{there exists } s > 0: A(-s, 0) > pB + Cs, A(-s, T_B) > qB + C(T_B + s)) \\ &\leq P(\text{there exists } s > 0: A(-s, T_B) > qB + C(T_B + s)) \\ &= P(\text{there exists } s > T_B: A(-s, 0) > qB + Cs). \end{aligned}$$

Above we saw that $P(\text{there exists } s > T_B: A(-s, 0) > qB + Cs)$ is $o(P(Q > pB) \times P(Q > qB))$ as B grows large. This observation completes the proof.

4. Heavy-tailed Lévy input

In this section we focus on the situation that the tail distribution of Q decays essentially polynomially.

Assumption 4.1. For some $\zeta > 0$, all $D > \varrho$, and some $K(\cdot) > 0$,

$$P(Q^D > B) \sim K(D)B^{-\zeta} \quad \text{as } B \rightarrow \infty.$$

In the sequel, we abbreviate $K := K(C)$.

Let us first check what Assumptions 3.1 and 3.2 look like in this situation.

- (i) Consider Assumption 3.1(ii). As has been noted in Remark 3.1, this assumption is valid under $T_B/B \rightarrow 0$ as $B \rightarrow \infty$.
- (ii) Now consider Assumption 3.2. It is readily checked that, under Assumption 4.1, this does *not* reduce to $T_B/B \rightarrow \infty$, but to $T_B/B^2 \rightarrow \infty$.

We mention here that, interestingly, Assumption 3.2 *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 decays subexponentially (but faster than polynomially); this is, for instance, the case when $\log P(Q^D > B)/B^\alpha \rightarrow -\kappa(D)$ as $B \rightarrow \infty$ for $\alpha \in (0, 1)$ and some $\kappa(\cdot) > 0$ (Weibullian decay). Interestingly, in the situation that $\log P(Q^D > B)/(\log B)^2 \rightarrow -\kappa(D)$ (which is a tail that resembles that of the lognormal distribution), Assumption 3.2 holds if

$$(\log(\eta T_B))^2 - (\log(pB))^2 - (\log(qB))^2 \rightarrow \infty;$$

with T_B of the type B^β , this simplifies to requiring that $\beta > \sqrt{2}$.

The above observations indicate that, for $P(Q > B)$ 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 note that Lemma 2.1 entails that we can decompose Π_B as

$$\Pi_B = P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2),$$

where

$$E_1 := \{\text{there exist } s \geq 0, t \in [0, T_B]: A(-s, 0) - Cs > pB, A(T_B - t, T_B) - Ct > qB\},$$

$$E_2 := \{\text{there exists } s \geq 0: A(-s, 0) - Cs > pB, A(-s, T_B) - C(s + T_B) > qB\}.$$

The following lemma and its corollary are useful in our proofs.

Lemma 4.1. *The following three statements hold under Assumption 4.1:*

(i) for any $B > 0$,

$$P(E_1) = P(Q > pB) P\left(\sup_{t \in [0, T_B]} (X_t - Ct) > qB\right); \tag{4.1}$$

(ii) as $B \rightarrow \infty$, given that $\lim_{B \rightarrow \infty} T_B/B = \infty$,

$$P(E_1) \sim P(Q > pB) P(Q > qB) \sim K^2(pq)^{-\zeta} B^{-2\zeta};$$

(iii) as $B \rightarrow \infty$,

$$P(E_2) \sim P(Q > \max\{pB, qB + (C - \varrho)T_B\});$$

(iv) if $T_B = RB^2$ for some $R > 0$ then $P(E_1 \cap E_2) = o(P(E_1))$ as $B \rightarrow \infty$.

Proof. Claim (i) follows directly from the independence of the increments of $(X_t)_t$.

Now concentrate on claim (ii). Recall that $T_B \rightarrow \infty$ as $B \rightarrow \infty$. Rewrite $P(E_1)$ as in (4.1).

Observe that

$$P\left(\sup_{t \in [0, T_B]} (X_t - Ct) > qB\right) = P\left(\sup_{t \geq 0} (X_t - Ct) > qB\right) - P\left(\sup_{t \geq 0} (X_t - Ct) > qB, \sup_{t \in [0, T_B]} (X_t - Ct) \leq qB\right). \tag{4.2}$$

The first of the two probabilities on the right-hand side of (4.2) is $P(Q > qB)$. The other one is majorized by

$$P(\text{there exists } t \geq T_B: (X_t - Ct) > qB).$$

Precisely as in the lower bound in Theorem 3.2, this probability is bounded from above by

$$P(Q \geq \delta CT_B) + P(Q^{C^*} > ((1 - \delta)C - C^*)T_B),$$

which is of the order of $(T_B)^{-\zeta}$, and, therefore, $o(P(Q > qB))$. This proves claim (ii).

We proceed by establishing claim (iii). To make the notation a bit lighter, we write T instead of T_B throughout the remainder of this proof. Let $\varepsilon \in (0, 1)$. Observe that $P(E_2) = P(E_{21}) + P(E_{22}) + P(E_{23})$, where

$$E_{21} := \{Q_0 > \max\{pB, qB + CT - X_T\}, X_T \in [\varrho T - \varepsilon(T + B), \varrho T + \varepsilon(T + B)]\},$$

$$E_{22} := \{Q_0 > \max\{pB, qB + CT - X_T\}, X_T < \varrho T - \varepsilon(T + B)\},$$

$$E_{23} := \{Q_0 > \max\{pB, qB + CT - X_T\}, X_T > \varrho T + \varepsilon(T + B)\}.$$

We first consider $P(E_{21})$. Due to Lemma 2.2, as $B \rightarrow \infty$, we have

$$\begin{aligned} P(E_{21}) &\leq P(Q > \max\{pB, qB + CT - \varrho T - \varepsilon(T + B)\}) \\ &\quad \times P(X_T \in [\varrho T - \varepsilon(T + B), \varrho T + \varepsilon(T + B)]) \\ &\leq P(Q > \max\{pB, (q - \varepsilon)B + (C - \varrho - \varepsilon)T\}) \\ &\sim K \min\{(pB)^{-\zeta}, ((q - \varepsilon)B + (C - \varrho - \varepsilon)T)^{-\zeta}\} \end{aligned} \tag{4.3}$$

and

$$\begin{aligned} P(E_{21}) &\geq P(Q > \max\{pB, qB + CT - \varrho T + \varepsilon(T + B)\}) \\ &\quad \times P(X_T \in [\varrho T - \varepsilon(T + B), \varrho T + \varepsilon(T + B)]) \\ &\gtrsim P(Q > \max\{pB, (q + \varepsilon)B + (C - \varrho + \varepsilon)T\}) \\ &\sim K \min\{(pB)^{-\zeta}, ((q + \varepsilon)B + (C - \varrho + \varepsilon)T)^{-\zeta}\}; \end{aligned} \tag{4.4}$$

here $g(B) \gtrsim h(B)$ means that $\liminf_{B \rightarrow \infty} g(B)/h(B) \geq 1$. Thus, letting ε go to 0 in (4.3) and (4.4), we obtain

$$\begin{aligned} P(E_{21}) &\sim K \min\{(pB)^{-\zeta}, (qB + (C - \varrho)T)^{-\zeta}\} \\ &\sim P(Q > \max\{pB, qB + (C - \varrho)T\}) \end{aligned} \tag{4.5}$$

as $B \rightarrow \infty$. Moreover, for each $\varepsilon \in (0, 1)$, as $B \rightarrow \infty$,

$$\begin{aligned} P(E_{22}) &\leq P(Q > \max\{pB, qB + CT - \varrho T + \varepsilon(T + B)\}) P(X_T < \varrho T - \varepsilon(T + B)) \\ &\leq P(Q > \max\{pB, qB + CT - \varrho T\}) P(X_T < \varrho T - \varepsilon(T + B)) \\ &\sim K \min\{(pB)^{-\zeta}, (qB + (C - \varrho)T)^{-\zeta}\} P(X_T < \varrho T - \varepsilon(T + B)) \\ &= o(P(E_{21})), \end{aligned} \tag{4.6}$$

by Lemma 2.2. It is standard that

$$\begin{aligned} P(X_T > \varrho T + \varepsilon(T + B)) &= P\left(X_T - \left(\varrho + \frac{\varepsilon}{2}\right)T > \frac{\varepsilon T}{2} + \varepsilon B\right) \\ &\leq P\left(\text{there exists } t \geq 0: X_t - \left(\varrho + \frac{\varepsilon}{2}\right)t > \frac{\varepsilon}{2}(T + B)\right) \\ &= P\left(Q^{\varrho + \varepsilon/2} > \frac{\varepsilon}{2}(T + B)\right). \end{aligned}$$

This means that there is, by virtue of Assumption 4.1, a constant $\tilde{K} > 0$ such that

$$\begin{aligned} P(E_{23}) &\leq P(Q > pB) P(X_T > \varrho T + \varepsilon(T + B)) \\ &\leq P(Q > pB) P\left(Q^{\varrho + \varepsilon/2} > \frac{\varepsilon}{2}(T + B)\right) \\ &\sim K (pB)^{-\zeta} \tilde{K} \left(\frac{\varepsilon}{2}(T + B)\right)^{-\zeta} \\ &= o(P(E_{21})). \end{aligned} \tag{4.7}$$

Upon combining (4.5) with (4.6) and (4.7), we establish claim (iii).

Finally, consider claim (iv). Let $\delta \in (0, \frac{1}{2})$ and $\varepsilon > 0$. We have

$$\begin{aligned} P(E_1 \cap E_2) &= P(E_1 \cap E_2, X_T \geq (\varrho + \varepsilon)T + T^{1-\delta}) \\ &\quad + P(E_1 \cap E_2, X_T \leq (\varrho + \varepsilon)T + T^{1-\delta}) \\ &\leq P(Q_0 > pB, X_T \geq (\varrho + \varepsilon)T + T^{1-\delta}) \\ &\quad + P\left(Q_0 > qB + (C - \varrho - \varepsilon)T - T^{1-\delta}, \sup_{t \in [0, T]} (X_t - X_{T-t} - Ct) > qB\right) \\ &= P(Q > pB) P(X_T \geq (\varrho + \varepsilon)T + T^{1-\delta}) \\ &\quad + P(Q > qB + (C - \varrho - \varepsilon)T - T^{1-\delta}) P\left(\sup_{t \in [0, T]} (X_t - Ct) > qB\right). \end{aligned} \tag{4.8}$$

Since $T = RB^2$ and $\delta \in (0, \frac{1}{2})$, for some constant $\bar{K} > 0$,

$$P(X_T \geq (\varrho + \varepsilon)T + T^{1-\delta}) \leq P\left(\sup_{t \geq 0} (X_t - (\varrho + \varepsilon)t) \geq T^{1-\delta}\right) \sim \bar{K}(T^{1-\delta})^{-\zeta} = o(B^{-\zeta})$$

(use Assumption 4.1). We thus conclude, using claim (ii) of this lemma, that

$$P(Q > pB) P(X_T \geq (\varrho + \varepsilon)T + T^{1-\delta}) = o(P(E_1)).$$

We also have, using the fact that claim (ii) implies that $P(E_1) = O(B^{-2\zeta})$,

$$\begin{aligned} P(Q > qB + (C - \varrho - \varepsilon)T - T^{1-\delta}) &\sim P(Q > (C - \varrho - \varepsilon)T) = O(B^{-2\zeta}) = O(P(E_1)), \\ P\left(\sup_{t \in [0, T]} (X_t - Ct) > qB\right) &\leq P\left(\sup_{t \geq 0} (X_t - Ct) > qB\right) = P(Q > qB) \rightarrow 0 \end{aligned}$$

as $B \rightarrow \infty$, which in view of (4.8) leads to $P(E_1 \cap E_2) = o(P(E_1))$. This completes the proof of claim (iv).

The following corollary directly follows from the proof of Lemma 4.1(ii).

Corollary 4.1. *Under Assumption 4.1, as $B \rightarrow \infty$, given that $\lim_{B \rightarrow \infty} T_B/B = \infty$,*

$$P\left(\sup_{t \in [0, T_B]} (X_t - Ct) > B\right) \sim P(Q > B).$$

We now present two propositions that, for the case that T_B is at least linear but slower than quadratic, express the asymptotics of Π_B in terms of the asymptotics of $P(Q > B)$, viz. Proposition 4.1 for the case in which $q \geq p$ and Proposition 4.2 for the case in which $p > q$. Corollaries 4.2 and 4.3, below, summarize the findings so far. In the rest of this section Assumption 4.1 is in force.

Proposition 4.1. *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*

$$\Pi_B \sim P(Q > qB + (C - \varrho)T_B).$$

(ii) *If $T_B = RB^2$ for some $R > 0$ then*

$$\Pi_B \sim P(Q > pB) P(Q > qB) + P(Q > (C - \varrho)T_B).$$

Proof. To prove claim (i), it suffices to show that $P(E_1) = o(P(E_2))$. From Lemma 4.1(i), it immediately follows that $P(E_1) \leq P(Q > pB)P(Q > qB)$. Since we have $q \geq p$, it follows by Lemma 4.1(iii) that $P(E_2) \sim P(Q > qB + (C - \varrho)T_B)$. It also holds that

$$P(Q > pB)P(Q > qB) = o(P(Q > qB + (C - \varrho)T_B))$$

as $B \rightarrow \infty$. This completes the proof of claim (i).

Now consider claim (ii). If $T_B = RB^2$ then, following Lemma 4.1(i)–(ii),

$$P(E_1) = P(Q > pB)P\left(\sup_{t \in [0, T_B]} (X(t) - Ct) > qB\right) \sim P(Q > pB)P(Q > qB)$$

and

$$P(E_2) \sim P(Q > qB + (C - \varrho)T_B) \sim P(Q > (C - \varrho)T_B),$$

as $B \rightarrow \infty$. Since $P(E_1) = O(P(E_2))$, it now suffices to recall that, owing to Lemma 4.1(iv), it holds that $P(E_1 \cap E_2) = o(P(E_1))$. We thus establish claim (ii).

The following corollary is an immediate consequence of Theorems 3.1 and 3.2, Proposition 4.1, Remark 3.1, and Corollary 4.1.

Corollary 4.2. *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 $\Pi_B \sim K(q + (C - \varrho)R)^{-\zeta}B^{-\zeta}$.*
- (iii) *If $T_B/B \rightarrow \infty$ and $T_B/B^2 \rightarrow 0$ as $B \rightarrow \infty$, then $\Pi_B \sim K((C - \varrho)T_B)^{-\zeta}$.*
- (iv) *If $T_B = RB^2$ for some $R > 0$ then $\Pi_B \sim (K^2(pq)^{-\zeta} + K((C - \varrho)R)^{-\zeta})B^{-2\zeta}$.*
- (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 in which $q < p$.

Proposition 4.2. *Let $q < p$.*

- (i) *If $T_B = RB$ with $R \leq (p - q)/(C - \varrho)$ then $\Pi_B \sim P(Q > pB)$.*
- (ii) *If $\liminf_{B \rightarrow \infty} T_B/B > (p - q)/(C - \varrho)$ and $T_B/B^2 \rightarrow 0$ as $B \rightarrow \infty$, then $\Pi_B \sim P(Q > qB + (C - \varrho)T_B)$.*
- (iii) *If $T = RB^2$ as $B \rightarrow \infty$ for some $R > 0$ then $\Pi_B \sim P(Q > pB)P(Q > qB) + P(Q > (C - \varrho)T_B)$.*

Proof. We only consider claim (i); the other claims can be proven as the corresponding statements in Proposition 4.1. Note that Lemma 4.1(iii) entails that $P(E_2) \sim P(Q > pB)$. Combining this with Lemma 4.1(i), we conclude that $P(E_1) = o(P(E_2))$. This implies that $\Pi_B \sim P(Q > pB)$, which completes the proof of (i).

Corollary 4.3. *Let $q < p$.*

- (i) *If $T_B/B \rightarrow 0$ as $B \rightarrow \infty$, or $T = RB$ with $R \leq (p - q)/(C - \varrho)$, then $\Pi_B \sim Kp^{-\zeta}B^{-\zeta}$.*
- (ii) *If $T_B = RB$ for $R > (p - q)/(C - \varrho)$ then $\Pi_B \sim K(q + (C - \varrho)R)^{-\zeta}B^{1-\zeta}$.*

- (iii) If $T_B/B \rightarrow \infty$ and $T_B/B^2 \rightarrow 0$ as $B \rightarrow \infty$, then $\Pi_B \sim K((C - \varrho)T_B)^{-\zeta}$.
- (iv) If $T_B = RB^2$ as $B \rightarrow \infty$ for some $R > 0$ then $\Pi_B \sim (K^2(pq)^{-\zeta} + K((C - \varrho)R)^{-\zeta})B^{-2\zeta}$.
- (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, which are included in the class of input processes that satisfies Assumption 4.1: (a) α -stable input and (b) compound Poisson input with polynomially decaying job size distribution.

(a) *α -stable input.* Let X_t be an α -stable Lévy process [18] with $\alpha \in (1, 2)$ and $\beta \in (-1, 1]$. We use the notation

$$\mathbb{B}(\alpha, \beta) := \frac{\Gamma(1 + \alpha)}{\pi} \sqrt{1 + \beta^2 \tan^2\left(\frac{\pi\alpha}{2}\right)} \sin\left(\frac{\pi\alpha}{2} + \arctan\left(\beta \tan\left(\frac{\pi\alpha}{2}\right)\right)\right).$$

Then, owing to [17], Assumption 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 λ) of independent and identically distributed jobs. Let the distribution of the jobs obey $P(J^r > x) \sim \kappa x^{-\zeta}$ for positive ζ and κ , where J^r denotes the residual job length:

$$P(J^r > x) = \frac{1}{E J} \int_x^\infty P(J > y) dy.$$

Note that $\varrho = \lambda E J$. Then [5], [6]

$$P(Q > x) \sim \frac{\varrho}{C - \varrho} \kappa x^{-\zeta}.$$

Again, we conclude that Assumption 4.1 (and, hence, the theory of this section) applies, with an obvious value for K .

5. Light-tailed input

In this section we derive the logarithmic asymptotics of Π_B as $B \rightarrow \infty$ for the case of light-tailed input. We impose the following assumption.

Assumption 5.1. *With*

$$\beta^* := \inf\{\beta : E(e^{-\beta X_1}) < \infty\},$$

assume that $\beta^ < 0$. Let $\varphi(\vartheta) := \log E \exp(-\vartheta X_1)$, and assume that there exists $\vartheta^* \in (\beta^*, 0)$ such that $\varphi(\vartheta^*) + C\vartheta^* = 0$.*

We first recall in Proposition 5.1, below, a result that is a special case of [11, Theorem 4], which states that the tail probabilities of the steady-state workload decay essentially exponentially. Bearing in mind Assumption 3.2, this means that Theorem 3.2 holds when $T_B/B \rightarrow \infty$. In Lemma 5.2, below, we check when Assumption 3.1 applies if $T_B/B \rightarrow 0$ as $B \rightarrow \infty$, so that the case $T_B/B \rightarrow 0$ is then covered by Theorem 3.1.

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 into these regimes.

In the following proposition we consider the logarithmic asymptotics of the stationary workload Q of a Lévy-driven queue; recall that Q is distributed as $\sup_{t>0}(X_t - Ct)$.

Proposition 5.1. *Under Assumption 5.1, it holds that*

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log P(Q > B) = \vartheta^*.$$

Remark 5.1. 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 5.2. Importantly, we obtain the uniform upper bound $P(Q > B) \leq e^{\vartheta^* B}$.

Under the assumption that $\varrho < C$, evidently the queueing system is stable under the measure P . We will now perform a change of measure, with which we associate Q , under which overflow occurs with probability 1, by application of an exponential twist ϑ^* . Under the light-tailed assumption, the Laplace exponent $\varphi(\vartheta)$ of X_t is well defined (for θ in $(\beta^*, 0]$) and characterized through, with $d, \sigma^2 > 0$ and a measure $\Pi_\varphi(\cdot)$ such that $\int_{\mathbb{R}} \min\{1, x^2\} \Pi_\varphi(dx) < \infty$,

$$\varphi(\vartheta) = -\vartheta d + \frac{1}{2} \vartheta^2 \sigma^2 + \int_{\mathbb{R}} (e^{-\vartheta x} - 1 + \vartheta x \mathbf{1}_{\{|x|<1\}}) \Pi_\varphi(dx).$$

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 Q , the Lévy process has Laplace exponent $\bar{\varphi}(\vartheta)$; from the convexity of $\varphi(\cdot)$, it is concluded that (in self-evident notation) $E_Q X_1 = -\bar{\varphi}'(\vartheta^*) > \varrho$, so that the system under the new measure is indeed unstable. (We can check that, under Q , the drift has increased to $d - \vartheta^* \sigma^2$, 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 that we compute $P(\sup_{t>0} X_t - Ct > B)$ by simulating under Q . There is the fundamental equality, with I denoting the indicator function of the event $\{\sup_{t>0} X_t - Ct > B\}$,

$$P\left(\sup_{t>0} X_t - Ct > B\right) = E_Q(LI)$$

(cf. [2, 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 P with respect to Q ; it is a standard result that at time t this likelihood ratio equals $e^{\vartheta^* X_t} e^{\varphi(\vartheta^*)t}$. Let τ_B be defined as the first epoch at which X_t exceeds $B + Ct$ (which is a stopping time); as $I = 1$ with Q -probability 1, we thus obtain

$$P\left(\sup_{t>0} X_t - Ct > B\right) = E_Q e^{\vartheta^* X_{\tau_B}} e^{\varphi(\vartheta^*)\tau_B} = E_Q e^{\vartheta^* X_{\tau_B}} e^{-C\vartheta^* \tau_B}.$$

As, by definition, $X_{\tau_B} \geq B + C\tau_B$, we thus find that $P(Q > B) \leq e^{\vartheta^* B}$.

In the next lemma we relate the decay rate ϑ^* to the large deviations rate function, defined through $I(r) := \sup_{\vartheta \geq 0} (\vartheta r - \varphi(-\vartheta))$, and an associated variational problem.

Lemma 5.1. *It holds that*

$$-\vartheta^* = \inf_{r>C} \frac{I(r)}{r - C}. \tag{5.1}$$

Proof. Let the minimizer on the right-hand side of (5.1) 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 $I(r) = \vartheta(r)r - \varphi(-\vartheta(r))$. Noting 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 ϑ^* 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.

As indicated at the beginning of this section, we would like to check whether Assumption 3.1 is valid if $T_B/B \rightarrow 0$ as $B \rightarrow \infty$. This is dealt with in the following lemma. We recall that it entails that then the only case left to analyze is the linear case, that is, $T_B = RB$ for some $R > 0$.

Lemma 5.2. *Under Assumption 5.1 and given that $\beta^* = -\infty$, Assumption 3.1(i) applies if $T_B/B \rightarrow 0$ as $B \rightarrow \infty$.*

Proof. Let $Q(\vartheta)$ be the probability measure obtained after exponentially twisting the original probability measure P with twist $\vartheta < 0$, as was done in Remark 5.1. In a similar fashion, it follows that

$$P(\text{there exists } t \in (0, T_B) : X_t - Ct > \eta B) \leq E_{Q(\vartheta)} e^{\vartheta(\eta B + C\tau_B)} e^{\varphi(\vartheta)\tau_B},$$

where τ_B is the minimum of T_B and the first epoch at which $X_t - Ct$ exceeds ηB (which is a stopping time). It then follows that, for all $\vartheta < 0$, bearing in mind that $\tau_B \leq T_B = o(B)$,

$$\begin{aligned} \limsup_{B \rightarrow \infty} \frac{1}{B} \log P(\text{there exists } t \in (0, T_B) : X_t - Ct > \eta B) \\ \leq \limsup_{B \rightarrow \infty} \frac{1}{B} \log E_{Q(\vartheta)} (e^{\vartheta(\eta B + C\tau_B)} e^{\varphi(\vartheta)\tau_B}) \\ = \eta\vartheta. \end{aligned}$$

This entails that $P(\text{there exists } t \in (0, T_B) : X_t - Ct > \eta B)$ decays *superexponentially*:

$$\limsup_{B \rightarrow \infty} \frac{1}{B} \log P(\text{there exists } t \in (0, T_B) : X_t - Ct > \eta B) \leq \eta \inf_{\vartheta < 0} \vartheta = -\infty.$$

Combining this with Proposition 5.1 completes the proof.

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 3.1. We define

$$\bar{R} := \max \left\{ \frac{p - q}{C - q}, \frac{q - p}{r^* - C} \right\}.$$

Proposition 5.2. *If $R < \bar{R}$ then*

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B = \max\{p, q\}\vartheta^*.$$

Proof. First suppose that $p > q > 0$. The upper bound follows immediately from Proposition 5.1:

$$\limsup_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B \leq \limsup_{B \rightarrow \infty} \frac{1}{B} \log P(Q > pB) = p\vartheta^*.$$

Now consider the lower bound, which we establish by applying the lower bound of a sample path large deviations result. We here rely on [7, Theorem 5.1], which can be applied to obtain

$$\liminf_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B \geq -\mathbb{I}(f)$$

for any absolute continuous $f \in \mathcal{A}$, where

$$\mathbb{I}(f) := \int_{-\infty}^{\infty} I(f'(\tau)) \, d\tau$$

and the set of paths \mathcal{A} is given by

$$\mathcal{A} := \{f : \text{there exist } (\sigma, \tau) \in \mathcal{E} : -f(-\sigma) \geq C\sigma + p, f(R) - f(R - \tau) \geq C\tau + q\}.$$

Now consider the continuous path f^* through the origin that has slope r^* between $-p/(r^* - C)$ and 0 and slope q elsewhere; clearly,

$$\mathbb{I}(f) := \int_{-p/(r^*-C)}^0 I(f'(\tau)) \, d\tau = \frac{p}{r^* - C} I(r^*) = -p\vartheta^*.$$

The statement 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 $R < (p - q)/(C - q)$,

$$f(R) - f\left(-\frac{p}{r^* - C}\right) = qR + \frac{pr^*}{r^* - C} > C\left(R + \frac{p}{r^* - C}\right) + q.$$

The case in which $q > p > 0$ can be dealt with along the same lines. The upper bound is identical, and in the lower bound we again use Theorem 5.1 of [7], but now with a path f^* that has slope r^* between $R - q/(r^* - C)$ and R and slope q elsewhere. The statement follows after checking that this path is in \mathcal{A} if $R < (q - p)/(r^* - C)$.

In the sequel we use the following lemma extensively; see [9, Lemma 1.2.15].

Lemma 5.3. *For any finite integer M and any array $(\alpha_{i,n})_{i=1,\dots,M;n \in \mathbb{N}}$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{i=1}^M \alpha_{i,n} \right) = \max_{i=1,\dots,M} \left(\lim_{n \rightarrow \infty} \frac{1}{n} \log \alpha_{i,n} \right),$$

provided that the limits on the right-hand side exist.

Proposition 5.3. *If $R > \bar{R}$ then*

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B = p\vartheta^* + \max\{q\vartheta^*, -\psi(R)\}, \quad \text{where } \psi(R) := RI\left(C + \frac{q-p}{R}\right).$$

Proof. First we establish the upper bound, which consists of five steps.

Step I. The probability of interest Π_B can be decomposed as $\Pi_B^{(1)} + \Pi_B^{(2)}$, with

$$\Pi_B^{(1)} := P(Q_0 > pB, Q_{RB} > qB, \text{ for all } t \in (0, RB): Q_t > 0),$$

$$\Pi_B^{(2)} := P(Q_0 > pB, Q_{RB} > qB, \text{ there exists } t \in (0, RB): Q_t = 0).$$

Step II. We first observe that we can bound $\Pi_B^{(2)}$ as follows:

$$\begin{aligned} \Pi_B^{(2)} &= P(Q_0 > pB, \text{ there exists } t \in (0, RB): A(RB - t, RB) - Ct \geq qB) \\ &= P(Q_0 > pB) P(\text{there exists } t \in (0, RB): A(RB - t, RB) - Ct \geq qB) \\ &\leq P(Q_0 > pB) P(\text{there exists } t \geq 0: A(RB - t, RB) - Ct \geq qB) \\ &= P(Q > pB) P(Q > qB), \end{aligned}$$

and, hence,

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B^{(2)} \leq \lim_{B \rightarrow \infty} \frac{1}{B} \log P(Q > pB) + \lim_{B \rightarrow \infty} \frac{1}{B} \log P(Q > qB) = (p + q)\vartheta^*.$$

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,

$$\Pi_B^{(1)} = P(\text{there exists } s \geq 0: A(-s, 0) - Cs > pB, A(-s, RB) - C(RB + s) > qB).$$

Let $\varepsilon > 0$ be picked arbitrarily, and let M be some natural number, whose value we specify later. Then $\Pi_B^{(1)}$ is majorized by

$$\begin{aligned} \sum_{k=0}^{M-1} P(\text{there exists } s \geq 0: A(-s, 0) - Cs \in ((p + k\varepsilon)B, (p + (k + 1)\varepsilon)B]; A(-s, RB) \\ - C(RB + s) > qB) + P(\text{there exists } s \geq 0: A(-s, 0) - Cs > (p + M\varepsilon)B). \end{aligned} \tag{5.2}$$

Now the k th term in the summation of the previous equation is bounded from above by

$$\begin{aligned} P(\text{there exists } s \geq 0: A(-s, 0) - Cs > (p + k\varepsilon)B) \\ \times P(A(0, RB) - CRB > (q - (p + (k + 1)\varepsilon))B), \end{aligned}$$

which we call $\zeta_B^{(k)}$. Owing to Proposition 5.1 and Cramér’s theorem,

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \zeta_B^{(k)} = (p + k\varepsilon)\vartheta^* - RI\left(C + \frac{q - p - (k + 1)\varepsilon}{R}\right).$$

We have now found that (5.2) is not larger than

$$\sum_{k=0}^{M-1} \zeta_B^{(k)} + P(Q > (p + M\varepsilon)B),$$

and, therefore, owing to Lemma 5.3,

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B^{(1)} \leq \max \left\{ \max_{k=0, \dots, M-1} \left\{ (p + k\varepsilon)\vartheta^* - RI \left(C + \frac{q - p - (k + 1)\varepsilon}{R} \right) \right\}, (p + M\varepsilon)\vartheta^* \right\}.$$

Step IV. We now study how $g_k := (p + k\varepsilon)\vartheta^* - RI(\Delta_k/R)$ 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 that $g_1 \leq g_0$ also yields $\max_{k=0, \dots, M-1} g_k = g_0$. To this end, first observe that, owing to the convexity of $I(\cdot)$ and using the fact 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 sufficiently small ε . We conclude, noting that we can take arbitrarily large M , that

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B^{(1)} \leq g_0 = p\vartheta^* - RI \left(C + \frac{q - p - \varepsilon}{R} \right). \tag{5.3}$$

Step V. By letting $\varepsilon \downarrow 0$ in (5.3), applying the upper bound on the decay rates of both $\Pi_B^{(1)}$ and $\Pi_B^{(2)}$, and Lemma 5.3 once more, we have

$$\lim_{B \rightarrow \infty} \frac{1}{B} \log \Pi_B \leq p\vartheta^* + \max \left\{ q\vartheta^*, -RI \left(C + \frac{q - p}{R} \right) \right\}.$$

This completes the upper bound.

The lower bound follows again from sample path large deviations arguments [7].

- Let us first consider the case in which

$$q\vartheta^* > -RI \left(C + \frac{q - p}{R} \right). \tag{5.4}$$

Condition (5.4) implies that $R \geq q/(r^* - C)$, as can be seen as follows. Supposing that $R < q/(r^* - C)$, and recalling that we have $R > (q - p)/(r^* - C)$, it follows that

$$RI \left(C + \frac{q - p}{R} \right) < \frac{q}{r^* - C} I(r) = -q\vartheta^*,$$

which is a contradiction; note that we also used the fact that $C + (q - p)/R > \varrho$.

Using that we know that (5.4) implies that $R \geq q/(r^* - C)$, 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 ϱ 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 (5.4) 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 ϱ elsewhere. It is easily seen that this path is feasible and, by applying the definition of $\mathbb{I}(\cdot)$,

$$\mathbb{I}(f^*) = -p\vartheta^* + RI\left(C + \frac{q - p}{R}\right),$$

as desired.

This concludes the proof of the lower bound.

Lemma 5.4. *For all $R > \bar{R}$, $\psi(R)$ is increasing. In addition, we have $\psi(\bar{R}) \leq -q\vartheta^*$.*

Proof. Observe, recalling that $I'(r) = \vartheta(r)$, that

$$\psi'(R) = -\frac{q - p}{R}\vartheta\left(C + \frac{q - p}{R}\right) + I\left(C + \frac{q - p}{R}\right).$$

First consider the case in which $p > q$, such that $\bar{R} = (p - q)/(C - \varrho)$. It then holds that $C + (q - p)/\bar{R} = \varrho$, so that

$$\psi'(\bar{R}) = -\frac{q - p}{R}I'(\varrho) + I(\varrho) = 0,$$

due to $I(\varrho) = I'(\varrho) = 0$. 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 nonnegative because of the convexity of $I(\cdot)$.

We now consider the case in which $q \geq p$, i.e. $\bar{R} = (q - p)/(r^* - C)$. It then holds that $C + (q - p)/\bar{R} = r^*$, so that

$$\psi'(\bar{R}) = (c - r^*)I'(r^*) + I(r^*) = 0;$$

see the proof of Lemma 5.1. 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 the case in which $p > q$, this equals 0, which is evidently below $-q\vartheta^*$. In the case in which $q \geq p$ 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 5.1. *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 $\psi(R) \leq -q\vartheta^*$; for all $R > \check{R}$, we have $\psi(R) > -q\vartheta^*$.*

Application of Propositions 5.2 and 5.3 and Corollary 5.1 immediately lead to the following theorem.

Theorem 5.1. (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 Π_B , and found three regimes for R . We dealt with this 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. [8], [15], and [19, Section 11.7]. The three regimes have an appealing intuitive explanation.

- For small values of R , the ‘tightest’ of the events $\{Q_0 > pB\}$ and $\{Q_{RB} > qB\}$ will essentially imply the other, thus leading to the decay rate $\max\{p, q\} \vartheta^*$.
- Then there is an intermediate range of values of R for which both $\{Q_0 > pB\}$ and $\{Q_{RB} > qB\}$ are tight, but 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 , both events are still 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 5.1 has made this heuristic rigorous. We finish this section with an example.

Example 5.1. Consider the Brownian case, that is, $\varphi(\vartheta) = -\varrho \vartheta + \frac{1}{2} \vartheta^2$. It is easy to derive that $I(a) = \frac{1}{2}(a - \varrho)^2$ and $\vartheta^* = -2(C - \varrho)$. The solution \check{R} (larger than \bar{R}) of $q \vartheta^* = -\psi(R)$ is

$$\check{R} = \frac{(\sqrt{p} + \sqrt{q})^2}{(C - \varrho)},$$

in line with Proposition 5.1 of [8].

6. Discussion and concluding remarks

In this paper we analyzed the asymptotics of Π_B for large B . We showed that, for T_B increasing sublinearly, the asymptotics reduce to those of the most demanding event; cf. (1.1). We also identified a criterion under which the events become asymptotically independent (‘decoupling’); cf. (1.2). The latter criterion reduces to $T_B/B \rightarrow \infty$ in many situations, a notable exception being the case that $P(Q > B)$ decays polynomially (in which case the condition is $T_B/B^2 \rightarrow \infty$).

While this paper gives a fairly complete picture of all possible regimes, a number of special cases are still open. For instance, when $P(Q > B)$ looks like $\exp(-B^\alpha)$ for some $\alpha \in (0, 1)$, the

abovementioned 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. In the case where $P(Q > B)$ decays like $B^{-\log B}$, an open question is: what do the asymptotics look like for T_B between linear (that is, proportional to B) and proportional to $B^{\sqrt{2}}$? Another topic for future research concerns the exact asymptotics for the light-tailed case and $T_B = RB$.

References

- [1] ASMUSSEN, S. (1998). Subexponential asymptotics for stochastic processes: extremal behavior, stationary distributions and first passage probabilities. *Ann. Appl. Prob.* **8**, 354–374.
- [2] ASMUSSEN, S. (2003). *Applied Probability and Queues*, 2nd edn. Springer, New York.
- [3] BERTOIN, J. AND DONEY, R. A. (1994). Cramér’s estimate for Lévy processes. *Statist. Prob. Lett.* **21**, 363–365.
- [4] BINGHAM, N. H. (1975). Fluctuation theory in continuous time. *Adv. Appl. Prob.* **7**, 705–766.
- [5] BOROVKOV, A. A. (1976). *Stochastic Processes in Queueing Theory*. Springer, New York.
- [6] COHEN, J. W. (1973). Some results on regular variation for distributions in queueing and fluctuation theory. *J. Appl. Prob.* **10**, 343–353.
- [7] DE ACOSTA, A. (1994). Large deviations for vector-valued Lévy processes. *Stoch. Process. Appl.* **51**, 75–115.
- [8] DEBICKI, K., ES-SAGHOUBANI, A. AND MANDJES, M. (2009). Transient characteristics of Gaussian queues. *Queueing Systems* **62**, 383–409.
- [9] DEMBO, A. AND ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd edn. Springer, New York.
- [10] ES-SAGHOUBANI, A. AND MANDJES, M. (2008). On the correlation structure of a Lévy-driven queue. *J. Appl. Prob.* **45**, 940–952.
- [11] GLYNN, P. AND WHITT, W. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. In *Studies in Applied Probability* (J. Appl. Prob. Spec. Vol. **31A**), Applied Probability Trust, Sheffield, pp. 131–156.
- [12] KELLA, O., BOXMA, O. AND MANDJES, M. (2006). A Lévy process reflected at a Poisson age process. *J. Appl. Prob.* **43**, 221–230.
- [13] KLÜPPELBERG, C., KYPRIANOU, A. E. AND MALLER, R. A. (2004). Ruin probabilities and overshoots for general Lévy insurance risk processes. *Ann. Appl. Prob.* **14**, 1766–1801.
- [14] KYPRIANOU, A. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin.
- [15] MANDJES, M. AND RIDDER, A. (2002). A large deviations analysis of the transient of a queue with many Markov fluid inputs: approximations and fast simulation. *ACM Trans. Model Comput. Simul.* **12**, 1–26.
- [16] MIKOSCH, T., RESNICK, S., ROOTZÉN, H. AND STEGEMAN, A. (2002). Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann. Appl. Prob.* **12**, 23–68.
- [17] PORT, S. C. (1989). Stable processes with drift on the line. *Trans. Amer. Math. Soc.* **313**, 805–841.
- [18] SAMORODNITSKY, G. AND TAQQU, M. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, London.
- [19] SHWARTZ, A. AND WEISS, A. (1995). *Large Deviations for Performance Analysis*. Chapman and Hall, London.
- [20] ZOLOTAREV, V. M. (1964). The first-passage time of a level and the behavior at infinity for a class of processes with independent increments. *Theory Prob. Appl.* **9**, 653–662.