Double Skorokhod Map and Reneging Real-Time Queues

Łukasz Kruk
Department of Mathematics
Maria Curie-Sklodowska University
Lublin, Poland
and
Institute of Mathematics
Polish Academy of Sciences
Warsaw, Poland
lkruk@hektor.umcs.lublin.pl

John Lehoczky*
Department of Statistics
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
jpl@stat.cmu.edu

Kavita Ramanan†
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
kramanan@math.cmu.edu

Steven Shreve‡
Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213
shreve@andrew.cmu.edu

November 15, 2006

Key Words. Skorokhod map, reflection map, double-sided reflection map.
AMS Subject Classification. Primary 60G07, 60G17; Secondary 90B05, 90B22.

*Partially supported by ONR/DARPA MURI under contract N000140-01-1-0576.
†Partially supported by the National Science Foundation under Grants No. 0405343 and 0406191.
‡Partially supported by the National Science Foundation under Grant No. DMS-0404682.

Abstract

An explicit formula for the Skorokhod map $\Gamma_{0,a}$ on $[0,a]$ for $a > 0$ is provided and related to similar formulas in the literature. Specifically, it is shown that on the space $\mathcal{D}[0,\infty)$ of right-continuous functions with left limits taking values in $\mathbb{R}$,

$$\Gamma_{0,a}(\psi)(t) = \psi(t) - \left[\left(\psi(0) - a\right)^{+} \wedge \inf_{u \in [0,t]} \psi(u) \right] \vee \sup_{s \in [0,t]} \left[\left(\psi(s) - a\right) \wedge \inf_{u \in [s,t]} \psi(u) \right].$$

is the unique function taking values in $[0,a]$ that is obtained from ψ by minimal "pushing" at the endpoints 0 and a. An application of this result to real-time queues with reneging is outlined.

1 Introduction

In 1961 A. V. Skorokhod [10] considered the problem of constructing solutions to stochastic differential equations on the half-line $\mathbb{R}_+$ with a reflecting boundary condition at 0. His construction implicitly used properties of a deterministic mapping on the space $\mathcal{C}[0,\infty)$ of continuous functions on $[0,\infty)$. This mapping was used more explicitly by Anderson and Orey in their study of large deviations properties of reflected diffusions on a half-space in $\mathbb{R}^N$ (see p. 194 of [1]). These authors exploited the fact that the mapping, which is now called the *Skorokhod map* and is denoted here by Γ_0, has the explicit representation

$$\Gamma_0(\psi)(t) \triangleq \psi(t) + \max_{s\in[0,t]} [-\psi(s)]^+ ,$$

and is consequently Lipschitz continuous (with constant 2). In fact, this formula easily extends to a mapping on $\mathcal{D}[0,\infty)$, the space of right-continuous functions with left limits mapping $[0,\infty)$ into $\mathbb{R}$. Given $\psi \in \mathcal{D}[0,\infty)$, define

$$\eta(t) = \sup_{s\in[0,t]} \big[-\psi(s)\big]^+ = -\inf_{s\in[0,t]} \big[\psi(s)\wedge 0\big] \tag{1.1}$$

and

$$\Gamma_0(\psi) = \psi + \eta \quad \forall \psi \in \mathcal{D}[0,\infty). \tag{1.2}$$

Then $\Gamma_0(\psi)$ is in $\mathcal{D}[0,\infty)$ and takes values in $\mathbb{R}_+$, η is in $\mathcal{D}[0,\infty)$ and is nondecreasing, and the pair of functions $(\Gamma_0(\psi),\eta)$ satisfies the *complementarity condition*

$$\int_0^\infty \mathbb{I}_{\{\Gamma_0(\psi)(s)>0\}}\, d\eta(s) = 0, \tag{1.3}$$

which says that η "pushes" only when $\Gamma_0(\psi)$ is zero. These properties uniquely characterize the pair of functions $(\Gamma_0(\psi),\eta)$, and this pair is said to solve the *Skorokhod problem for* ψ *on* $[0,\infty)$.

Let $z<a$ be a real number. The *double Skorokhod map* $\Gamma_{z,a}$ is the mapping from $\mathcal{D}[0,\infty)$ into itself such that for $\psi\in\mathcal{D}[0,\infty)$, $\Gamma_{z,a}(\psi)$ takes values in $[z,a]$ and has the decomposition

$$\Gamma_{z,a}(\psi) = \psi + \eta_\ell - \eta_u, \tag{1.4}$$

where η_ℓ and η_u are nondecreasing functions in $\mathcal{D}[0,\infty)$ so that the triple $(\Gamma_{z,a}(\psi),\eta_\ell,\eta_u)$ satisfies the *complementarity conditions*

$$\int_0^\infty \mathbb{I}_{\{\Gamma_{z,a}(\psi)(s)>z\}}\, d\eta_\ell(s) = 0, \quad \int_0^\infty \mathbb{I}_{\{\Gamma_{z,a}(\psi)(s)<a\}}\, d\eta_u(s) = 0. \tag{1.5}$$

The function η_ℓ "pushes" only when $\Gamma_{z,a}(\psi)$ is at the lower boundary z, and η_u "pushes" only when $\Gamma_{z,a}(\psi)$ is at the upper boundary a. Existence and uniqueness of η_ℓ and η_u, and hence the validity of the definition of $\Gamma_{z,a}(\psi)$ for continuous functions ψ as well as step functions in $\mathcal{D}[0,\infty)$, follow directly from Tanaka [11], Lemmas 2.1, 2.3 and 2.6. In fact, it is well known that for every $\psi\in\mathcal{D}[0,\infty)$, there exist unique nondecreasing η_ℓ and η_u in $\mathcal{D}[0,\infty)$ so that $\Gamma_{z,a}(\psi)$ is a function in $\mathcal{D}[0,\infty)$ taking values in $[z,a]$ and (1.5) is satisfied (see, e.g., [2], [13]). The triple $(\Gamma_{z,a}(\psi),\eta_\ell,\eta_u)$ is said to solve the *Skorokhod problem for* ψ *on* $[z,a]$.

In contrast to the formulas (1.1), (1.2) for the Skorokhod map on $[0,\infty)$, no explicit formula for the Skorokhod map $\Gamma_{z,a}$ was known until recently. Such a formula was provided for $\Gamma_{0,a}$ by [8] as a composition of maps with explicit formulas, and a formula for $\Gamma_{z,a}$ with $z \neq 0$ can be immediately obtained by translation. Given $\phi \in \mathcal{D}[0,\infty)$, define

$$\Lambda_a(\phi)(t) \triangleq \phi(t) - \sup_{s\in[0,t]} \left[\left(\phi(s) - a\right)^+ \wedge \inf_{u\in[s,t]} \phi(u) \right]. \tag{1.6}$$

It is shown in [8] that

$$\Gamma_{0,a} = \Lambda_a \circ \Gamma_0. \tag{1.7}$$

It has recently come to the attention of the authors of this paper that there are formulas in the literature for $\Gamma_{0,a}$ restricted to various subsets of $\mathcal{D}[0,\infty)$. One of these, equation (15) in Cooper, Schmidt and Serfozo [3], provides a formula for $\Gamma_{0,a}$ restricted to the set of bounded-variation functions in $\mathcal{D}[0,\infty)$. Although the authors do not discuss its relation to the Skorohod map, Lemma 5.6 of Ganesh, O'Connell and Wischik [5] give a formula for $\Gamma_{0,a}$ restricted to a subset of the bounded-variation functions. Equations (4) and (5) of Toomey [12] give an analogue of $\Gamma_{0,a}$ for functions on the integers rather than on $[0,\infty)$. Functions defined on the integers can be regarded as piecewise constant functions in $\mathcal{D}[0,\infty)$. Therefore, in [3], [5] and [12], the formulas provided can be interpreted as a definition of $\Gamma_{0,a}$ on subsets of $\mathcal{D}[0,\infty)$. In the case of [3], the subset on which the mapping is defined is dense in $\mathcal{D}[0,\infty)$ in the Skorokhod metric. Because the formulas provided by [3], [5] and [12] are Lipschitz continuous in the Skorokhod metric, their continuous extensions to the closures of the sets on the which they are specified must be given by the same formulas. A separate argument (see [13]) can be used to show that $\Gamma_{0,a}$ is also Liptschitz continuous on $\mathcal{D}[0,\infty)$. Therefore, these extended formulas must agree with $\Gamma_{0,a}$ on their domains of definition. This suggests that all these formulas are closely related. One purpose of this paper is to work out these relationships. In doing so, we discover another formulation of (1.7) that avoids the need to compose mappings. In particular, we show that

$$\Gamma_{0,a}(\psi)(t) = \psi(t) - \left[\left(\psi(0) - a\right)^+ \wedge \inf_{u\in[0,t]} \psi(u) \right] \vee \sup_{s\in[0,t]} \left[\left(\psi(s) - a\right) \wedge \inf_{u\in[s,t]} \psi(u) \right] \tag{1.8}$$

for all $\psi \in \mathcal{D}[0,\infty)$.

There is a second purpose for this paper. The derivation of the explicit formula for $\Gamma_{0,a}$ developed in [8] was motivated by the analysis of real-time queues with reneging (see [9]). Because the derivation in [9] is highly technical, we provide here a non-rigorous but more accessible explanation of that application.

2 Alternative formula for $\Gamma_{0,a}$

We begin with the proof of (1.8).

Theorem 2.1 *For $\psi \in \mathcal{D}[0,\infty)$, define*

$$\Xi_a(\psi)(t) \triangleq \psi(t) - \left[\left(\psi(0) - a\right)^+ \wedge \inf_{u\in[0,t]} \psi(u) \right] \vee \sup_{s\in[0,t]} \left[\left(\psi(s) - a\right) \wedge \inf_{u\in[s,t]} \psi(u) \right]. \tag{2.1}$$

Then $\Xi_a = \Lambda_a \circ \Gamma_0 = \Gamma_{0,a}$.

PROOF: We show that $\Xi_a = \Lambda_a \circ \Gamma_0$ and then appeal to (1.7). Let $\psi \in \mathcal{D}[0,\infty)$ be given. We have immediately from (2.1) that

$$\Xi_a(\psi)(0) = \psi(0) - \big(\psi(0)-a\big)^+ \wedge \psi(0) = \begin{cases} a & \text{if } \psi(0) \geq a, \\ \psi(0) & \text{if } 0 \leq \psi(0) \leq a, \\ 0 & \text{if } \psi(0) \leq 0. \end{cases}$$

which agrees with $\Lambda_a(0) = \Gamma_{0,a}(\psi)(0)$.

Now let $t > 0$ be fixed. Let us define $\phi = \psi + \eta$, where η is given by (1.1). In other words, $\phi = \Gamma_0(\psi)$. Let us next define $\overline{\phi} = \Lambda_a(\phi)$. We must show that $\overline{\phi}(t) = \Xi_a(\psi)(t)$.

<u>Case I</u>: $\eta(t) = 0$.

Because η is nondecreasing, in this case we have $\eta(s) = 0$ and $\phi(s) = \psi(s)$ for all $s \in [0,t]$. In particular, ψ is nonnegative on $[0,t]$. Therefore, $\big(\psi(0)-a\big)^+ \wedge \inf_{u\in[0,t]} \psi(u) \geq 0$, and, in fact, this expression is $0 \vee \big[\big(\psi(0)-a\big) \wedge \inf_{u\in[0,t]} \psi(u)\big]$. It follows that

$$\begin{aligned} \Xi_a(\psi)(t) &= \psi(t) - 0 \vee \sup_{s\in[0,t]} \left[\big(\psi(s)-a\big) \wedge \inf_{u\in[s,t]} \psi(u)\right] \\ &= \psi(t) - \sup_{s\in[0,t]} \left[\big(\psi(s)-a\big)^+ \wedge \inf_{u\in[s,t]} \psi(u)\right] \\ &= \Lambda_a(\psi)(t) \ = \ \Lambda_a(\phi)(t) \ = \ \overline{\phi}(t). \end{aligned}$$

<u>Case II</u>: $\eta(t) > 0$.

In this case, (1.1) becomes $\eta(t) = -\inf_{u\in[0,t]} \psi(u)$. Using $\psi = \phi - \eta$, we write (2.1) as

$$\begin{aligned} \Xi_a(\psi)(t) &= \phi(t) - \Bigg\{\eta(t) + \left[\big(\phi(0)-a-\eta(0)\big)^+ \wedge \big(-\eta(t)\big)\right] \\ &\qquad \vee \sup_{s\in[0,t]} \left[\big(\phi(s)-a-\eta(s)\big) \wedge \inf_{u\in[s,t]} \big(\phi(u)-\eta(u)\big)\right]\Bigg\} \\ &= \phi(t) - \left[\big((\phi(0)-a-\eta(0))^+ + \eta(t)\big) \wedge 0\right] \\ &\qquad \vee \sup_{s\in[0,t]} \left[\big(\phi(s)-a+\eta(t)-\eta(s)\big) \wedge \inf_{u\in[s,t]} \big(\phi(u)+\eta(t)-\eta(u)\big)\right]. \end{aligned}$$

The term $\big(\phi(0)-a-\eta(0)\big)^+ + \eta(t)$ is nonnegative, so $\left[\big(\phi(0)-a-\eta(0)\big)^+ + \eta(t)\right] \wedge 0 = 0$. Therefore,

$$\Xi_a(\psi)(t) = \phi(t) - \sup_{s\in[0,t]} \left[\big(\phi(s)-a+\eta(t)-\eta(s)\big) \wedge \inf_{u\in[s,t]} \big(\phi(u)+\eta(t)-\eta(u)\big)\right]^+.$$

We conclude the proof that this last expression is $\overline{\phi}(t) \triangleq \Lambda_a(\phi)(t)$ by showing that

$$\left[\big(\phi(s)-a+\eta(t)-\eta(s)\big) \wedge \inf_{u\in[s,t]} \big(\phi(u)+\eta(t)-\eta(u)\big)\right]^+ = \big(\phi(s)-a\big)^+ \wedge \inf_{u\in[s,t]} \phi(u). \qquad (2.2)$$

There are two possibilities. The first is that $\phi(u) \triangleq \Gamma_0(\psi)(u) > 0$ for every $u \in [s,t]$. According to the complementarity condition (1.3), η is constant on $[s,t]$, and the left-hand side of (2.2) becomes $\left[\big(\phi(s)-a\big) \wedge \inf_{u\in[s,t]} \phi(u)\right]^+$, which agrees with the right-hand side.

The other possibility is that $\phi(u)=0$ for some $u\in[s,t]$. Define $u_*=\sup\{u\in[s,t]:\phi(u)=0\}$. According to the complementarity condition (1.3), either $\phi(u_*)=0$ and η is constant on $[u_*,t]$ or else $u_*>s$, $\phi(u_*-)=0$, $\phi(u_*)>0$, η is constant on $[u_*,t]$, and η is continuous at u_*. In either case $\inf_{u\in[s,t]}(\phi(u)+\eta(t)-\eta(u))=0$ and $\inf_{u\in[s,t]}\phi(u)=0$, so (2.2) holds with both sides equal to zero. □

Remark 2.2 *If $\psi(0)\le 0$, then*

$$\big(\psi(0)-a\big)^+\wedge\inf_{u\in[0,t]}\psi(u)=\inf_{u\in[0,t]}\psi(u),$$

and

$$\begin{aligned}\Xi_a(\psi)(t) &= \psi(t)-\inf_{u\in[0,t]}\psi(u)\vee\sup_{s\in[0,t]}\left[(\psi(s)-a)\wedge\inf_{u\in[s,t]}\psi(u)\right]\\ &= \psi(t)-\sup_{s\in[0,t]}\left[\Big((\psi(s)-a)\vee\inf_{u\in[0,t]}\psi(u)\Big)\wedge\Big(\inf_{u\in[s,t]}\psi(u)\vee\inf_{u\in[0,t]}\psi(u)\Big)\right]\\ &= \psi(t)-\sup_{s\in[0,t]}\left[\Big((\psi(s)-a)\vee\inf_{u\in[0,t]}\psi(u)\Big)\wedge\inf_{u\in[s,t]}\psi(u)\right]. \qquad (2.3)\end{aligned}$$

Example 2.3 We provide here an example illustrating the fact that $\Xi_a=\Lambda_a\circ\Gamma_0$, and demonstrating in addition that $\Xi_a\neq\Lambda_a$. Let $a=1$ and

$$\psi(t)=\begin{cases}-2+t, & 0\le t\le 4,\\ 6-t, & 4\le t\le 6.\end{cases} \qquad (2.4)$$

For $0\le t\le 6$, we have $\inf_{u\in[0,t]}\psi(u)=\psi(0)=-2$. It is straightforward to compute

$$\sup_{s\in[0,t]}\left[\Big(\big(\psi(s)-1\big)\vee(-2)\Big)\wedge\inf_{u\in[s,t]}\psi(u)\right]=\begin{cases}-2, & 0\le t\le 1,\\ -3+t, & 1\le t\le 4,\\ 1, & 4\le t\le 5,\\ 6-t, & 5\le t\le 6.\end{cases}$$

According to (2.3),

$$\Xi_a(\psi)(t)=\psi(t)-\sup_{s\in[0,t]}\left[\Big(\big(\psi(s)-1\big)\vee(-2)\Big)\wedge\inf_{u\in[s,t]}\psi(u)\right]=\begin{cases}t, & 0\le t\le 1,\\ 1, & 1\le t\le 4,\\ 5-t, & 4\le t\le 5,\\ 0, & 5\le t\le 6.\end{cases}$$

We see that $\Xi_a(\psi)\neq\Lambda_a(\psi)$ because

$$\sup_{s\in[0,t]}\left[\big(\psi(s)-1\big)^+\wedge\inf_{u\in[s,t]}\psi(u)\right]=\begin{cases}-2+t, & 0\le t\le 2,\\ 0, & 2\le t\le 3,\\ -3+t, & 3\le t\le 4,\\ 1, & 4\le t\le 5,\\ 6-t, & 5\le t\le 6,\end{cases}$$

so

$$\Lambda_a(\psi)(t)=\psi(t)-\sup_{s\in[0,t]}\left[\big(\psi(s)-1\big)^+\wedge\inf_{u\in[s,t]}\psi(u)\right]=\begin{cases}0, & 0\le t\le 2,\\ -2+t, & 2\le t\le 3,\\ 1, & 3\le t\le 4,\\ 5-t, & 4\le t\le 5,\\ 0, & 5\le t\le 6.\end{cases}$$

The discrepancy between $\Xi_a(\psi)$ and $\Lambda_a(\psi)$ is due to the fact that ψ can take negative values. If ψ is nonnegative, then $\Gamma_0(\psi) = \psi$, and Theorem 2.1 implies $\Xi_a(\psi) = \Lambda_a \circ \Gamma_0(\psi) = \Lambda_a(\psi)$. For ψ given by (2.4),

$$\phi(t) \triangleq \Gamma_0(\psi)(t) = \psi(t) + 2 = \begin{cases} t, & 0 \le t \le 4, \\ 8 - t, & 4 \le t \le 6, \end{cases}$$

and

$$\sup_{s\in[0,t]} \left[\big(\phi(s) - 1\big)^+ \wedge \inf_{u\in[s,t]} \phi(u) \right] = \begin{cases} 0, & 0 \le t \le 1, \\ -1 + t, & 1 \le t \le 4, \\ 3, & 4 \le t \le 5, \\ 8 - t, & 5 \le t \le 6. \end{cases}$$

Therefore,

$$\Lambda_a(\phi)(t) = \phi(t) - \sup_{s\in[0,t]} \left[\big(\phi(s) - 1\big)^+ \wedge \inf_{u\in[s,t]} \phi(u) \right] = \begin{cases} t, & 0 \le t \le 1, \\ 1, & 1 \le t \le 4, \\ 5 - t, & 4 \le t \le 5, \\ 0, & 5 \le t \le 6. \end{cases}$$

This illustrates the result $\Xi_a(\psi) = \Lambda_a \circ \Gamma_0(\psi) = \Lambda_a(\phi)$ of Theorem 2.1. □

3 The formula of Cooper, Schmidt and Serfozo [3]

Following Cooper, Schmidt and Serfozo [3], we let H be a signed measure on the Borel subsets of $\mathbb{R}_+$ whose total variation on each compact interval is finite. The function $t \mapsto H(0,t]$ is right-continuous with left limits and is of bounded variation. We denote this function by $H(0,\cdot]$. Let a be a positive number, and let $x \in [-a, 0]$ be given. Cooper et. al. [3] (equation [15]) define

$$X(t) \triangleq \sup_{s\in[0,t]} \inf_{u\in[s,t]} \left[x\mathbb{I}_{\{s=u=0\}} + H(u,t] - a\mathbb{I}_{\{s=u>0\}} \right] \tag{3.1}$$

and show that $X = \Gamma_{-a,0}\big(x + H(0,\cdot]\big)$. In particular, $X(0) = x$.

Negating (3.1), we obtain

$$-X(t) = -\sup_{s\in[0,t]} \inf_{u\in[s,t]} \left[x\mathbb{I}_{\{s=u=0\}} + H(u,t] - a\mathbb{I}_{\{s=u>0\}} \right], \tag{3.2}$$

and the result in [3] implies that

$$-X = \Gamma_{0,a}\big(-x - H(0,\cdot]\big). \tag{3.3}$$

In particular, $-X(0) = -x$.

To relate (3.2) to Ξ_a, we let ψ be a bounded variation function in $\mathcal{D}[0,\infty)$. We can then define the signed measure H by

$$H(u,t] = \psi(u) - \psi(t), \quad 0 \le u \le t. \tag{3.4}$$

The number $-x$ in (3.3) must be taken to be in the interval $[0,a]$. We define $-x$ in terms of ψ by

$$-x = \Gamma_{0,a}(\psi)(0) = \big[\psi(0)\big]^+ \wedge a. \tag{3.5}$$

It is then easily verified that

$$x + \psi(0) = \big(\psi(0) - a\big)^+ \wedge \psi(0). \tag{3.6}$$

With the choices of H and $-x$ given by (3.4) and (3.5), (3.2) becomes

$$\begin{aligned}
-X(t) &= -\sup_{s\in[0,t]}\inf_{u\in[s,t]}\left[x\mathbb{I}_{\{s=u=0\}}+\psi(u)-\psi(t)-a\mathbb{I}_{\{s=u>0\}}\right]\\
&= \psi(t)-\inf_{u\in[0,t]}\left[x\mathbb{I}_{\{u=0\}}+\psi(u)\right]\vee\sup_{s\in(0,t]}\inf_{u\in[s,t]}\left[\psi(u)-a\mathbb{I}_{\{s=u\}}\right]\\
&= \psi(t)-\left[\big(x+\psi(0)\big)\wedge\inf_{u\in(0,t]}\psi(u)\right]\vee\sup_{s\in(0,t]}\left[\big(\psi(s)-a\big)\wedge\inf_{u\in(s,t]}\psi(u)\right]\\
&= \psi(t)-\left[\big(\psi(0)-a\big)^+\wedge\psi(0)\wedge\inf_{u\in(0,t]}\psi(u)\right]\vee\sup_{s\in(0,t]}\left[\big(\psi(s)-a\big)\wedge\inf_{u\in(s,t]}\psi(u)\right]\\
&= \psi(t)-\left[\big(\psi(0)-a\big)^+\wedge\inf_{u\in[0,t]}\psi(u)\right]\vee\sup_{s\in(0,t]}\left[\big(\psi(s)-a\big)\wedge\inf_{u\in(s,t]}\psi(u)\right]. \qquad (3.7)
\end{aligned}$$

Because ψ is right continuous, for $0\le s\le t$,

$$\big(\psi(s)-a\big)\wedge\inf_{u\in(s,t]}\psi(u)=\big(\psi(s)-a\big)\wedge\inf_{u\in[s,t]}\psi(u),$$

where we adopt the convention that $\inf_{u\in(t,t]}\psi(u)=\infty$ to handle the case $s=t$. Furthermore, this expression is right-continuous in s. Therefore, for $t>0$,

$$\sup_{s\in(0,t]}\left[\big(\psi(s)-a\big)\wedge\inf_{u\in(s,t]}\psi(u)\right]=\sup_{s\in[0,t]}\left[\big(\psi(s)-a\big)\wedge\inf_{u\in[s,t]}\psi(u)\right].$$

We thus conclude from (3.7) that

$$-X(t)=\psi(t)-\left[\big(\psi(0)-a\big)^+\wedge\inf_{u\in[0,t]}\psi(u)\right]\vee\sup_{s\in[0,t]}\left[\big(\psi(s)-a\big)\wedge\inf_{u\in[s,t]}\psi(u)\right]=\Xi_a(\psi)(t). \quad (3.8)$$

To deal with the case that $t=0$, it is easily verified that

$$\Xi_a(\psi)(0)=\psi(0)-(\psi(0)-a)^+\wedge\psi(0)=\big[\psi(0)\big]^+\wedge a=\Gamma_{0,a}(\psi)(0)=\Lambda_a\circ\Gamma_0(\psi)(0)=-X(0). \quad (3.9)$$

□

4 The formula of Ganesh, O'Connell and Wischik [5]

Section 5.7 of [5] records the size of a finite-buffer queue at time zero under the assumption that the queue was empty at time $-t$, where $t>0$. The buffer size of the queue is a, a positive number. We adjust the formula in [5] by relabeling time; our queue is empty at time zero and we record its size at time t. In [5], the arrivals and departures take place at discrete times, so the cumulative arrivals, the cumulative offered service, and the so-called netput, the difference between cumulative arrivals and cumulative offered service, is piecewise continuous and of bounded variation. (Offered service is service received, unless the queue is empty, in which case offered service is wasted.) We call the netput process ψ. The queue length is then $\Gamma_{0,a}(\psi)$.

We thus begin with a bounded-variation function $\psi\in\mathcal{D}[0,\infty)$ satisfying $\psi(0)=0$. Following [5], we define

$$M(s,t)\triangleq\inf\nolimits_{u\in[s,t]}\big(\psi(t)-\psi(u)\big)=\psi(t)-\sup\nolimits_{u\in[s,t]}\psi(u),$$
$$N(s,t)\triangleq\sup\nolimits_{u\in[s,t]}\big(\psi(t)-\psi(u)\big)=\psi(t)-\inf\nolimits_{u\in[s,t]}\psi(u).$$

We note that like ψ itself, $M(s,t)$ and $N(s,t)$ are right-continuous with left-hand limits in s. Here and elsewhere, we adopt the notational conventions

$$\sup_{u\in[s-,t]}\psi(u)\triangleq\lim_{v\uparrow s}\sup_{u\in[v,t]}\psi(u),\qquad \inf_{u\in[s-,t]}\psi(u)\triangleq\lim_{v\uparrow s}\inf_{u\in[v,t]}\psi(u), \tag{4.1}$$

We shall also use the notation $(s-,t]\triangleq[s,t]$ and $0-=0$.

Lemma 4.1 *For all $t\ge 0$,*

$$\sup_{s\in[0,t]}\big[N(s,t)\wedge\big(M(s,t)+a\big)\big]\le\inf_{s\in[0,t]}\big[N(s,t)\vee\big(M(s,t)+a\big)\big]. \tag{4.2}$$

PROOF: For $t\ge 0$,

$$\begin{aligned}\sup_{s\in[0,t]}\big[N(s,t)\wedge\big(M(s,t)+a\big)\big] &= \psi(t)+\sup_{s\in[0,t]}\left[-\inf_{u\in[s,t]}\psi(u)\wedge\big(a-\sup_{u\in[s,t]}\psi(u)\big)\right]\\ &= \psi(t)-\inf_{s\in[0,t]}\left[\inf_{u\in[s,t]}\psi(u)\vee\sup_{u\in[s,t]}\big(\psi(u)-a\big)\right]. \end{aligned}\tag{4.3}$$

The infimum over $s\in[0,t]$ in the last line of (4.3) is either attained by some $s_1\in[0,t]$, or else there is an s_1 in $(0,t]$ for which the infimum is (in the notation (4.1))

$$\inf_{u\in[s_1-,t]}\psi(u)\vee\sup_{u\in[s_1-,t]}\big(\psi(u)-a\big).$$

In the former case, we let s_1' denote s_1; in the latter case, s_1' denotes s_1-. Capturing both cases, we say that $s_1'\in[0,t]$ satisfies

$$\inf_{u\in[s_1',t]}\psi(u)\vee\sup_{u\in[s_1',t]}\big(\psi(u)-a\big)=\inf_{s\in[0,t]}\left[\inf_{u\in[s,t]}\psi(u)\vee\sup_{u\in[s,t]}\big(\psi(u)-a\big)\right].$$

Continuing in this way, we observe that if $s_1'=s_1$, then there is an $s_*\in[s_1,t]$ that attains $\inf_{u\in[s_1,t]}\psi(u)$ or else there is an $s_*\in(s_1,t]$ for which the infimum is $\psi(s_*-)$. If $s_1'=s_1-$, then either there is an $s_*\in[s_1,t]$ that attains $\inf_{[s_1-,t]}\psi(u)$ or else there is $s_*\in[s_1,t]$ for which the infimum is $\psi(s_*-)$. If $\psi(s_*)=\inf_{u\in[s_1',t]}\psi(u)$, we set $s_*'=s_*$; if $\psi(s_*-)=\inf_{u\in[s_1',t]}\psi(u)$, s_*' denotes s_*-. Capturing all these cases, we say that $s_*'\in[s_1',t]$ satisfies $\psi(s_*')=\inf_{u\in[s_1',t]}\psi(u)$. With these conventions, we have

$$\begin{aligned}\inf_{s\in[0,t]}\left[\inf_{u\in[s,t]}\psi(u)\vee\sup_{u\in[s,t]}\big(\psi(u)-a\big)\right] &= \inf_{u\in[s_1',t]}\psi(u)\vee\sup_{u\in[s_1',t]}\big(\psi(u)-a\big)\\ &\ge \psi(s_*')\vee\sup_{u\in[s_*',t]}\big(\psi(u)-a\big)\\ &\ge \inf_{s\in[0,t]}\left[\psi(s)\vee\sup_{u\in[s,t]}\big(\psi(u)-a\big)\right].\end{aligned}$$

The reverse inequality

$$\inf_{s\in[0,t]}\left[\inf_{u\in[s,t]}\psi(u)\vee\sup_{u\in[s,t]}\big(\psi(u)-a\big)\right]\le\inf_{s\in[0,t]}\left[\psi(s)\vee\sup_{u\in[s,t]}\big(\psi(u)-a\big)\right]$$

obviously holds. Returning to (4.3), we see that

$$\sup_{s\in[0,t]} \big[N(s,t) \wedge \big(M(s,t)+a\big)\big] = \psi(t) - \inf_{s\in[0,t]} \left[\psi(s) \vee \sup_{u\in[s,t]} \big(\psi(u)-a\big)\right]. \tag{4.4}$$

An analogous argument shows that

$$\inf_{s\in[0,t]} \big[N(s,t) \vee \big(M(s,t)+a\big)\big] = \psi(t) - \sup_{s\in[0,t]} \left[\big(\psi(s)-a\big) \wedge \inf_{u\in[s,t]} \psi(u)\right]. \tag{4.5}$$

Now choose $s_3 \in [0,t]$, where s_3 attains the infimum on the right-hand side of (4.4) (in which case we write $s_3' = s_3$) or if no such s_3 exists, then choose $s_3 \in (0,t]$, where s_3- attains the infimum on the right-hand side of (4.4) (in which case we write $s_3' = s_3-$). Let s_4' be defined analogously in connection with the supremum on the right-hand side of (4.5). If $s_3' \le s_4'$ (this means either that $s_3 < s_4$ or else that $s_3 = s_4$ and it is not the case that $s_3' = s_3$, $s_4' = s_4-$), we have $\sup_{u\in[s_3',t]} \big(\psi(u)-a\big) \ge \psi(s_4') - a$, and so

$$\begin{aligned}
\inf_{s\in[0,t]} \left[\psi(s) \vee \sup_{u\in[s,t]} \big(\psi(u)-a\big)\right] &= \psi(s_3') \vee \sup_{u\in[s_3',t]} \big(\psi(u)-a\big) \\
&\ge \big(\psi(s_4')-a\big) \wedge \inf_{u\in[s_4',t]} \psi(u) \\
&= \sup_{s\in[0,t]} \left[\big(\psi(s)-a\big) \wedge \inf_{u\in[s,t]} \psi(u)\right].
\end{aligned} \tag{4.6}$$

Relation (4.2) follows from (4.4) and (4.5). On the other hand, if $s_3' \ge s_4'$, then $\psi(s_3') \ge \inf_{u\in[s_4',t]} \psi(u)$, and again relation (4.6) and hence relation (4.2) hold. $\square$

For $x \in \mathbb{R}$ and $\alpha \le \beta$, define $[x]_\alpha^\beta = (x \vee \alpha) \wedge \beta$. On a subset of bounded-variation functions in $\mathcal{D}[0,\infty)$ whose initial condition is zero, in [5] a mapping Φ_a is defined by the formula

$$\Phi_a(\psi)(t) \triangleq \big[\psi(t)\big]_{\sup_{s\in[0,t]}[N(s,t)\wedge(M(s,t)+a)]}^{\inf_{s\in[0,t]}[N(s,t)\vee(M(s,t)+a)]}. \tag{4.7}$$

According to this definition and relations (4.4) and (4.5),

$$\begin{aligned}
&\Phi_a(\psi)(t) \\
&\quad= \left(\psi(t) \vee \sup_{s\in[0,t]} \big[N(s,t) \wedge \big(M(s,t) \wedge a\big)\big]\right) \wedge \inf_{s\in[0,t]} \big[N(s,t) \vee \big(M(s,t)+a\big)\big] \\
&\quad= \left\{\psi(t) \vee \left(\psi(t) - \inf_{s\in[0,t]} \left[\psi(s) \vee \sup_{u\in[s,t]} \big(\psi(u)-a\big)\right]\right)\right\} \\
&\qquad\quad \wedge \left(\psi(t) - \sup_{s\in[0,t]} \left[\big(\psi(s)-a\big) \wedge \inf_{u\in[s,t]} \psi(u)\right]\right) \\
&\quad= \psi(t) - \left(0 \wedge \inf_{s\in[0,t]} \left[\psi(s) \vee \sup_{u\in[s,t]} \big(\psi(u)-a\big)\right]\right) \vee \sup_{s\in[0,t]} \left[\big(\psi(s)-a\big) \wedge \inf_{u\in[s,t]} \psi(u)\right].
\end{aligned} \tag{4.8}$$

Theorem 4.2 *Let $\psi \in \mathcal{D}[0,\infty)$ satisfy $\psi(0)=0$. Then $\Phi_a(\psi) = \Xi_a(\psi)$, where $\Xi_a(\psi)$ is given by (2.1).*

Proof: According to (4.8),

$$\psi(t) - \Phi_a(\psi)(t) = \big(0 \wedge A(t)\big) \vee B(t),$$

where

$$A(t) = \inf_{s\in[0,t]} \left[\psi(s) \vee \sup_{u\in[s,t]} \big(\psi(u)-a\big)\right], \quad B(t) = \sup_{s\in[0,t]} \left[\big(\psi(s)-a\big) \wedge \inf_{u\in[s,t]} \psi(u)\right].$$

Using $\psi(0)=0$, we obtain from equation (2.1) that

$$\psi(t) - \Xi_a(\psi)(t) = C(t) \vee B(t),$$

where

$$C(t) = \inf_{s\in[0,t]} \psi(s).$$

To prove the theorem, we must show that

$$\big(0 \wedge A(t)\big) \vee B(t) = C(t) \vee B(t). \tag{4.9}$$

Clearly, $A(t) \geq C(t)$, and because $\psi(0)=0$, we have also $0 \geq C(t)$. It follows that $\big(0\wedge A(t)\big) \geq C(t)$, and thus

$$\big(0 \wedge A(t)\big) \vee B(t) \geq C(t) \vee B(t). \tag{4.10}$$

If $A(t) = C(t)$, then $0 \wedge A(t) = C(t)$ and so equality holds in (4.10). To complete the proof, we establish the implication

$$A(t) > C(t) \quad \Longrightarrow \quad A(t) \leq B(t). \tag{4.11}$$

If $A(t) > C(t)$, (4.11) will imply $\big(0 \wedge A(t)\big) \vee B(t) \leq B(t) \leq C(t) \vee B(t)$, and we have the reverse of (4.10).

Assume

$$A(t) > C(t). \tag{4.12}$$

Using the notation developed for Lemma 4.1, we choose $s_1 \in [0,t]$ so that $\psi(s_1) = C(t)$ or $s_1 \in (0,t]$ so that $\psi(s_1-) = C(t)$. We use s_1' to denote s_1 in the former case and s_1- in the latter case. We capture both cases by the equation

$$\psi(s_1') = C(t) \leq \psi(s) \quad \forall s \in [0,t]. \tag{4.13}$$

We next define

$$s_2 \triangleq \sup\left\{ s\in[0,t] \,\middle|\, \left[\psi(s) \vee \sup_{u\in[s,t]}\big(\psi(u)-a\big)\right] \wedge \left[\psi(s-) \vee \sup_{u\in[s-,t]}\big(\psi(u)-a\big)\right] = A(t) \right\}.$$

Either $s_2 \in [0,t]$ and $\psi(s_2) \vee \sup_{u\in[s_2,t]}\big(\psi(u)-a\big) = A(t)$, in which case we denote s_2 by s_2', or else $s_2 \in (0,t]$, $\psi(s_2) \vee \sup_{u\in[s_2,t]}\big(\psi(u)-a\big) > A(t)$, and $\psi(s_2-) \vee \sup_{u\in[s_2-,t]}\big(\psi(u)-a\big) = A(t)$, in which case we denote s_2- by s_2'. We capture both cases by the equation

$$\psi(s_2') \vee \sup_{u\in[s_2',t]} \big(\psi(u)-a\big) = A(t). \tag{4.14}$$

We cannot have $s_2' < s_1'$ (which means $s_2 < s_1$ or $s_2 = s_1$, $s_2' = s_2-$, $s_1' = s_1$), for then we would have, using (4.13) and the definition of $A(t)$,

$$A(t) \le \psi(s_1') \vee \sup_{u\in[s_1',t]} \big(\psi(u) - a\big) \le \psi(s_2') \vee \sup_{u\in[s_2',t]} \big(\psi(u) - a\big) = A(t),$$

a contradiction to the maximality property of s_2'. Therefore, $s_2' \ge s_1'$.

We must have

$$\psi(s) \ge \psi(s_2') \quad \forall s \in [s_2, t]. \tag{4.15}$$

If this were not the case, then we would have $\psi(s) < \psi(s_2')$ for some $s \in (s_2', t]$, and then

$$A(t) \le \psi(s) \vee \sup_{u\in[s,t]} \big(\psi(u) - a\big) \le \psi(s_2') \vee \sup_{u\in[s_2',t]} \big(\psi(u) - a\big) = A(t),$$

which also contradicts the maximality property of s_2'.

Case I: $A(t) = \sup_{u\in[s_2',t]} \big(\psi(u) - a\big) \ge \psi(s_2')$.

Define

$$u_2 \triangleq \sup\big\{\, u \in [s_2, t] \big|\, \big(\psi(u) - a\big) \vee \big(\psi(u-) - a\big) = A(t) \big\}\,.$$

Then either $\psi(u_2) - a = A(t)$ or else $u_2 > 0$, $\psi(u_2) - a < A(t)$, $\psi(u_2-) - a = A(t)$. Let us consider first the case that $\psi(u_2) - a = A(t)$, in which case we denote u_2 by u_2'. There cannot exist $u_3 \in (u_2, t]$ for which $\psi(u_3) < A(t)$, for if such a u_3 were to exist, we would have $\psi(u_3) \vee \sup_{u\in[u_3,t]} \big(\psi(u) - a\big) < A(t)$. Therefore,

$$\psi(u) \ge A(t) \quad \forall u \in [u_2, t]. \tag{4.16}$$

Let us next consider the case that $u_2 > 0$, $\psi(u_2) - a < A(t)$ and $\psi(u_2-) - a = A(t)$, in which case we denote u_2- by u_2'. There cannot exist $u_3 \in [u_2, t]$ such that $\psi(u_3) < A(t)$, for if such a u_3 were to exist, we would again have $\psi(u_3) \vee \sup_{u\in[u_3,t]} \big(\psi(u) - a\big) < A(t)$. Once again, (4.16) holds. From (4.16) and the fact that $\psi(u_2') - a = A(t)$, we have immediately

$$B(t) \ge \big(\psi(u_2') - a\big) \wedge \inf_{u\in[u_2',t]} \psi(u) = A(t).$$

This completes the proof of (4.11) in Case I.

Case II: $A(t) = \psi(s_2') > \sup_{u\in[s_2',t]} \big(\psi(u) - a\big)$.

Define

$$u_1 \triangleq \sup\big\{\, u \in [s_1, s_2] \big|\, \big(\psi(u) - a\big) \vee \big(\psi(u-) - a\big) \ge \psi(s_2') \big\}\,.$$

If no such u_1 were to exist, then we would have $\big(\psi(u)-a\big) \vee \big(\psi(u-)-a\big) < \psi(s_2')$ for all $u \in [s_1, s_2]$, in which case we would have from (4.12), (4.13), and the case assumption that

$$\psi(s_1') \vee \sup_{u\in[s_1',t]} \big(\psi(u) - a\big) = \psi(s_1') \vee \sup_{u\in[s_1',s_2]} \big(\psi(u) - a\big) \vee \sup_{u\in[s_2,t]} \big(\psi(u) - a\big) < \psi(s_2').$$

But according to the definition of $A(t)$, it is dominated by the left-hand side of this expression. We have a contradiction to the case assumption, which shows that $u_1 \in [s_1, s_2]$ is well defined.

If $\psi(u_1) - a \ge \psi(s_2')$, we denote u_1 by u_1'. If this is not the case, then $u_1 > 0$, $\psi(u_1) - a < \psi(s_2')$, $\psi(u_1-) - a = \psi(s_2')$, and we denote u_1- by u_1'. We capture both cases by the equation

$$\psi(u_1') - a \ge \psi(s_2') = A(t). \tag{4.17}$$

The maximality property of u_1' implies that

$$\sup_{u\in(u_1',s_2]} \psi(u) - a < \psi(s_2').$$

If $\psi(u_3) < \psi(s_2')$ for some $u_3 \in (u_1', s_2]$, then we would have

$$A(t) \le \psi(u_3) \vee \sup_{u\in[u_3,t]} \big(\psi(u) - a\big) < \psi(s_2'),$$

a contradiction to the case assumption. Therefore,

$$\psi(u) \ge \psi(s_2') \quad \forall u \in (u_1', s_2]. \tag{4.18}$$

It follow that

$$\begin{aligned} B(t) &\ge \big(\psi(u_1') - a\big) \wedge \inf_{u\in[u_1',t]} \psi(u) \\ &\ge \big(\psi(u_1') - a\big) \wedge \psi(u_1') \wedge \inf_{u\in(u_1',s_2]} \psi(u) \wedge \inf_{u\in[s_2,t]} \psi(u), \end{aligned}$$

and each of these terms dominates $\psi(s_2') = A(t)$ by (4.17), (4.18), and (4.15). This completes the proof of (4.11) in Case II, and thus completes the proof of Theorem 4.2. □

5 The formula of Toomey [12]

Toomey [12] records the size of a finite-buffer queue at time $-k$ under the assumption that the queue was of size q_m at time $-m < -k$. The buffer size of the queue is a, a positive number, and q_m is assumed to be in $[0, a]$. There are two formulas, (4) and (5), in [12], each obtained from the other by reversing the spatial axis. We deal with (5), mapping $-m$ into time zero and mapping $-k$ into time $t > 0$ and writing the formula for piecewise constant functions in $\mathcal{D}[0,\infty)$ rather than functions defined on the integers.

The netput process, cumulative arrivals minus offered service, over the time interval $-k$ to $-m$ is denoted U_{km} by [12] and by $\psi(t) - \psi(0)$ here. We take $\psi(0) = q_m \in [0, a]$, the initial queue length. In our notation, formula (5) in [12] is

$$\begin{aligned} &\inf_{s\in(0,t]} \sup_{u\in(s,t]} \Big[\big(a + \psi(t) - \psi(s)\big) \vee \big(\psi(t) - \psi(u)\big)\Big] \\ &\quad \wedge \sup_{u\in(0,t]} \Big[\big(q_m + \psi(t) - \psi(0)\big) \vee \big(\psi(t) - \psi(u)\big)\Big] \\ &\quad\quad = \psi(t) - \sup_{s\in(0,t]} \inf_{u\in(s,t]} \Big[\big(\psi(s) - a) \wedge \psi(u)\Big] \vee \inf_{u\in(0,t]} \left[0 \wedge \psi(u)\right]. \end{aligned}$$

Because ψ is right-continuous and $(\psi(0) - a)^+ = 0$, this expression can be rewritten as

$$\psi(t) - \sup_{s\in[0,t]} \left[\big(\psi(s) - a\big) \wedge \inf_{u\in[s,t]} \psi(u)\right] \vee \left[\big(\psi(0) - a\big)^+ \wedge \inf_{u\in[0,t]} \psi(u)\right],$$

which is $\Xi_a(\psi)(t)$ given by (2.1).

6 Application to real-time queues with reneging

6.1 Heavy-traffic convergence

Consider a sequence of single station queueing systems indexed by the positive integers. In the n-th system, the the *interarrival times* are a sequence of positive, independent, identically distributed random variables $u_1^{(n)}, u_2^{(n)}, \ldots$ and the *service times* are likewise a sequence of positive, independent, identically distributed random variables $v_1^{(n)}, v_2^{(n)}, \ldots$. The *arrival rate* in the n-th system is $\lambda^{(n)} = 1/\mathbb{E}u_i^{(n)}$, the *service rate* is $\mu^{(n)} = 1/\mathbb{E}v_i^{(n)}$, and the *traffic intensity* is $\rho^{(n)} \triangleq \lambda^{(n)}/\mu^{(n)}$. We assume that $\lambda^{(n)}$ has a positive limit λ as $n \to \infty$, $\mu^{(n)}$ also has a positive limit μ as $n \to \infty$, $u_i^{(n)}$ has a limiting positive variance α^2 as $n \to \infty$, and $v_i^{(n)}$ has a limiting variance β^2 as $n \to \infty$. We make the *heavy traffic assumption*

$$\rho^{(n)} = 1 - \frac{\gamma}{\sqrt{n}} \tag{6.1}$$

for some nonzero constant γ. This implies $\lambda = \mu$.

For the n-th system, the *customer arrival times* are

$$S_k^{(n)} \triangleq \sum_{i=1}^{k} u_i^{(n)}$$

and the *customer arrival process* is

$$A^{(n)}(t) \triangleq \max\left\{k \,\middle|\, S_k^{(n)} \le t\right\}.$$

The *work arrival process* is

$$V^{(n)}(k) \triangleq \sum_{j=1}^{k} v_j^{(n)}.$$

The *netput process*

$$N^{(n)}(t) \triangleq V^{(n)}\big(A^{(n)}(t)\big) - t$$

represents the work that would be present in queue at time t if the server were never idle between times 0 and t. We are taking the queue to be empty at time zero. However, the queue may be idle prior to time t, and thus the work that is actually present at time t is given by the *workload process*

$$W^{(n)} \triangleq \Gamma_0(N^{(n)}), \tag{6.2}$$

where Γ_0 is defined by (1.2). The *idleness process*

$$I(t) \triangleq -\inf_{s\in[0,t]} N^{(n)}(s)$$

plays the role of η of (1.1).

The *scaled workload process* is

$$\widehat{W}^{(n)}(t) \triangleq \frac{1}{\sqrt{n}} W^{(n)}(nt), \quad t \ge 0. \tag{6.3}$$

It is well known that the following heavy traffic convergence result holds under assumption (6.1) and the Lindeberg condition

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{E}\left[\left(u_j^{(n)} - (\lambda^{(n)})^{-1}\right)^2 \mathbb{I}_{\left\{\left|u_j^{(n)}-(\lambda^{(n)})^{-1}\right|>c\sqrt{n}\right\}}\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\left(v_j^{(n)} - (\mu^{(n)})^{-1}\right)^2 \mathbb{I}_{\left\{\left|v_j^{(n)}-(\mu^{(n)})^{-1}\right|>c\sqrt{n}\right\}}\right] = 0 \quad \forall c > 0.
\end{aligned}$$

Theorem 6.1 (Kingman [7], Iglehart and Whitt [6]) *As $n \to \infty$,*

$$\widehat{W}^{(n)} \Rightarrow W^*, \tag{6.4}$$

where W^* is a Brownian motion with drift $-\gamma$ and variance per unit time $\lambda(\alpha^2+\beta^2)$, reflected at the origin so as to always be nonnegative. More precisely, define $N^*(t) \triangleq -\gamma t + \sqrt{\lambda(\alpha^2+\beta^2)}\,B(t)$, where B is a standard Brownian motion. Then

$$W^* \triangleq \Gamma_0(N^*). \tag{6.5}$$

6.2 Lead times

In *real-time queues*, customer deadlines are taken into account. We introduce a sequence of *lead times*, which are positive, independent and identically distributed random variables $L_1^{(n)}, L_2^{(n)}, \ldots$. Customer k arrives in system n at time $S_k^{(n)}$ with deadline $S_k^{(n)} + L_k^{(n)}$. The lead time of customer k, which is the time until the customer's deadline elapses, is $L_k^{(n)}$ upon arrival of customer k and then decreases at rate one thereafter, becoming negative when the customer becomes late.

Under the heavy traffic assumption (6.1), delay in the n-th system will be of order $\sqrt{n}$, so the lead times must also be of order $\sqrt{n}$ to avoid trivialities. We assume therefore that there is a cumulative distribution function G independent of n such that

$$\mathbb{P}\left\{\frac{L_j^{(n)}}{\sqrt{n}} \le y\right\} = G(y). \tag{6.6}$$

For technical reasons, we also assume that there exists a finite y^* for which $G(y^*) = 1$ and $G(y) < 1$ for $y < y^*$, i.e., we cannot have a lead time in the n-th system larger than $\sqrt{n}\,y^*$ but we can have lead times equal to or at least arbitrarily close to $\sqrt{n}\,y^*$.

We serve the customers using the *earliest deadline first (EDF)* protocol. The customer in service may be preempted by the arrival of a more urgent customer. When service eventually resumes on the preempted customer, the service begins where it left off, i.e., the work already done on that customer is not lost. We wish to determine the heavy traffic limit of the distribution of the lead times of customers in queue.

We define two measure-valued processes, $\mathcal{W}^{(n)}$ and $\mathcal{V}^{(n)}$, by specifying for every Borel subset B of $\mathbb{R}$ that

$$\mathcal{W}^{(n)}(t)(B) \triangleq \left\{\begin{array}{l}\text{Work associated with customers in}\\ \text{queue at time } t \text{ with lead times in } B\end{array}\right\}, \tag{6.7}$$

$$\mathcal{V}^{(n)}(t)(B) \triangleq \left\{\begin{array}{l}\text{Work associated with customers arrived}\\ \text{by time } t \text{ with lead times in } B, \text{ whether}\\ \text{or not still present at time } t\end{array}\right\}. \tag{6.8}$$

We define the *frontier* to be

$$F^{(n)}(t) \triangleq \left\{ \begin{array}{l} \text{Largest lead time of any customer who has ever been in} \\ \text{service, whether or not that customer is still present, or} \\ \sqrt{n}\,y^* - t \text{ if this quantity is larger than the former one} \end{array} \right\}. \tag{6.9}$$

We also define the *scaled measure-valued workload process*, *scaled measure-valued work arrival process*, and *scaled frontier*, respectively:

$$\widehat{\mathcal{W}}^{(n)}(t)(B) \triangleq \frac{1}{\sqrt{n}}\mathcal{W}^{(n)}(nt)(\sqrt{n}\,B), \tag{6.10}$$

$$\widehat{\mathcal{V}}^{(n)}(t)(B) \triangleq \frac{1}{\sqrt{n}}\mathcal{V}^{(n)}(nt)(\sqrt{n}\,B), \tag{6.11}$$

$$\widehat{F}^{(n)}(t) \triangleq \frac{1}{\sqrt{n}}F^{(n)}(nt). \tag{6.12}$$

At time $t > 0$, work whose lead time is less or equal to $F^{(n)}(t)$ receives priority. However, the arrival rate of work in this category is less than the full arrival rate, since $F^{(n)}(t) < \sqrt{n}\,y^*$. Therefore, this work is not in heavy traffic, which leads to the following result, proved in [4].

Lemma 6.2 (Crushing) *As $n \to \infty$, $\widehat{\mathcal{W}}^{(n)}(-\infty, \widehat{F}^{(n)}] \Rightarrow 0$.*

Lemma 6.2 says that in order to understand the limiting distribution of lead times, it is enough to consider only work whose lead time exceeds $F^{(n)}$. However, this work has never been in service, so we can restrict attention to the measure-valued work arrival process $\mathcal{V}^{(n)}$ rather than the more complicated measure-valued workload process $\mathcal{W}^{(n)}$. The following limit for the scaled version of this process is obtained in [4].

Theorem 6.3 *For all $y \in \mathbb{R}$,*

$$\widehat{\mathcal{V}}^{(n)}(\cdot)(y,\infty) \Rightarrow H(y) \triangleq \int_y^\infty \big(1 - G(x)\big)\,dx. \tag{6.13}$$

In fact, the convergence in (6.13) is weak convergence of a sequence of measure-valued processes to a measure on $\mathbb{R}$, not just weak convergence of a real-valued process for each fixed y.

Theorem 6.3 can be explained by the following heuristic. To have scaled lead time x at scaled time t, a customer must have entered the system scaled time units s earlier with scaled lead time $x + s$. Given that a customer arrives at scaled time $t - s$, the density at time t for the lead time at x of this customer is $G'(x + s)$. We must integrate this density over all possible values of $s \geq 0$ and multiply by the limiting arrival rate λ of customers to obtain the density of customers with scaled lead time x at scaled time t, which is therefore

$$\lambda \int_0^\infty G'(x+s)\,ds = \lambda\big(G(\infty) - G(x)\big) = \lambda\big(1 - G(x)\big).$$

The limiting work brought by each customer is $1/\mu = 1/\lambda$. Therefore, to find the density of work (as opposed to customers) with scaled lead time x at scaled time t, we divide the expression above by λ. Finally, to obtain the amount of work in (y,∞) at time t, we integrate the resulting expression from y to ∞ and obtain $H(y)$ defined in (6.13).

The limiting scaled work in the system is W^* defined by (6.5), and according to Lemma 6.2, as $n \to \infty$, this work is increasingly concentrated to the right of the frontier $\widehat{F}^{(n)}$. One can use these observations to show that $\widehat{F}^{(n)}$ has a limit, and this limit must be $F^* = H^{-1}(W^*)$ so that

$$\lim_{n\to\infty} \widehat{\mathcal{V}}^{(n)}(t)\big(F^{(n)}(t),\infty\big) = H\big(F^*(t)\big) = W^*(t).$$

In the limit, arrived work to the right of the frontier accounts for all work in the system. We summarize with the principal conclusions from [4].

Theorem 6.4 *As $n \to \infty$,*

$$\widehat{F}^{(n)} \;\Rightarrow\; F^* \triangleq H^{-1}(W^*), \tag{6.14}$$

$$\widehat{\mathcal{W}}^{(n)}(\cdot)(y,\infty) \;\Rightarrow\; H(y \vee F^*) \quad \forall y \in \mathbb{R}. \tag{6.15}$$

The convergence in (6.15) is weak convergence of a sequence of measure-valued processes to a measure-valued process. In other words, the density of the limit of the measure-valued workload processes $\widehat{\mathcal{W}}^{(n)}(t)$ is $\big(1-G(x)\big)\mathbb{I}_{\{x\geq F^*(t)\}}$, which is the density of the limit of $\mathcal{V}^{(n)}(t)$ truncated at the random process $F^*(t)$.

6.3 Reneging

We modify the real-time queueing system of the previous subsection by assuming that customers renege when they become late, i.e., a customer whose lead time reaches zero disappears from the queue never to return. This system has the same customer arrival process $A^{(n)}$, work arrival process $V^{(n)}$, and netput process $N^{(n)}$ as the system without reneging. However, its workload process, denoted $W_R^{(n)}$, is less than or equal to the workload process $W^{(n)}$ of (6.2). For the reneging system, we scale the workload process to obtain (cf. (6.3))

$$\widehat{W}_R^{(n)}(t) = \frac{1}{\sqrt{n}} W_R^{(n)}(nt), \quad t \geq 0.$$

For the reneging system, we define the *measure-valued workload process* (cf. (6.7))

$$\mathcal{W}_R^{(n)}(t)(B) = \left\{ \begin{array}{l} \text{Work associated with customers in the reneg-} \\ \text{ing system at time } t \text{ with lead times in } B \end{array} \right\}$$

and the *scaled measure-valued workload process* (cf. (6.10))

$$\widehat{\mathcal{W}}_R^{(n)}(t)(B) = \frac{1}{\sqrt{n}} \mathcal{W}_R^{(n)}(nt)(\sqrt{n}\, B).$$

Here B is an arbitrary Borel subset of $\mathbb{R}$. The *measure-valued work arrival process* $\mathcal{V}^{(n)}$ and *scaled measure-valued work arrival process* $\widehat{\mathcal{V}}^{(n)}$ for the reneging system are the same as for the non-reneging system; these are given by (6.8) and (6.11). For the reneging system, the frontier is (cf. (6.9))

$$F_R^{(n)}(t) \triangleq \left\{ \begin{array}{l} \text{Largest lead time of any customer who has ever been in service in} \\ \text{the reneging system, whether or not that customer is still present,} \\ \text{or } \sqrt{n}\, y^* - t \text{ if this quantity is larger than the former one} \end{array} \right\}.$$

Because the reneging system serves customers that have not yet been in service in the non-reneging system, we have $F^{(n)} \leq F_R^{(n)}$. The scaled frontier for the reneging system is (cf. (6.12))

$$\widehat{F}_R^{(n)}(t) = \frac{1}{\sqrt{n}} F_R^{(n)}(t).$$

Recall the definition (1.6) of Λ_a and the reflected Brownian motion W^* of (6.5). The principal result of [9] is the following.

Theorem 6.5 *As $n \to \infty$,*

$$\widehat{W}_R^{(n)} \Rightarrow W_R^* \stackrel{\Delta}{=} \Lambda_{H(0)}(W^*), \tag{6.16}$$

which is a Brownian motion with drift $-\gamma$ and variance per unit time $\lambda(\alpha^2+\beta^2)$, doubly reflected to stay in the interval $\big[0, H(0)\big]$. As $n \to \infty$,

$$\widehat{F}_R^{(n)} \Rightarrow F_R^* \stackrel{\Delta}{=} H^{-1}(W_R^*), \tag{6.17}$$

$$\widehat{\mathcal{W}}_R^{(n)}(\cdot)(y,\infty) \Rightarrow H(y \vee F_R^*) \quad \forall y \in \mathbb{R}. \tag{6.18}$$

We sketch the proof of Theorem 6.5. For this we introduce $\mathcal{M}$, the set of finite measures on the Borel subsets of $\mathbb{R}$. We endow $\mathcal{M}$ with the topology of weak convergence. We denote by $\mathcal{D}_{\mathcal{M}}[0,\infty)$ the set of functions from $[0,\infty)$ to $\mathcal{M}$ that are right-continuous and have left limits. We further define a mapping $\overline{\Lambda}\colon \mathcal{D}_{\mathcal{M}}[0,\infty) \to \mathcal{D}_{\mathcal{M}}[0,\infty)$ by

$$\overline{\Lambda}(\mu)(t)(-\infty,y] \stackrel{\Delta}{=} \left(\mu(t)(-\infty,y] - \sup_{s\in[0,t]}\left[\mu(s)(-\infty,0] \wedge \inf_{u\in[s,t]}\mu(u)(\mathbb{R})\right]\right)^+. \tag{6.19}$$

Consideration of (6.19) reveals that $\overline{\Lambda}(\mu)(t)$ is the measure on $\mathbb{R}$ that agrees with $\mu(t)$ except that it has all mass removed to the left of some point, no mass removed to the right of that point, and perhaps some of the mass removed at that point if there is a point mass there. The point in question is the supremum of those y for which

$$\mu(t)(-\infty,y] \leq \sup_{s\in[0,t]}\left[\mu(s)(-\infty,0] \wedge \inf_{u\in[s,t]}\mu(u)(\mathbb{R})\right].$$

The total amount of mass removed is almost the largest amount of "lateness" prior to time t, by which we mean $\sup_{s\in[0,t]}\mu(s)(-\infty,0]$, but this is tempered by the fact that at some time between t and the prior time s when this maximal lateness was obtained, the system may have become empty. For example, if there is an $s_1 \in [0,t]$ and a $u_1 \in [s_1,t]$ such that

$$\sup_{s\in[0,t]}\left[\mu(s)(-\infty,0] \wedge \inf_{u\in[s,t]}\mu(u)(\mathbb{R})\right] = \mu(s_1)(-\infty,0] \wedge \inf_{u\in[s_1,t]}\mu(u)(\mathbb{R}) = \mu(u_1)(\mathbb{R}),$$

then

$$\overline{\Lambda}(\mu)(u_1)(\mathbb{R}) = \left(\mu(u_1)(\mathbb{R}) - \sup_{s\in[0,u_1]}\left[\mu(s)(-\infty,0] \wedge \inf_{u\in[s,u_1]}\mu(u)(\mathbb{R})\right]\right)^+ = 0;$$

the system is empty at time u_1 and rather than subtracting mass $\mu(s_1)(-\infty,0]$ from $\mu(t)$ to obtain $\overline{\Lambda}(\mu)(t)$, we subtract only $\mu(u_1)(\mathbb{R})$, the amount removed at time u_1 in order to create the empty system. We conclude the paper with the detailed Example 6.7 of the operation of $\overline{\Lambda}$ on a path of $\mathcal{W}^{(n)}$.

Unlike Λ_a of (1.6), which maps real-valued functions to real-valued function, $\overline{\Lambda}$ maps measure-valued functions to measure-valued functions. To obtain a real-valued process, we define

$$\begin{aligned}\mathcal{U}^{(n)}(t) &\triangleq \overline{\Lambda}(\mathcal{W}^{(n)})(t),\\ U^{(n)}(t) &\triangleq \mathcal{U}^{(n)}(\mathbb{R})(t)\\ &= \left(W^{(n)}(t) - \sup_{s\in[0,t]}\left[\mathcal{W}^{(n)}(s)(-\infty,0] \wedge \inf_{u\in[s,t]} W^{(n)}(u)\right]\right)^+.\end{aligned}$$

Scaling these relations, we obtain

$$\begin{aligned}\widehat{U}^{(n)}(t) &\triangleq \frac{1}{\sqrt{n}}U^{(n)}(nt)\\ &= \left(\widehat{W}^{(n)}(t) - \sup_{s\in[0,t]}\left[\widehat{\mathcal{W}}^{(n)}(s)(-\infty,0] \wedge \inf_{u\in[s,t]} \widehat{W}^{(n)}(u)\right]\right)^+. \end{aligned} \tag{6.20}$$

Note that the processes $\widehat{W}^{(n)}(\cdot) = \widehat{\mathcal{W}}^{(n)}(\cdot)(\mathbb{R})$ and $\widehat{\mathcal{W}}^{(n)}$ appearing in (6.20) are for the non-reneging system. We take the limit as $n \to \infty$. Although $\overline{\Lambda}$ is not continuous on the set of all measure-valued processes, it is continuous on the set of processes that can result as the limit of $\mathcal{W}^{(n)}$. Therefore, we can use Theorems 6.1 and 6.4 and the continuous mapping theorem to obtain

$$\widehat{U}^{(n)}(t) \Rightarrow \left(W^*(t) - \sup_{s\in[0,t]}\left[\big(W^*(s) - H(0\vee F^*(s)\big) \wedge \inf_{u\in[s,t]} W^*(u)\right]\right)^+, \tag{6.21}$$

where we have used the fact that

$$\widehat{\mathcal{W}}^{(n)}(s)(-\infty,0] = \widehat{\mathcal{W}}^{(n)}(s)(\mathbb{R}) - \widehat{\mathcal{W}}^{(n)}(s)(0,\infty) \Rightarrow W^*(s) - H\big(0\vee F^*(s)\big).$$

Because H is nonincreasing,

$$H\big(0\vee F^*(s)\big) = H(0)\wedge H\big(F^*(s)\big) = H(0)\wedge W^*(s).$$

Therefore,

$$W^*(s) - H\big(0\vee F^*(s)\big) = W^*(s) - \big(H(0)\wedge W^*(s)\big) = \big(W^*(s) - H(0)\big)^+.$$

Making this substitution in (6.21), we see that

$$\widehat{U}^{(n)} \Rightarrow \Lambda_{H(0)}(W^*). \tag{6.22}$$

In conclusion, we have defined

$$\widehat{U}^{(n)}(t) = \overline{\Lambda}(\widehat{\mathcal{W}}^{(n)})(t)(\mathbb{R}) \quad \forall t\geq 0, \tag{6.23}$$

taken the limit as $n\to\infty$, and obtained (6.22).

The following lemma implies that the processes $\widehat{U}^{(n)}$ and $\widehat{W}_R^{(n)}$ have the same limit. In particular, (6.22) yields (6.16).

Lemma 6.6 *Let $D^{(n)}(t)$ denote the work that arrives to the reneging system that has lead time upon arrival less than or equal to the frontier at the time of arrival and that ultimately reneges. Define $\widehat{D}^{(n)}(t) = \frac{1}{\sqrt{n}} D^{(n)}(nt)$. Then*

$$0 \leq \widehat{U}^{(n)} - \widehat{W}_R^{(n)} \leq \widehat{D}^{(n)} \Rightarrow 0, \tag{6.24}$$

where the limit in (6.24) is taken as $n \to \infty$.

Just as with the non-reneging system, as $n \to \infty$, work in the reneging system concentrates to the right of the frontier $\widehat{F}_R^{(n)}$. The remainder of the argument follows as in the derivation of (6.14) and (6.15). We know that the limiting scaled work in the system is W_R^*, that this work must be concentrated to the right of the limiting frontier, and that work to the right of the frontier has never been is service and hence is just the work that has arrived. The limit of arrived work is given by Theorem 6.3, and (6.17) and (6.18) follow.

We do not attempt to prove Lemma 6.6 here. Instead, we illustrate it with the following example.

Example 6.7 Consider a system realization in which

$$\begin{aligned}
u_1^{(n)} = 1,\ v_1^{(n)} = 4,\ L_1^{(n)} = 3,\ S_1^{(n)} = 1,\\
u_2^{(n)} = 1,\ v_2^{(n)} = 4,\ L_2^{(n)} = 5,\ S_2^{(n)} = 2,\\
u_3^{(n)} = 3,\ v_3^{(n)} = 2,\ L_3^{(n)} = 1,\ S_3^{(n)} = 5,\\
u_4^{(n)} = 2,\ v_4^{(n)} = 1,\ L_4^{(n)} = 4,\ S_4^{(n)} = 7,\\
u_5^{(n)} = 2,\ v_5^{(n)} = 1,\ L_5^{(n)} = 1,\ S_5^{(n)} = 9.
\end{aligned}$$

Then using δ_s to denote a unit of mass at the point s, we have

$$\mathcal{W}^{(n)}(t) = \begin{cases} 0, & 0 \leq t < 1, \\ (5-t)\delta_{4-t}, & 1 \leq t < 2, \\ (5-t)\delta_{4-t} + 4\delta_{7-t}, & 2 \leq t < 5, \\ (7-t)\delta_{6-t} + 4\delta_{7-t}, & 5 \leq t < 7, \\ (11-t)\delta_{7-t} + \delta_{11-t}, & 7 \leq t < 9, \\ 2\delta_{-2} + \delta_1 + \delta_2, & t = 9. \end{cases}$$

The measure $\mathcal{W}^{(n)}(t)$ is shown for integer values of t ranging between 1 and 9 in Figure 1.

We have $W^{(n)}(u) \geq 4$ for all $u \in [2, 8]$ and hence

$$\begin{aligned}
K^{(n)}(t) &\triangleq \sup_{s \in [0,t]} \left[\mathcal{W}^{(n)}(s)(-\infty, 0] \wedge \inf_{u \in [s,t]} W^{(n)}(u) \right] \\
&= \sup_{s \in [0,t]} \mathcal{W}^{(n)}(s)(-\infty, 0] \\
&= \begin{cases} 0, & 0 \leq t < 4, \\ 1, & 4 \leq t < 7, \\ 4, & 7 \leq t \leq 8. \end{cases}
\end{aligned}$$

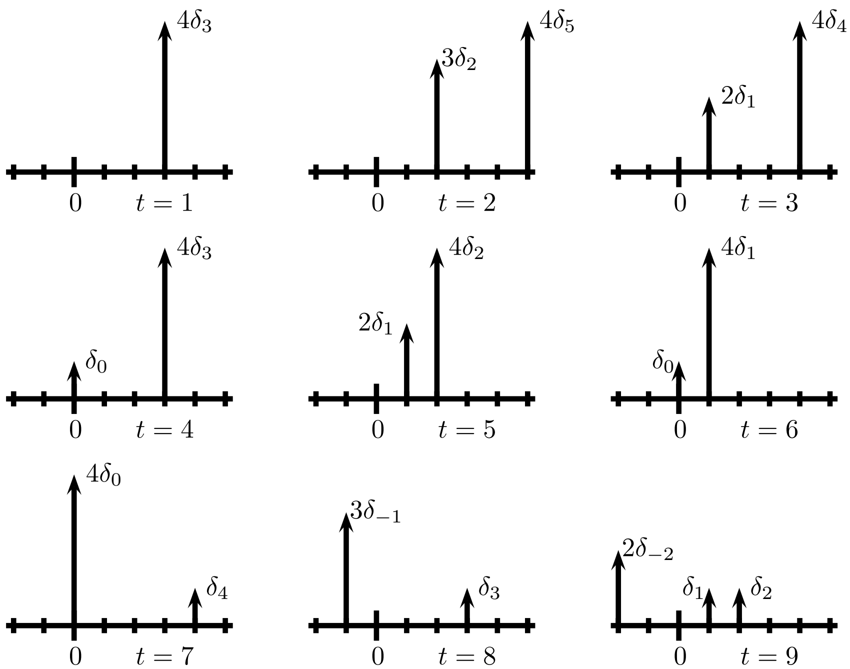

Figure 1: Evolution of $\mathcal{W}^{(n)}$

However, for $8 \le t < 9$, we have $W^{(n)}(t) = 12 - t \le 4$. For t in this range, the supremum in the definition of $K^{(n)}(t)$ is attained at $s = 7$, and

$$K^{(n)}(t) = \mathcal{W}^{(n)}(7)(-\infty, 0] \wedge \inf_{u \in [7,t]} W^{(n)} = 4 \wedge (12 - t) = 12 - t.$$

For $t = 9$, we have $W^{(n)}(9) = 4 \ne 12 - t$. Nonetheless, the supremum in the definition of $K^{(n)}(9)$ is still attained at $s = 7$. Indeed,

$$K^{(n)}(9) = \mathcal{W}^{(n)}(u)(-\infty, 0] \wedge \inf_{u \in [7,9]} W^{(n)}(u) = 4 \wedge \left[\inf_{u \in [7,9)} (12 - t) \wedge 4 \right] = 3.$$

In summary,

$$K^{(n)}(t) = \begin{cases} 0, & 0 \le t < 4, \\ 1, & 4 \le t < 7, \\ 4, & 7 \le t \le 8, \\ 12 - t, & 8 \le t \le 9. \end{cases}$$

The measure $\mathcal{U}^{(n)}(t)$ is obtained by subtracting mass $K^{(n)}(t)$ from the measure $\mathcal{W}^{(n)}(t)$, working

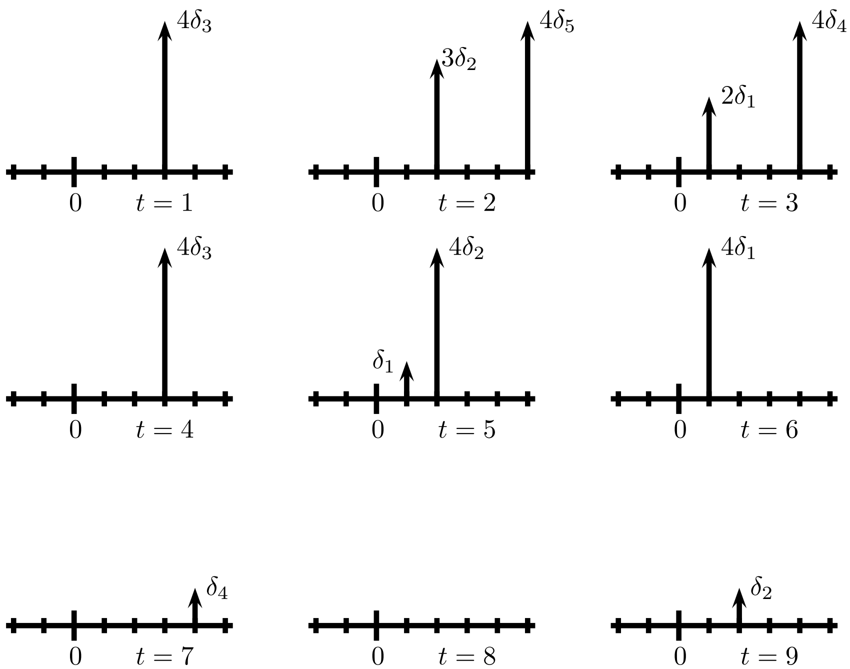

Figure 2: Evolution of $\mathcal{U}^{(n)}$

from left to right. This results in the formula

$$\mathcal{U}^{(n)}(t) = \begin{cases} 0, & 0 \leq t < 1, \\ (5-t)\delta_{4-t}, & 1 \leq t < 2, \\ (5-t)\delta_{4-t} + 4\delta_{7-t}, & 2 \leq t < 4, \\ (8-t)\delta_{7-t}, & 4 \leq t < 5, \\ (6-t)\delta_{6-t} + 4\delta_{7-t}, & 5 \leq t < 6, \\ (10-t)\delta_{7-t}, & 6 \leq t < 7, \\ (8-t)\delta_{11-t}, & 7 \leq t < 8, \\ 0, & 8 \leq t < 9, \\ \delta_2, & t = 9. \end{cases}$$

The measure $\mathcal{U}^{(n)}(t)$ is shown for integer values of t ranging between 1 and 9 in Figure 2. The total mass in the $\mathcal{U}^{(n)}$ system is

$$U^{(n)}(t) = \begin{cases} 0, & 0 \leq t < 1, \\ 5-t, & 1 \leq t < 2, \\ 9-t, & 2 \leq t < 4, \\ 8-t, & 4 \leq t < 5, \\ 10-t, & 5 \leq t < 7, \\ 8-t, & 7 \leq t < 8, \\ 0, & 8 \leq t < 9, \\ 1, & t = 9. \end{cases}$$

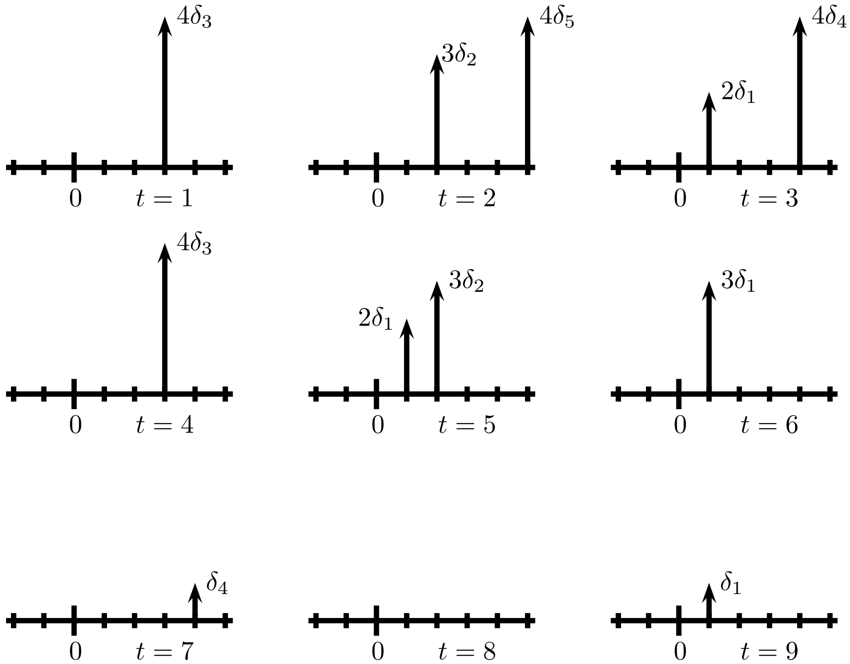

Figure 3: Evolution of the reneging system $\mathcal{W}_R^{(n)}$

This total mass path has jumps $\Delta U^{(n)}(1) = 4$, $\Delta U^{(n)}(2) = 4$, $\Delta U^{(n)}(4) = -1$, $\Delta U^{(n)}(5) = 2$, $\Delta U^{(n)}(7) = -2$ (the result of an arrival of mass 1 and the deletion of mass -3), and $U^{(n)}(9) = 1$.

We see that arriving mass to $\mathcal{U}^{(n)}$ is not always placed at the lead time of the arriving customer. In particular, $\mathcal{U}^{(n)}(5-) = 3\delta_2$, but $\mathcal{U}^{(n)}(5) = \delta_1 + 4\delta_2$. The mass $v_3^{(n)} = 2$ arriving at time 5 is distributed with one unit at $L_3^{(n)} = 1$ and one unit at 2. Furthermore, the mass $v_5^{(n)} = 1$ arriving at time $t = 9$, which begins a new busy period for $\mathcal{U}^{(n)}$, is placed at 2 rather than at $L_5^{(n)} = 1$.

Because of the failures of $\mathcal{U}^{(n)}$ to place all arriving masses at their lead times, the reneging system measure $\mathcal{W}_R^{(n)}(t)$ is not $\mathcal{U}^{(n)}(t)$ for $5 \le t < 7$ and $t = 9$. The full formula for the reneging system is

$$\mathcal{W}_R^{(n)}(t) = \begin{cases} 0, & 0 \le t < 1, \\ (5-t)\delta_{4-t}, & 1 \le t < 2, \\ (5-t)\delta_{4-t} + 4\delta_{7-t}, & 2 \le t < 4, \\ (8-t)\delta_{7-t}, & 4 \le t < 5, \\ (7-t)\delta_{6-t} + 3\delta_{7-t}, & 5 \le t < 6, \\ (9-t)\delta_{7-t}, & 6 \le t < 7, \\ (8-t)\delta_{11-t}, & 7 \le t < 8, \\ 0, & 8 \le t < 9, \\ \delta_1, & t = 9. \end{cases}$$

The measure $\mathcal{W}_R^{(n)}(t)$ is shown for integer values of t ranging between 1 and 9 in Figure 3.

Beginning at time $t = 4$, the reneging system begins serving the customer with lead time 3, and thus by time $t = 5$, this customer, whose lead time is now 2, requires only three remaining units of service. The customer arriving at time $t = 5$ with lead time 1 brings an additional two units of work. At time $t = 5$, the reneging system thus has five units of work, which agrees with $U^{(n)}(5) = 5$, but the mass in the reneging system is not distributed according to the measure $\mathcal{U}^{(n)}(5)$. At time $t = 6$, an additional unit of work is deleted from the reneging system but not from the $\mathcal{U}^{(n)}$ system, and so $W_R^{(n)}(6) = 3$, whereas $U^{(n)}(6) = 4$. This discrepancy is due to the deletion in the reneging system at time 6 of the customer who arrived at time $t = 5$, a customer who upon arrival was more urgent than the customer in service in the reneging system. The work associated with this customer upon arrival is counted in the process $D^{(n)}$ in Lemma 6.6.

Lemma 6.6 asserts that we always have $W_R^{(n)}(t) \le U^{(n)}(t)$, and the inequality can be strict due to work that preempts the customer in service in the reneging system, but the difference between $W_R^{(n)}(t)$ and $U^{(n)}(t)$ is never more than the amount of such work deleted by the reneging system up to time t. □

Acknowledgment

We are grateful to Søren Asmussen for pointing out [3] and to Neil O'Connell for telling us about [5] and [12].

References

[1] Anderson, R. and Orey. S., Small random perturbations of dynamical systems with reflecting boundary. *Nagoya Math. J.*, **60**:189–216, 1976.

[2] Anulova, S. V. and Liptser, R. Sh., Diffusion approximation for processes with normal reflection. *Theory Probab. Appl.*, **35**, 3:411–423, 1990.

[3] Cooper, W., V. Schmidt and R. Serfozo, Skorohod-Loynes characterizations of queueing, fluid, and inventory processes. *Queueing Systems* **37**:233–257, 2001.

[4] Doytchinov, B., J. Lehoczky, and S. Shreve, Real-time queues in heavy traffic with earliest-deadline-first queue discipline. *Annals of Applied Probability,* **11**:332–378, 2001.

[5] Ganesh, A., N. O'Connell and D. Wischik, *Big Queues*, Lectures Notes in Mathematics 1838, Springer, New York, 2004.

[6] Iglehart, D. and W. Whitt, Multiple channel queues in heavy traffic I. *Adv. Appl. Probab.,* **2**:150–177, 1970.

[7] Kingman, J. F. C., A single server queue in heavy traffic. *Proc. Cambridge Phil. Soc.,* **48**:277-289, 1961.

[8] Kruk, Ł., Lehoczky, J., Ramanan, K. and Shreve. S., An explicit formula for the Skorokhod map on $[0, a]$. *Ann. Probab.*, to appear.

[9] Kruk, Ł., Lehoczky, J., Ramanan, K. and Shreve. S., Diffusion approximation for an earliest-deadline-first queue with reneging. *In preparation.*

[10] Skorokhod, A. V., Stochastic equations for diffusions in a bounded region. *Theor. of Prob. and Its Appl.*, **6**:264–274, 1961.

[11] Tanaka. H., Stochastic differential equations with reflecting boundary conditions in convex regions. *Hiroshima Math. J.*, **9**:163–177, 1979.

[12] Toomey, T., Bursty traffic and finite capacity queues. *Ann. Oper. Research*, **79**: 45–62, 1998.

[13] Whitt, W., *An Introduction to Stochastic-Process Limits.* Springer, 2002.