Transient characteristics of Gaussian queues

Dębicki, K.; Es-Saghouani, A.; Mandjes, M.

DOI
10.1007/s11134-009-9120-2

Publication date
2009

Document Version
Final published version

Published in
Queueing Systems

Queueing Syst (2009) 62: 383–409
DOI 10.1007/s11134-009-9120-2

Transient characteristics of Gaussian queues

Krzysztof Dębicki · Abdelghafour Es-Saghouani · Michel Mandjes

Received: 11 September 2008 / Revised: 20 April 2009 / Published online: 11 June 2009

Abstract This paper analyzes transient characteristics of Gaussian queues. More specifically, we determine the logarithmic asymptotics of $\mathbb{P}(Q_0 > pB, Q_{TB} > qB)$, where Q_t denotes the workload at time t. For any pair (p, q), three regimes can be distinguished: (A) For small values of T, one of the events $\{Q_0 > pB\}$ and $\{Q_{TB} > qB\}$ will essentially imply the other. (B) Then 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 that none of them essentially implies the other), but that the time epochs 0 and T lie in the same busy period with overwhelming probability. (C) Finally, for large T, still both events are tight, but now they occur in different busy periods with overwhelming probability. For the short-range dependent case, explicit calcula-

K. Dębicki was partially supported by MNiSW Grant No. 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.

Part of this work was done while M. Mandjes was at Stanford University, Stanford, CA 94305, US.

K. Dębicki
Instytut Matematyczny, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland
e-mail: Krzysztof.Debicki@math.uni.wroc.pl

A. Es-Saghouani · M. Mandjes (✉)
Korteweg-de Vries Institute for Mathematics, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
e-mail: mmandjes@science.uva.nl

A. Es-Saghouani
e-mail: aessagho@science.uva.nl

M. Mandjes
CWI, Amsterdam, The Netherlands

M. Mandjes
EURANDOM, Eindhoven, The Netherlands

tions are presented, whereas for the long-range dependent case, structural results are proven.

Keywords Gaussian queues · Large deviations · Transient behavior

Mathematics Subject Classification (2000) 60G15 · 60F10 · 60K25

1 Introduction

Over the past decade a substantial research effort has been devoted to the analysis of queues with Gaussian input [14, 17, 21]. 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 [9, 11, 18, 20]. Importantly, however, so far hardly any attention has been paid to *transient* properties. A notable exception is the recent paper [10], where asymptotics of transient probabilities under a so-called many-sources scaling were found (for specific Gaussian inputs).

In more detail, in [10] 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_t^n denoting the buffer content at time t, the logarithmic asymptotics

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(Q_0^n > np, Q_T^n > nq\big)$$

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 [10] is that the correlation structure of the input process essentially carries over to the workload process.

In the present paper 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 paper is to determine the decay rate

$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log\mathbb{P}(Q_0 > pB, Q_{TB} > qB). \tag{1}$$

Interestingly, in view of earlier work, see, e.g., [15] and [23, Sect. 11.7], multiple regimes are envisaged. 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 (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 $\{Q_0 > pB\}$ 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; we refer to this regime as regime (C). A second goal of the paper is to make the above statements rigorous.

This paper is organized as follows. In Sect. 2 we present the model and give a problem description. Then Sect. 3 introduces additional notation, and we establish a useful reduction property. Our first main result, namely an explicit representation of the decay rate (1), is given in Sect. 4. The cases of short-range dependent and long-range dependent input are dealt with in Sect. 5; in both cases the regimes (A), (B), and (C) are studied.

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 be *centered*, i.e., $\mathbb{E}X(t) = 0$ for any t. Furthermore, the variance function is given through $v(t) := \mathbb{V}\mathrm{ar}(X(t))$.

Throughout the paper we impose the following assumption.

Assumption 2.1 $v(\cdot)$ *is continuous and regularly varying* (*at* ∞) *of index* $\alpha \in (0, 2)$.

In this paper 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:

$$Q_t = \sup_{s\geq 0}\bigl(A(t-s,t) - Cs\bigr), \tag{2}$$

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

As mentioned in the introduction, this paper focuses on analyzing transient properties of the buffer content process, or more specifically, we wish to determine, under Assumption 2.1, the asymptotics of

$$\begin{aligned} N(B) \equiv N_{p,q,T}(B) &:= \mathbb{P}(Q_0 > pB, Q_{TB} > qB) \\ &= \mathbb{P}\bigl(\exists s \geq 0 : A(-s,0) > pB + Cs, \exists t \geq 0 : A(TB-t,TB) > qB + Ct\bigr) \end{aligned}$$

for B large and $p, q, T > 0$ given (the latter identity follows from a direct interpretation of the definition of the supremum in (2)).

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

$$\lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > B) = -\frac{1}{2}\left(\frac{2}{2-\alpha}\right)^{2-\alpha}\left(\frac{2C}{\alpha}\right)^{\alpha}. \tag{3}$$

We refer to, e.g., [4]; studies on the accuracy of the resulting approximations are, e.g., [1, 16].

In Sect. 4 it will turn out that the nature of the decay rate (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 (1) reduce to the one-dimensional asymptotics; in light of (3) this means that for $p > q$ and T small, we have

$$\lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > pB, Q_{TB} > qB) = -\frac{1}{2}\left(\frac{2p}{2-\alpha}\right)^{2-\alpha}\left(\frac{2C}{\alpha}\right)^{\alpha},$$

while for $q > p$, we have the same result but with p replaced by q. We will, for any pair (p, q), show in Sect. 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 Sect. 5 we will explicitly give a threshold above which time 0 and time TB lie in separate busy periods (with overwhelming probability).

3 Notation and preliminaries

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

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

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

Lemma 3.1 *For any $p, q, T > 0$,*

$$\mathbb{P}\big(\exists s \geq 0, t \geq 0 : A(-s,0) - Cs > p, A(T-t,T) - Ct > q\big)$$
$$= \mathbb{P}\big(\exists (s,t) \in \mathcal{E}_T : A(-s,0) - Cs > p, A(T-t,T) - Ct > q\big).$$

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

$$\mathcal{A}_T := \big\{\exists (s,t) \in \mathcal{E}_T : A(-s,0) - Cs > p, A(T-t,T) - Ct > q\big\},$$
$$\mathcal{A} := \big\{\exists (s,t) \in \mathbb{R}_+^2 : A(-s,0) - Cs > p, A(T-t,T) - Ct > q\big\}.$$

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

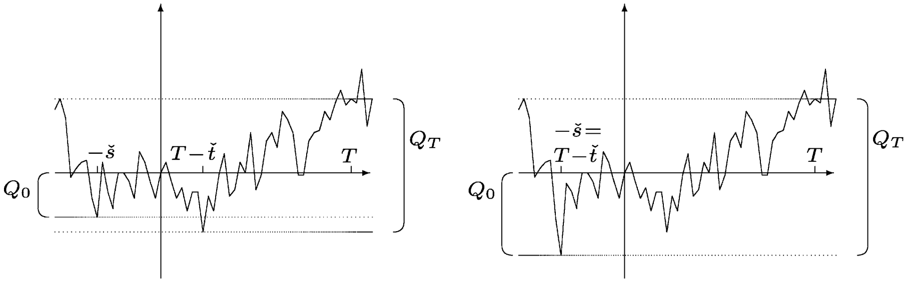

Fig. 1 Proof of Lemma 3.1. In the *left* picture the busy period in which T is contained starts after time 0; in the *right* picture the busy periods in which 0 and T are contained start at the same moment. Here $Q_u := \sup_{v\le u} A(v,u) - C(u-v)$

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

$$\begin{aligned}A(-\check{s},T) - C(T+\check{s}) &= \big(A(-\check{s},0) - C\check{s}\big) + \big(A(0,T) - CT\big)\\ &\geq \big(A(T-t,0) - C(t-T)\big) + \big(A(0,T) - CT\big)\\ &= A(T-t,T) - Ct > q.\end{aligned}$$

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

Remark 3.2 An alternative, more intuitive but essentially equivalent, line of reasoning is the following. Let $\check{t}$ be the optimizer in $\sup_{t\ge 0} A(T-t,T) - Ct$. 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 Fig. 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 nonempty 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 $\check{s}$).

The following corollary is an immediate consequence of Lemma 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 *With* $\mathcal{D}_B := \mathcal{E}_{TB}$,

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

3.2 Notation

In the sequel we extensively use the following Gaussian processes:

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

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

$$\sigma_Y(s) := \sqrt{\mathbb{V}\mathrm{ar} Y_B(s)} = \frac{\sigma(s)}{pB + Cs}; \qquad \sigma_Z(t) := \sqrt{\mathbb{V}\mathrm{ar} Z_B(t)} = \frac{\sigma(t)}{qB + Ct}.$$

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:

$$r(s,t) \equiv r_{B,T}(s,t) = \mathbb{C}\mathrm{orr}\big(Y_B(s), Z_B(t)\big) = \frac{\mathbb{C}\mathrm{ov}(A(-s,0), A(TB-t,TB))}{\sigma(s)\sigma(t)}.$$

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

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

with $I(s,t) := 1_{\{r(s,t)<\gamma(s,t)\}}$. As will appear later on, it turned out to be practical to add the subscript "X" that indicates the underlying Gaussian process (that in turn defines the processes Y_B and Z_B).

4 General results

The following general result can be deduced. It is a generalization of the one-dimensional logarithmic asymptotics of [4] and extension of [22], 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 ∞. 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 4.1 *Assume that* $\{X(t) : t \in \mathbb{R}\}$ *satisfies Assumption* 2.1 *with* $\alpha \in (0, 2)$. *Then for all* $p, q, T > 0$,

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

Notice that the above theorem entails that, under Assumption 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 4.1. We start by establishing a lemma that is also of independent interest.

Lemma 4.2 *For arbitrary* $0 < \underline{\varepsilon} < \overline{\varepsilon} < \infty$,

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

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

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

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

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

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

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

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

Proof The proof of Lemma 4.2 follows straightforwardly from Assumption 2.1, combined with standard properties of regularly varying functions. □

Lemma 4.3 *For each* $0 < \underline{\varepsilon} < \overline{\varepsilon} < \infty$,

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

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

Proof The claim follows from applying Lemma 4.2 to the definition of $\xi_{X;B}(s,t)$. □

Lemma 4.4 *For all* $0 < \underline{\varepsilon} < \overline{\varepsilon} < \infty$,

$$\lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}\big(A(-sB, 0) - CsB > pB; A(TB - tB, TB) - CtB > qB\big)$$
$$= -\xi_{\mathbb{B}_\alpha;1}(s,t)$$

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

Proof Follows from the combination of classical asymptotics of the bivariate Normal random variable, in conjunction with Lemma 4.3; see (3) in [22] and also Example 4.1.9 in [14]. □

Corollary 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 $0 < \underline{\varepsilon} < \overline{\varepsilon}$ (where $\underline{\varepsilon} < T$), we let

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

Lemma 4.5 *There exist* $\overline{\varepsilon} > \underline{\varepsilon} > 0$ *such that*

$$N(B) = \mathbb{P}\big(\exists(s,t) \in \mathcal{C}_B : A(-s,0) - Cs > pB, A(TB-t,TB) - Ct > qB\big) \\ \times \big(1 + o(1)\big)$$

as $B \to \infty$.

Proof In view of Corollary 3.3, it suffices to establish an upper bound. An obvious inequality is

$$\mathbb{P}\big(\exists(s,t) \in \mathcal{D}_B : A(-s,0) - Cs > pB, A(TB-t,TB) - Ct > qB\big) \le \pi_1 + \pi_2,$$

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

$$\pi_1 := \mathbb{P}\big(\exists(s,t) \in \mathcal{C}_B : A(-s,0) - Cs > pB, A(TB-t,TB) - Ct > qB\big);$$
$$\pi_2 := \mathbb{P}\big(\exists(s,t) \in \mathcal{D}_B \setminus \mathcal{C}_B : A(-s,0) - Cs > pB, A(TB-t,TB) - Ct > qB\big).$$

Observe that it suffices to show that $\pi_2 = o(\pi_1)$ as $B \to \infty$. We do so by bounding π_1 from below and π_2 from above as follows.

Let $\overline{\varepsilon} > \underline{\varepsilon} > 0$ be such that $\bar{s} := \alpha p/((2-\alpha)C) \in [\underline{\varepsilon}, \overline{\varepsilon}]$. Then, by virtue of Lemma 4.4, we have

$$\begin{aligned} \log \pi_1 &\ge \log \mathbb{P}\big(A(-\bar{s}B,0) - C\bar{s} > pB;\, A(-\bar{s}B,TB) - C(\bar{s}+T)B > qB\big) \\ &= -\frac{B^2}{v(B)}\xi_{\mathbb{B}_\alpha;1}\big(\bar{s},\bar{s}+T\big)\big(1+o(1)\big) \end{aligned} \tag{4}$$

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

$$\pi_3 \equiv \pi_3(B) := \mathbb{P}\Bigg(\sup_{s \in [0,\underline{\varepsilon}B]} \big(A(-s,0) - Cs\big) > pB\Bigg);$$

$$\pi_4 \equiv \pi_4(B) := \mathbb{P}\Bigg(\sup_{s \in [\overline{\varepsilon}B,\infty)} \big(A(-s,0) - Cs\big) > pB\Bigg).$$

By applying Borell's inequality—see, e.g., Adler [2, Theorem 2.1], or, alternatively, see the remark on p. 147, combined with Theorem 1 of [13, Sect. 12]—we can bound both probabilities from above. Let us first focus on π_3. For $B \to \infty$,

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

this is due to the fact that $v(\cdot)$ is regularly varying and 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 $\zeta \le (2-\alpha)/2$ and B sufficiently large,

$$\log \pi_4 \le -\frac{1}{2} \inf_{s\in[\overline{\varepsilon}B,\infty)} \frac{(Cs+pB)^2}{v(s)}\big(1+o(1)\big)$$
$$= -\frac{1}{2} \inf_{s\in[\overline{\varepsilon},\infty)} (Cs+p)^2 \frac{v(B)}{v(sB)}\frac{B^2}{v(B)}\big(1+o(1)\big)$$
$$\le -\frac{1}{2} \inf_{s\in[\overline{\varepsilon},\infty)} (1-\zeta)\frac{(Cs+p)^2}{s^{\alpha+\zeta}}\frac{B^2}{v(B)}\big(1+o(1)\big).$$

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

$$\inf_{s\in[\overline{\varepsilon},\infty)} \frac{(Cs+p)^2}{s^{\alpha+\zeta}} \to \infty$$

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

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

This completes the proof. □

Before proving Theorem 4.1, we first prove a useful lemma.

Lemma 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

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

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

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

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. □

Proof of Theorem 4.1 In this proof (and in the sequel), we choose $\underline{\varepsilon}$ and $\overline{\varepsilon}$ as indicated in Lemma 4.5. We subsequently prove the lower bound and upper bound.

Lower bound We use the argumentation of [21]. An evident lower bound is

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

Hence, due to Lemma 4.4, we have

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

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

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

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

Upper bound The upper bound is considerably more involved than the lower bound. Due to Lemma 4.5, we have

$$\begin{aligned}N(B) &= \mathbb{P}\big(\exists (s,t) \in \mathcal{C}_B : A(-s,0) - Cs > pB, A(TB-t, TB) - Ct > qB\big) \\ &\quad \times \big(1 + o(1)\big) \\ &= \mathbb{P}\big(\exists (s,t) \in \mathcal{C}_B : Y_B(s) > 1, Z_B(t) > 1\big)\big(1 + o(1)\big);\end{aligned}$$

recall that $\mathcal{C}_B \subseteq \mathcal{D}_B$. In this proof we need the following additional notation:

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

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

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

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

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

We now prove that

$$\mathbb{E}\left(\sup_{(s,t)\in\mathcal{D}_B^{(1)}} \frac{\min\{\sigma_Y(s), \sigma_Z(t)\}}{\theta(s,t) + \beta(s,t)\gamma(s,t)}\left(\frac{\theta(s,t)Y_B(s)}{\sigma_Y(s)} + \frac{\beta(s,t)Z_B(t)}{\sigma_Z(t)}\right)\right) \to 0 \quad (5)$$

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

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

where

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

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 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 $B\to\infty$, and following Lemma 2.2 in [4],

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

as $B\to\infty$. Hence $\psi_1\to 0$ as $B\to\infty$. Analogously, $\psi_2\to 0$ as $B\to\infty$. Hence, we have proved (5).

The fact that (5) applies means that Borell's inequality [2, Theorem 2.1] yields (B large)

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

the last step is due to Lemma 4.6. The latter expression equals, by virtue of Lemma 4.3, as $B\to\infty$,

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

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

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

We conclude that, as $B \to \infty$,

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

This completes the proof. □

Remark 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 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 4.1. First observe that

$$\begin{aligned}
N(B) &= \mathbb{P}\big(\exists s \geq 0 : A(-sB, 0) > pB + CsB, \\
&\qquad\quad \exists t \geq 0 : A(TB - tB, TB) > qB + CtB\big) \\
&= \mathbb{P}\left(\exists s \geq 0 : \frac{A(-sB,0)}{B} > p + Cs, \exists t \geq 0 : \frac{A(TB - tB, TB)}{B} > q + Ct\right) \\
&\overset{\text{(i)}}{=} \mathbb{P}\left(\exists s \geq 0 : \frac{A(-s,0)}{B^{1-\alpha/2}} > p + Cs, \exists t \geq 0 : \frac{A(T-t,T)}{B^{1-\alpha/2}} > q + Ct\right) \\
&= \mathbb{P}\left(\exists s \geq 0 : \frac{A(-s,0)}{p + Cs} > B^{1-\alpha/2}, \exists t \geq 0 : \frac{A(T-t,T)}{q + Ct} > B^{1-\alpha/2}\right),
\end{aligned}$$

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 [3, 14]. 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:

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

where $\mathcal{S}^0_s := \{f \mid -f(-s) > p + Cs\}$ 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:

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

Now let $X(t)$ satisfy Assumption 2.1 with $\alpha \in (1,2)$. Schilder's theorem combined with Theorem 4.1 entails the following result (as $B \to \infty$):

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

Here $\mathbb{I}(f)$ is the rate function of a path f; for a detailed introduction and a formal framework, see, e.g., [1, 3, 17]. The last equality is due to Lemma 3.1. Now consider the evaluation of the inner infimum (for fixed s,t). The key observation is that

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

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 [14, Exercise 4.1.9] can be applied, and three cases are to be distinguished:

- If $r(s,t)\geq\gamma(s,t)$ and $\sigma^2_Y(s)\leq\sigma^2_Z(t)$, then only the first requirement is "tight" and $\xi(s,t)$ is independent of t:

$$\xi(s,t)=\frac{1}{2}\frac{1}{\sigma^2_Y(s)}=\frac{1}{2}\frac{(p+Cs)^2}{v(s)}. \tag{6}$$

- If $r(s,t)\geq\gamma(s,t)$ and $\sigma^2_Y(s)>\sigma^2_Z(t)$, then only the first requirement is "tight" and $\xi(s,t)$ is independent of s:

$$\xi(s,t)=\frac{1}{2}\frac{1}{\sigma^2_Z(t)}=\frac{1}{2}\frac{(q+Ct)^2}{v(t)}. \tag{7}$$

- If $r(s,t)<\gamma(s,t)$, then, with $\Gamma(s,t):=\mathbb{C}\mathrm{ov}(A(-s,0),A(T-t,T))$, both requirements are "tight":

$$\xi(s,t)=\frac{1}{2}(p+Cs,q+Ct)\begin{pmatrix}v(s)&\Gamma(s,t)\\\Gamma(s,t)&v(t)\end{pmatrix}^{-1}\begin{pmatrix}p+Cs\\q+Ct\end{pmatrix}$$
$$=\frac{1}{2}\frac{1}{1-r^2(s,t)}\left(\frac{(p+Cs)^2}{v(s)}-2\frac{\Gamma(s,t)(p+Cs)(q+Ct)}{v(t)v(s)}+\frac{(q+Ct)^2}{v(t)}\right). \tag{8}$$

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

$$\frac{\Gamma(s,t)}{(p+Cs)(q+Ct)}<\min\{\sigma^2_Y(s),\sigma^2_Z(t)\}.$$

We thus retrieve

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

Remark 4.8 It is noted that Theorem 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

$$\mathbb{P}(Q_0 > p_0 B, Q_{T_1} > p_1 B, \ldots, Q_{T_n B} > p_n B)$$

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

5 Special cases

In this section we apply Theorem 4.1 to two special cases, viz.

- Gaussian input processes which possess a *short-range dependent* structure (SRD), by which we mean that $v(\cdot)$ is regularly varying with parameter $\alpha = 1$;
- Gaussian input processes which possess a *long-range dependent* structure (LRD), 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 $X(t) = \int_0^t Z(s)\,\mathrm{d}s$, where $Z(\cdot)$ is a centered stationary Gaussian process with continuous covariance function $R(t) := \mathbb{C}\mathrm{ov}(Z(s), Z(s+t)) > 0$. Note that if $\int_0^\infty R(v)\,\mathrm{d}v < \infty$, then

$$\mathbb{V}\mathrm{ar}\big(X(t)\big) = v(t) = 2\left(\int_0^\infty R(v)\,\mathrm{d}v\right) \cdot t\big(1 + o(1)\big)$$

as $t \to \infty$, and hence $X(\cdot)$ has an SRD structure. If $R(t)$ is regularly varying at ∞ with index $\alpha - 2$ for $\alpha \in (1, 2)$, then $\mathbb{V}\mathrm{ar}(X(t))$ is regularly varying at ∞ with index α, 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., [6, 7]; see also [5, 19]. The use of fractional Brownian motions in modeling input processes has been advocated by, e.g., [21, 24].

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 $\mathbb{V}\mathrm{ar}(X(t)) = v(t)$ is regularly varying at infinity with index $\alpha = 1$.

Proposition 5.1 *Assume that* $\{X(t) : t \in \mathbb{R}\}$ *satisfies Assumption* 2.1 *with* $\alpha = 1$.

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

$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B) = -\begin{cases} 2pC & \text{if } T \le \frac{p-q}{C};\\ 2pC + \frac{(CT+q-p)^2}{2T} & \text{if } \frac{p-q}{C} < T \le \frac{(\sqrt{p}+\sqrt{q})^2}{C};\\ 2pC + 2qC & \text{if } T > \frac{(\sqrt{p}+\sqrt{q})^2}{C}.\end{cases} \tag{9}$$

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

$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B) = -\begin{cases} 2pC + \frac{C^2T}{2} & \text{if } T \le \frac{4p}{C};\\ 4pC & \text{if } T > \frac{4p}{C}.\end{cases} \tag{10}$$

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

$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B) = -\begin{cases} 2qC & \text{if } T \le \frac{q-p}{C};\\ 2pC + \frac{(CT+q-p)^2}{2T} & \text{if } \frac{q-p}{C} < T \le \frac{(\sqrt{p}+\sqrt{q})^2}{C};\\ 2pC + 2qC & \text{if } T > \frac{(\sqrt{p}+\sqrt{q})^2}{C}.\end{cases} \tag{11}$$

Proof By virtue of Theorem 4.1, we analyze

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

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

$$\begin{aligned}\inf_{s\ge 0}\inf_{t\in[0,T)}\xi_{\mathbb{B}_1;1}(s,t) &= \inf_{s\ge 0}\inf_{t\in[0,T)}\frac{1}{2}\left(\frac{(p+Cs)^2}{s}+\frac{(q+Ct)^2}{t}\right)\\ &= 2pC + \frac{1}{2}\frac{(q+C\min\{T,q/C\})^2}{\min\{T,q/C\}}.\end{aligned} \tag{12}$$

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

$$\xi_{\mathbb{B}_1;1}(s,t) \ge \frac{1}{2\min\{\sigma_Y^2(s),\sigma_Z^2(t)\}} \ge \frac{1}{2\sigma_Y^2(s^\star)} = 2pC$$

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

$$\mathcal{S}_1 := \left\{s\ge 0 : \sigma_Y(s) \le \sigma_Z(s+T)\right\}, \qquad \mathcal{S}_2 := \left\{s\ge 0 : \sigma_Y(s) > \sigma_Z(s+T)\right\}.$$

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

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

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

$$\begin{aligned}\inf_{s \geq 0} \xi_{\mathbb{B}_1;1}(s, s+T) &= \inf_{s \geq 0} \frac{1}{2} \frac{(p+Cs)^2}{s} + \frac{1}{2} \frac{(CT+q-p)^2}{T} \\ &= 2pC + \frac{1}{2} \frac{(CT+q-p)^2}{T}.\end{aligned} \tag{13}$$

Finally, in order to complete the proof of (i), it suffices to check that combination of (12) with (13) leads to

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

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

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

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

$$\mathcal{S}_{21} := \left\{s \geq 0 : \sigma_Y(s) > \sigma_Z(s+T), r(s, s+T) < \gamma(s, s+T)\right\},$$
$$\mathcal{S}_{22} := \left\{s \geq 0 : \sigma_Y(s) > \sigma_Z(s+T), r(s, s+T) \geq \gamma(s, s+T)\right\}.$$

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

$$\xi_{\mathbb{B}_1;1}(s, s+T) = \frac{1}{2}\frac{(p+Cs)^2}{s} + \frac{1}{2}\frac{(CT+q-p)^2}{T}. \tag{14}$$

- If $s \in \mathcal{S}_{21}$, then standard calculation leads to the same formula as in (14), i.e.,

$$\begin{aligned}\xi_{\mathbb{B}_1;1}(s, s+T) &= \frac{1}{2}\frac{(q+C(T+s))^2}{T+s} + \frac{1}{2}\frac{(pT+s(p-q))^2}{sT(s+T)} \\ &= \frac{1}{2}\frac{(p+Cs)^2}{s} + \frac{1}{2}\frac{(CT+q-p)^2}{T}. \end{aligned} \tag{15}$$

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

$$\inf_{s\in\mathcal{S}_1\cup\mathcal{S}_{21}} \xi_{\mathbb{B}_1;1}(s, s+T) = 2pC + \frac{1}{2}\frac{(CT+q-p)^2}{T}. \tag{16}$$

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

$$\xi_{\mathbb{B}_1;1}(s, s+T) = \frac{1}{2\min\{\sigma_Y^2(s), \sigma_Z^2(s+T)\}} = \frac{(q+C(s+T))^2}{2(s+T)}. \tag{17}$$

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

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

  We conclude that

$$\inf_{s\in\mathcal{S}_{22}} \xi_{\mathbb{B}_1;1}(s, s+T) = \xi_{\mathbb{B}_1;1}\left(\frac{pT}{q-p}, \frac{pT}{q-p} + T\right) = \frac{1}{2}\frac{q(CT+q-p)^2}{(q-p)T}. \tag{18}$$

The comparison of (16) with (18) now implies that

$$\inf_{s\geq 0} \xi_{\mathbb{B}_1;1}(s, s+T) = 2pC + \frac{1}{2}\frac{(CT+q-p)^2}{T}. \tag{19}$$

Analogously to the proof of (i), the combination of (12) with (19) completes the proof. □

Remark 5.2 Related results for queues fed by Brownian motion have recently been obtained in [12]. There also emphasis was put on the nature of the decay rates and the shape of the *most likely path* towards the rare event [1, 14]. In accordance with Proposition 5.1, it was found that for T up to some threshold, the decay rate of the joint probability equals the decay rate of $\mathbb{P}(Q > \max\{p, q\}B)$, with Q 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 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

$$\big\{Q_0 > pB, A(0, TB) \geq qB + CT - pB\big\};$$

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.

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 given $\alpha \in (1, 2)$ and $p, q, C > 0$,

$$s^\star := \arg\max_{s\geq 0}\left\{\frac{s^{\alpha/2}}{p+Cs}\right\} = \frac{p}{C}\frac{\alpha}{2-\alpha},$$

$$t^\star := \arg\max_{t\geq 0}\left\{\frac{t^{\alpha/2}}{q+Ct}\right\} = \frac{q}{C}\frac{\alpha}{2-\alpha},$$

and

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

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

$$\max_{s\geq 0} \mathbb{V}\mathrm{ar}\big(Y_1(s)\big) = \mathbb{V}\mathrm{ar}\big(Y_1(s^\star)\big) = \frac{1}{2R(p)},$$

$$\max_{t\geq 0} \mathbb{V}\mathrm{ar}\big(Z_1(t)\big) = \mathbb{V}\mathrm{ar}\big(Z_1(t^\star)\big) = \frac{1}{2R(q)}.$$

The following general bounds hold. The upper bound in (20) 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 (20) says that the joint probability is larger than the product of the individual probabilities (which makes sense in view of the positive correlation).

Proposition 5.3 *Assume that* $\{X(t) : t \in \mathbb{R}\}$ *satisfies Assumption* 2.1 *with* $\alpha \in (1, 2)$. *Then*

$$-\max\big\{R(p), R(q)\big\} \geq \lim_{B\to\infty} \frac{v(B)}{B^2} \log N(B) > -\big(R(p) + R(q)\big). \tag{20}$$

Proof The upper bound follows immediately from

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

The lower bound is due to the fact that, due to Lemma 4.5, for some $\overline{\varepsilon} > \underline{\varepsilon} > 0$,

$$\begin{aligned}
&\inf_{s\geq 0}\inf_{t\in[0,T)\cup\{T+s\}} \xi_{\mathbb{B}_\alpha;1}(s,t)\\
&\quad = \min_{s\in[\underline{\varepsilon},\overline{\varepsilon}]}\min_{t\in[\underline{\varepsilon},T)\cup\{T+s\}} \xi_{\mathbb{B}_\alpha;1}(s,t)\\
&\quad = \min_{s\in[\underline{\varepsilon},\overline{\varepsilon}]}\min_{t\in[\underline{\varepsilon},T)\cup\{T+s\}} \frac{1}{2}\frac{1}{\min\{\sigma_Y^2(s), \sigma_Z^2(t)\}}\left(1 + \frac{(\gamma(s,t) - r(s,t))^2}{1 - r^2(s,t)} I(s,t)\right).
\end{aligned} \tag{21}$$

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 all $s, t > 0$, and therefore (21) is majorized by

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

This completes the proof. □

In the following we determine the values of T for which the lower bound in (20) is tight.

Proposition 5.4 *Assume that* $\{X(t):t\in\mathbb{R}\}$ *satisfies Assumption* 2.1 *with* $\alpha\in(1,2)$.

(i) *If* $p>q>0$, *then there exists a unique* $T^\star$ *solving the equation*

$$\gamma(s^\star,s^\star+T^\star)=r(s^\star,s^\star+T^\star)\tag{22}$$

such that

$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B)=-R(p)\quad\text{for } T\le T^\star;$$
$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B)<-R(p)\quad\text{for } T>T^\star.$$

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

$$\gamma(t^\star-T_\star,t^\star)=r(t^\star-T_\star,t^\star)\tag{23}$$

such that

$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B)=-R(q)\quad\text{for } T\le T_\star;$$
$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B)<-R(q)\quad\text{for } T>T_\star.$$

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

$$\lim_{B\to\infty}\frac{v(B)}{B^2}\log N(B)=-R(p),$$

we need the following two conditions to be satisfied:

$$\gamma(s^\star,s^\star+T)\le r(s^\star,s^\star+T),\tag{24}$$
$$\sigma_Y(s^\star)\le\sigma_Z(s^\star+T).\tag{25}$$

Under (25) we have

$$r(s^\star, s^\star+T) = \frac{(T+s^\star)^\alpha - T^\alpha + (s^\star)^\alpha}{2(s^\star(s^\star+T))^{\alpha/2}}$$
$$= \frac{1}{2}\left(\frac{s^\star}{s^\star+T}\right)^{\alpha/2}\left(\left(\frac{T+s^\star}{s^\star}\right)^\alpha - \left(\frac{T}{s^\star}\right)^\alpha + 1\right);$$
$$\gamma(s^\star, s^\star+T) = \left(\frac{s^\star}{s^\star+T}\right)^{\alpha/2}\frac{q+Cs^\star+CT}{p+Cs^\star}$$
$$= \left(\frac{s^\star}{s^\star+T}\right)^{\alpha/2}\left(1+\frac{q-p}{p+Cs^\star}+\frac{Cs^\star}{p+Cs^\star}\frac{T}{s^\star}\right).$$

Noticing that

$$\frac{Cs^\star}{p+Cs^\star} = \alpha/2; \qquad -1 < \frac{q-p}{p+Cs^\star} = \frac{q-p}{p}\left(1-\frac{\alpha}{2}\right) < 0,$$

inequalities (24) and (25) are equivalent to respectively

$$1+2\frac{q-p}{p}(1-\alpha/2)+\alpha\frac{T}{s^\star} \le \left(1+\frac{T}{s^\star}\right)^\alpha - \left(\frac{T}{s^\star}\right)^\alpha, \tag{26}$$
$$1+\frac{q-p}{p}(1-\alpha/2)+\frac{\alpha}{2}\frac{T}{s^\star} \le \left(1+\frac{T}{s^\star}\right)^{\alpha/2}. \tag{27}$$

Interestingly, however, we have that inequality (26) implies inequality (27). This can be shown as follows. First rewrite inequality (26) as

$$1+\frac{q-p}{p}(1-\alpha/2)+\frac{\alpha}{2}\frac{T}{s^\star} \le \frac{1}{2}\left(1+\left(1+\frac{T}{s^\star}\right)^\alpha - \left(\frac{T}{s^\star}\right)^\alpha\right). \tag{28}$$

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

$$\frac{\mathbb{C}\mathrm{ov}(\check{A}(0,s^\star), \check{A}(0,s^\star+T))}{\mathbb{V}\mathrm{ar}(\check{A}(0,s^\star))} = \frac{1}{2}\left(1+\left(1+\frac{T}{s^\star}\right)^\alpha - \left(\frac{T}{s^\star}\right)^\alpha\right);$$
$$\sqrt{\frac{\mathbb{V}\mathrm{ar}(\check{A}(0,s^\star+T))}{\mathbb{V}\mathrm{ar}(\check{A}(0,s^\star))}} = \left(1+\frac{T}{s^\star}\right)^{\alpha/2}.$$

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

$$0 < \frac{1}{2}\left(1+\left(1+\frac{T}{s^\star}\right)^\alpha - \left(\frac{T}{s^\star}\right)^\alpha\right)\Big/\left(1+\frac{T}{s^\star}\right)^{\alpha/2}$$
$$= \frac{\mathbb{C}\mathrm{ov}(\check{A}(0,s^\star), \check{A}(0,s^\star+T))}{\sqrt{\mathbb{V}\mathrm{ar}(\check{A}(0,s^\star+T))\mathbb{V}\mathrm{ar}(\check{A}(0,s^\star))}} = \mathbb{C}\mathrm{orr}\big(\check{A}(0,s^\star), \check{A}(0,s^\star+T)\big) < 1.$$

Hence the right-hand side of inequality (28) is smaller then the right-hand side of inequality (27), and we indeed have that inequality (26) implies inequality (27).

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 $x^\star > 0$. 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 $x^\star > 0$. We have thus found that there exists a unique $T^\star \geq 0$ such that for all $T \leq T^\star$, we have that inequality (24) 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. □

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

Proposition 5.5 *(i) If $p > q > 0$, then $T^\star \geq (p-q)/C$. (ii) If $q > p > 0$, then $T_\star \geq (q-p)/C$.*

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 (24).

First notice that (under the notation used in the proof of Proposition 5.4)

$$g\left(\frac{p-q}{Cs^\star}\right) = 1 + 2\frac{q-p}{p+Cs^\star} + \alpha\frac{p-q}{Cs^\star} = 1,$$

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

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

This proves the claim in part (i). □

Remark 5.6 Conditions $T < T^\star$ and $T < T_\star$ have interesting interpretations. Consider, for instance, $T < T^\star$. Elementary computations with the conditional distribution of Normal random variables yield that $T < T^\star$ is equivalent to

$$\mathbb{E}\big(A(0,T) \mid A(-s^\star, 0) = p + Cs^\star\big) \geq q - p + CT.$$

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 $T < T_\star$.

Proposition 5.4 says that, just as in the SRD case, if and only if T is smaller than some threshold, 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: $T^\star$ (in case $p > q$) or $T_\star$ (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, 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 5.8. It says that for T large enough, we are in regime (C).

Lemma 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 result. □

Due to Proposition 5.3, for $\alpha \in (1, 2)$, we have

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

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 5.8 *For*

$$T > T^\sharp := \left(\frac{2\xi^\star}{C^2}\right)^{1/(2-\alpha)},$$

we have that

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

Remark 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 $-\max\{R(p), R(q)\}$; the value of this threshold can be determined as in Remark 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

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

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.

6 Discussion and concluding remarks

Exact asymptotics This paper analyzed the logarithmic asymptotics of $\mathbb{P}(Q_0 > pB, Q_{TB} > qB)$. 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

$$\mathbb{P}(Q_0 > pB, Q_{TB} > qB) \cdot \varphi(B) \to 1$$

as $B \to \infty$? It is noted that for the single-dimensional case, this was already a highly nontrivial task [11, 18, 20], 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 $T^\star$ for $p > q$ and $T_\star$ 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 $T^\star$ ($T_\star$, 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 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 2.1.

In view of the results in [8], it is anticipated that if $T_B/B \to 0$ as $B \to \infty$, then

$$\begin{aligned}
&\lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > pB, Q_{T_B} > qB) \\
&\quad = \min\left\{ \lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > pB), \lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > qB) \right\} \\
&\quad = -\max\big\{R(p), R(q)\big\};
\end{aligned}$$

if $T_B/B \to \infty$ as $B \to \infty$, then

$$\begin{aligned}
&\lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > pB, Q_{T_B} > qB) \\
&\quad = \lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > pB) + \lim_{B\to\infty} \frac{v(B)}{B^2} \log \mathbb{P}(Q_0 > qB) \\
&\quad = -\big(R(p) + R(q)\big),
\end{aligned}$$

using the function $R(\cdot)$ introduced in Sect. 5.2.

References

1. Addie, R., Mannersalo, P., Norros, I.: Most probable paths and performance formulae for buffers with Gaussian input traffic. Eur. Trans. Telecommun. **13**, 183–196 (2002)
2. Adler, R.: An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Inst. Math. Statist. Lecture Notes-Monograph Series, vol. 12. Inst. Math. Statist., Hayward (1990)
3. Bahadur, R., Zabell, S.: Large deviations of the sample mean in general vector spaces. Ann. Probab. **7**, 587–621 (1979)
4. Dębicki, K.: A note on LDP for supremum of Gaussian processes over infinite horizon. Stat. Probab. Lett. **44**, 211–219 (1999)
5. Dębicki, K., Mandjes, M.: Traffic with an fBm limit: Convergence of the stationary workload process. Queueing Syst. **46**, 113–127 (2004)
6. Dębicki, K., Palmowski, Z.: On-off fluid models in heavy traffic environment. Queueing Syst. **33**, 327–338 (1999)
7. Dębicki, K., Rolski, T.: A Gaussian fluid model. Queueing Syst. **20**, 433–452 (1995)
8. Dębicki, K., Es-Saghouani, A., Mandjes, M.: Transient asymptotics of Lévy-driven queues. CWI Report PNA-E0809
9. Duffield, N., O'Connell, N.: Large deviations and overflow probabilities for the general single-server queue, with applications. Math. Proc. Camb. Philos. Soc. **118**, 363–374 (1995)
10. Es-Saghouani, A., Mandjes, M.: On the correlation structure of Gaussian queues. CWI Report PNA-R0706 (2007)
11. Hüsler, J., Piterbarg, V.: Extremes of a certain class of Gaussian processes. Stoch. Proc. Appl. **83**, 257–271 (1999)
12. Lieshout, P., Mandjes, M.: Transient analysis of Brownian queues. CWI Report PNA-R0705 (2007)
13. Lifshits, M.: Gaussian Random Functions. Kluwer, Dordrecht (1995)
14. Mandjes, M.: Large Deviations for Gaussian Queues. Wiley, Chichester (2007)

15. Mandjes, M., Ridder, A.: 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 (2002)
16. Mandjes, M., van de Meent, R.: Resource dimensioning through buffer sampling. IEEE/ACM Trans. Netw. (2009, to appear). CWI Report PNA-E0610
17. Mannersalo, P., Norros, I.: A most probable path approach to queueing systems with general Gaussian input. Comput. Netw. **40**, 399–412 (2002)
18. Massoulié, L., Simonian, A.: Large buffer asymptotics for the queue with FBM input. J. Appl. Probab. **36**, 894–906 (1999)
19. Mikosch, T., Resnick, S., Rootzén, H., Stegeman, A.: Is network traffic approximated by stable Lévy motion or fractional Brownian motion? Ann. Appl. Probab. **12**, 23–68 (2002)
20. Narayan, O.: Exact asymptotic queue length distribution for fractional Brownian traffic. Adv. Perf. Anal. **1**, 39–63 (1998)
21. Norros, I.: A storage model with selfsimilar input. Queueing Syst. **16**, 387–396 (1994)
22. Piterbarg, V.I., Stamatović, B.: Crude asymptotics of the probability of simultaneous high extrema of two Gaussian processes: the dual action functional. Russ. Math. Surv. **60**, 167–168 (2005)
23. Shwartz, A., Weiss, A.: Large Deviations for Performance Analysis. Chapman & Hall, New York (1995)
24. Taqqu, M., Willinger, W., Sherman, R.: Proof of a fundamental result in self-similar traffic modeling. Comput. Commun. Rev. **27**, 5–23 (1997)