J. Appl. Prob. **50,** 16–28 (2013)
Printed in England

ON THE GENERALIZED DRIFT SKOROKHOD PROBLEM IN ONE DIMENSION

JOSH REED,* *New York University*
AMY WARD ** AND
DONGYUAN ZHAN,** *University of Southern California*

Abstract

We show how to write the solution to the generalized drift Skorokhod problem in one-dimension in terms of the supremum of the solution of a tractable unrestricted integral equation (that is, an integral equation with no boundaries). As an application of our result, we equate the transient distribution of a reflected Ornstein–Uhlenbeck (OU) process to the first hitting time distribution of an OU process (that is *not* reflected). Then, we use this relationship to approximate the transient distribution of the GI/GI/1 + GI queue in conventional heavy traffic and the M/M/N/N queue in a many-server heavy traffic regime.

Keywords: Skorokhod map; reflected Ornstein–Uhlenbeck process; abandonment; queueing

2010 Mathematics Subject Classification: Primary 90B22
Secondary 90B15; 60G17; 60J60

1. Introduction

The Skorokhod problem was originally introduced by Skorokhod [15] in order to study continuous solutions to stochastic differential equations with a reflecting boundary at zero.

Definition 1.1. (*Skorokhod problem.*) Given a process $X \in D([0,\infty),\mathbb{R})$, we say that the pair of processes $(Z, L) \in D^2([0,\infty),\mathbb{R})$ satisfy the Skorokhod problem for X if the following four conditions are satisfied:

1. $Z(t) = X(t) + L(t)$ for $t \geq 0$,
2. $Z(t) \geq 0$ for $t \geq 0$,
3. L is nondecreasing with $L(0-) = 0$,
4. $\int_0^\infty \mathbf{1}\{Z(t) > 0\}\, \mathrm{d}L(t) = 0$.

It is well known that, for each $X \in D([0,\infty),\mathbb{R})$, the unique solution $(Z, L) = (\Phi(X), \Psi(X))$ to the Skorokhod problem is

$$Z(t) = X(t) + \sup_{0\leq s\leq t} -X(s) \vee 0 \quad \text{and} \quad L(t) = \sup_{0\leq s\leq t} -X(s) \vee 0.$$

Received 7 January 2011; revision received 11 July 2012.
* Postal address: Stern School of Business, New York University, New York, NY 10012, USA.
Email address: jreed@stern.nyu.edu
** Postal address: Marshall School of Business, University of Southern California, Los Angeles, CA 90089-0809, USA.

https://doi.org/10.1239/jap/1363784421 Published online by Cambridge University Press

In subsequent papers, the Skorokhod problem has been extended to multiple dimensions and also to include both smooth and nonsmooth domains (see, for example, [4], [6], [8], [13], and [17]), although we do not treat such cases in the present paper. There is a useful integral representation of the one-dimensional Skorokhod problem solution (see [2]). There is also an explicit solution to the (one-dimensional) Skorokhod problem when there is an upper boundary (see [9] and [10])) and to the (one-dimensional) Skorokhod problem in a time-dependent interval (see [3]).

In this paper we study a generalization of the one-dimensional Skorokhod problem that incorporates a state-dependent drift.

Definition 1.2. (*Generalized drift Skorokhod problem in one dimension.*) Given a process $X \in D([0,\infty),\mathbb{R})$ with $X(0)=0$ and a Lipschitz continuous function $f\colon \mathbb{R}_+ \to \mathbb{R}$, we say that the pair of processes $(Z,L) \in D^2([0,\infty),\mathbb{R})$ satisfy the Skorokhod problem for X with state-dependent drift function f if the following four conditions are satisfied:

1. $Z(t) = X(t) - \int_0^t f(Z(s))\,\mathrm{d}s + L(t)$ for $t \ge 0$,
2. $Z(t) \ge 0$ for $t \ge 0$,
3. L is nondecreasing with $L(0-)=0$,
4. $\int_0^\infty \mathbf{1}\{Z(t) > 0\}\,\mathrm{d}L(t) = 0$.

The unique solution to the generalized drift Skorokhod problem in one dimension can be written in terms of the solution to the Skorokhod problem following a standard construction; see, for example, [22]. Specifically, set

$$(Z,L) = (\Phi(\mathcal{M}(X)), \Psi(\mathcal{M}(X))) \tag{1.1}$$

for $\mathcal{M}\colon D([0,\infty),\mathbb{R}) \to D([0,\infty),\mathbb{R})$, the mapping that sets $\mathcal{M}(X) = V$ for V that solves the integral equation

$$V(t) = X(t) - \int_0^t f(\Phi(V)(s))\,\mathrm{d}s \quad \text{for all } t \ge 0. \tag{1.2}$$

Note that, since $f\colon \mathbb{R}_+ \to \mathbb{R}$ is a Lipschitz continuous function, a standard Picard iteration shows that there exists a unique solution to (1.2). The fact that $(\Phi(\mathcal{M}(X)), \Psi(\mathcal{M}(X)))$ solves the Skorokhod problem for $\mathcal{M}(X)$ (and so satisfies conditions 1–4 of Definition 1.1) shows that conditions 1–4 of Definition 1.2 are satisfied. The uniqueness of representation (1.1) follows from the uniqueness of the mappings $\mathcal{M}$ and (Φ,Ψ).

Next, we observe that the solution Z can be represented in terms of an unrestricted integral equation (that is, an integral equation with no boundaries). Specifically, note that, from (1.2),

$$V(t) - V(s) = X(t) - X(s) - \int_s^t f(\Phi(V)(u))\,\mathrm{d}u.$$

Since, when $X(0)=0$,

$$\Phi(V)(u) = \sup_{0 \le r \le u} V(u) - V(r),$$

if we define

$$R(s,t) = V(t) - V(s)$$

https://doi.org/10.1239/jap/1363784421 Published online by Cambridge University Press

then

$$R(s,t) = X(t) - X(s) - \int_s^t f\Big(\sup_{0\le r\le u} R(r,u)\Big)\,\mathrm{d}u. \tag{1.3}$$

Finally, it follows from (1.1) and the above displays that

$$Z(t) = \sup_{0\le s\le t} R(s,t). \tag{1.4}$$

However, the integral equation (1.3) is not tractable.

In this paper we establish how to represent Z in terms of the solution to a *tractable* unrestricted integral equation. Specifically, we establish that

$$Z(t) = \sup_{0\le s\le t} Z_s(t-s), \qquad t\ge 0, \tag{1.5}$$

for $Z_s = \{Z_s(t),\ t\ge 0\}$ that solves

$$Z_s(t) = X(s+t) - X(s) - \int_0^t f_{\mathrm{e}}(Z_s(u))\,\mathrm{d}u, \tag{1.6}$$

where $f_{\mathrm{e}}\colon \mathbb{R}\to\mathbb{R}$ is any extension of $f\colon \mathbb{R}_+\to\mathbb{R}$ that preserves the Lipschitz continuity of f. For one example, let $f_{\mathrm{e}}(x) = f(0)$ if $x<0$ and $f_{\mathrm{e}}(x) = f(x)$ if $x\ge 0$. It is interesting to observe that it follows from (1.4) and (1.5) that

$$\sup_{0\le s\le t} R(s,t) = \sup_{0\le s\le t} Z_s(t-s).$$

As an application of representation (1.5), we show how to use (1.5) to write the transient distribution of a reflected Ornstein–Uhlenbeck (OU) process in terms of the first hitting time distribution of an unreflected OU process, which additionally yields a uniform integrability result for reflected OU processes. Such a result can also be derived using duality theory (see, for example, [5] and [14]); however, the proof methodology is much different, because there is no sample path representation that is equivalent to (1.5) in either [5] or [14]. Because the reflected OU process has been shown to approximate the GI/GI/1+GI and M/M/N/N queues (see [20] and [16]), we see that the transient distribution of the number-in-system process for the GI/GI/1+GI and M/M/N/N queues can be approximated by the first hitting time distribution of an OU process (that is *not* reflected).

The remainder of this paper is organized as follows. In Section 2 we prove (1.5). In Section 3 we apply (1.5) in the context of a reflected OU process. In Section 4 we perform simulation studies that support approximating the transient distribution of the number-in-system process for the GI/GI/1 + GI and M/M/N/N queues with the first hitting time distribution of an OU process (that is *not* reflected).

2. The generalized drift Skorokhod problem solution (in one dimension)

In this section we establish (1.5).

Theorem 2.1. *Let (Z, L) be the unique solution to the generalized Skorokhod problem for X with $X(0) = 0$, and with state-dependent drift function f that is Lipschitz continuous. For each $s\ge 0$, let Z_s be defined as in (1.6). Then, for each $t\ge 0$,*

$$Z(t) = \sup_{0\le s\le t} Z_s(t-s).$$

Proof. We first claim that, for each $0 \le s \le t$,

$$Z_s(t-s) \le Z(t).$$

To see this, first recall from (1.6) that $Z_s(t-s)$ is the solution to the equation

$$Z_s(u) = X(s+u) - X(s) - \int_0^u f_e(Z_s(v))\,dv, \tag{2.1}$$

evaluated at the point $u = t - s$, where $f_e \colon \mathbb{R} \mapsto \mathbb{R}$ is an arbitrary Lipschitz extension of $f \colon \mathbb{R}_+ \mapsto \mathbb{R}$. Next, it is straightforward to see from condition 1 of Definition 1.1 that $Z(t)$ is the unique solution to the equation

$$Z(s+u) = Z(s) + (X(s+u) - X(s) + L(s+u) - L(s)) - \int_0^u f_e(Z(s+v))\,dv \tag{2.2}$$

for $u \ge 0$, also evaluated at the point $u = t - s$ (note that in (2.2) we have replaced f by f_e). Subtracting (2.1) from (2.2) we therefore obtain

$$(Z(s+u) - Z_s(u)) = Z(s) + L(s+u) - L(s) - \int_0^u (f_e(Z(s+v)) - f_e(Z_s(v)))\,dv$$

for $u \ge 0$. Note also that by the Lipschitz continuity of f_e we have, for some constant $K > 0$,

$$(Z(s+u) - Z_s(u)) \ge Z(s) + L(s+u) - L(s) - K\int_0^u |Z(s+v) - Z_s(v)|\,dv \tag{2.3}$$

for $u \ge 0$. Now consider the solution $W_s = \{W_s(u),\ u \ge 0\}$ to the ordinary differential equation

$$W_s(u) = Z(s) + L(s+u) - L(s) - K\int_0^u |W_s(v)|\,dv, \qquad u \ge 0. \tag{2.4}$$

We claim that

$$W_s(u) = Z(s)e^{-Ku} + \int_0^u e^{K(v-u)}\,dL(s+v), \qquad u \ge 0.$$

This may be verified by noting that $W_s(u) \ge 0$ for $u \ge 0$, since $Z(s) \ge 0$ and L is a nondecreasing function. Subtracting (2.4) from (2.3) and using Gronwall's inequality, it follows that $Z(s+u) - Z_s(u) \ge W_s(u) \ge 0$, and so $Z(s+u) \ge Z_s(u)$, which, evaluating at $u = t-s$, yields $Z_s(t-s) \le Z(t)$, our desired result. We have therefore shown that

$$Z(t) \ge \sup_{0 \le s \le t} \{Z_s(t-s)\}. \tag{2.5}$$

It now remains to reverse the direction of the inequality in (2.5). In order to do so, it suffices to show that there exists at least one point $s^\star$ such that $Z_{s^\star}(t-s^\star) = Z(t)$. Let $s^\star = \sup\{s \le t \colon Z(s) = 0\}$ be the last time at which the process Z hit zero. Note that $s^\star$ is well defined since $Z(0) = 0$. Also, note that $L(s) = L(s^\star)$ for $s \ge s^\star$. Thus, by (2.2) we have

$$Z(s^\star + u) = X(s^\star + u) - X(s^\star) - \int_0^u f_e(Z(s^\star + v))\,dv, \qquad u \ge 0,$$

and so, $Z(s^\star + u) = Z_{s^\star}(u)$ for $0 \le u \le t - s^\star$, and, in particular, $Z(t) = Z_{s^\star}(t - s^\star)$, which completes the proof.

https://doi.org/10.1239/jap/1363784421 Published online by Cambridge University Press

3. Reflected OU processes

In this section we let the process X in the definition of the generalized Skorokhod problem be a Brownian motion with constant drift θ and infinitesimal variance σ^2 defined on a suitable probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We also set $f(x) = \gamma x$ for $x \geq 0$ and some $\gamma \in \mathbb{R}$. The resulting process Z, defined sample pathwise as the solution to the generalized Skorokhod problem for X and f, is referred to as a (σ, θ, γ) reflected OU process, which has initial condition $Z(0) = 0$. It is immediate that the following definition of a reflected OU process is equivalent to the prescription given above.

Definition 3.1. (*Reflected OU process.*) Let $B = \{B(t),\ t \geq 0\}$ be a standard Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\sigma > 0$, and $\theta, \gamma \in \mathbb{R}$. We say that the process Z is a (σ, θ, γ) reflected OU process if the following four conditions are satisfied $\mathbb{P}$-almost-surely:

1. $Z(t) = \sigma B(t) + \theta t - \gamma \int_0^t Z(s)\,\mathrm{d}s + L(t)$ for $t \geq 0$,
2. $Z(t) \geq 0$ for $t \geq 0$,
3. L is nondecreasing with $L(0-) = 0$,
4. $\int_0^\infty \mathbf{1}\{Z(t) > 0\}\,\mathrm{d}L(t) = 0$.

Now, for each $s \geq 0$, recall from (1.6) the definition of the associated unreflected processes

$$Z_s(u) = (\sigma B(s+u) + \theta(s+u)) - (\sigma B(s) + \theta s) - \gamma \int_0^u Z_s(v)\,\mathrm{d}v$$

for $u \geq 0$, where here we have set $X(t) = \sigma B(t) + \theta t$, and we take the natural extension $f_{\mathrm{e}}(x) = \gamma x$ for $x \in \mathbb{R}$. For clarity of exposition in the sequel, we now hold $t \geq 0$ fixed and define the new process

$$Y_t(u) = Z_{t-u}(u), \qquad 0 \leq u \leq t.$$

Since $\{Y_t(u),\ 0 \leq u \leq t\}$ is just the process $\{Z_s(t-s),\ 0 \leq s \leq t\}$ run backwards in time, it follows that

$$\sup_{0 \leq u \leq t} \{Y_t(u)\} = \sup_{0 \leq s \leq t} \{Z_s(t-s)\},$$

and so it follows from Theorem 2.1 that if Z is a (σ, θ, γ) reflected OU process then

$$Z(t) = \sup_{0 \leq u \leq t} \{Y_t(u)\}. \tag{3.1}$$

In preparation for our next result, we now say that a process X is a (σ, θ, γ) OU process starting from $X(0)$ (note the absence of reflection here) if it is the unique strong solution to the stochastic differential equation

$$X(t) = X(0) + \sigma B(t) + \theta t - \int_0^t \gamma X(s)\,\mathrm{d}s$$

for $t \geq 0$, where B is a standard Brownian motion. We then make the following claim regarding the process $\{Y_t(u),\ 0 \leq u \leq t\}$.

Proposition 3.1. *The process* $\{\mathrm{e}^{\gamma u} Y_t(u),\ 0 \leq u \leq t\}$ *is equal in distribution to a* $(\sigma, \theta, -\gamma)$ *OU process on* $[0, t]$ *which starts from* 0.

https://doi.org/10.1239/jap/1363784421 Published online by Cambridge University Press

Proof. First note that, since $X(t) = \sigma B(t) + \theta t$ is a Brownian motion with infinitesimal variance σ^2 and constant drift θ, it follows that, for each $s \geq 0$, the process $X_s = \{X(s+t) - X(s),\ t \geq 0\}$ is also Brownian motion with the same parameters and so, for each $s \geq 0$, the process $Z_s = \{Z_s(u),\ u \geq 0\}$ is an OU process whose explicit solution is given by

$$Z_s(u) = \frac{\theta}{\gamma}(1 - \mathrm{e}^{-\gamma u}) + \int_0^u \sigma \mathrm{e}^{\gamma(v-u)}\, \mathrm{d}B_s(v), \qquad u \geq 0,$$

where $B_s = \{B(s+t) - B(s),\ t \geq 0\}$.

Setting $Y_t(u) = Z_{t-u}(u)$, it therefore follows that

$$Y_t(u) = \frac{\theta}{\gamma}(1 - \mathrm{e}^{-\gamma u}) + \int_0^u \sigma \mathrm{e}^{\gamma(v-u)}\, \mathrm{d}B_{t-u}(v).$$

However, since $\mathrm{d}B_{t-u}(v) = \mathrm{d}B(t-u+v)$, the change with respect to v, it follows that making the change of variable $\zeta = u - v$, the above becomes

$$\begin{aligned} Y_t(u) &= \frac{\theta}{\gamma}(1 - \mathrm{e}^{-\gamma u}) + \int_0^u \sigma \mathrm{e}^{\gamma(v-u)}\, \mathrm{d}B(t-u+v) \\ &= \frac{\theta}{\gamma}(1 - \mathrm{e}^{-\gamma u}) + \int_u^0 \sigma \mathrm{e}^{-\gamma\zeta}\, \mathrm{d}B(t-\zeta) \\ &= \frac{\theta}{\gamma}(1 - \mathrm{e}^{-\gamma u}) - \int_0^u \sigma \mathrm{e}^{-\gamma\zeta}\, \mathrm{d}B(t-\zeta). \end{aligned}$$

However, it is clear that the above, as a process, is also equal in distribution to

$$\frac{\theta}{\gamma}(1 - \mathrm{e}^{-\gamma u}) + \int_0^u \sigma \mathrm{e}^{-\gamma t}\, \mathrm{d}B(t), \qquad u \geq 0.$$

Multiplying both sides of the above by $\mathrm{e}^{\gamma u}$, we then obtain

$$\mathrm{e}^{\gamma u} Y_t(u) = -\frac{\theta}{\gamma}(1 - \mathrm{e}^{\gamma u}) + \int_0^u \sigma \mathrm{e}^{-\gamma(t-u)}\, \mathrm{d}B(t),$$

which is just an OU process on $[0, t]$ with infinitesimal variance σ^2, constant drift θ, and linear drift $-\gamma$.

The following is our main result of this section, relating the distribution of the supremum appearing in (3.1) to the first hitting distribution of an OU process. Let

$$\sigma_x = \inf\{t \geq 0 \colon U(t) = x\},$$

where $U = \{U(t),\ t \geq 0\}$ is an OU process with parameters $(\sigma, -\gamma x + \theta, -\gamma)$ and started from 0. In other words, σ_x is the first hitting time of x by U. We then have the following proposition.

Proposition 3.2. *For each* $t \geq 0$*,*

$$\mathbb{P}(Z(t) \geq x) = \mathbb{P}(\sigma_x \leq t).$$

Proof. Note that, for each $x \geq 0$,

$$\Big\{\sup_{0\leq u\leq t} Y_t(u) \geq x\Big\} = \{\inf\{u\colon Y_t(u) \geq x\} \leq t\}$$
$$= \{\inf\{u\colon \mathrm{e}^{\gamma u} Y_t(u) \geq \mathrm{e}^{\gamma u} x\} \leq t\}$$
$$= \{\inf\{u\colon x(1-\mathrm{e}^{\gamma u}) + \mathrm{e}^{\gamma u} Y_t(u) \geq x\} \leq t\}.$$

Now, by Proposition 3.1, $\{x(1-\mathrm{e}^{\gamma u}) + \mathrm{e}^{\gamma u} Y_t(u),\ u \geq 0\}$ is simply an OU process with infinitesimal variance σ^2, constant drift $-\gamma x + \theta$, and linear drift $-\gamma$. The result then follows immediately.

Sigman and Ryan [14] established an equivalent result to Proposition 3.2; however, their proof methodology is much different. In particular, Sigman and Ryan related the transient distribution of any continuous-time, real-valued stochastic process that can be defined recursively (either explicitly in discrete time or implicitly in continuous time, through the use of an integral equation) to the ruin time of a dual risk process. There is no result in [14] that is equivalent to Theorem 2.1, which is the basis for our proof of Proposition 3.2.

3.1. Computing the first hitting time

In order to use Proposition 3.2 to compute $\mathbb{P}(Z(t) \geq x)$, it is necessary that the distribution of σ_x is known. Fortunately, there are various results in the literature available for computing the first hitting time distributions of OU processes. Linetsky [11] provided a spectral expansion for the first hitting time of OU processes and the results of Alili *et al.* [1] provide three different means to compute various probabilities associated with this hitting time. In what follows, we use the results in [1].

Let $p^{(\sigma,\theta,\gamma)}_{x_0\to x}$ denote the density of the distribution of σ_x for a (σ, θ, γ) OU process, so that we may write

$$\mathbb{P}(\sigma_x \leq t) = \int_0^t p^{(\sigma,\theta,\gamma)}_{x_0\to x}(s)\,\mathrm{d}s, \qquad t \geq 0.$$

Alili *et al.* [1] showed how to calculate $p^{(1,0,\gamma)}_{x_0\to x}$ when $\gamma > 0$. Since we are interested in the more general case, we first express $p^{(\sigma,\theta,\gamma)}_{x_0\to x}$ in terms of $p^{(1,0,\gamma)}_{x_0\to x}$. In order to do this, note that, since a (σ, θ, γ) OU process starting from x_0 has the same distribution as a $(1, \theta/\sigma, \gamma)$ OU process starting from x_0/σ, it follows that

$$p^{(\sigma,\theta,\gamma)}_{x_0\to x}(t) = p^{(1,\theta/\sigma,\gamma)}_{x_0/\sigma\to x/\sigma}(t), \qquad t \geq 0. \tag{3.2}$$

Next, Remark 2.5 of [1] shows that

$$p^{(1,\theta/\sigma,\gamma)}_{x_0/\sigma\to x/\sigma}(t) = p^{(1,0,\gamma)}_{x_0/\sigma-\theta/\sigma\gamma\to x/\sigma-\theta/\sigma\gamma}(t), \qquad t \geq 0. \tag{3.3}$$

When $x - \theta/\gamma = 0$, the above expression may be immediately evaluated because

$$p^{(1,0,\gamma)}_{\zeta\to 0}(t) = \frac{|\zeta|}{\sqrt{2\pi}}\bigg(\frac{\lambda}{\sinh(\lambda t)}\bigg)^{3/2} \exp\bigg(-\frac{\lambda\zeta^2\mathrm{e}^{-\lambda t}}{2\sinh(\lambda t)} + \frac{\lambda t}{2}\bigg), \tag{3.4}$$

as is found in [12] and reproduced in Equation (2.8) of [1]. Otherwise, when $x - \theta/\gamma \neq 0$, one must appeal to one of the three representations in [1] (one that hinges on an eigenvalue expansion, one that is an integral representation, and one that is given in terms of a functional of a three-dimensional Bessel bridge) in order to compute $\mathbb{P}(\sigma_x \leq t)$.

To compute the transient distribution of the (σ, θ, γ) reflected OU process Z, we first apply Proposition 3.2, and then use the distributional equalities (3.2) and (3.3) as follows:

$$\begin{aligned}\mathbb{P}(Z(t) \geq x) &= \mathbb{P}(\sigma_x \leq t) \\ &= \int_0^t p_{0\to x}^{(\sigma,-\gamma x+\theta,-\gamma)}(s)\,\mathrm{d}s \\ &= \int_0^t p_{0\to x/\sigma}^{(1,(-\gamma x+\theta)/\sigma,-\gamma)}(s)\,\mathrm{d}s \\ &= \int_0^t p_{\theta/\sigma\gamma - x/\sigma\to\theta/\sigma\gamma}^{(1,0,-\gamma)}(s)\,\mathrm{d}s.\end{aligned} \tag{3.5}$$

We double check the calculation (3.5) by recalling that it also follows [14]. Specifically, Proposition 4.3 of [14] establishes that

$$\mathbb{P}(Z(t) \geq x) = \mathbb{P}(\sigma^R \leq t), \tag{3.6}$$

where σ^R is the first time a $(\sigma, -\theta, -\gamma)$ OU process with initial point $x > 0$ becomes negative. To see that (3.5) and (3.6) are equivalent, first observe that

$$\begin{aligned}\mathbb{P}(\sigma^R \leq t) &= \int_0^t p_{x\to 0}^{(\sigma,-\theta,-\gamma)}(s)\,\mathrm{d}s \\ &= \int_0^t p_{x/\sigma\to 0}^{(1,-\theta/\sigma,-\gamma)}(s)\,\mathrm{d}s \\ &= \int_0^t p_{x/\sigma-\theta/\sigma\gamma\to-\theta/\sigma\gamma}^{(1,0,-\gamma)}(s)\,\mathrm{d}s,\end{aligned}$$

where the second and third equalities follow from (3.2) and (3.3). Then, since symmetry implies that

$$p_{\theta/\sigma\gamma-x/\sigma\to\theta/\sigma\gamma}^{(1,0,-\gamma)}(s) = p_{x/\sigma-\theta/\sigma\gamma\to-\theta/\sigma\gamma}^{(1,0,-\gamma)}(s),$$

we conclude that $\mathbb{P}(\sigma_x \leq t) = \mathbb{P}(\sigma^R \leq t)$.

3.2. Uniform integrability

It is well known (see, for example, Proposition 1 of [19]) that if $\gamma > 0$ then, for a (σ, θ, γ) reflected OU process, $Z(t) \Rightarrow Z(\infty)$ as $t \to \infty$, where $Z(\infty)$ is a normal random variable with mean θ/γ and variance $\sigma^2/(2\gamma)$ conditioned to be positive. We now show that the sequence of random variables $\{Z(t),\ t \geq 0\}$ is uniformly integrable as well.

Proposition 3.3. *If $\gamma > 0$ then, for a (σ, θ, γ) reflected OU process started at the origin, the sequence of random variables $\{Z(t),\ t \geq 0\}$ is uniformly integrable.*

Proof. First note that, without loss of generality, we may assume that $\sigma = 1$ since otherwise we may rescale. Now recall that, by Proposition 3.2, it follows that $\mathbb{P}(Z(t) \geq x) = \mathbb{P}(\sigma_x \leq t)$, where $\sigma_x = \inf\{t \geq 0\colon U_t = x\}$ and U_t is an OU process with parameters $(1, -\gamma x + \theta, -\gamma)$ started from 0. Hence, is suffices to show that there exists a function g integrable on $\mathbb{R}^+$ such that $\mathbb{P}(\sigma_x \leq t) \leq g(x)$ for all $x, t \geq 0$.

Next, it follows from (3.5) that $\mathbb{P}(\sigma_x \leq t) = \int_0^t p_{\theta/\gamma-x\to\theta/\gamma}^{(1,0,-\gamma)}(s)\,\mathrm{d}s$. Remark 2.4 of [1] shows that

$$p_{\theta/\gamma-x\to\theta/\gamma}^{(1,0,-\gamma)}(s) = \exp\bigg(\gamma\bigg(\frac{\theta^2}{\gamma^2} - \bigg(\frac{\theta}{\gamma} - x\bigg)^2 - s\bigg)\bigg)p_{\theta/\gamma-x\to\theta/\gamma}^{(1,0,\gamma)}(s).$$

https://doi.org/10.1239/jap/1363784421 Published online by Cambridge University Press

(We note that there is a missing negative sign in the display appearing in Remark 2.4 of [1]; specifically, the correct equation is $p^{(\lambda)}_{x\to a}(t) = \exp(-\lambda(a^2 - x^2 - t))p^{(-\lambda)}_{x\to a}(t)$.) Hence,

$$\mathbb{P}(\sigma_x \le t) = \exp\left(-\gamma\left(x^2 - 2\frac{\theta}{\gamma}x\right)\right)\int_0^t \exp(-\gamma s)p^{(1,0,\gamma)}_{\theta/\gamma - x\to\theta/\gamma}(s)\,\mathrm{d}s$$
$$\le \exp\left(-\gamma\left(x^2 - 2\frac{\theta}{\gamma}x\right)\right),$$

where the last inequality follows since

$$\int_0^\infty p^{(1,0,\gamma)}_{\theta/\gamma - x\to\theta/\gamma}(s)\,\mathrm{d}s = 1.$$

Finally, since, for $\gamma > 0$,

$$\int_0^\infty \exp\left(-\gamma\left(x^2 - 2\frac{\theta}{\gamma}x\right)\right)\mathrm{d}x < \infty,$$

the proof is complete.

4. Approximating the transient distribution of the GI/GI/1 + GI and M/M/N/N queues

In this section we perform simulation studies that support using the first hitting time distribution of an OU process (that is *not* reflected) to approximate the transient distribution of the number-in-system process for the GI/GI/1 + GI queue (Section 4.1) and the M/M/N/N queue (Section 4.2).

4.1. The GI/GI/1 + GI queue

The M/M/1 + M queueing model assumes that customers arrive according to a Poisson process with rate λ to an infinite waiting room service facility, that their service times form an independent and identically distributed (i.i.d.) sequence of exponential random variables having mean $1/\mu > 0$, and that each customer independently reneges if his/her service has not begun within an exponentially distributed amount of time that has mean $1/\gamma > 0$. Theorem 2 of [18] supports approximating the number-in-system process $Q = \{Q(t),\ t \ge 0\}$ by a $(\sqrt{2\lambda}, \lambda-\mu, \gamma)$ reflected OU process Z.

The more general GI/GI/1 + GI queueing model assumes that the customer arrival process is a renewal process with rate λ, that the service time distribution is general with mean $1/\mu$, and that each customer independently reneges if his/her service has not begun within an amount of time that is distributed according to some probability distribution function F. In the case that F has a density and $F'(0) > 0$ is finite, Theorem 3 of [20] combined with the arguments in the proof of Theorem 2 of [18] shows that Q may be approximated by a $(\sqrt{2\lambda}, \lambda - \mu, F'(0))$ reflected OU process. Note that this is consistent with the approximation for Q in the previous paragraph since the value of the density of an exponential random variable at 0 is equal to its rate.

Our results in Section 3 (specifically, Proposition 3.2 and (3.5)) then imply for the M/M/1+ M case that

$$\mathbb{P}(Q(t) \ge x) \approx \mathbb{P}(Z(t) \ge x) = \int_0^t p^{(1,0,-\gamma)}_{(\lambda-\mu-\gamma x)/\gamma\sqrt{2\lambda}\to(\lambda-\mu)/\gamma\sqrt{2\lambda}}(s)\,\mathrm{d}s, \tag{4.1}$$

https://doi.org/10.1239/jap/1363784421 Published online by Cambridge University Press

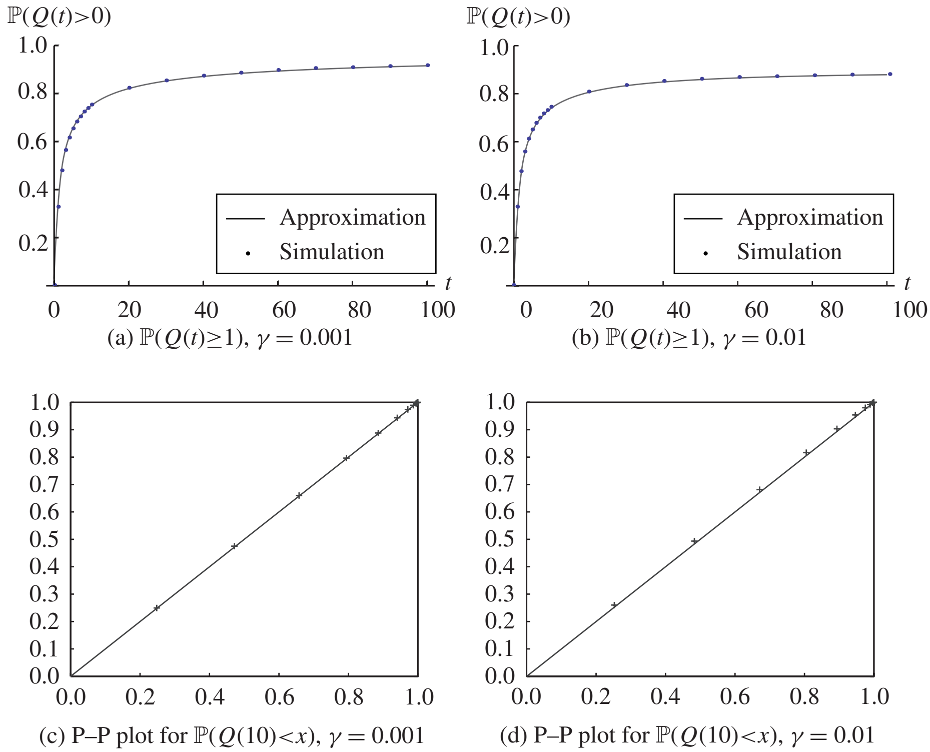
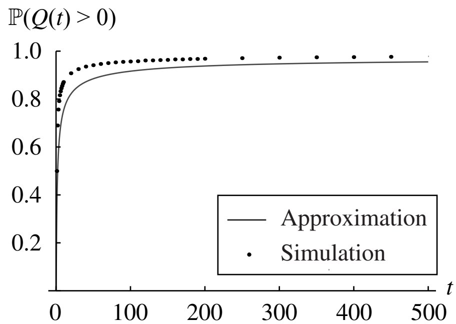


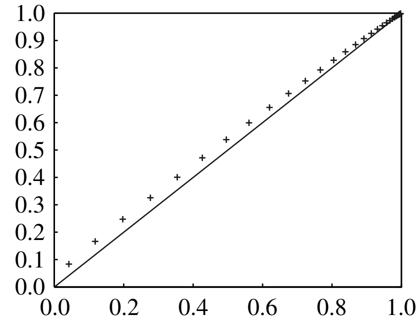
FIGURE 1: Simulated and approximated results for the $M/M/1+M$ queueing model when $\lambda = \mu = 0.5$, $\gamma = 0.001$, so that $P_R = 3.41\%$, and $\lambda = \mu = 0.5$, $\gamma = 0.01$, so that $P_R = 9.79\%$. P_R is the steady-state percentage of arriving customers that renege.

when $Q(0) = 0$. For the $GI/GI/1+GI$ case, we may replace γ with $F'(0)$ in the above. Hence, we have an approximation for the transient distribution for the number-in-system process in a $GI/GI/1+GI$ queue. Note that the theory in [18] and [20] suggests that the approximation in (4.1) will be good when λ and μ are close, and when γ is small compared to λ and μ (that is, the percentage of customers reneging is not too large). For related work, we refer the interested reader to Fralix [7], who derived the time-dependent moments of an $M/M/1+M$ queue, and then used those to obtain the time-dependent moment expressions for a reflected OU process.

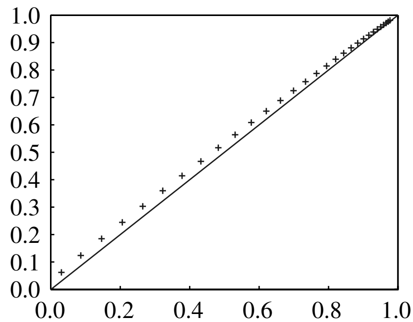
We now proceed to verify approximation (4.1) in an $M/M/1+M$ model via simulation. Note that even in the case of an $M/M/1+M$ model, the problem of finding an exact expression for its transient distribution appears to be very difficult (as is suggested by the computations in [21], which provide some performance measure expressions in terms of transforms for a many-server model with reneging). The plots in Figure 1 reveal that approximation (4.1) is very accurate, both for calculating the probability that the system is nonempty for a range of t values, and for finding the entire distribution of $Q(t)$ for a fixed t. The simulation results shown are averaged over 10 000 runs, stopped at the relevant time value. Note that we chose $\lambda = \mu$ so that we could use the very simple expression (3.4) when computing $\mathbb{P}(Z(t) \geq x)$. When $\lambda \neq \mu$, there is another source of error that comes into approximation (4.1) that is due to the methodology in [1] for computing the hitting time density function of an OU process.



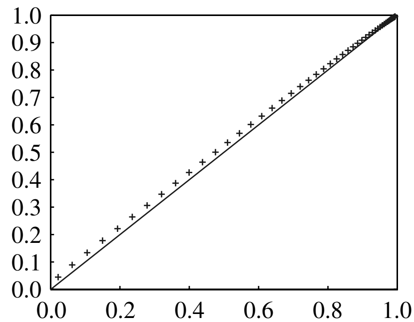
(a) $\mathbb{P}(Q(t) \geq 1)$



(b) P–P plot for $\mathbb{P}(Q(100) < x)$



(c) P–P plot for $\mathbb{P}(Q(200) < x)$



(d) P–P plot for $\mathbb{P}(Q(500) < x)$

FIGURE 2: Simulated and approximated results for the $\mathrm{GI}/\mathrm{GI}/1+\mathrm{GI}$ queueing model when the interarrival and service time distributions are gamma(2,2) and the reneging distribution F is uniform on [0, 1000].

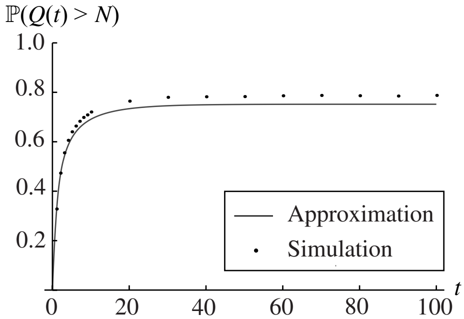
The plots in Figure 2 verify approximation (4.1) in a $\mathrm{GI}/\mathrm{GI}/1 + \mathrm{GI}$ queueing model. Note that the relevant approximating reflected OU process is exactly the same as in the $\mathrm{M}/\mathrm{M}/1+\mathrm{M}$ queueing model in Figure 1(a) and (c). We observe that the transient distribution approximation is good for 'medium' t but not for 'small' t. (The simulation results in [20] imply that the approximation is good for 'large' t, when the system is close to its steady state.) The $\mathrm{GI}/\mathrm{GI}/1+\mathrm{GI}$ queue that we simulated had simulated steady-state mean number-in-system 18.12, and simulated mean number-in-system at times $t = 100$, $t = 200$, and $t = 500$ of 7.73, 10.43, and 14.43, respectively. Then, the displayed P–P plots for $\mathbb{P}(Q(t) < x)$ in Figure 2 are such that the transient distribution is relevant (and not the steady-state distribution).

4.2. The $\mathrm{M}/\mathrm{M}/N/N$ queue

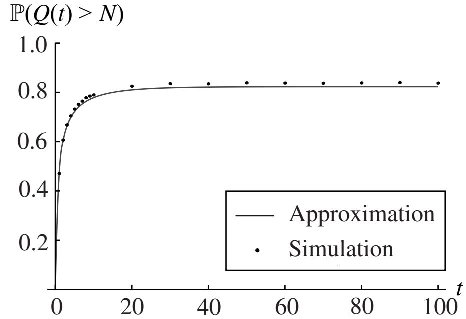
The $\mathrm{M}/\mathrm{M}/N/N$ queueing model assumes that customers arrive at rate $\lambda > 0$ in accordance with a Poisson process to a service facility with N servers and no additional place for waiting, and that their service times form an i.i.d. sequence of exponential random variables with mean $1/\mu$. Any arriving customer that finds N customers in the system is blocked from receiving service, and so is lost. Suppose that we let the number of servers in the system be a function of the arrival rate λ, and assume that

$$N^{\lambda} = \frac{\lambda + \beta\sqrt{\lambda}}{\mu} \quad \text{for } \beta \in \mathbb{R}. \tag{4.2}$$

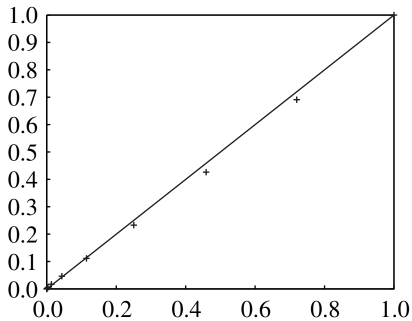
https://doi.org/10.1239/jap/1363784421 Published online by Cambridge University Press



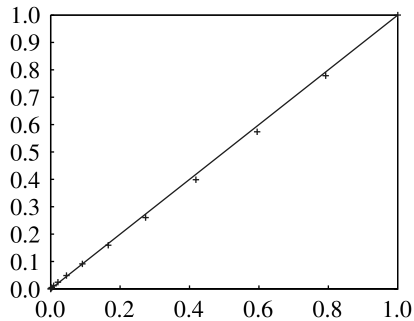
(a) $\mathbb{P}(Q(t) < N)$, $\lambda = 0.5$, $\mu = 0.05$, and $N = 10$



(b) $\mathbb{P}(Q(t) < N)$, $\lambda = 1$, $\mu = 0.05$, and $N = 20$



(c) P–P plot for $\mathbb{P}(Q(10) \leq x)$, $\lambda = 0.5$, $\mu = 0.05$, and $N = 10$



(d) P–P plot for $\mathbb{P}(Q(10) \leq x)$, $\lambda = 1$, $\mu = 0.05$, and $N = 20$

FIGURE 3: Simulated and approximated results for the M/M/N/N queueing model.

Then, Srikant and Whitt [16] showed that

$$\frac{N^\lambda - Q^\lambda}{\sqrt{\lambda}} \Rightarrow Z \quad \text{as } \lambda \to \infty,$$

where Z is a $(\sqrt{2}, \beta, \mu)$ reflected OU process. Hence, our results in Section 3 (specifically, Proposition 3.2 and (3.5)) imply that

$$\begin{aligned} \mathbb{P}(Q(t) \leq x) &= \mathbb{P}\left(\frac{N - Q(t)}{\sqrt{\lambda}} \geq \frac{N - x}{\sqrt{\lambda}} \right) \\ &\approx \mathbb{P}\left(Z(t) \geq \frac{N - x}{\sqrt{\lambda}} \right) \\ &= \int_0^t p^{(1,0,-\mu)}_{(N-2\lambda/\mu+x)/2\sqrt{\lambda} \to (\mu N - \lambda)/\sqrt{\lambda}}(s)\, \mathrm{d}s, \end{aligned} \tag{4.3}$$

when $Q(0) = N$.

In Figure 3 we compare simulated results for the M/M/N/N queue to values obtained using the approximation in (4.3). We see that the approximation becomes more accurate as N becomes larger, which is as expected. Note that, by (4.2), this also implies that the utilization is close to 1. The simulation results shown are the average over 10 000 runs, stopped at the relevant time value.

References

[1] ALILI, L., PATIE, P. AND PEDERSEN, J. L. (2010). Representations of the first hitting time density of an Ornstein-Uhlenbeck process. Preprint.

[2] ANANTHARAM, V. AND KONSTANTOPOULOS, T. (2011). Integral representation of Skorokhod reflection. *Proc. Amer. Math. Soc.* **139,** 2227–2237.

[3] BURDZY, K., KANG, W. AND RAMANAN, K. (2009). The Skorokhod problem in a time-dependent interval. *Stoch. Process. Appl.* **119,** 428–452.

[4] CHALEYAT-MAUREL, M., EL KAROUI, N. AND MARCHAL, B. (1980). Reflexion discontinue et systèmes stochastiques. *Ann. Prob.* **8,** 1049–1067.

[5] COX, J. T. AND ROSLER, U. (1984). A duality relation for entrance and exit laws for Markov processes. *Stoch. Process. Appl.* **16,** 141–156.

[6] DUPUIS, P. AND ISHII, H. (1991). On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications. *Stoch. Stoch. Reports* **35,** 31–62.

[7] FRALIX, B. H. (2012). On the time-dependent moments of Markovian queues with reneging. *Queueing Systems* 20pp.

[8] HARRISON, J. M. AND REIMAN, M. I. (1981). Reflected Brownian motion on an orthant. *Ann. Prob.* **9,** 302–308.

[9] KRUK, Ł., LEHOCZKY, J., RAMANAN, K. AND SHREVE, S. (2007). An explicit formula for the Skorokhod map on $[0, a]$. *Ann. Prob.* **35,** 1740–1768.

[10] KRUK, Ł., LEHOCZKY, J., RAMANAN, K. AND SHREVE, S. (2008). Double Skorokhod map and reneging real-time queues. In *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz* (Inst. Math. Statist. Collect. **4**), Institute of Mathematical Statistics, Beachwood, OH, pp. 169–193.

[11] LINETSKY, V. (2004). Computing hitting time densities for OU and CIR processes: applications to mean-reverting models. *J. Comput. Finance* **7,** 1–22.

[12] PITMAN, J. AND YOR, M. (1981). Bessel processes and infinitely divisible laws. In *Stochastic Integrals* (Lecture Notes Math. **851**), Springer, Berlin, pp. 285–370.

[13] RAMANAN, K. (2006). Reflected diffusions via the extended Skorokhod map. *Electron. J. Prob.* **36,** 934–992.

[14] SIGMAN, K. AND RYAN, R. (2000). Continuous-time monotone stochastic recursions and duality. *Adv. Appl. Prob.* **32,** 426–445.

[15] SKOROKHOD, A. V. (1961). Stochastic equations for diffusions in a bounded region. *Theoret. Prob. Appl.* **6,** 264–274.

[16] SRIKANT, R. AND WHITT, W. (1996). Simulation run lengths to estimate blocking probabilities. *ACM Trans. Model. Comput. Simul.* **6,** 7–52.

[17] TANAKA, H. (1979). Stochastic differential equations with reflecting boundary conditions in convex regions. *Hiroshima Math J.* **9,** 163–177.

[18] WARD, A. R. AND GLYNN, P. W. (2003). A diffusion approximation for a Markovian queue with reneging. *Queueing Systems* **43,** 103–128.

[19] WARD, A. R. AND GLYNN, P. W. (2003). Properties of the reflected Ornstein–Uhlenbeck process. *Queueing Systems* **44,** 109–123.

[20] WARD, A. R. AND GLYNN, P. W. (2005). A diffusion approximation for a GI/GI/1 queue with balking or reneging. *Queueing Systems* **50,** 371–400.

[21] WHITT, W. (1999). Improving service by informing customers about anticipated delays. *Manag. Sci.* **45,** 192–207.

[22] ZHANG, T. S. (1994). On the strong solutions of one-dimensional stochastic differential equations with reflecting boundary. *Stoch. Process. Appl.* **50,** 135–147.