Heavy-traffic single-server queues and the transform method

Available online at www.sciencedirect.com

ScienceDirect

Indagationes Mathematicae 34 (2023) 1014–1037

indagationes mathematicae

www.elsevier.com/locate/indag

Special Issue dedicated to the memory of J.W. Cohen

Heavy-traffic single-server queues and the transform method

M.A.A. Boon[a,*], A.J.E.M. Janssen[a], J.S.H. van Leeuwaarden[b]

[a] *Department of Mathematics and Computer Science, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands*

[b] *Department of Econometrics and Operations Research, Tilburg University, 5000 LE Tilburg, The Netherlands*

Abstract

Heavy-traffic limit theory is concerned with queues that operate close to criticality and face severe queueing times. Let W denote the steady-state waiting time in the GI/G/1 queue. Kingman (1961) showed that W, when appropriately scaled, converges in distribution to an exponential random variable as the system's load approaches 1. The original proof of this famous result uses the transform method. Starting from the Laplace transform of the pdf of W (Pollaczek's contour integral representation), Kingman showed convergence of transforms and hence weak convergence of the involved random variables. We apply and extend this transform method to obtain convergence of moments with error assessment. We also demonstrate how the transform method can be applied to so-called nearly deterministic queues in a Kingman-type and a Gaussian heavy-traffic regime. We demonstrate numerically the accuracy of the various heavy-traffic approximations.

Keywords: Queueing theory; Transform method; G/G/1 queue

1. Introduction and results

The title of this contribution to the memorial issue for J.W. Cohen refers to the The Single Server Queue, the monumental book [12] in which J.W. Cohen teaches the reader how to use complex analysis and transform methods to obtain rigorous results for the general GI/G/1 queue and its many extensions. In turn, J.W. Cohen admired the work of F. Pollaczek, in particular for the analytic treatment of queues by means of complex function theory and integral equations [13], techniques that also feature prominently in The Single Server Queue, and in this paper on the GI/G/1 queue in heavy traffic.

* Corresponding author.

E-mail addresses: m.a.a.boon@tue.nl (M.A.A. Boon), a.j.e.m.janssen@tue.nl (A.J.E.M. Janssen), J.S.H.vanLeeuwaarden@tilburguniversity.edu (J.S.H. van Leeuwaarden).

https://doi.org/10.1016/j.indag.2023.02.004

F. Pollaczek initiated the analysis of the GI/G/1 queue in the 1940s and 1950s, and obtained a contour integral representation for the Laplace transform of the steady-state waiting time. J.F.C. Kingman introduced heavy-traffic analysis in the 1960s [20,21]. For the GI/G/1 queue in a regime where the system load tends to 1, Kingman showed that, under certain conditions, the Laplace transform of the pdf of an appropriately normalized steady-state waiting time (Pollaczek's contour integral representation) converges to the Laplace transform of an exponential distribution. We will refer to this technique–to show convergence in distribution by convergence of transforms–as the transform method. To explain Kingman's result in more detail, let W denote the steady-state waiting time in the GI/G/1 queue, which solves the stochastic equation

$$W \overset{d}{=} \Big(W + V - \frac{1}{\rho}\,U\Big)^{+}, \tag{1}$$

with $x^{+} = \max\{0, x\}$. Here, V is the generic service time with mean 1 and variance $\sigma_V^2 \in (0, \infty)$, U is the generic inter-arrival time with mean 1 and variance $\sigma_U^2 \in (0, \infty)$ and $\rho \in (0, 1)$. It is assumed that V and U are independent. Since convergence of transforms implies convergence of distributions, Kingman effectively showed using the transform method that, for $W = W_\alpha$ solving (1) with $\alpha = 1 - \rho$,

$$P(\alpha\, W_\alpha \le t) \to 1 - e^{-2t/\sigma^2}, \qquad \alpha \downarrow 0, \tag{2}$$

for all $t \ge 0$ with $\sigma^2 = \sigma_U^2 + \sigma_V^2$.

Kingman's observation that heavily loaded systems admit a simple scaling limit triggered a surge of research in the 1960s and 1970s; see [6,8,16,22,23,25,29], among others. In the decades that followed, heavy-traffic analysis, and more generally, stochastic-process limits, developed into popular topics in the applied probability community, with queueing theory as one of its many applications. The general idea remained to consider a parametrized set of systems and to identify the limiting system as the parameter converges to a limiting value yielding criticality (e.g. $\alpha \downarrow 0$ as in (2)).

Kingman's transform method is thus largely based on Pollaczek's contour integral that we introduce next, see [1]. Assume analyticity of $\psi(s) = \mathbb{E}\,[\exp(s(V - \frac{1}{\rho}\,U))]$ in an open strip containing $|\mathrm{Re}(s)| \le \delta$ for some $\delta > 0$; in particular, all moments of $V - \frac{1}{\rho}\,U$ are finite. Then, Pollaczek's integral reads

$$\mathbb{E}\,[e^{-sW}] = \exp\Big\{\frac{-1}{2\pi i}\int_C \frac{s}{z(s-z)}\,\log(1-\psi(-z))dz\Big\}, \tag{3}$$

where C is a contour to the left of, and parallel to, the imaginary axis, and to the right of the singularities of $\log(1-\psi(-z))$ in the left half-plane, and s is any complex number to the right of C. Kingman uses (3) to prove that $\mathbb{E}\,[\exp(-s\alpha W)] \to (1+\sigma^2 s/2)^{-1}$ in $\mathrm{Re}(s) \ge 0$ as $\alpha \downarrow 0$, yielding (2).

Being tailor-made for the steady-state GI/G/1 queue, the transform method that uses contour integral representations has rarely been applied to other queueing models. A notable exception is the heavy-traffic analysis of O.J. Boxma and J.W. Cohen [7] for the GI/G/1 queue with heavy-tailed distributions, so when the second moment of the service time and/or interarrival time is infinite. Boxma and Cohen apply the transform method to an extended form of Pollaczek's integral (3) to identify the heavy-traffic limit. Other studies that apply this transform method for heavy-traffic analysis are [22,23] on the GI/G/s queue and [4,5] on the fixed-cycle traffic-light queue, a variation of the GI/G/1 queue.

More probabilistic methods for proving heavy-traffic results developed later use functional limit theorems, and typically establish that the sample path of a scaled waiting time converges to some limiting stochastic process. One is then faced with the problem of showing that the steady-state of the limiting process corresponds to the limiting steady-state of the queueing model in heavy traffic. This requires an interchange-of limits argument, which is often challenging as it involves proving tightness of sequences of probability measures. The transform method works directly with the steady-state random variables, and thus avoids the problem of interchanging limits (cf. [9,15]).

1.1. Nearly deterministic queues and the transform method

Next to the classical heavy-traffic setting, we will consider so-called nearly deterministic $\mathrm{GI}_n/\mathrm{G}_n/1$ queues, whose heavy-traffic behavior has been investigated in [27,28] using sample-path methods. Nearly-deterministic queues are motivated by cycling thinning. To explain this, denote for all $n = 1, 2, \ldots$

$$W_n \stackrel{d}{=} \Big(W_n + V_n - \frac{1}{\rho_n}\, U_n\Big)^+, \tag{4}$$

with

$$V_n = \frac{1}{n} \sum_{k=1}^{n} V_{n,k}, \qquad U_n = \frac{1}{n} \sum_{k=1}^{n} U_{n,k}, \tag{5}$$

where $V_{n,k}$ are i.i.d. copies of V and $U_{n,k}$ are i.i.d. copies of U, with V and U as before, and $\rho_n \in (0, 1)$. The cyclic thinning thus regards each interarrival (service) time as the nth occurrence in a sequence of i.i.d. random variables, which mitigates the random fluctuations. For instance, if this sequence consists of exponential random variables, the interarrival or service time would follow an Erlang distribution, and if n is large, this becomes 'increasingly deterministic'.

Interesting heavy-traffic regimes now arise when $\rho_n \to 1$ as $n \to \infty$. In [27,28], two heavy-traffic regimes are considered. The first (Kingman-type) regime assumes that $(1 - \rho_n)n \to \beta$ as $n \to \infty$ for some fixed $\beta > 0$. In this case, W_n converges in distribution to an exponential random variable with mean $\sigma^2/2\beta$, where $\sigma^2 = \sigma_U^2 + \sigma_V^2$. The second (Gaussian) regime assumes $(1 - \rho_n)\sqrt{n} \to \beta$ as $n \to \infty$ for some fixed $\beta > 0$. In this case $\sqrt{n}\, W_n/\sigma$ converges in distribution to the all-time maximum M_β of a one-dimensional directed random walk, starting at 0, with normally distributed step sizes of mean $-\beta$. This Gaussian random walk and M_β are well studied, see [2,10,17,18,26].

1.2. Main results

In the present paper we apply the transform method for heavy-traffic analysis of the GI/G/1 and $\mathrm{GI}_n/\mathrm{G}_n/1$ queues. We first consider the classical heavy-traffic regime, and provide a detailed proof of a version of Kingman's original result using the transform method. We do this with service times V and interarrival times U that do not depend on $\rho = 1 - \alpha$ (in Kingman's original result, a controlled dependence of V and U on α is allowed). In this more restricted setting, we show the following.

Theorem 1. *With $\sigma_\alpha^2 = (\sigma_V^2 + \rho^{-2}\sigma_U^2)\rho$,*

$$\mathbb{E}\,[e^{-\alpha s W}] = (1+\sigma_\alpha^2\, s/2)^{-1} + O(\alpha\,\log(1/\alpha)), \qquad \alpha \downarrow 0, \tag{6}$$

uniformly in any bounded set of s with $\mathrm{Re}(s) \geq -1/2\sigma_\alpha^2$.

As a consequence of Theorem 1, we have in terms of the moments of W for $k = 1, 2, \ldots$

$$\mathbb{E}\,[(\alpha W)^k] = k!(\tfrac{1}{2}\sigma_\alpha^2)^k + O(\alpha\log(1/\alpha)), \qquad \alpha \downarrow 0, \tag{7}$$

where the leading term in (7) is the kth moment of an exponentially distributed random variable. We observe that for $k = 1$ in (7) we get

$$\mathbb{E}\,[\alpha W] = \tfrac{1}{2}\sigma_\alpha^2 + O(\alpha\log(1/\alpha)), \qquad \alpha \downarrow 0. \tag{8}$$

It turns out that, after appropriate identifications, the quantity $\frac{1}{2}\sigma_\alpha^2$ at the right-hand side of (8) coincides with the right-hand side of (6) in [11].

We next apply the transform method to the $\mathrm{GI}_n/\mathrm{G}_n/1$ queues described in Section 1.1; in the sequel we use for two positive sequences a_n, b_n, $n = 1, 2, \ldots,$ the notation $a_n \asymp b_n$ to indicate that there are β_1, β_2 with $0 < \beta_1 \leq \beta_2 < \infty$ such that $a_n/b_n \in [\beta_1, \beta_2]$.

Theorem 2. *Assume that $(1-\rho_n) \asymp 1/n$. With $\gamma_n = (\sigma_V^2 + \rho_n^{-2}\sigma_U^2)\rho_n/(2n(1-\rho_n))$,*

$$\mathbb{E}\,[e^{-tW_n}] = (1+\gamma_n t)^{-1} + O\Big(\frac{\log n}{\sqrt{n}}\Big), \qquad n \to \infty, \tag{9}$$

uniformly in any bounded set of t with $\mathrm{Re}(t) \geq -1/4\gamma_n$.

From Theorem 2 we obtain for any $k = 1, 2, \ldots$

$$\mathbb{E}\,[W_n^k] = k!\,\gamma_n^k + O\Big(\frac{\log n}{\sqrt{n}}\Big), \qquad n \to \infty. \tag{10}$$

We then proceed to apply the transform method to the $\mathrm{GI}_n/\mathrm{G}_n/1$ when we let $(1-\rho_n) \asymp 1/\sqrt{n}$. In this regime, the integration contour C occurring in (3) can be chosen to pass through the saddle point $z = \zeta_{sp}$ of $h(z) = \log(\psi(-z))$ on the negative real axis, allowing a saddle point analysis (under an additional assumption). We show the following, with M_β as in Section 1.1.

Theorem 3. *Assume that $(1-\rho_n) \asymp 1/\sqrt{n}$, and let $\sigma_n = (h''(\zeta_{sp}))^{1/2}$ and $\beta_n = -\zeta_{sp}\sigma_n\sqrt{n}$. Then $\sigma_n \asymp 1$, $\beta_n \asymp 1$ as $n \to \infty$, and*

$$\mathbb{E}\Big[\exp\Big(-s\,\frac{\sqrt{n}}{\sigma_n}\,W_n\Big)\Big] = \mathbb{E}\,[\exp(-sM_{\beta_n})] + O\Big(\frac{1}{\sqrt{n}}\Big), \qquad n \to \infty, \tag{11}$$

uniformly in any bounded set of s with $\mathrm{Re}(s) \geq -\frac{1}{2}\,\beta_n$.

From Theorem 3 we get for any $k = 1, 2, \ldots$

$$\mathbb{E}\Big[\Big(\frac{\sqrt{n}}{\sigma_n}\,W_n\Big)^k\Big] = m_k(\beta_n) + O\Big(\frac{1}{\sqrt{n}}\Big), \qquad n \to \infty, \tag{12}$$

where $m_k(\beta) = \mathbb{E}\,[M_\beta^k]$. Theorem 3 can be refined by taking account of $h'''(\zeta_{sp})$ in the saddle point analysis. This yields Theorem 4, see Section 5 for its precise formulation, where the right-hand side of (11) is replaced by $\mathbb{E}\,[\exp(-R_n(s)\,M_{B_n})] + O(1/n)$, with appropriate non-linear transforms $R_n(s)$ and $B_n(\beta) = B_n$ of s and β. Theorem 3 and its refinement Theorem 4 result from an adaptation of the saddle point method developed in [19,24]. Theorem 5 in Section 5 gives a consequence of Theorem 4 on the level of moments.

Theorems 1–4 generalize and refine some classical heavy-traffic results. Theorem 1 recovers Kingman's weak convergence result (2), and generalizes this to a heavy-traffic limit theorem for all moments of W. The refined heavy-traffic approximations not only provide the heavy-traffic limit, but also contain pre-limit information (for ρ away from 1) and shed light on the rate of convergence (the speed at which the limit is attained, as a function of the scaling parameter). Similarly, Theorems 2–4 recover results in [27,28] for convergence in distribution, and convergence of the first two moments. In the paper we show that all normalized moments of W_n converge to those of M_β, again with rate of convergence and refinements. As a consequence, the theoretical results in Theorems 1–4 give sharp approximations, not only in heavy traffic, but also in more moderate conditions.

We demonstrate the accuracy of the approximations by comparing with exact results. We also address the computational aspects of numerically calculating complex contour integrals, which is required for both the exact and approximate performance analysis. In particular, we explain how to calculate reliably the Pollaczek contour integrals with numerical integration. Since we operate in heavy-traffic regimes, numerical integration can become cumbersome, with integration contours closely passing the origin, but we show how to deal with this.

1.3. Organization of the paper

In Section 2 we present assumptions and preliminaries on the function $\psi(-z) = \mathbb{E}\,[\exp(-z(V - \frac{1}{\rho}U))]$ that occurs in the various Pollaczek integrals in Sections 3–5. Section 2 also contains information on the function $h(z) = \log\psi(-z)$ that is heavily used in the special saddle point method of Section 5; we avoid using saddle points in Sections Section 3, 4 on the Kingman-type results. In Section 3 we present the formulation and proof of our version of Kingman's classical result (Theorem 1), yielding convergence (with error assessment) of the mgf and all moments of those of an exponentially distributed random variable with a tailored α-dependent mean. In Section 4, we consider the nearly deterministic queue in the regime $(1-\rho) \asymp 1/n$, and prove that the mgf and all moments of W_n converge to those of a specifically designed exponentially distributed random variable (Theorem 2). In Section 5 we present Theorem 3 and its refinement Theorem 4, with consequences for moment convergence, for the nearly deterministic queue when $(1-\rho) \asymp 1/\sqrt{n}$. This requires an additional assumption on the function ψ, allowing a saddle point approach to Pollaczek's integral around the saddle point ζ_{sp}, at the expense of an exponentially small error. The proof of the refinement Theorem 4, and its consequence (62) for moment convergence, is rather involved and technical, so that we have deferred details to Appendices A and B. In Section 6 we summarize in detail the computational schemes for the quantities we want to calculate via Pollaczek's formula (3). In Section 7 we present our conclusions.

2. Preliminaries

The convergence results of the Kingman type given in the present paper will be shown under the condition that the function ψ is analytic in an open set containing a strip $|\mathrm{Re}(z)| \le \delta$ with $\delta > 0$. For the convergence results for nearly deterministic queues related to the maximum of the Gaussian random walk, we use a saddle point method, see [14,19,24]. This method requires an assumption that guarantees one can confine attention to the immediate vicinity of the saddle point on the negative real axis when conducting an asymptotic analysis on the relevant Pollaczek integral when $n \to \infty$. We refer to Section 5 for the technical details.

We use in the sequel the letter δ to denote a generic positive number that may take case-dependent values. We consider random variables $V, U \geq 0$ that are independent with $\mathbb{E}[V] = 1 = \mathbb{E}[U]$ and $0 < \sigma_V^2 + \sigma_U^2 < \infty$. We furthermore assume that there is a $\delta > 0$ such that $\mathbb{E}[\exp(zV)]$, $\mathbb{E}[\exp(zU)]$ are analytic in an open strip containing $-\delta \leq \mathrm{Re}(z) \leq \delta$. For $\rho \in (0, 1)$, we let

$$\psi(-\zeta) = \psi(-\zeta\,;\,\rho) = \mathbb{E}[e^{-\zeta(V-\frac{1}{\rho}U)}]. \tag{13}$$

Then $\psi(-\zeta)$ is analytic in an open strip containing $-\delta \leq \mathrm{Re}(\zeta) \leq \delta$ for some $\delta > 0$. Since

$$\psi(-\zeta) \equiv \int_{-\infty}^{\infty} e^{-\zeta t}\, d\,G(t), \tag{14}$$

with $G(t)$ the cumulative distribution function of $V - \frac{1}{\rho}U$, we have that $\psi(-\zeta)$, $-\delta \leq \zeta \leq \delta$, is logarithmically convex. We have, uniformly in $\rho \in [\frac{1}{2}, 1]$,

$$\begin{aligned}\psi(-\zeta) &= \mathbb{E}\Big[1 - \zeta\Big(V - \frac{1}{\rho}U\Big) + \frac{1}{2}\zeta^2\Big(V - \frac{1}{\rho}U\Big)^2\Big] + O(\zeta^3)\\ &= 1 + \Big(\frac{1}{\rho} - 1\Big)\zeta + \frac{1}{2}\Big(\sigma_V^2 + \frac{1}{\rho^2}\sigma_U^2 + \Big(1 - \frac{1}{\rho}\Big)^2\Big)\zeta^2 + O(\zeta^3), \quad |\zeta| \leq \delta,\end{aligned} \tag{15}$$

for some $\delta > 0$. Therefore, there is a $\delta > 0$ such that $\psi(-\zeta) \geq 1/2$ when $\rho \in [\frac{1}{2}, 1]$ and $-\delta \leq \zeta \leq 0$. Hence, by continuity, there is a $\delta > 0$ such that for $\rho \in [\frac{1}{2}, 1]$

$$h(\zeta) = \log\psi(-\zeta) = \log(\mathbb{E}[e^{-\zeta(V-\frac{1}{\rho}U)}]) \tag{16}$$

is well-defined and analytic as a principal value logarithm in an open set containing the rectangle $-\delta \leq \mathrm{Re}(\zeta) \leq 0$, $|\mathrm{Im}(\zeta)| \leq \delta$.

We have for $\rho \in [\frac{1}{2}, 1]$

$$h(0) = 0, \qquad h'(0) = \frac{1}{\rho} - 1, \qquad h''(0) = \sigma_V^2 + \frac{1}{\rho^2}\sigma_U^2, \tag{17}$$

and there is a $\delta > 0$ such that

$$h(\zeta) = \Big(\frac{1}{\rho} - 1\Big)\zeta + \frac{1}{2}\Big(\sigma_V^2 + \frac{1}{\rho^2}\sigma_U^2\Big)\zeta^2 + O(\zeta^3) \tag{18}$$

in an open set containing the rectangle $-\delta \leq \mathrm{Re}(\zeta) \leq 0$, $|\mathrm{Im}(\zeta)| \leq \delta$. There is a $\delta > 0$ such that the function $h(\zeta)$, $-\delta \leq \zeta \leq 0$, is convex. Furthermore, h has, for ρ sufficiently close to 1, a unique saddle point $\zeta_{sp} \in [-\delta, 0]$. We have

$$\zeta_{sp} = -\frac{1}{\rho}\,\frac{1-\rho}{\sigma_V^2 + \frac{1}{\rho^2}\sigma_U^2} + O((1-\rho)^2), \tag{19}$$

and

$$h(\zeta_{sp}) = -\frac{1}{\rho^2}\,\frac{(1-\rho)^2}{2(\sigma_V^2 + \frac{1}{\rho^2}\sigma_U^2)} + O((1-\rho)^3), \qquad h''(\zeta_{sp}) = \sigma_V^2 + \frac{1}{\rho^2}\sigma_U^2 + O(1-\rho). \tag{20}$$

3. Classical heavy-traffic result of the Kingman type

In this section we prove Theorem 1, Kingman's classical result that the scaled waiting time converges to an exponentially distributed random variable. Novel results include error assessments and statements about the rate of convergence.

With V and U as in Section 2, we let

$$W \stackrel{d}{=} \Big(W + V - \frac{1}{\rho}\, U\Big)^{+}, \tag{21}$$

where $\rho = 1 - \alpha$ and $\alpha \downarrow 0$. We shall show that

$$\log(\mathbb{E}\,[e^{-\alpha s W}]) = -\log(1 + \tfrac{1}{2}\,\sigma_\alpha^2 s) + O(\alpha\,\log(1/\alpha)), \qquad \alpha \downarrow 0, \tag{22}$$

uniformly in any bounded set of s with $\mathrm{Re}(s) \geq \frac{1}{2}\,\mu_0$, where

$$\sigma_\alpha^2 = \frac{-1}{\mu_0} = \Big(\sigma_V^2 + \frac{1}{\rho^2}\,\sigma_U^2\Big)\,\rho. \tag{23}$$

The Laplace–Stieltjes transform $\mathbb{E}\,[\exp(-t\ \ \mathrm{Exp}(\theta))]$ of an exponentially distributed random variable $\mathrm{Exp}(\theta)$ with density $\vartheta\, e^{-\vartheta x}$, $x \geq 0$ and mean $1/\vartheta$ is given by $(1+t/\vartheta)^{-1}$, $\mathrm{Re}(t) > -\vartheta$. Hence, the moments of αW equal the moments of an exponentially distributed random variable with mean $2/\sigma_\alpha^2$, up to an error $O(\alpha\,\log(1/\alpha))$ as $\alpha \downarrow 0$.

We show this result by using the Pollaczek result for W, in which we follow the argumentation as given by Kingman in [20]. Observe that V and U are independent of α, which allows us to be explicit about the error term in (22).

From Pollaczek's result, we have

$$\log(\mathbb{E}\,[e^{-\alpha s W}]) = \frac{-1}{2\pi i}\int_C \frac{\alpha s}{\zeta(\alpha s - \zeta)}\,\log(1 - \psi(-\zeta))d\zeta, \tag{24}$$

where $\psi(-\zeta)$ is as in (13), and C is a line parallel to, and to the left, of the imaginary axis, and to the right of the singularities of $\log(1 - \psi(-\zeta))$ in the open left-half-plane, and αs in (24) lies to the right of C. To be more detailed about the choice of C, we observe from (15) that there is a $\delta > 0$ such that

$$\psi(-\zeta) = 1 + \alpha\zeta/\rho + \tfrac{1}{2}\,\sigma_\alpha^2\zeta^2/\rho + O(\alpha^2\zeta^2) + O(\zeta^3), \qquad |\zeta| \leq \delta, \tag{25}$$

where σ_α^2 is as in (23). The leading part $1 + \alpha\zeta/\rho + \sigma_\alpha^2\zeta^2/2\rho$, considered for $\zeta \leq 0$, in (25) equals 1 for $\zeta = 0$ and $\zeta = 2\alpha\mu_0$, and is minimal for $\zeta = \alpha\mu_0$, where $\mu_0 < 0$ is as in (23), with minimal value $1 - \alpha^2/(2\rho\sigma_\alpha^2)$. By (14) we have

$$|\psi(-\alpha\mu_0 - i\eta)| \leq \psi(-\alpha\mu_0) = 1 - \frac{\alpha^2}{2\rho\sigma_\alpha^2} + O(\alpha^3), \qquad \eta \in \mathbb{R}, \tag{26}$$

and so

$$|\psi(-\alpha\mu_0 - i\eta)| \leq 1 - \frac{\alpha^2}{4\rho\sigma_\alpha^2}, \qquad \eta \in \mathbb{R}, \tag{27}$$

when α is sufficiently small. For such α, we can therefore choose $C = \{\alpha\mu_0 + i\eta \mid \eta \in \mathbb{R}\}$, with principal value of the log in (24), and $\mathrm{Re}(s) \geq \frac{1}{2}\,\mu_0$.

In (24) we substitute $\zeta = \alpha z$, with $z \in \{\mu_0 + i\eta \mid \eta \in \mathbb{R}\} = C_0$ and $d\zeta = \alpha\, dz$, to get

$$\log(\mathbb{E}\,[e^{-\alpha s W}]) = \frac{-1}{2\pi i}\int_{C_0} \frac{s}{z(s - z)}\,\log(1 - \psi(-\alpha z))\,dz. \tag{28}$$

From (25) we have when $|\alpha z| \leq \delta$,

$$\psi(-\alpha z) = 1 + \alpha^2 z/\rho + \sigma_\alpha^2\alpha^2 z^2/2\rho + O(\alpha^4 z^2) + O(\alpha^3 z^3), \tag{29}$$

and so

$$\frac{1 - \psi(-\alpha z)}{-\alpha^2 z/\rho} = 1 + \tfrac{1}{2}\,\sigma_\alpha^2 z + O(\alpha^2 z) + O(\alpha z^2). \tag{30}$$

We have from (27) that both $1-\psi(-\alpha z)$ and $-\alpha z^2/\rho$ lie in the open right half-plane when $z \in C_0$ and α is sufficiently small. Hence, with principal-value logarithms,

$$\log(1-\psi(-\alpha z)) = \log\Big(\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho}\Big) + \log(-\alpha^2 z/\rho), \qquad z \in C_0. \tag{31}$$

Since $\mathrm{Re}(s) \geq \frac{1}{2}\mu_0$, we have by Cauchy's theorem

$$\frac{-1}{2\pi i}\int_{C_0} \frac{s}{z(s-z)} \log(-\alpha^2 z/\rho)\,dz = 0, \tag{32}$$

and so

$$\log(\mathbb{E}\,[e^{-\alpha s W}]) = \frac{-1}{2\pi i}\int_{C_0} \frac{s}{z(s-z)} \log\Big(\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho}\Big)\,dz. \tag{33}$$

We shall show now that

$$\log\Big(\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho}\Big) = \log(1+\tfrac{1}{2}\sigma_\alpha^2 z) + O(\alpha^2) + O(\alpha z) \tag{34}$$

when $|\alpha z| \leq c$ and α and c are sufficiently small. We have from (30)

$$\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho} = (1+\tfrac{1}{2}\sigma_\alpha^2 z)\Big(1+\frac{O(\alpha^2 z)+O(\alpha z^2)}{1+\frac{1}{2}\sigma_\alpha^2 z}\Big). \tag{35}$$

Now when $z=\mu_0+i\eta \in C_0$ with $\eta \in \mathbb{R}$, we have by (23)

$$\begin{aligned} |1+\tfrac{1}{2}\sigma_\alpha^2 z|^2 &= |\tfrac{1}{2}+\tfrac{1}{2}i\sigma_\alpha^2\eta|^2 = \tfrac{1}{4}+\tfrac{1}{4}\sigma_\alpha^4\eta^2 \\ &= \tfrac{1}{4}\sigma_\alpha^4(\mu_0^2+\eta^2) = \tfrac{1}{4}\sigma_\alpha^4\,|z|^2. \end{aligned} \tag{36}$$

Hence

$$\frac{O(\alpha^2 z)+O(\alpha z^2)}{1+\frac{1}{2}\sigma_\alpha^2 z} = O(\alpha^2)+O(\alpha z), \tag{37}$$

and this has modulus $\leq\ 1/2$ when α is sufficiently small and $|\alpha z| \leq c$ with c determined by the implicit constants in the O's in (30). This gives (34).

To finish the proof of (22), we split the integral in (33) into the parts $|z| \leq c/\alpha$ and $|z| \geq c/\alpha$. We have by (34)

$$\begin{aligned} \log(\mathbb{E}\,[e^{-\alpha s W}]) &= \frac{-1}{2\pi i}\int_{\substack{z\in C_0,\\ |z|\leq c/\alpha}} \frac{s}{z(s-z)} \log(1+\tfrac{1}{2}\sigma_\alpha^2 z)\,dz \\ &+ \int_{\substack{z\in C_0,\\ |z|\geq c/\alpha}} \frac{s}{z(s-z)}(O(\alpha^2)+O(\alpha z))\,dz - \frac{1}{2\pi i}\int_{\substack{z\in C_0,\\ |z|\leq c/\alpha}} \frac{s}{z(s-z)} \log\Big(\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho}\Big)\,dz. \end{aligned} \tag{38}$$

For the first integral on the second line of (38), we have

$$\begin{aligned} &\Big|\int_{\substack{z\in C_0,\\ |z|\leq c/\alpha}} \frac{s}{z(s-z)}(O(\alpha^2)+O(\alpha z))\,dz\Big| \\ &= O(\alpha^2) + \Big(\int_{\substack{z\in C_0,\\ |z|\leq c/\alpha}} \frac{\alpha}{|z|}\,|dz|\Big) = O(\alpha^2)+O(\alpha\,\log(1/\alpha)), \end{aligned} \tag{39}$$

uniformly in any bounded set of s with $\mathrm{Re}(s) \geq \frac{1}{2}\mu_0$. For the second integral on the second line of (38), we use (27), so that

$$\frac{1}{4\sigma_\alpha^2|z|} \leq \Big|\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho}\Big| \leq \frac{2\rho}{\alpha^2|z|}, \qquad z \in C_0. \tag{40}$$

Therefore, for $z \in C_0$ and $|z| \geq c/\alpha$,

$$-\log|z| + \log\Big(\frac{1}{4\sigma_\alpha^2}\Big) \leq \log\Big|\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho}\Big| \leq \log\Big(\frac{2\rho}{c\alpha}\Big). \tag{41}$$

Using that $(1-\psi(-\alpha z))/(-\alpha^2 z/\rho)$ lies in the cut plane $\mathbb{C}\backslash(-\infty, 0]$, so that its argument is between $-\pi$ and π, we get

$$\Big|\log\Big(\frac{1-\psi(-\alpha z)}{-\alpha^2 z/\rho}\Big)\Big| = O(\log|z|) + O(\log(1/\alpha)) + O(1). \tag{42}$$

From (42) it follows that the second integral on the second line of (38) is $O(\alpha\,\log(1/\alpha))$, uniformly in any bounded set of s with $\mathrm{Re}(s) \geq \frac{1}{2}\,\mu_0$.

We conclude that

$$\log(\mathbb{E}\,[e^{-\alpha s W}]) = \frac{-1}{2\pi i}\int_{\substack{z\,\in\,C_0,\\ |z|\,\leq\,c/\alpha}} \frac{s}{z(s-z)}\,\log(1+\tfrac{1}{2}\,\sigma_\alpha^2 z)\,dz + O(\alpha\,\log(1/\alpha)), \tag{43}$$

uniformly in any bounded set of s with $\mathrm{Re}(s) \geq \frac{1}{2}\,\mu_0$. We extend the integration range in the integral in (43) to all $z \in C_0$, at the expense of an error $O(\alpha\,\log(1/\alpha))$ uniformly in any bounded set of s with $\mathrm{Re}(s) \geq \frac{1}{2}\,\mu_0$. Finally,

$$\frac{-1}{2\pi i}\int_{C_0} \frac{s}{z(s-z)}\,\log(1+\tfrac{1}{2}\,\sigma_\alpha^2 z)\,dz = -\log(1+\tfrac{1}{2}\,\sigma_\alpha^2 s), \qquad \mathrm{Re}(s) \geq \tfrac{1}{2}\,\mu_0, \tag{44}$$

by Cauchy's theorem, and we get (22).

4. Kingman-type result for nearly deterministic queues

In nearly deterministic regimes (see [27,28]), random fluctuations in the interarrival and service times are reduced using the cyclic thinning principle discussed earlier. In this section, we consider the Kingman-type regime, where $(1-\rho_n)n \to \beta$ as $n \to \infty$ for some fixed $\beta > 0$.

In more detail, we consider

$$W_n \stackrel{d}{=} \Big(W_n + V_n - \frac{1}{\rho_n}\,U_n\Big)^+, \tag{45}$$

with

$$V_n = \frac{1}{n}\sum_{k=1}^{n} V_{n,k}, \qquad U_n = \frac{1}{n}\sum_{k=1}^{n} U_{n,k}, \tag{46}$$

where $0 < \rho_n < 1$, $V_{n,k}$ are i.i.d. copies of V and $U_{n,k}$ are i.i.d. copies of U, with V and U as in Section 2, and $V_{n,k}$ and $U_{n,k}$ are independent. We shall show that, when $(1-\rho_n) \asymp 1/n$ (so that there are fixed β_1, β_2 with $0 < \beta_1 \leq \beta_2 < \infty$ such that $(1-\rho_n)n \in [\beta_1, \beta_2]$ for $n = 1, 2, \ldots$),

$$\log(\mathbb{E}\,[e^{-tW_n}]) = -\log(1+\gamma_n t) + O\Big(\frac{\log n}{\sqrt{n}}\Big), \tag{47}$$

uniformly in any bounded set of t with $\mathrm{Re}(t) \geq \frac{1}{2}\,x_0$. Here

$$\gamma_n = \frac{\sigma_V^2 + \rho^{-2}\sigma_U^2}{2n(1-\rho_n)}\,\rho_n, \qquad x_0 = \frac{-1}{2\gamma_n}, \tag{48}$$

with γ_n bounded away from 0 and ∞.

We use again Pollaczek's result, so that

$$\log(\mathbb{E}\,[e^{-tW_n}]) = \frac{-1}{2\pi i}\int_C \frac{t}{z(t-z)}\,\log(1-\varphi(-z))\,dz, \tag{49}$$

where for $-n\delta \le \mathrm{Re}(z) \le n\delta$ (with δ as in the second paragraph of Section 2)

$$\varphi(-z) = \varphi_n(-z) = \mathbb{E}\,[e^{-z(V_n-\frac{1}{\rho_n}U_n)}] = (\mathbb{E}\,[e^{-\frac{z}{n}(V-\frac{1}{\rho_n}U)}])^n = \psi^n(-z/n), \tag{50}$$

with ψ as in (13) with $\rho=\rho_n$, and C is a line parallel to, and to the left of, the imaginary axis, and to the right of the singularities of $\log(1-\varphi(-z))$ in the open left half-plane, and lies to the right of C.

Assume that $1-\rho_n \asymp 1/n$, and delete n from ρ_n and γ_n below for conciseness. We use (15) with $\zeta = z/n$ and $z = O(\sqrt{n})$, so that

$$\psi(-z/n) = 1 + \Big(\frac{1}{\rho}-1\Big)\frac{z}{n} + \Big(\frac{1}{\rho}-1\Big)\gamma\,\frac{z^2}{n} + O\Big(\frac{z^2}{n^4}\Big) + O\Big(\frac{z^3}{n^3}\Big). \tag{51}$$

With $\varphi(-z)$ as in (50) and using the expansion $(1+X)^n = 1+nX+\frac{1}{2}n^2X^2+O(n^3X^3)$, valid for $X=O(\frac{1}{n})$, we get

$$\varphi(-z) = 1 + \Big(\frac{1}{\rho}-1\Big)z + \Big(\frac{1}{\rho}-1\Big)\gamma z^2 + O\Big(\frac{|z|^2+|z|^3+|z|^4}{n^2}\Big). \tag{52}$$

In the O in (52) terms like z^6/n^4, which is $O(z^2/n^2)$ when $z=O(\sqrt{n})$, have been collected. The leading part of the right-hand side of (52), considered for $z\le 0$, equals 1 for $z=0$ and $z=-1/\gamma$, and is minimal for $z=-1/2\gamma = x_0$, with minimum value $1-(1-\rho)/4\rho\gamma$. Therefore, for large n,

$$|\varphi(-x_0-iy)| \le \varphi(-x_0) \le 1-(1-\rho)/8\gamma, \qquad y\in\mathbb{R}. \tag{53}$$

Hence, we can use $C=\{x_0+iy \mid y\in\mathbb{R}\}$ in (49), so that $1-\varphi(-z)$ has positive real part when $z\in C$ and n is large, with principal-value logarithm for the log in the integral.

We have for $z\in C$, $z=O(\sqrt{n})$ from (52) and $(1-\rho_n)\asymp 1/n$,

$$\frac{1-\varphi(-z)}{-(\frac{1}{\rho}-1)\,z} = 1+\gamma z + O\Big(\frac{|z|+|z|^2+|z|^3}{n}\Big). \tag{54}$$

We are now in a quite similar position as in Section 3 from (30) onwards. In particular, using $|1+\gamma z| = \gamma|z|$ for $z\in C$, we have

$$\begin{aligned}\log\Big(\frac{1-\varphi(-z)}{-(\frac{1}{\rho}-1)\,z}\Big) &= \log(1+\gamma z) + O\Big(\frac{|z|+|z|^2+|z|^3}{n(1+\gamma z)}\Big)\\ &= \log(1+\gamma z) + O\Big(\frac{1+|z|+|z|^2}{n}\Big),\end{aligned} \tag{55}$$

when $|z|\le c\sqrt{n}$ and $c>0$ is small enough to ensure that the O-terms in (55) do not exceed 1/2. Furthermore,

$$\begin{aligned}&\int_{\substack{z\in C,\\ |z|\le c\sqrt{n}}} \Big|\frac{t}{z(t-z)}\Big|\,\frac{1+|z|+|z|^2}{n}\,|dz|\\ &= O\Big(\int_{\substack{z\in C,\\ |z|\le c\sqrt{n}}} \frac{1+|z|+|z|^2}{n\,|z|^2}\,|dz|\Big) = O\Big(\frac{1}{\sqrt{n}}\Big),\end{aligned} \tag{56}$$

uniformly in any bounded set of t with $\mathrm{Re}(t) \geq \frac{1}{2}x_0$. Proceeding then as in Section 3 from (38) onwards, with (56) as substitute for (39) and $\sqrt{n}$ instead of $1/\alpha$, we get (47).

5. Gaussian regime for nearly deterministic queues

We consider the same setting as in Section 4, but now we assume that $(1-\rho_n) \asymp 1/\sqrt{n}$, so that there are fixed β_1, β_2 with $0 < \beta_1 \leq \beta_2 < \infty$ such that $(1-\rho_n)\sqrt{n} \in [\beta_1, \beta_2]$ for $n = 1, 2, \ldots$. The precise formulation of our results requires quantities derived from $h(\zeta) = \log(\psi(-\zeta))$ at the saddle point ζ_{sp} of h on the negative real axis, and an additional condition discussed below.

We shall show the following (which proves Theorem 3). Let

$$\sigma_n = (h''(\zeta_{sp}))^{1/2}, \qquad \beta_n = -\zeta_{sp}\,\sigma_n\sqrt{n}. \tag{57}$$

Then we have $\sigma_n \asymp 1$, $\beta_n \asymp 1$ as $n \to \infty$, and

$$\log\Big(\mathbb{E}\Big[\exp\Big(-s\,\frac{\sqrt{n}}{\sigma_n}\,W_n\Big)\Big]\Big) = \log(\mathbb{E}\,[e^{-sM_{\beta_n}}]) + O\Big(\frac{1}{\sqrt{n}}\Big), \qquad n \to \infty, \tag{58}$$

uniformly in any bounded set of s with $\mathrm{Re}(s) \geq -\frac{1}{2}\beta_n$. We shall also show the following refinement of Theorem 3.

Theorem 4. *Let σ_n and β_n as in (57), and let*

$$d_2 = -\frac{h'''(\zeta_{sp})}{6h''(\zeta_{sp})}. \tag{59}$$

Then $d_2 = O(1)$, $n = 1, 2, \ldots$, and, uniformly in any bounded set of s with $\mathrm{Re}(s) \geq -\frac{1}{2}\beta_n$*,*

$$\log\Big(\mathbb{E}\Big[\exp\Big(-s\,\frac{\sqrt{n}}{\sigma_n}\,W_n\Big)\Big]\Big) = \log(\mathbb{E}\,[e^{-R_n M_{B_n}}]) + O\Big(\frac{1}{n}\Big), \qquad n \to \infty, \tag{60}$$

where

$$B_n = \frac{\beta_n}{1+\beta_n\varphi_n}, \qquad R_n = R_n(s) = \frac{s}{(1+\beta_n\varphi_n)(1+(s+\beta_n)\varphi_n)}, \tag{61}$$

with $\varphi_n = d_2/(\sigma_n\sqrt{n}) = O(1/\sqrt{n})$.

From (58) the moments of $\frac{\sqrt{n}}{\sigma_n}W_n$ are approximated with error $O(1/\sqrt{n})$ by the moments $m_k(\beta_n)$ of M_{β_n} as in (12).

As a consequence of Theorem 4, we have the following result.

Theorem 5. *For $k = 1, 2, \ldots$*

$$\begin{aligned}\mathbb{E}\Big[\Big(\frac{\sqrt{n}}{\sigma_n}\,W_n\Big)^k\Big] &= (-1)^k\Big(\frac{d}{ds}\Big)^k(\mathbb{E}\,[e^{-R_n(s)M_{B_n}}])\Big|_{s=0} + O\Big(\frac{1}{n}\Big)\\ &= \frac{m_k(B_n)}{(1+\beta_n\varphi_n)^{2k}} + \frac{k(k-1)m_{k-1}(B_n)\varphi_n}{(1+\beta_n\varphi_n)^{2k-1}} + O\Big(\frac{1}{n}\Big),\end{aligned} \tag{62}$$

with φ_n as in Theorem 4 and $m_k(B_n) = \mathbb{E}\,[M_{B_n}^k]$.

We shall prove Theorem 3, and we present the proofs of Theorems 4 and 5 in Appendices A and B. For all these proofs, we use again Pollaczek's formula (49) in which we substitute

$\zeta = z/n$. Thus, we have

$$\log(\mathbb{E}\,[e^{-tW_n}]) = \frac{-1}{2\pi i}\int_{C_n} \frac{t/n}{\zeta(t/n-\zeta)}\,\log(1-\psi^n(-\zeta))\,d\zeta, \tag{63}$$

where $C_n = \frac{1}{n}\,C$ is a line parallel to, and to the left of, the imaginary axis, and to the right of the singularities of $\log(1-\psi^n(-\zeta))$, and t/n lies to the right of C_n. We now need the following assumption, because it allows us to integrate in (63) over ζ with $|\mathrm{Im}(\zeta)| \le \delta$, at the expense of an error of exponential decay as $n \to \infty$.

Assumption 6. There is a $\delta > 0$, $\varepsilon > 0$ such that for all $\rho \in [\frac{1}{2}, 1)$ and $x \in [-\delta, 0)$ and $y \in \mathbb{R}$, $|y| > \delta$, we have $|\psi(-x+iy)| < \psi(-x) - \varepsilon$.

With reference to Section 2, we can assume that $\delta > 0$ is such that $h(\zeta) = \log\psi(-\zeta)$ is analytic in the rectangle $-\delta \le \mathrm{Re}(\zeta) \le 0$, $|\mathrm{Im}(\zeta)| \le \delta$, and, by taking n sufficiently large with $(1-\rho) = (1-\rho_n) \asymp 1/\sqrt{n}$ in (19), that the saddle point ζ_{sp} of h lies in this rectangle. We then have, with exponentially small error as $n \to \infty$,

$$\log(\mathbb{E}\,[e^{-tW_n}]) = \frac{-1}{2\pi i}\int_{\zeta_{sp}-i\delta}^{\zeta_{sp}+i\delta} \frac{t/n}{\zeta(t/n-\zeta)}\,\log(1-e^{nh(\zeta)})\,d\zeta \tag{64}$$

when $\mathrm{Re}(t/n) \ge \frac{1}{2}\,\zeta_{sp}$.

We have from (18)–(20) and $(1-\rho_n) \asymp 1/\sqrt{n}$ that

$$\zeta_{sp} \asymp \frac{1}{\sqrt{n}}, \qquad h(\zeta_{sp}) \asymp \frac{1}{n}, \qquad \sigma_n^2 = h''(\zeta_{sp}) \asymp 1, \tag{65}$$

and this shows that $\sigma_n \asymp 1$, $\beta_n \asymp 1$ and also that $d_2 = O(1)$, see (57) and (58).

For both Theorems 3 and 4, we shall work from the integral in (64) towards the integral representation of the Laplace–Stieltjes transform $\mathbb{E}\,[\exp(-s\,M_\beta)]$ of the maximum M_β of the Gaussian random walk with drift $-\beta$. For the latter we have (from Pollaczek's formula, applied to $W = (W+V-U)^+$ with V and U having pdf's $\frac{1}{\sqrt{2\pi}}\,\exp(-\frac{1}{2}\,(x-A)^2)\,\chi_{[0,\infty)}(x)$ and $\delta_\beta(x-A)$, respectively, and letting $A \to \infty$)

$$\log(\mathbb{E}\,[e^{-sM_\beta}]) = \frac{-1}{2\pi i}\int_C \frac{s}{z(s-z)}\,\log(1-e^{\beta z+\frac{1}{2}z^2})\,dz, \tag{66}$$

where C is a line parallel to, and to the left of, the imaginary axis, and $s \in C$ is to the right of C. Substituting $z = -\beta + iy$, $-\infty < y < \infty$, we get for $s \in \mathbb{C}$, $\mathrm{Re}(s) > -\beta$

$$\log(\mathbb{E}\,[e^{-sM_\beta}]) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{s}{(\beta-iy)(s+\beta-iy)}\,\log(1-e^{-\frac{1}{2}\beta^2-\frac{1}{2}y^2})\,dy. \tag{67}$$

We make in the integral in (64) a substitution that brings $\exp(nh(\zeta))$ in Gaussian form. We thus let $\zeta = \zeta(v)$ be the solution of the equation

$$h(\zeta(v)) = h(\zeta_{sp}) - \tfrac{1}{2}\,v^2\,h''(\zeta_{sp}) \tag{68}$$

that satisfies $\zeta(v) = \zeta_{sp} + iv + O(v^2)$ as $v \to 0$. By Lagrange's theorem, there is an $r > 0$, independent of n, such that

$$\zeta(v) = \zeta_{sp} + iv + \sum_{l=2}^{\infty} d_l (iv)^l, \qquad |v| \le r, \tag{69}$$

with real d_l, and d_2 given by (59), see [19], end of Section 3 for more details about such a substitution. With the substitution $\zeta = \zeta(v)$, $-r \le v \le r$, in (64), we get with exponentially small error

$$\log(\mathbb{E}\,[e^{-tW_n}]) = \frac{-1}{2\pi i}\int_{-r}^{r} \frac{\zeta'(v)\,t/n}{\zeta(v)(t/n-\zeta(v))}\,\log(1-e^{nh(\zeta_{sp})-\frac{1}{2}v^2h''(\zeta_{sp})})\,dv \tag{70}$$

when $n\to\infty$ and $\mathrm{Re}(t)\geq \frac{1}{2}n\zeta_{sp}$. We write for conciseness $\sigma=\sigma_n$ in the sequel, and we take $t=s\sqrt{n}/\sigma$ in (71) and substitute $y=v\sigma\sqrt{n}$, see (57). Then, for s in a bounded set contained in $\mathrm{Re}(s)\geq -\frac{1}{2}\beta_n$ (so that $\mathrm{Re}(t)\geq -\frac{1}{2}\beta_n\sqrt{n}/\sigma = \frac{1}{2}n\,\zeta_{sp}$, see (57)), we have with exponentially small error

$$\begin{aligned}&\log\Big(\mathbb{E}\Big[\exp\Big(-\frac{s\sqrt{n}}{\sigma}\,W_n\Big)\Big]\Big)\\ &= \frac{1}{2\pi}\int_{-R}^{R}\frac{is\zeta'(y/\sigma\sqrt{n})\,\log(1-e^{nh(\zeta_{sp})-\frac{1}{2}y^2})}{\sigma\sqrt{n}\,\zeta(y/\sigma\sqrt{n})(s-\sigma\sqrt{n}\,\zeta(y/\sigma\sqrt{n}))}\,dy,\end{aligned} \tag{71}$$

where $R=r\sigma\sqrt{n}$.

The remainder of the proofs of (58) and Theorem 4 consists now of approximating $nh(\zeta_{sp})$ by $-\frac{1}{2}\beta_n^2$ and $-\frac{1}{2}B_n^2$, respectively, and approximating the front factor

$$FF = \frac{is\zeta'(y/\sigma\sqrt{n})}{\sigma\sqrt{n}\,\zeta(y/\sigma\sqrt{n})(s-\sigma\sqrt{n}\,\zeta(y/\sigma\sqrt{n}))}, \tag{72}$$

using a linear and quadratic approximation, respectively, from the power series in (69).

For (58) we thus use in (71)

$$\zeta'(y/\sigma\sqrt{n}) = i + O(y/\sqrt{n}),\qquad \sigma\sqrt{n}\,\zeta(y/\sigma\sqrt{n}) = -\beta_n + iy + O(y^2/\sqrt{n}), \tag{73}$$

and obtain, uniformly in any bounded set of s such that $\mathrm{Re}(s)\geq -\frac{1}{2}\beta_n$,

$$\begin{aligned}&\log\Big(\mathbb{E}\Big[\exp\Big(-\frac{s\sqrt{n}}{\sigma_n}\,W_n\Big)\Big]\Big)\\ &= \frac{1}{2\pi}\int_{-R}^{R}\frac{s\,\log(1-e^{-\frac{1}{2}\beta_n^2-\frac{1}{2}y^2})}{(\beta_n-iy)(s+\beta_n-iy)}\,dy\Big(1+O\Big(\frac{1}{\sqrt{n}}\Big)\Big),\end{aligned} \tag{74}$$

where we have restored $\sigma=\sigma_n$ on the left-hand side of (74). Recalling that $R=r\,\sigma_n\sqrt{n}$, with $\sigma_n\asymp 1$ and $r>0$ independent of n, we see that the integral in (74) equals the integral in (67), with β_n instead of β, within exponentially small error when $\mathrm{Re}(s)\geq -\frac{1}{2}\beta_n$. This completes the proof of Theorem 3.

The proof of Theorem 4 is similar, but the details require somewhat more elaboration, and are therefore given in Appendix A. We show Theorem 5 in Appendix B and also that the $O(1/\sqrt{n})$ at the right-hand side of (11) can be replaced by $O(1/n)$ when the third cumulants of V and U are equal.

6. Computational issues

In this section we present computational schemes for exact and approximate values of the (properly scaled) moments of the steady-state waiting time W via Pollaczek's formula (3),

$$\log\big(\mathbb{E}[e^{-sW}]\big) = \frac{-1}{2\pi i}\int_C \frac{s}{z(s-z)}\log(1-\psi(-z))\,dz, \tag{75}$$

with $\psi(-z)=\mathbb{E}[\exp(-z(V-\frac{1}{\rho}U))]$, as earlier.

6.1. Moments and cumulants

Assume that X is a random variable with finite moments of all order. We compute the moments

$$m_k(X) = \mathbb{E}[X^k] = \left(\frac{d}{ds}\right)^k \left(\mathbb{E}[e^{sX}]\right)_{s=0}, k = 0, 1, \ldots, \tag{76}$$

of X from the cumulants

$$c_l(X) = \left(\frac{d}{ds}\right)^l \log\left(\mathbb{E}[e^{sX}]\right)_{s=0}, l = 0, 1, \ldots, \tag{77}$$

of X recursively according to

$$m_0(X) = 1\,;\; m_k(X) = \sum_{l=1}^{k} \binom{k-1}{l-1} c_l(X) m_{k-l}(X),\; k = 1, 2, \ldots. \tag{78}$$

6.2. Moments in the classical heavy-traffic Kingman case

With $\alpha = 1 - \rho \downarrow 0$, we compute the moments $m_k(\alpha W) = \mathbb{E}[(\alpha W)^k]$ of αW from the cumulants of αW according to (78) in Section 6.1. The latter are obtained from the appropriate version (28) of Pollaczek's formula, so that

$$\begin{aligned} c_l(\alpha W) &= (-1)^l \left(\frac{d}{ds}\right)^l \left(\frac{-1}{2\pi i}\int_{C_0} \frac{z}{z(s-z)} \log(1-\psi(-\alpha z))dz\right)_{s=0} \\ &= \frac{(-1)^l l!}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{\log(1-\psi(-\alpha z))}{z^{l+1}}\right] dy, \end{aligned} \tag{79}$$

where $C_0 = \{z = \mu_0 + iy \mid -\infty < y < \infty\}$ with $\mu_0 = -1/\sigma_\alpha^2 = -1/[(\sigma_V^2 + \rho^{-2}\sigma_U^2)\rho]$, compare [1, Equation (15)].

From the result of Section 3, see (7) in Theorem 1, we have for $k = 1, 2, \ldots$

$$\mathbb{E}[(\alpha W)^k] = k!(\frac{1}{2}\sigma_\alpha^2)^k + O(\alpha \log(1/\alpha)), \quad \alpha \downarrow 0, \tag{80}$$

where $\sigma_\alpha^2 = (\sigma_V^2 + \rho^{-2}\sigma_U^2)\rho$.

6.3. Moments in the nearly deterministic heavy-traffic Kingman case

With $1 - \rho_n \asymp 1/n$ and W_n as in Section 4, we compute the moments $m_k(W_n) = \mathbb{E}[W_n^k]$ of W_n from the cumulants $c_l(W_n)$ of W_n according to (78) in Section 6.1. The latter are obtained from the appropriate version (49) of Pollaczek's formula, so that

$$\begin{aligned} c_l(W_n) &= (-1)^l \left(\frac{d}{ds}\right)^l \left(\frac{-1}{2\pi i}\int_{C} \frac{z}{z(s-z)} \log(1-\psi^n(-z/n))dz\right)_{s=0} \\ &= \frac{(-1)^l l!}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{\log(1-\psi^n(-z/n))}{z^{l+1}}\right] dy, \end{aligned} \tag{81}$$

where $C = \{z = \frac{-1}{2\gamma_n} + iy \mid -\infty < y < \infty\}$ and

$$\gamma_n = (\sigma_V^2 + \rho_n^{-2}\sigma_U^2)\rho_n/(2n(1-\rho_n)). \tag{82}$$

From the result of Section 4, see (10) in Theorem 2, we have for $k = 1, 2, \ldots$

$$\mathbb{E}[W_n^k] = k!\gamma_n^k + O\left(\frac{\log n}{\sqrt{n}}\right), \quad n \to \infty. \tag{83}$$

6.4. Moments in the nearly deterministic heavy-traffic Gaussian case

We consider the special saddle point method with $1 - \rho_n \asymp 1/\sqrt{n}$ and W_n as given in Section 5. Thus ζ_{sp} is the unique saddle point ζ on the negative real axis of $h(\zeta) = \log \psi(-\zeta)$, characterized by $h'(\zeta) = \psi'(-\zeta) = 0$, see Section 2. In some specific cases, ζ_{sp} can be found in closed form; in general a Newton iteration can be used, using the leading term at the right-hand side of (19) as initial value. We then let

$$\sigma_n = (h''(\zeta_{sp}))^{1/2}, \quad \beta_n = -\zeta_{sp}\sigma_n\sqrt{n}. \tag{84}$$

The moments $m_k\left(\frac{\sqrt{n}}{\sigma_n}W_n\right) = \mathbb{E}\left[\left(\frac{\sqrt{n}}{\sigma_n}W_n\right)^k\right]$ of $\frac{\sqrt{n}}{\sigma_n}W_n$ can be computed from the cumulants $c_l\left(\frac{\sqrt{n}}{\sigma_n}W_n\right)$ according to (78) in Section 6.1. The latter are obtained from the appropriate version (63) of Pollaczek's formula, with $t = s\sqrt{n}/\sigma_n$, so that

$$\begin{aligned} c_l\left(\frac{\sqrt{n}}{\sigma_n}W_n\right) &= (-1)^l\left(\frac{d}{ds}\right)^l\left(\frac{-1}{2\pi i}\int_{C_n}\frac{(s/\sigma_n\sqrt{n})\log(1-\psi^n(-\zeta))}{\zeta((s/\sigma_n\sqrt{n})-\zeta)}d\zeta\right)_{s=0} \\ &= \frac{(-1)^l l!}{\pi}\left(\frac{1}{\sigma_n\sqrt{n}}\right)^l\int_0^\infty \mathrm{Re}\left[\frac{\log(1-\psi^n(-\zeta))}{\zeta^{l+1}}\right]dy, \end{aligned} \tag{85}$$

where $C_n = \{\zeta = \zeta_{sp} + iy \mid -\infty < y < \infty\}$.

From the result in Section 5, see (12) in Theorem 3, we have for $k = 1, 2, \ldots$

$$\mathbb{E}\left[\left(\frac{\sqrt{n}}{\sigma_n}W_n\right)^k\right] = m_k(\beta_n) + O\left(\frac{1}{\sqrt{n}}\right), \quad n \to \infty, \tag{86}$$

where $m_k(\beta)$ is the kth moment of the maximum M_β of the Gaussian random walk with drift $-\beta$. These $m_k(\beta)$ can be computed from the cumulants $c_l(M_\beta)$ of M_β using (78) in Section 6.1. The latter can be computed by numerical integration, using (67), so that

$$c_l\left(M_\beta\right) = \frac{(-1)^l l!}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{\log(1-e^{\beta z+\frac{1}{2}z^2})}{z^{l+1}}\right]dy, \tag{87}$$

where $z = -\beta + iy, -\infty < y < \infty$. Alternatively, we have from Theorem 1 in [18] for $0 < \beta < 2\sqrt{\pi}$ and $l = 1, 2, \ldots$

$$\begin{aligned} c_l\left(M_\beta\right) &= \frac{(l-1)!}{(2\beta)^l} + \frac{1}{\sqrt{2\pi}}\sum_{j=0}^{l}\binom{l}{j}\Gamma\left(\frac{l-j+1}{2}\right)\zeta\left(-\frac{1}{2}l-\frac{1}{2}j+1\right)2^{\frac{l-j-1}{2}}(-\beta)^j \\ &+ \frac{(-1)^{l+1}l!}{\sqrt{2\pi}}\sum_{r=0}^{\infty}\frac{\zeta(-l-r+\frac{1}{2})(-\frac{1}{2})^r\beta^{2r+l+1}}{r!(2r+1)\cdot\ldots\cdot(2r+l+1)}, \end{aligned} \tag{88}$$

where ζ denotes here the Riemann zeta function (not to be confused with the function $\zeta(v)$ in

(68)–(69) pertaining to the saddle point method).

From Theorems 4 and 5 in Section 5, we can refine the result in (86) according to

$$\mathbb{E}\Big[\Big(\frac{\sqrt{n}}{\sigma_n}\,W_n\Big)^k\Big]=\frac{m_k(B_n)}{(1+\beta_n\varphi_n)^{2k}}+\frac{k(k-1)\,m_{k-1}(B_n)\varphi_n}{(1+\beta_n\varphi_n)^{2k-1}}+O\Big(\frac{1}{n}\Big), \tag{89}$$

where

$$B_n=\beta_n/(1+\beta_n\varphi_n),\qquad \varphi_n=-\frac{h'''(\zeta_{sp})}{6h''(\zeta_{sp})\sigma_n\sqrt{n}}, \tag{90}$$

with σ_n and β_n given in (84).

In several cases, the quantities $h''(\zeta_{sp})$ and $h'''(\zeta_{sp})$ as required in (84) and (90) can be computed in closed form. In general, one has

$$h''(\zeta_{sp})\;=\;\frac{f_V''}{f_V}-\Big(\frac{f_V'}{f_V}\Big)^2+\left(\frac{f_U''}{f_U}-\Big(\frac{f_U'}{f_U}\Big)^2\right)\frac{1}{\rho^2}, \tag{91}$$

$$h'''(\zeta_{sp})\;=\;-\frac{f_V'''}{f_V}+3\frac{f_V''f_V'}{f_V^2}+\left(\frac{f_U'''}{f_U}-3\frac{f_U''f_U'}{f_U^2}\right)\frac{1}{\rho^3}, \tag{92}$$

where $f_V(z)=\mathbb{E}[\exp(zV)]$, $f_U(z)=\mathbb{E}[\exp(zU)]$, and where all $f_V^{(l)}$ and $f_U^{(l)}$ appearing at the right-hand sides of (91)–(92) have to be evaluated at $z=-\zeta_{sp}$ and $z=\zeta_{sp}/\rho$, respectively.

We end this section with a note on the computational issues encountered when evaluating the contour integrals in expressions such as (79), (81), (85) and (87). Although modern computer algebra software can numerically evaluate these integrals, despite one of the limits being infinity, one has to carefully choose a sufficiently high numerical accuracy in order to obtain accurate results. Not all software packages may support this feature, which is the reason why we used Wolfram Mathematica for the numerical computations in this paper. Unsurprisingly, this may lead to long computation times, in particular when the load of the system is close to one. In contrast, the numerical evaluation of the approximations in this paper takes practically zero time and does not suffer from numerical issues.

6.5. *Numerical example: U and V are Gamma distributed*

We now demonstrate the results for a specific example. In this numerical example, we consider the case that both V and U have a Gamma distribution, with means $k_U\vartheta_U=k_V\vartheta_V=1$, variances $\sigma_V^2=k_V\vartheta_V^2=\vartheta_V$ and $\sigma_U^2=k_U\vartheta_U^2=\vartheta_U$, and pdfs

$$\frac{x^{k_V-1}\,e^{-x/\vartheta_V}}{\Gamma(k_V)\,\vartheta_V^{k_V}}\,,\quad x\ge 0\,;\qquad \frac{x^{k_U-1}\,e^{-x/\vartheta_U}}{\Gamma(k_U)\,\vartheta_U^{k_U}}\,,\quad x\ge 0, \tag{93}$$

respectively, with $0<\vartheta_V,\vartheta_U<\infty$. Note that Assumption 6 in Section 5 is easily checked. The third cumulants of U and V are $c_3(U)=2\vartheta_U^2$ and $c_3(V)=2\vartheta_V^2$, respectively. Then

$$\psi(-\zeta)=\mathbb{E}\,[e^{-\zeta(V-\frac{1}{\rho}U)}]=(1+\vartheta_V\zeta)^{-k_V}\,(1-\vartheta_U\zeta/\rho)^{-k_U} \tag{94}$$

for $-\vartheta_V^{-1}<\mathrm{Re}(\zeta)<\rho\,\vartheta_U^{-1}$. In our numerical experiments we take three different parameter sets $(\vartheta_U,\vartheta_V)=(5/2,1/2)$, $(1/2,5/2)$ and $(3/2,3/2)$. As a consequence of U and V both being Gamma distributed, all approximations can be obtained in closed-form. Further numerical examples are provided in the extended version of this paper available at [3], where we also include combinations of interarrival and service-time distributions that do not yield closed-form

Table 1
Numerical example — Classical HT Kingman: Comparison of exact and asymptotic results for $m_k(\alpha W)$.

$\vartheta_U = 5/2,\ \vartheta_V = 1/2$						
k	$\alpha = 1/10$		$\alpha = 1/100$		$\alpha = 1/1000$	
	Exact	Asymp	Exact	Asymp	Exact	Asymp
1	1.396	1.614	1.490	1.510	1.499	1.501
2	4.092	5.209	4.459	4.561	4.496	4.506
3	17.987	25.222	20.021	20.663	20.227	20.291
4	105.426	162.819	119.860	124.814	121.336	121.825
5	772.403	1313.861	896.946	942.427	909.816	914.295
$\vartheta_U = 1/2,\ \vartheta_V = 5/2$						
k	$\alpha = 1/10$		$\alpha = 1/100$		$\alpha = 1/1000$	
	Exact	Asymp	Exact	Asymp	Exact	Asymp
1	1.331	1.403	1.483	1.490	1.498	1.499
2	4.083	3.936	4.459	4.440	4.496	4.494
3	18.806	16.562	20.111	19.849	20.236	20.210
4	115.505	92.932	120.933	118.300	121.444	121.176
5	886.802	651.817	909.028	881.352	911.030	908.217
$\vartheta_U = 3/2,\ \vartheta_V = 3/2$						
k	$\alpha = 1/10$		$\alpha = 1/100$		$\alpha = 1/1000$	
	Exact	Asymp	Exact	Asymp	Exact	Asymp
1	1.367	1.508	1.487	1.500	1.499	1.500
2	4.100	4.550	4.460	4.500	4.496	4.500
3	18.452	20.589	20.071	20.253	20.232	20.250
4	110.711	124.223	120.427	121.525	121.393	121.500
5	830.332	936.845	903.203	911.480	910.446	911.252

approximations. The focus in these examples lies on the asymptotics of higher-order cumulants, which distinguishes them from, for example, numerical examples in [1,28]. In all our numerical examples, also those in [3], we found that numerical evaluation of the asymptotic bounds is significantly faster than computation of the exact values.

6.5.1. *Classical regime*

We approximate the moments of αW, for $\alpha \downarrow 0$, by (80), with

$$\sigma_\alpha^2 = (\sigma_V^2 + \rho^{-2}\sigma_U^2)\,\rho = (k_V\vartheta_V^2 + \rho^{-2}k_U\vartheta_U^2)\,\rho = (\vartheta_V + \rho^{-2}\vartheta_U)\,\rho, \tag{95}$$

and $\rho = 1-\alpha$, where we take $\alpha = 1/10,\ 1/100,\ 1/1000$. The results, for the first five moments of W, are presented in Table 1.

We observe that the error behaves practically linearly with α for a fixed k. Thus, for this case, the error estimate $O(\alpha\ \log(1/\alpha))$ in Theorem 1 seems a factor $\log(1/\alpha)$ too pessimistic. To confirm this, we included a plot for the difference between the approximation (based on the asymptotic result) and the exact values for the first three moments. Indeed, Fig. 1 confirms that the absolute error is almost completely linear in α, meaning that the factor $\log(1/\alpha)$ is negligible here. Higher moments show the same type of behavior, but for reasons of compactness we have omitted the corresponding figures.

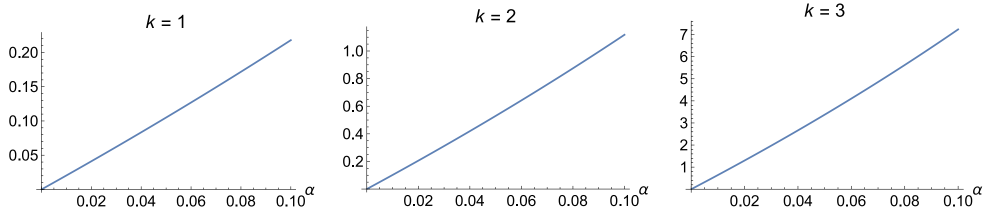

Fig. 1. Absolute error $k!(\frac{1}{2}\sigma_\alpha^2)^k - m_k(\alpha W)$ for the classical Kingman heavy-traffic regime.

It is notable that for the first moment ($k = 1$), the approximations systematically overestimate the exact values, which is known in the literature as Kingman's bound, see below (8). For higher moments, this is still the case for $\vartheta_U = 5/2$, $\vartheta_V = 1/2$, but choosing parameter values $\vartheta_U = 1/2$, $\vartheta_V = 5/2$ leads to underestimations. The k-behavior of the error for the three considered cases is markedly different, especially when α is not small.

6.5.2. Nearly-deterministic Kingman regime

Take a fixed $\beta > 0$, and let $1 - \rho_n = \beta/n$. From Theorem 2, we can approximate $\mathbb{E}[W_n^k]$ by $k!\gamma_n^k$, with

$$\gamma_n = (\sigma_V^2 + \rho_n^{-2}\sigma_U^2)\,\rho_n/(2n(1-\rho_n)) = (\vartheta_V + \rho_n^{-2}\vartheta_U)\,\rho_n/(2n(1-\rho_n)). \tag{96}$$

In Table 2, we choose $\beta = 1$. Note that the approximation, based on the asymptotic results, for $n = 10, 100, 1000$ yields exactly the same numerical values as in the classical HT Kingman case with $\alpha = 1/n$. This is due to the fact that accidentally $\frac{1}{2}\sigma_\alpha^2 = \gamma_n$ for our chosen parameter values.

We observe that the error, the difference between the exact result and the asymptotic result, decays like $1/\sqrt{n}$ for a fixed k (this decay behavior becomes more manifest when n is further increased). This is in reasonable agreement with the error estimate $O(\log n/\sqrt{n})$ that is given in Theorem 2, (10). The k-behavior of the error for the three considered cases is markedly different, especially for low n.

6.5.3. Nearly-deterministic Gaussian regime

We have now to invoke the whole machinery of the special saddle point method. We have

$$h(\zeta) = -k_V\,\log(1+\vartheta_V\zeta) - k_U\,\log(1-\vartheta_U\zeta/\rho), \qquad \frac{-1}{\vartheta_V} < \mathrm{Re}(\zeta) < \frac{\rho}{\vartheta_U}, \tag{97}$$

$$\zeta_{sp} = -\frac{1-\rho}{\vartheta_V+\vartheta_U}, \qquad h''(\zeta_{sp}) = \frac{(\vartheta_U+\vartheta_V)^3}{(\vartheta_U+\rho\vartheta_V)^2}, \qquad d_2 = -\frac{\vartheta_U^2-\vartheta_V^2}{3(\vartheta_U+\rho\vartheta_V)}. \tag{98}$$

Take a fixed $\beta > 0$, and let $1-\rho_n = \beta/\sqrt{n}$. We approximate the moments of the scaled waiting times $\frac{\sqrt{n}}{\sigma_n}W_n$ by the moments of the maximum of the Gaussian random walk with drift $-\beta$. These $m_k(\beta)$ can be computed from the cumulants $c_l(M_\beta)$ of M_β, using (78) and (87) or (88).

The results, shown in Table 3, indicate that the absolute error behaves quite accurately as $O(1/\sqrt{n})$ and $O(1/n)$, respectively, for the two asymptotic estimates. It is also interesting to see that the two asymptotic estimates perform equally well for low k and n, while the refined asymptotic estimate based on (86) outperforms the asymptotic estimate based on (89) in all other cases.

Table 2
Example 1 — Nearly-deterministic HT Kingman: Comparison of exact and asymptotic results for $m_k(W_n)$.

$\vartheta_U = 5/2, \vartheta_V = 1/2$

k	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$		$n = 100000$	
	Exact	Asymp	Exact	Asymp	Exact	Asymp	Exact	Asymp	Exact	Asymp
1	1.212	1.614	1.404	1.510	1.469	1.501	1.490	1.500	1.497	1.500
2	3.572	5.209	4.205	4.561	4.405	4.506	4.470	4.501	4.490	4.500
3	15.711	25.222	18.880	20.663	19.819	20.291	20.114	20.254	20.207	20.250
4	92.084	162.819	113.030	124.814	118.888	121.825	120.680	121.532	121.241	121.503
5	674.652	1313.861	845.833	942.427	891.460	914.295	905.080	911.554	909.307	911.280

$\vartheta_U = 1/2, \vartheta_V = 5/2$

k	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$		$n = 100000$	
	Exact	Asymp	Exact	Asymp	Exact	Asymp	Exact	Asymp	Exact	Asymp
1	1.151	1.403	1.397	1.490	1.468	1.499	1.490	1.500	1.497	1.500
2	3.551	3.936	4.204	4.440	4.405	4.494	4.470	4.499	4.490	4.500
3	16.371	16.562	18.962	19.849	19.828	20.210	20.115	20.246	20.207	20.250
4	100.564	92.932	114.027	118.300	118.993	121.176	120.691	121.468	121.242	121.497
5	772.098	651.817	857.114	881.352	892.646	908.217	905.200	910.946	909.320	911.220

$\vartheta_U = 3/2, \vartheta_V = 3/2$

k	$n = 10$		$n = 100$		$n = 1000$		$n = 10000$		$n = 100000$	
	Exact	Asymp	Exact	Asymp	Exact	Asymp	Exact	Asymp	Exact	Asymp
1	1.183	1.508	1.401	1.500	1.468	1.500	1.490	1.500	1.497	1.500
2	3.568	4.550	4.205	4.500	4.405	4.500	4.470	4.500	4.490	4.500
3	16.065	20.589	18.922	20.253	19.823	20.250	20.114	20.250	20.207	20.250
4	96.396	124.223	113.532	121.525	118.940	121.500	120.685	121.500	121.242	121.500
5	722.972	936.845	851.487	911.480	892.054	911.252	905.140	911.250	909.314	911.250

Notice that the two approximations yield the same results when $\vartheta_U = \vartheta_V = 3/2$ in Table 3. This is caused by U and V having the same Gamma distribution, resulting in $h'''(\zeta_{sp}) = 0$. It follows that $\phi_n = 0$ and $B_n = \beta_n$ in Theorem 4. Consequently, the refined approximation and the original approximation are equal here.

7. Conclusions

We have presented several heavy-traffic limit theorems, using Kingman's transform method based on Pollaczek's formula for the Laplace transform of the steady-state waiting time distribution of the GI/G/1 queue. Under the assumption that the distribution of both the service time and the interarrival time have a Laplace transform, analytic in an open strip containing the imaginary axis, these heavy-traffic limit results can be shown to be valid on the level of transforms in a full neighborhood of the origin, with error assessment. As a consequence, there is convergence for all moments, with a corresponding error assessment. We have considered the classical heavy-traffic regime in which the transform of the steady-state waiting time distribution converges, after appropriate scaling, to the transform of an exponentially distributed random variable as the system load $\rho = 1 - \alpha$ tends to 1 (Kingman-type result), with error shown to be bounded as $O(\alpha \log(1/\alpha))$ as $\alpha \downarrow 0$. We also have considered nearly deterministic queues (obtained through cyclic thinning) in two different regimes, viz. where the system's

Table 3
Example 1 — Nearly-deterministic HT Gaussian: Comparison of exact results for $m_k(\frac{\sqrt{n}}{\sigma_n} W_n)$ and the asymptotic results for the case that $\vartheta_U = 5/2$, $\vartheta_V = 1/2$ with $\beta = 1$, $n = 10, 100, 1000$ and $k = 1, 2, 3, 4, 5$. The two entries in the Asymp-columns give, for a particular k, the asymptotic result from (86) and (89) in that order.

$\vartheta_U = 5/2, \vartheta_V = 1/2$

k	$n = 10$			$n = 100$			$n = 1000$		
	Exact	Asymp 1	Asymp 2	Exact	Asymp 1	Asymp 2	Exact	Asymp 1	Asymp 2
1	0.3674	0.3748	0.3776	0.4021	0.4015	0.4030	0.4105	0.4100	0.4106
2	0.5898	0.6499	0.6136	0.7071	0.7202	0.7094	0.7396	0.7432	0.7398
3	1.3565	1.6288	1.3756	1.7936	1.8703	1.7963	1.9283	1.9514	1.9286
4	4.1049	5.3717	3.8574	5.9908	6.3985	5.9714	6.6235	6.7521	6.6218
5	15.4834	22.0568	12.5010	24.9315	27.2673	24.6639	28.3460	29.1063	28.3202

$\vartheta_U = 1/2, \vartheta_V = 5/2$

k	$n = 10$			$n = 100$			$n = 1000$		
	Exact	Asymp 1	Asymp 2	Exact	Asymp 1	Asymp 2	Exact	Asymp 1	Asymp 2
1	0.2227	0.2259	0.2282	0.3506	0.3523	0.3514	0.3938	0.3943	0.3938
2	0.3242	0.3139	0.3367	0.6006	0.5931	0.6025	0.7039	0.7009	0.7041
3	0.6903	0.6223	0.7043	1.4922	1.4410	1.4947	1.8236	1.8032	1.8239
4	1.9359	1.6078	1.8922	4.8802	4.6016	4.8685	6.2235	6.1091	6.2220
5	6.7472	5.1462	6.2056	19.8621	18.2869	19.6959	26.4527	25.7788	26.4304

$\vartheta_U = 3/2, \vartheta_V = 3/2$

k	$n = 10$			$n = 100$			$n = 1000$		
	Exact	Asymp 1	Asymp 2	Exact	Asymp 1	Asymp 2	Exact	Asymp 1	Asymp 2
1	0.2919	0.2985	0.2985	0.3761	0.3768	0.3768	0.4021	0.4022	0.4022
2	0.4504	0.4660	0.4660	0.6532	0.6550	0.6550	0.7217	0.7219	0.7219
3	1.0062	1.0449	1.0449	1.6410	1.6462	1.6462	1.8757	1.8763	1.8763
4	2.9556	3.0712	3.0712	5.4270	5.4442	5.4442	6.4226	6.4245	6.4245
5	10.7979	11.2169	11.2169	22.3476	22.4176	22.4176	27.3938	27.4020	27.4020

load ρ_n satisfies $1 - \rho_n \asymp 1/n$ and $1 - \rho_n \asymp 1/\sqrt{n}$, respectively, as the thinning factor n tends to infinity. The first regime allows a result of the Kingman-type, viz. convergence in terms of transforms to an exponential distribution, with error bounded as $O(\frac{1}{\sqrt{n}} \log n)$. The second regime allows, after appropriate scaling, convergence in terms of transforms to the maximum of the Gaussian random walk with a specific negative drift. For this case, we have shown $O(1/\sqrt{n})$ error behavior of transforms and moments. The latter result is refined, so as to yield $O(1/n)$ errors, by judicious choice of the drift parameter, as well as by using a weakly non-linear transformation of the Laplace variable in the transform of the maximum of the Gaussian random walk. For all regimes, the heavy-traffic limits for the moments of the waiting time distribution were also found to be good asymptotic approximations for the exact values.

Acknowledgments

The work of J.S.H. van Leeuwaarden was partly supported by the Netherlands Organization for Scientific Research (NWO) through VICI [Grant 202.068].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Finishing the proof of Theorem 4

We approximate the front factor FF in (72) by using a linear approximation of $\zeta'(v)$ and a quadratic approximation of $\zeta(v)$ from (69). Writing $\sigma = \sigma_n$ as we did in Section 5, we have

$$\zeta'(y/(\sigma\sqrt{n})) = i - 2d_2 y/(\sigma\sqrt{n}) + O(y^2/n), \tag{99}$$

$$\sigma\sqrt{n}\,\zeta(y/(\sigma\sqrt{n})) = -\beta + iy - \varphi y^2 + O(y^3/n), \tag{100}$$

where we have written $\beta = \beta_n$ and $\varphi = \varphi_n = d_2/\sigma_n\sqrt{n}$. Using this in (72), we get

$$\begin{aligned} FF &= \frac{-s(1 + 2i\varphi y + O(y^2/n))}{(-\beta + iy - \varphi y^2 + O(y^3/n))(s + \beta - iy + \varphi y^2 + O(y^3/n))} \\ &= -s\,\frac{1 + i\varphi y}{-\beta + iy - \varphi y^2}\,\frac{1 + i\varphi y}{s + \beta - iy + \varphi y^2}\,(1 + O(y^2/n)). \end{aligned} \tag{101}$$

Now

$$\begin{aligned} \frac{-\beta + iy - \varphi y^2}{1 + i\varphi y} &= (-\beta + iy - \varphi y^2)(1 - i\varphi y) + O(y^2/n) \\ &= -\beta + iy(1 + \beta\varphi) + O(y^2/n) \\ &= (1 + \beta\varphi)(-B + iy) + O(y^2/n), \end{aligned} \tag{102}$$

with $B = B_n = \beta_n/(1 + \beta_n\varphi_n)$, see (61). Similarly,

$$\begin{aligned} \frac{s + \beta - iy + \varphi y^2}{1 + i\varphi y} &= s + \beta - iy(1 + (s + \beta)\varphi) + O(y^2/n) \\ &= (1 + (s + \beta)\varphi)\left(\frac{s + \beta}{1 + (s + \beta)\varphi} - iy\right) + O(y^2/n)\,. \end{aligned} \tag{103}$$

We further set, see (61),

$$\frac{s + \beta}{1 + (s + \beta)\varphi} = R + \frac{\beta}{1 + \beta\varphi} = R + B, \tag{104}$$

and we compute

$$R = \frac{s + \beta}{1 + (s + \beta)\varphi} - \frac{\beta}{1 + \beta\varphi} = \frac{s}{(1 + (s + \beta)\varphi)(1 + \beta\varphi)}. \tag{105}$$

Therefore,

$$\begin{aligned} FF &= \frac{-s}{(1 + \beta\varphi)(1 + (s + \beta)\varphi)}\,\frac{1 + O(y^2/n)}{(-B + iy)(R + B - iy)} \\ &= \frac{R}{(B - iy)(R + B - iy)}\,(1 + O(y^2/n)), \end{aligned} \tag{106}$$

and this takes care of the factor in the integral in (71).

We finally relate $n\,h(\zeta_{sp})$, occurring in the exponential in the integrand in (71), and B. We have with $\beta=\beta_n$, $\sigma=\sigma_n$ given in (57) and d_2 given by (59)

$$B=\frac{\beta}{1+\beta\varphi}=\frac{\beta_n}{1+\beta_n\,\frac{d_2}{\sigma\sqrt{n}}}=\frac{-\zeta_{sp}\sqrt{n\,h''(\zeta_{sp})}}{1+\frac{\zeta_{sp}\,h'''(\zeta_{sp})}{6h''(\zeta_{sp})}}$$
$$=-\zeta_{sp}\sqrt{n\,h''(\zeta_{sp})}\Big(1-\frac{\zeta_{sp}\,h'''(\zeta_{sp})}{6h''(\zeta_{sp})}+O\Big(\frac{1}{n}\Big)\Big). \tag{107}$$

On the other hand, from $h'(\zeta_{sp})=0=h(0)$ and

$$h(0)=h(\zeta_{sp})-\zeta_{sp}\,h'(\zeta_{sp})+\tfrac{1}{2}\,\zeta_{sp}^2\,h''(\zeta_{sp})-\tfrac{1}{6}\,\zeta_{sp}^3\,h'''(\zeta_{sp})+O\Big(\frac{1}{n^2}\Big), \tag{108}$$

we get

$$-\,n\,h(\zeta_{sp})=\tfrac{1}{2}\,n\,\zeta_{sp}^2\,h''(\zeta_{sp})-\tfrac{1}{6}\,n\,\zeta_{sp}^3\,h'''(\zeta_{sp})+O\Big(\frac{1}{n}\Big)$$
$$=\tfrac{1}{2}\,n\,\zeta_{sp}^2\,h''(\zeta_{sp})\Big(1-\frac{\zeta_{sp}\,h'''(\zeta_{sp})}{3h''(\zeta_{sp})}\Big)+O\Big(\frac{1}{n}\Big)$$
$$=\tfrac{1}{2}\,B^2+O\Big(\frac{1}{n}\Big). \tag{109}$$

Using (106) and (109), valid uniformly in any bounded set of s with $\mathrm{Re}(s)\geq-\frac{1}{2}\,\beta_n$, in (71), we get

$$\log(\mathbb{E}\Big[\exp\Big(-\frac{s\sqrt{n}}{\sigma_n}\,W_n\Big)\Big])$$
$$=\frac{1}{2\pi i}\int_{-R}^{R}\frac{R_n\,\log(1-e^{-\frac{1}{2}B_n^2-\frac{1}{2}y^2})}{(B_n-iy)(R_n+B_n-iy)}\,dy\,\Big(1+O\Big(\frac{1}{n}\Big)\Big), \tag{110}$$

where we have restored the n in σ_n, R_n and B_n. Then the proof of Theorem 4 can be finished in the same way as the proof of Theorem 3 was finished.

Appendix B. Proof of Theorem 5

The first line of (62) is an immediate consequence of Theorem 4, and can be rewritten as

$$\mathbb{E}\Big[\Big(\frac{\sqrt{n}}{\sigma_n}\,W_n\Big)^k\Big]=\Big(\frac{d}{ds}\Big)^k\,[F(T_n(s))]_{s=0}+O\Big(\frac{1}{n}\Big), \tag{111}$$

where $F(s)=\mathbb{E}\,[\exp(s\,M_{B_n})]$ and $T_n(s)=-R_n(-s)$ with B_n and R_n given in (61). We have from (61), deleting the index n from β_n and φ_n,

$$T_n(s)=\frac{s}{1+\beta\varphi}\;\frac{1}{1+\beta\varphi-\varphi s}=\sum_{r=1}^{\infty}\frac{\varphi^{r-1}}{(1+\beta\varphi)^{r+1}}\,s^r. \tag{112}$$

Therefore, $T_n(0)=0$ and

$$T_n^{(r)}(0)=\frac{r!\,\varphi^{r-1}}{(1+\beta\varphi)^{r+1}},\qquad r=1,2,\ldots, \tag{113}$$

so that, in particular,

$$T_n'(0)=\frac{1}{(1+\beta\varphi)^2},\qquad T_n''(0)=\frac{2\varphi}{(1+\beta\varphi)^3}, \tag{114}$$

$$T_n^{(r)}(0) = O(\varphi^{r-1}) = O(n^{-\frac{1}{2}(r-1)}), \tag{115}$$

see below (61).

We compute

$$V_k = \left(\frac{d}{ds}\right)^k [F(T_n(s))](0), \tag{116}$$

noting that $F^{(k)}(0) = m_k(B_n)$ for $k = 1, 2, \ldots$, within absolute error $O(1/n)$. We have, writing $B = B_n$,

$$V_1 = F'(T_n(s))\, T_n'(s)\Big|_{s=0} = F'(0)\, T_n'(0) = \frac{m_1(B)}{(1+\beta\varphi)^2}, \tag{117}$$

$$\begin{aligned} V_2 &= \frac{d}{ds}\,(F'(T_n(s))\, T_n'(s))\Big|_{s=0} \\ &= (F''(T_n(s))(T_n'(s))^2 + F'(T_n(s))T_n''(s))\Big|_{s=0} \\ &= F''(0)(T_n'(0))^2 + F'(0)\, T_n''(0) = \frac{m_2(B)}{(1+\beta\varphi)^4} + \frac{2\varphi\, m_1(B)}{(1+\beta\varphi)^3}, \end{aligned} \tag{118}$$

and similarly

$$\begin{aligned} V_3 &= F'''(0)(T_n'(0))^3 + 3F''(0)\, T_n'(0)\, T_n''(0) + F'(0)\, T_n'''(0) \\ &= \frac{m_3(B)}{(1+\beta\varphi)^6} + \frac{6\varphi\, m_2(B)}{(1+\beta\varphi)^5} + O\Big(\frac{1}{n}\Big), \end{aligned} \tag{119}$$

where (114)–(115) has been used. In general, one finds inductively

$$\begin{aligned} V_k &= (F^{(k)}(T_n(s))(T_n'(s))^k + \tfrac{1}{2}\, k(k-1)\, F^{(k-1)}(T_n(s))(T_n'(s))^{k-1} T_n''(s) + \cdots)\Big|_{s=0} \\ &= \frac{m_k(B)}{(1+\beta\varphi)^{2k}} + \frac{k(k-1)\,\varphi\, m_{k-1}(B)}{(1+\beta\varphi)^{2k-1}} + O\Big(\frac{1}{n}\Big), \end{aligned} \tag{120}$$

and this gives the expression on the second line of (62) after restoring the n in B, β and φ.

We finally show that Theorem 3 gives qualitatively the same accurate estimates of the moments as Theorem 5 does when the third cumulants of V and U are equal. We have, abbreviating "kth cumulant of" by "c_k",

$$\begin{aligned} h(\zeta) &= \log\left(\mathbb{E}[e^{-\zeta V}]\right) + \log\left(\mathbb{E}[e^{\zeta U/\rho}]\right) \\ &= \sum_{k=1}^{\infty} \frac{\rho^{-k} c_k(U) + (-1)^k c_k(V)}{k!}\zeta^k. \end{aligned} \tag{121}$$

With $1-\rho = 1-\rho_n \asymp 1/\sqrt{n}$, we have that $\zeta_{sp} = O(1/\sqrt{n})$. Assume that $c_3(V) = c_3(U)$. Then from (121)

$$h'''(\zeta_{sp}) = O(\rho^{-3} c_3(U) - c_3(V)) + O(\zeta_{sp}) = O(1/\sqrt{n}). \tag{122}$$

Therefore, ϕ_n in Theorem 4 is $O(1/n)$, and $B_n = \beta_n + O(1/n)$. Hence, the leading term in (62) and the leading term in (11) agree with one another within an error $O(1/n)$, and so the first order term in (62) approximates the scaled moment at the left-hand side of (62) within an error $O(1/n)$. We conclude that the leading term in (11) approximates the scaled moment at the left-hand side of (62) and (11) within an error $O(1/n)$ as well.

References

[1] J. Abate, G.L. Choudhury, W. Whitt, Calculation of the GI/G/1 waiting time distribution and its cumulants from Pollaczek's formulas, Arch. Elektron. Uebertrag.tech. 47 (5/6) (1993) 311–321.

[2] J. Blanchet, P.W. Glynn, Complete corrected diffusion approximations for the maximum of a random walk, Ann. Appl. Probab. 16 (2) (2006) 951–983.

[3] M.A.A. Boon, A.J.E.M. Janssen, J.S.H. van Leeuwaarden, Heavy-traffic single-server queues and the transform method, 2022, arXiv preprint arXiv:2206.09844.

[4] M.A.A. Boon, A.J.E.M. Janssen, J.S.H. van Leeuwaarden, R.W. Timmerman, Pollaczek contour integrals for the fixed-cycle traffic-light queue, Queueing Syst. 91 (1) (2019) 89–111.

[5] M.A.A. Boon, A.J.E.M. Janssen, J.S.H. van Leeuwaarden, R.W. Timmerman, Optimal capacity allocation for heavy-traffic fixed-cycle traffic-light queues and intersections, Transp. Res. B: Methodol. 167 (2023) 79–98.

[6] A.A. Borovkov, Some limit theorems in the theory of mass service, Theory Probab. Appl. 10 (3) (1965) 375–400.

[7] O.J. Boxma, J.W. Cohen, Heavy-traffic analysis for the GI/G/1 queue with heavy-tailed distributions, Queueing Syst. 33 (1999) 177–204.

[8] S.L. Brumelle, Some inequalities for parallel-server queues, Oper. Res. 19 (2) (1971) 402–413.

[9] A. Budhiraja, C. Lee, Stationary distribution convergence for generalized Jackson networks in heavy traffic, Math. Oper. Res. 34 (1) (2009) 45–56.

[10] J.T. Chang, Y. Peres, Ladder heights, Gaussian random walks and the Riemann zeta function, Ann. Probab. 25 (2) (1997) 787–802.

[11] Y. Chen, W. Whitt, Algorithms for the upper bound mean waiting time in the GI/GI/1 queue, Queueing Syst. 94 (2020) 327–356.

[12] J.W. Cohen, The Single Server Queue, second ed., North-Holland, Elsevier, 1982.

[13] J.W. Cohen, Complex functions in queueing theory, AEU. Arch. Elektron. Uebertrag.tech. 47 (5–6) (Pollaczek memorial volume, 1993) 300–310.

[14] N.G. de Bruijn, Asymptotic Methods in Analysis, third edition, Dover Publications, New York, 1981.

[15] D. Gamarnik, A. Zeevi, Validity of heavy traffic steady-state approximations in generalized Jackson networks, Ann. Appl. Probab. 16 (1) (2006) 56–90.

[16] D.L. Iglehart, W. Whitt, Multiple channel queues in heavy traffic, Adv. Appl. Probab. 2 (2) (1970) 355–369.

[17] A.J.E.M. Janssen, J.S.H. van Leeuwaarden, On Lerch's transcendent and the Gaussian random walk, Ann. Appl. Probab. 17 (2) (2006) 421–439.

[18] A.J.E.M. Janssen, J.S.H. van Leeuwaarden, Cumulants of the maximum of the Gaussian random walk, Stochastic Process. Appl. 117 (2007) 1928–1959.

[19] A.J.E.M. Janssen, J.S.H. van Leeuwaarden, B.W.J. Mathijsen, Novel heavy-traffic regimes for large-scale service systems, SIAM J. Appl. Math. 78 (2) (2015) 787–812.

[20] J.F.C. Kingman, The single server queue in heavy traffic, Math. Proc. Camb. Phil. Soc. 57 (4) (1961) 902–904.

[21] J.F.C. Kingman, On queues in heavy traffic, J. R. Stat. Soc. Ser. B Stat. Methodol. 24 (2) (1962) 383–392.

[22] J. Köllerström, Heavy traffic theory for queues with several servers. I, J. Appl. Probab. 11 (3) (1974) 544–552.

[23] J. Köllerström, Heavy traffic theory for queues with several servers. II, J. Appl. Probab. 16 (2) (1979) 393–401.

[24] B.W.J. Mathijsen, A.J.E.M. Janssen, J.S.H. van Leeuwaarden, B. Zwart, Robust heavy-traffic approximations for service systems facing overdispersed demand, Queueing Syst. 90 (2018) 257–289.

[25] G.F. Newell, Approximate Stochastic Behavior of N-Server Service Systems with Large N, in: Lecture Notes in Economics and Mathematical Systems, No. 87, Springer-Verlag, 1973.

[26] D. Siegmund, Corrected diffusion approximations in certain random walk problems, Adv. Appl. Probab. 11 (4) (1979) 701–719.

[27] K. Sigman, W. Whitt, Heavy-traffic limits for nearly deterministic queues, J. Appl. Probab. 48 (3) (2011a) 657–678.

[28] K. Sigman, W. Whitt, Heavy-traffic limits for nearly deterministic queues: Stationary distributions, Queueing Syst. 69 (2011b) 145–173.

[29] W. Whitt, Heavy Traffic Limit Theorems for Queues: A Survey, in: Lecture Notes in Economics and Mathematical Systems, No. 98, Springer-Verlag, 1974.