EURANDOM PREPRINT SERIES
2011-024

A new Wiener-Hopf identity for a general class of reflected processes

Brian H. Fralix, Johan S.H. Leeuwaarden, Onno J. Boxma
ISSN 1389-2355

A new Wiener-Hopf identity for a general class of reflected processes

Brian H. Fralix, Johan S.H. van Leeuwaarden and Onno J. Boxma *

April 11, 2011

Abstract

We derive a new "Wiener-Hopf identity" for a class of preemptive-resume queueing systems, with batch arrivals and catastrophes that, whenever they occur, eliminate multiple customers present in the system. These processes are quite general, as they can be used to approximate Lévy processes, diffusion processes, and certain types of growth-collapse processes: thus, all of the processes mentioned above also satisfy this type of Wiener-Hopf identity. In the Lévy case, this identity simplifies to the well-known Wiener-Hopf factorization. We also show how the ideas can be used to derive transforms for some well-known state-dependent/inhomogeneous birth-death processes and diffusion processes.

Keywords: Lévy processes, Palm distribution, random walks, time-dependent behavior, Wiener-Hopf factorization
2010 MSC: 60G50, 60G51, 60G55, 60K25

1 Introduction

The Wiener-Hopf factorization is a classical result in both the theory of random walks and the theory of Lévy processes. One useful aspect of this factorization is that it gives much-needed insight into the time-dependent behavior of both a random walk and a Lévy process, reflected at a fixed level. These processes appear naturally in many areas of applied probability theory: hence, not only is the Wiener-Hopf factorization interesting in itself, it is also highly useful.

This paper presents a study of the Wiener-Hopf factorization, from a queueing perspective. Indeed, we will show that a certain type of production system that processes work under the Last-Come-First-Served preemptive-resume discipline also satisfies the classical Wiener-Hopf factorization. In addition, we will show that the one-sided reflected version of such a production system satisfies *another* type of factorization that is very much analogous to the Wiener-Hopf factorization of the original production system. This factorization, for some specific special cases, has appeared in the literature as a transform factorization, but we are not aware of a reference that successfully identifies each of the factors found in the transform with its corresponding random variable.

We also show that a very general type of identity can be derived for production systems with state-dependent arrivals and services: we refer to this identity as the Wiener-Hopf identity, which is the main result of this paper. When the process is no longer state-dependent, our identity immediately simplifies to the classical Wiener-Hopf factorization. While our identity, in its most general form, is not a true factorization, it can still play a valuable role towards computing many time-dependent quantities of interest, for a wide variety of production systems. Furthermore, our class of production systems can be used to approximate many classical types of Markov processes found in the literature, such as Lévy processes, diffusion processes, and Markovian Growth-Collapse models: hence, even

*BF (corresponding author) is with Clemson University, Department of Mathematical Sciences, O-110 Martin Hall, Box 340975, Clemson, SC 29634, USA. Email: bfralix@clemson.edu. JvL and OB are with Eindhoven University of Technology, Department of Mathematics and Computer Science, Eindhoven, The Netherlands. Emails: j.s.h.v.leeuwaarden@tue.nl, o.j.boxma@tue.nl

though our production systems may appear, at first glance, to be rather unorthodox, they can be used to approximate many types of stochastic processes that play prominent roles in the literature. To demonstrate how our identity can be useful in a state-dependent setting, we will use it to derive quantities which help describe the time-dependent behavior of some specific types of birth-death processes and diffusion processes.

We also feel that our *proof technique* should be of some interest to the applied probability community. The overall approach towards proving the Wiener-Hopf identity is to first use a basic sample-path argument to derive an infinite system of equations, and then argue that at most one probability mass function may be a solution to such a system. In the continuous-time setting, Palm measures are used to derive the system of equations: however, only their basic properties are used, and the argument should be relatively simple to follow for those who are not aware of the theory of Palm distributions. We also note that proving this identity in this manner does not make use of increasing or decreasing ladder height random variables, contrary to the other standard proofs of the Wiener-Hopf factorization that have appeared in the literature for random walks and Lévy processes.

This type of proof can be simplified considerably, when it is used to establish the Wiener-Hopf factorization for a random walk. In this case, Palm measures are no longer needed, and neither is the strong Markov property. Hence, we show here that it is possible to prove the classical independence property found in the Wiener-Hopf factorization for random walks, without having to introduce the notions of stopping times and the strong Markov property. This could be especially interesting to instructors that wish to introduce students to the Wiener-Hopf factorization at a relatively early moment in a course on applied stochastic processes.

Our proof seems to be similar in flavor to the older approaches toward proving the Wiener-Hopf factorization. In the works of Percheskii and Rogozin [36] and Gusak and Korolyuk [27], the factorization is first established for a special type of stochastic process, and then limiting arguments are used to extend the factorization result to an arbitrary Lévy process. Neither of these works, however, takes advantage of a discrete-state approximation, and this is what allows us to give a proof that both minimizes the mathematical machinery involved, and shows how the factorization result extends to more general settings. A discussion on the history of the Wiener-Hopf factorization, along with related results, can be found in the recent paper of Kuznetsov [30].

The outline of this paper is as follows. Section 2 introduces our production system, which we refer to as a Preemptive-Resume Production system, or PRP system. In Section 3, we establish the main result, which is a Wiener-Hopf identity for a PRP system. Section 4 illustrates how our class of PRP systems can be sufficiently scaled, so that they can be used to approximate an arbitrary Lévy process, in a manner so that our Wiener-Hopf identity for the PRP system can successfully carry over to the classical Lévy setting. Section 5 shows how the proof technique simplifies, when used to derive the Wiener-Hopf identity for random walks. In Section 6, we illustrate how the Wiener-Hopf identity may be used to study the time-dependent behavior of birth-death processes and diffusions. The necessary facts from the theory of Palm distributions are included, for the reader's convenience, in an appendix.

2 Preemptive-Resume Production systems

We now define what we refer to as a Preemptive-Resume Production system, or PRP system. We assume that at time zero there are a countably infinite number of customers present, which are labeled $0, -1, -2, -3, \ldots$. The system then begins, at time zero, to process the work of the customer that possesses the highest label, or number, which at time zero is customer 0. Moreover, at all times, the server always devotes its full attention to the customer that has the highest label. All customers possess a random amount of work, which has distribution function S, and the amount of work possessed by a given customer is independent of the amounts of work of all other customers that will visit, or have visited the system. We are interested in studying the process $Q := \{Q(t); t \geq 0\}$, where $Q(t)$ represents the label of the customer being served by the server at time t.

The dynamics of this system are as follows. We assume that single customers arrive to the system according to a modulated Poisson process $A_0 := \{A_0(t); t \geq 0\}$, with rate function $\lambda_{0,Q(t-)}$, so that the rate is dependent on the current level of the PRP system. Each customer brings with it to the system a random amount of work, which again has distribution function S. Furthermore, we also allow the same type of customers to arrive in batches, where batches arrive according to a modulated Poisson process A_1 with rate $\lambda_{1,Q(t-)}$. The distribution of the number of customers found in a batch is dependent on the highest-labeled customer present in the system, immediately before the batch arrival. The labels of each customer in the new batch are assigned in the following manner: suppose for instance that a batch arrives at time t, and finds that $Q(t-) = n$. Each of the customers in the batch are then given the labels $n+1$, $n+2$, and so on, until each newly arriving customer has a label. The server then immediately begins to process the work of the customer with the highest label, and will not return to customer n until all higher-labeled customers have left the system. Once the server returns to customer n, it resumes working at the point where it previously left-off. Therefore, the server processes work in a Last-Come-First-Served Preemptive-Resume manner. We assume that the rate at which the server processes work also depends on Q, in that at time t, the service rate is $\mu_{Q(t-)} > 0$.

We further assume that catastrophes occur according to a modulated Poisson process $D := \{D(t); t \geq 0\}$, with rate $\delta_{Q(t-)}$. At the time of a catastrophe, a random number of customers are removed from the system: in particular, if $Q(t-) = n$, and a catastrophe occurs at time t, which eliminates k customers, then customers $n, n-1, n-2, \ldots, n-k+1$ are immediately removed from the system, and at time t the server begins to process the remaining amount of work possessed by customer $n-k$, and so $Q(t) = n-k$. We assume that the distribution function of the number of removals at time t depends on $Q(t-)$, so that the downward jump distribution of the process may depend on the level of the process, immediately before a jump.

This type of model can be used to approximate the sample paths of many types of Markov processes of interest. For instance, we will show that Lévy processes can be approximated by PRP systems. Also, the paths of Markovian growth-collapse processes can be approximated. Such growth-collapse processes are fairly prominent in many studies in applied probability, particularly in the TCP literature: examples include Guillemin et al. [26], Boxma et al. [11], and Löpker and van Leeuwaarden [32]. It is also interesting to realize that this model is a more general version of a continuous-time Markov chain on the integers. Hence, many types of diffusion processes may be approximated by PRP systems as well.

3 A useful system of equations

In this section, we will show that PRP systems exhibit a variant of the Wiener-Hopf factorization, which we will refer to as the "Wiener-Hopf identity". The key to deriving our result is to understand conditional probabilities of the form

$$P\left(Q(e_q) = k+l \mid \inf_{0 \leq s \leq e_q} Q(s) = l\right)$$

for integers k, l, where $k \geq 0$, and where e_q is an exponential random variable with rate $q > 0$, which is independent of the PRP system Q.

Throughout this section, we will work with a collection of random variables $\tau_{m,k}(s)$, where

$$\tau_{m,k}(s) = \inf\{t \geq s : Q(t) < k\}$$

under the condition that $Q(s) = m$. Usage of such notation is mainly due to visual appeal, along with our desire to save space: throughout, we could easily replace statements involving $\tau_{m,k}(s)$ with statements containing $\inf_{u \in [s,t]} Q(u)$.

We begin by first presenting the following identity, which is satisfied by the sample paths of our PRP system: for each $t \geq 0$, we see that for any two integers k, l with $k \geq 1$,

$$\mathbf{1}(Q(t) \geq k+l, \inf_{0 \leq u \leq t} Q(u) = l) = \int_0^t \mathbf{1}(Q(s-) = k-1+l, \inf_{0 \leq u \leq s} Q(u) = l)\mathbf{1}(\tau_{k+l,k+l}(s) > t)A_{0,k-1+l}(ds)$$

$$+ \sum_{j=0}^{k-1} \sum_{m=k}^{\infty} \int_0^t \mathbf{1}(Q(s-) = j+l, \inf_{0 \le u < s} Q(u) = l) \mathbf{1}(\tau_{m+l,k+l}(s) > t) A_{1,j+l,m+l}(ds) \quad (1)$$

where $A_{0,k+l}$ is a Poisson process with rate $\lambda_{0,k+l}$, and $A_{1,j+l,m+l}$ is a Poisson process with rate $\lambda_{1,j+l} P(Z_{1,j+l} = m + l - (j+l))$, with $Z_{1,j+l}$ representing an arbitrary jump size from state $j+l$. Here, we are making heavy use of a classical thinning property of Poisson processes, in that a Poisson process can always be expressed in terms of an independent sum of other Poisson processes. In fact, the arrival process of our PRP system is actually governed by an infinite collection of Poisson processes, where at each moment of time only a finite subset of them are active, and this finite subset depends on the current state of the PRP system. This type of modeling trick may seem a bit artificial, but it can be quite useful: a nice example of its use in Markov chain theory can be found in Chapter 9 of Brémaud [13], and another application of this idea can be found in [19].

The identity (1) says that, in order for $Q(t) \ge k+l$, exactly one of two things must happen: if the infimum of the process over $[0,t]$ is l, either (i) there exists a time point $s \le t$ such that $Q(s-) = k-1+l$, $Q(s) = k+l$ (due to the arrival of a customer from A_0 at time s), and the process stays at or above level $k+l$ in $[s,t]$, or (ii) there exists a time point $s \le t$ such that, due to a batch of customers arriving at time s (which is contributed by A_1), the process crosses level $k+l$, reaching some level at or above $k+l$ at time s, and stays at or above $k+l$ during $[s,t]$.

After taking expected values of both sides of (1), we get

$$P(Q(t) \ge k+l, \inf_{0 \le u \le t} Q(u) = l) = E\left[\int_0^t \mathbf{1}(Q(s-) = k-1+l, \inf_{0 \le u < s} Q(u) = l) \mathbf{1}(\tau_{k+l,k+l}(s) > t) A_{0,k-1+l}(ds)\right]$$
$$+ \sum_{j=0}^{k-1} \sum_{m=k}^{\infty} E\left[\int_0^t \mathbf{1}(Q(s-) = j+l, \inf_{0 \le u < s} Q(u) = l) \mathbf{1}(\tau_{m+l,k+l}(s) > t) A_{1,j+l,m+l}(ds)\right]. \quad (2)$$

We can use the Campbell-Mecke formula to evaluate the expected values found on the right-hand side of Equation (2). Notice first that

$$E\left[\int_0^t \mathbf{1}(Q(s-) = k-1+l, \inf_{0 \le u < s} Q(u) = l) \mathbf{1}(\tau_{k+l,k+l}(s) > t) A_{0,k-1+l}(ds)\right]$$
$$= \lambda_{0,k-1} \int_0^t \mathcal{P}_s(Q(s-) = k-1+l, \inf_{0 \le u < s} Q(u) = l, \tau_{k+l,k+l}(s) > t) ds,$$

where $\mathcal{P}$ represents the Palm kernel induced by $A_{0,k-1+l}$. Furthermore, since the server processes work in a preemptive-resume manner, we can also use the Campbell-Mecke formula to establish that

$$\mathcal{P}_s(\tau_{k+l,k+l}(s) > t, Q(s-) = k-1+l, \inf_{0 \le u < s} Q(u) = l)$$
$$= P(\tau_{k+l,k+l} > t-s) \mathcal{P}_s(Q(s-) = k-1+l, \inf_{0 \le u < s} Q(u) = l)$$

where $\tau_{k+l,k+l} \stackrel{d}{=} \tau_{k+l,k+l}(0)$, assuming that at time 0 the process has just made an upward transition into state $k+l$. Moreover, if we let $\{\mathcal{F}_t; t \ge 0\}$ represent the minimal filtration induced by Q and our Poisson processes, we see that the event $\{Q(s-) = k-1+l, \inf_{0 \le u < s} Q(u) = l\} \in \mathcal{F}_{s-}$, and so by Proposition A.1 (see the Appendix) we find that

$$\mathcal{P}_s(Q(s-) = k-1+l, \inf_{0 \le u < s} Q(u) = l) = P(Q(s) = k-1+l, \inf_{0 \le u \le s} Q(u) = l).$$

An analogous argument can be used to evaluate the second type of expectation found in (2). Plugging these expressions into (2) gives

$$P(Q(t) \ge k+l, \inf_{0 \le u \le t} Q(u) = l) = \lambda_{0,k-1+l} \int_0^t P(Q(s) = k-1+l, \inf_{0 \le u \le s} Q(u) = l) P(\tau_{k+l,k+l} > t-s) ds$$
$$+ \sum_{j=0}^{k-1} \sum_{m=k}^{\infty} \lambda_{1,j+l} P(Z_{1,j+l} = m-j) \int_0^t P(\tau_{m+l,k+l} > t-s) P(Q(s-) = j+l, \inf_{0 \le u \le s} Q(u) = l) ds. \quad (3)$$

After integrating both sides of (3) with respect to an exponential density with rate $q > 0$, we get

$$P(Q(e_q) \geq k+l, \inf_{0\leq u\leq e_q} Q(u) = l) = \lambda_{0,k-1+l}\frac{(1-\phi_{k+l,k+l}(q))}{q}P(Q(e_q) = k-1+l, \inf_{0\leq u\leq e_q} Q(u) = l)$$
$$+ \sum_{j=0}^{k-1}\sum_{m=k}^{\infty}\lambda_{1,j+l}P(Z_{1,j+l} = m-j)\frac{(1-\phi_{m+l,k+l}(q))}{q}P(Q(e_q) = j+l, \inf_{0\leq u\leq e_q} Q(u) = l)$$

where $\phi_{m+l,k+l}$ represents the Laplace-Stieltjes transform of $\tau_{m+l,k+l}(0)$ (with $Q(0) = m+l$). Notice that we can also rewrite this equality in the following way: by dividing by $P(\inf_{0\leq u\leq e_q} Q(u) = l)$, we get

$$P(Q(e_q) \geq k+l \mid \inf_{0\leq u\leq e_q} Q(u) = l) = \lambda_{0,k-1+l}\frac{(1-\phi_{k+l,k+l}(q))}{q}P(Q(e_q) = k-1+l \mid \inf_{0\leq u\leq e_q} Q(u) = l)$$
$$+ \sum_{j=0}^{k-1}\sum_{m=k}^{\infty}\lambda_{1,j+l}P(Z_{1,n} = m-j)\frac{(1-\phi_{m+l,k+l}(q))}{q}P(Q(e_q) = j+l \mid \inf_{0\leq u\leq e_q} Q(u) = l). \quad (4)$$

The reader should also notice how the LCFS preemptive-resume discipline is used to derive Equality (3). Indeed, $P_s(\tau_{m+l,k+l}(s) > t-s \mid Q(s-) = k-1+l, \inf_{0\leq u<s} Q(u) = l)$ is equal to the amount of time it takes our production system, starting at level $m+l$ with new customers, to drop below level $k+l$.

This fact follows from the semi-regenerative properties exhibited by our production system at arrival times: it is this semi-regenerative feature of preemptive-resume systems that makes them highly tractable, when compared to queues that operate under other types of service disciplines. Readers wishing to learn more about the LCFS preemptive-resume discipline are referred to the work of Núñez-Queija [34], which also mentions many other useful references on the LCFS preemptive-resume discipline. An especially notable demonstration of how the LCFS discipline can be used to prove other results in queueing can be found in the work of Fuhrmann and Cooper [23]: their result also establishes a factorization, in that it shows that important classes of queueing systems with vacations have stationary distributions that can be factorized in a very useful manner. Indeed, one could consider our work to be another example of an instance where the LCFS discipline can be used to prove a result that, at first glance, appears to have little to do with this discipline.

The next observation we make is crucial, and is stated in the form of a lemma.

Lemma 3.1 *The set of equations generated by* (4) *has a unique probability measure among its set of solutions.*

Proof This follows from the fact that for a fixed integer l, the equations given by (4) can now be iteratively solved, if we also make use of the fact that

$$\sum_{k=0}^{\infty} P(Q(e_q) = k+l \mid \inf_{0\leq u\leq e_q} Q(u) = l) = 1.$$

Indeed, notice that

$$1 - P(Q(e_q) = l \mid \inf_{0\leq u\leq e_q} Q(u) = l) = P(Q(e_q) \geq l+1 \mid \inf_{0\leq u\leq e_q} Q(s) = l)$$
$$= \lambda_{0,l}\frac{(1-\phi_{l+1,l+1}(q))}{q}P(Q(e_q) = l \mid \inf_{0\leq u\leq e_q} Q(u) = l)$$
$$+ \sum_{m=1}^{\infty}\lambda_{1,l}P(Z_{1,l} = m)\frac{(1-\phi_{m+l,1+l}(q))}{q}P(Q(e_q) = l \mid \inf_{0\leq u\leq e_q} Q(u) = l)$$

which allows us to determine $P(Q(e_q) = l \mid \inf_{0\leq u\leq e_q} Q(u) = l)$, and all other probabilities can be determined in a similar, iterative manner. Hence, there is a unique probability measure on the integers that satisfies these equations. $\diamondsuit$

We will now introduce, for a fixed integer l, a new process $Q_l := \{Q_l(t); t \geq 0\}$, where $Q_l(0) = l$. This process behaves in the same way as the original process Q, except that it never drops to a level lower than l. In particular, if the process is currently at a level k, it makes upward transitions in the same way as the original process Q, but potential downward jumps from k to any level lower than l are instead to level l. Notice that in the case where the arrival rates and the jump distributions are not state-dependent, Q_l is merely the reflection of Q, with a reflecting barrier at level l.

The reason why we are interested in Q_l is that its sample paths also satisfy the system of equations given by (4). In particular, we can perform the same type of analysis as above to show that

$$P(Q_l(e_q) \geq k+l) = \lambda_{0,k+l-1}\frac{1-\phi_{k+l,k+l}(q)}{q}P(Q_l(e_q) = k+l-1)$$
$$+ \sum_{j=0}^{k-1}\sum_{m=k}^{\infty}\lambda_{1,l+j}P(Z_{1,l+j} = m-j)\frac{1-\phi_{m+l,k+l}(q)}{q}P(Q_l(e_q) = l+j).$$

Therefore, since Lemma 3.1 tells us that there is a unique probability mass function on the set $\{l, l+1, l+2, \ldots\}$ which satisfies these equations, we conclude that the "Wiener-Hopf identity" holds, which is the main result of our paper.

Theorem 3.1 *(Wiener-Hopf identity) For any two integers k, l, where $k \geq 0$ and $l \leq 0$,*

$$\begin{aligned} P(Q_l(e_q) = k+l) &= P(Q(e_q) = k+l \mid \inf_{0\leq u\leq e_q} Q(u) = l) \\ &= P(Q(e_q) - \inf_{0\leq u\leq e_q} Q(u) = k \mid \inf_{0\leq u\leq e_q} Q(u) = l). \end{aligned}$$

Remark We believe that there are two very interesting aspects of this proof. First, this argument makes heavy use of the fact that the state space of our process forms a lattice on the real line. The fact allows us to derive an infinite system of linear equations, and then conclude that only one probability measure can be a solution to such a system. Second, notice also that our proof of this identity makes virtually no use of ascending and descending ladder epochs. These epochs will, however, come into play when we show that $\inf_{0\leq s\leq e_q} Q(s)$ is infinitely divisible, in the case where the jump distribution of Q is not state-dependent.

3.1 Infinite divisibility

Proposition 3.1 $\inf_{0\leq s\leq e_q} Q(s)$ *is an infinitely divisible random variable.*

Proof If we let $\psi(q)$ denote the Laplace-Stieltjes transform of the busy period of the model of Section 3 (induced by one customer), and let $\{U_k\}_{k\geq 1}$ denote a sequence of i.i.d. random variables, with U_1 being equal in distribution to the number of customers in a batch departing immediately at the end of an arbitrary busy period (i.e. U represents the undershoot), then it is not difficult to see that

$$P(\inf_{0\leq u\leq e_q} Q(u) = 0) = 1 - \psi(q)$$

$$P(\inf_{0\leq u\leq e_q} Q(u) = -1) = \psi(q)P(U=1)(1-\psi(q))$$

$$P(\inf_{0\leq u\leq e_q} Q(u) = -2) = (\psi(q))P(U=2)(1-\psi(q)) + \psi(q)P(U=1)\psi(q)P(U=1)(1-\psi(q))$$

and so on. To see this, note that for the event $\{\inf_{0\leq u\leq e_q} Q(u) = -1\}$ to occur, customer 0 must leave before time e_q, and when he leaves, no other customers leave with him. Moreover, at this departure time, both the system and e_q regenerate, since e_q is exponential, and customer -1 has not received any attention from the server. Hence, conditional on only customer 0 leaving, the probability that customer -1 does not leave before time e_q is just $1-\psi(q)$. The other probabilities can be computed in an analogous manner, by considering all possible scenarios at each busy period completion instant contained in $[0, e_q]$.

In general, for each $k \geq 0$,

$$P(\inf_{0\leq u\leq e_q} Q(u) = -k) = \sum_{j=1}^{k}(1-\psi(q))\psi(q)^j \sum_{r_i\geq 1, r_1+\ldots+r_j=k} \prod_{l=1}^{j} P(U=r_l).$$

However, after some thought, it is easy to see that this expression tells us that

$$P\left(\inf_{0\leq u\leq e_q} Q(u) = -k\right) = P(\sum_{j=1}^{N} U_j = k)$$

where $P(N=n) = (1-\psi(q))\psi(q)^n$, for $n \geq 0$, and N is independent of the sequence $\{U_i\}_{i\geq 1}$. Moreover, $\sum_{j=1}^{N} U_j$ is an infinitely divisible random variable: this can be observed via elementary arguments, but to save space we will simply refer the interested reader to Chapter 2 of Steutel [38]. $\diamondsuit$

The reader should also observe that we are now making use of descending ladder height epochs to establish that $\inf_{0\leq u\leq e_q} Q(u)$ is infinitely divisible.

This argument, of course, does not allow us to conclude that $Q(e_q) - \inf_{0\leq s\leq e_q} Q(s)$ is infinitely divisible: however, if Q is also a Lévy process, we can use a duality argument (see e.g. Chapter 3 of [31]) to conclude that

$$Q(e_q) - \inf_{0\leq u\leq e_q} Q(u) \stackrel{d}{=} \sup_{0\leq u\leq e_q} Q(u)$$

and similar techniques can be used to establish the fact that this random variable is infinitely divisible as well.

3.2 A new type of factorization

We will now show that the process $\{Q_0(t); t \geq 0\}$ also exhibits a very interesting variant of our Wiener-Hopf identity. Suppose that, at time zero, $Q_0(0) = n_0 \geq 0$, for some nonnegative integer n_0. Notice that a sample-path identity that is completely analogous to (1) can be established for Q_0: for each $l \geq 0$, $k \geq 1$,

$$\begin{aligned}
&\mathbf{1}(Q_0(t) \geq k+l, \inf_{0\leq u\leq t} Q_0(u) = l) = \int_0^t \mathbf{1}(Q_0(s-) = k-1+l, \inf_{0\leq u\leq s} Q_0(u) = l)\mathbf{1}(\tau_{k+l,k+l}(s) > t)A_{0,k-1+l}(ds) \\
&+ \sum_{j=0}^{k-1}\sum_{m=k}^{\infty}\int_0^t \mathbf{1}(Q_0(s-) = j+l, \inf_{0\leq u<s} Q_0(u) = l)\mathbf{1}(\tau_{m+l,k+l}(s) > t)A_{1,j+l,m+l}(ds). \qquad (5)
\end{aligned}$$

After taking expectations of both sides of (5), integrating with respect to an exponential density with rate $q > 0$, and conditioning, we may conclude that

$$\begin{aligned}
&P(Q_0(e_q) \geq k+l \mid \inf_{0\leq u\leq e_q} Q_0(u) = l) = \lambda_{0,k-1+l}\frac{(1-\phi_{k+l,k+l}(q))}{q}P(Q_0(e_q) = k-1+l \mid \inf_{0\leq u\leq e_q} Q_0(u) = l) \\
&+ \sum_{j=0}^{k-1}\sum_{m=k}^{\infty}\lambda_{1,j+l}P(Z_{1,j+l} = m-j)\frac{(1-\phi_{m+l,k+l}(q))}{q}P(Q_0(e_q) = j+l \mid \inf_{0\leq u\leq e_q} Q_0(u) = l). \qquad (6)
\end{aligned}$$

For our fixed l, we notice that the equations that form system (4) are the same as the equations found in (6). Hence, we see from Lemma 3.1 that, by uniqueness, it must be the case that

$$P(Q(e_q) \geq k + l \mid \inf_{0\leq u\leq e_q} Q(u) = l) = P(Q_0(e_q) \geq k + l \mid \inf_{0\leq u\leq e_q} Q_0(u) = l).$$

This results in the following theorem.

Theorem 3.2 *Suppose Q is a PRP system with $Q(0) = n_0$, and let Q_0 be the reflected version of Q at level zero, with $Q_0(0) = n_0$. Then for each integer $l \geq 0$, and each integer $k \geq 1$,*

$$P(Q(e_q) - \inf_{0\leq u\leq e_q} Q(u) = k \mid \inf_{0\leq u\leq e_q} Q(u) = l) \quad = \quad P(Q_0(e_q) - \inf_{0\leq u\leq e_q} Q_0(u) = k \mid \inf_{0\leq u\leq e_q} Q_0(u) = l).$$

3.3 State-independent jumps

Two interesting factorization results can be derived, when the jump sizes of both Q and Q_0 are state-independent. What is especially interesting about this result is the following: suppose that the arrival rates, service rates and jump distributions are not dependent on the level of the Q process. Then, for each $k \geq 0$ and l, $P(Q_l(e_q) = k + l) = P(Q_0(e_q) = k)$, and since Q_0 is the reflection of Q at level 0, we also find that

$$Q_0(e_q) \stackrel{d}{=} Q(e_q) - \inf_{0\leq u\leq e_q} Q(u).$$

Hence, Theorem 3.1 tells us that

$$P(Q(e_q) - \inf_{0\leq u\leq e_q} Q(u) = k) = P(Q(e_q) - \inf_{0\leq u\leq e_q} Q(u) = k \mid \inf_{0\leq u\leq e_q} Q(u) = l).$$

In other words, the Wiener-Hopf factorization holds for our PRP system. We state this in the form of a corollary.

Corollary 3.1 *Suppose that $\{Q(t); t \geq 0\}$ represents a PRP system, with state-independent jumps, and let e_q be an exponential random variable with rate $q > 0$, independent of Q. Then for each $\omega \in \mathbb{R}$,*

$$E_0[e^{i\omega Q(e_q)}] = E_0[e^{i\omega \inf_{0\leq u\leq e_q} Q(u)}]E_0[e^{i\omega(Q(e_q)-\inf_{0\leq u\leq e_q} Q(u))}].$$

Here, we let P_x represent a probability measure, under the condition that our process starts at level x.

This factorization has been well-known for Lévy processes since the late 60's, due to Percheskii and Rogozin [36], and the first probabilistic proof of this result was given in Greenwood and Pitman [25].

Moreover, we can also conclude from Theorem 3.2 that for $l \geq 0$,

$$\begin{aligned} P(Q_0(e_q) - \inf_{0\leq u\leq e_q} Q_0(u) = k \mid \inf_{0\leq u\leq e_q} Q_0(u) = l) &= P(Q(e_q) - \inf_{0\leq u\leq e_q} Q(u) = k) \\ &= P(Q_0(e_q) - \inf_{0\leq u\leq e_q} Q_0(u) = k) \end{aligned}$$

where the second equality follows from the simple fact that the reflection of Q_0 at level 0 is equal in distribution to the reflection of Q at level 0. Hence, we see that $Q_0(e_q) - \inf_{0\leq u\leq e_q} Q_0(u)$ is actually independent of $\inf_{0\leq u\leq e_q} Q_0(u)$, which gives us another interesting corollary.

Corollary 3.2 *Suppose that $\{Q_0(t); t \geq 0\}$ is a reflected version of our PRP system, reflected at 0. Then for each $\omega \in \mathbb{R}$, and each integer $n_0 \geq 0$,*

$$E_{n_0}[e^{i\omega Q_0(e_q)}] = E_{n_0}[e^{i\omega \inf_{0\leq u\leq e_q} Q_0(u)}]E_0[e^{i\omega Q_0(e_q)}].$$

Such a factorization result is useful when we are interested in studying processes that start in an arbitrary initial state. The classical Wiener-Hopf factorization tells us that, since $\inf_{0\le u\le e_q} Q(u)$ is independent of $Q(e_q) - \inf_{0\le u\le e_q} Q(u)$, we can use information about the transforms of $Q(e_q)$ and $\inf_{0\le u\le e_q} Q(u)$ to derive the transform of $Q(e_q) - \inf_{0\le u\le e_q} Q(u)$, which represents the distribution of the reflected process, starting in level zero. Theorem 3.2 can then be used to find the distribution of the reflected process, starting in any initial state, since it is clearly equal in distribution to a convolution of the reflected PRP system starting in level zero, and a truncated version of $\inf_{0\le u\le e_q} Q(u)$.

More remarks will be made about both of these factorizations in the next section, within the context of Lévy processes.

4 Connections to Lévy processes

In this section, we will show how our PRP systems can be used to approximate the sample paths of Lévy processes consisting of a Brownian part and a compound Poisson part. Once this approximation procedure has been established, we will be able to show how the Wiener-Hopf identity for PRP systems carries over flawlessly to this special type of Lévy process. Next, the well-established methods of deriving the Lévy-Khintchine representation of Lévy processes will be used to show how the Wiener-Hopf factorization for an arbitrary Lévy process also follows from our results.

4.1 Lévy processes

We begin with a quick definition of a Lévy process. We say that $X := \{X(t); t \ge 0\}$ is a Lévy process if (i) $X(0) = 0$, (ii) X has stationary increments, in that for $0 \le s \le t$, $X(t) - X(s) \stackrel{d}{=} X(t-s)$, and (iii) X has independent increments, in that for $0 < t_1 < t_2 < \ldots < t_n$, the increments $X(t_1)$, $X(t_2) - X(t_1)$, $\ldots$, $X(t_n) - X(t_{n-1})$ are all independent of one another.

Notice that these properties imply that $X(t)$ is an infinitely divisible random variable, for each $t \ge 0$. Moreover, the process X can be sufficiently characterized by the characteristic triple (θ, σ, ν), where $\theta \in \mathbb{R}$, $\sigma > 0$, and ν is a measure on $\mathbb{R} \setminus \{0\}$ such that

$$\int_{\mathbb{R}} (1 \wedge x^2)\nu(dx) < \infty. \tag{7}$$

In particular, we can say that

$$E[e^{-i\omega X(t)}] = e^{-t\Psi(\omega)}$$

where

$$\begin{aligned}\Psi(\omega) &= -i\theta\omega + \frac{\sigma^2\omega^2}{2} \\ &+ \nu(\mathbb{R}\setminus(-1,1))\int_{|x|\ge 1}(1-e^{i\omega x})\frac{1}{\nu(\mathbb{R}\setminus(-1,1))}\nu(dx) \\ &+ \int_{0<|x|<1}(1-e^{i\omega x}+i\omega x)\nu(dx).\end{aligned}$$

Here Ψ is referred to as the characteristic exponent of the Lévy process X. This exponent illustrates that this process consists of an independent sum of three processes: (i) a Brownian motion with drift θ and diffusion coefficient σ^2, (ii) a compound Poisson process with arrival rate $\nu(\mathbb{R}\setminus(-1,1))$ and jump sizes having distribution $\nu(dx)/(\nu(\mathbb{R}\setminus(-1,1)))$, and (iii) a third process, which can be interpreted as a countable sum of independent compound Poisson processes with drift. This expression for Ψ is classical, and is known as the Lévy-Khintchine representation: we refer readers that are not familiar with this representation to Chapter 2 of Kyprianou [31].

4.2 Our approximations

When using PRP systems to approximate Lévy processes that consist of a Brownian part and a compound Poisson part, three types of PRP systems will be relevant:

- If our underlying Lévy process has a Brownian component, our approximation will be a diffusion-scaled PRP system, with exponential services.
- If our underlying Lévy process does not have a Brownian component, but it has negative drift, our approximation will be a fluid-scaled PRP system, with exponential services.
- Finally, if the underlying Lévy process does not have a Brownian component, and has positive drift, then the service times of each customer will be infinite, so the only way a customer can leave the PRP system is through a catastrophic event (i.e. a "large" downward jump). This approximation is analogous to the non-Brownian negative-drift case, in that a fluid scaling is used here as well.

Once we have established the Wiener-Hopf factorization for these processes, the factorization will follow for an arbitrary Lévy process: this extension depends on the arguments used to derive the Lévy-Itô decomposition for a Lévy process, as we will explain below.

4.2.1 The Brownian case

We begin by reviewing some classical concepts from heavy-traffic theory. This theory will show us exactly how to approximate a Brownian motion with drift θ and diffusion coefficient σ^2, with a scaled difference of Poisson processes. In other words, this theory illustrates how a Brownian motion can be thought of as a pathwise limit of a sequence of PRP systems. Readers who wish to consult a recent textbook treatment on the subject are referred to Chapters 5 and 6 of Chen and Yao [14].

Suppose $\{A_0^n(t); t \geq 0\}$ and $\{S_0^n(t); t \geq 0\}$ represent two independent Poisson processes, with rates $\lambda_n^* := n\sigma^2\lambda_n/2$ and $\mu_n^* := n\sigma^2\mu_n/2$, respectively. It is well-known (e.g. [14, 39]) that, from the Functional Central Limit Theorem, the entire process

$$\frac{A_0^n(t) - \lambda_n^* t}{\sqrt{n}}$$

converges weakly as $n \to \infty$ (under the usual Skorohod J_1 topology) to a driftless Brownian motion, with diffusion coefficient $\sigma^2/2$. Similarly, the process

$$\frac{S_0^n(t) - \mu_n^* t}{\sqrt{n}}$$

converges weakly as $n \to \infty$ to the same type of Brownian motion.

If we now impose the following *heavy-traffic* condition, in that

$$\lim_{n\to\infty} \sqrt{n}(\lambda_n - \mu_n) = \frac{2\theta}{\sigma^2}$$

then it follows from these observations that the process

$$B_0^n(t) = \frac{A_0^n(t) - S_0^n(t)}{\sqrt{n}}$$

converges to a Brownian motion, with drift θ and diffusion coefficient σ^2. In our heavy-traffic assumption, we will also assume that $\mu_n \to \mu$, and quite often $\mu = 1$. Thus, the heavy-traffic assumption says that, as n gets large, the system (if stable) becomes more and more congested, and the rate at which it does this is on the order of the square root of n.

A similar approximation scheme can be used to approximate an independent sum of a Brownian motion, and a compound Poisson process. A compound Poisson process is a process $\{C_1(t); t \geq 0\}$, which has jumps at Poisson times: in particular,

$$C_1(t) = \sum_{k=1}^{A_1(t)} Z_k$$

where $A_1(\cdot)$ is a Poisson process with rate γ, and $\{Z_k; k \geq 1\}$ is an i.i.d. sequence of jumps, which is independent of all other random elements. Here, the support of the distribution of the jumps is arbitrary, and could include both positive and negative values.

It is easy to see how to approximate C_1 with a discrete process: just let $Z_{n,k} = \inf\{j \in \mathbb{Z} : \frac{j}{\sqrt{n}} > Z_k\}$. Thus, the process

$$C_1^n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{A_1(t)} Z_{n,k}$$

converges uniformly on compact sets to C_1, as $n \to \infty$. We use $\sqrt{n}$ as our scaling factor, so that it matches the space-scaling factor found in our approximation of Brownian motion.

Due to the independence of B_0^n and C_1^n, we can again apply a continuous mapping argument to conclude that $X^n := B_0^n + C_1^n$ converges weakly (under the Skorohod metric) to a Lévy process X, which consists of both a Brownian motion with drift θ and diffusion coefficient σ^2, and a compound Poisson process with jumps that occur at rate γ, and jump sizes that have distribution $P(Z_1 \in dx)$.

It is not difficult to see that the $B_0^n + C_1^n$ process is just a scaled PRP system, where single customers arrive according to the Poisson process A_0^n, and each brings an exponentially distributed amount of work. Moreover, positive jumps of the C_1^n process correspond to batch arrivals, and negative jumps of the C_1^n process correspond to catastrophes.

4.2.2 The compound Poisson with negative drift case, without a Brownian part

We will now assume that our Lévy process X is of the form

$$X(t) = C_1(t) - \theta t,$$

where $\theta > 0$, and C_1 is again a compound Poisson process with jumps $\{Z_k; k \geq 0\}$ that occur according to a Poisson process A_1.

To derive a sequence of PRP systems whose sample paths converge in a suitable manner to the sample paths of X, we will make use of a classical fluid scaling result. Suppose $\{D(t); t \geq 0\}$ represents a Poisson process with rate $\theta > 0$. Define, for each $n \geq 1$, and each $t \geq 0$

$$D^{n,*}(t) = \frac{1}{n} D(nt).$$

It is well-known that $D(t)/t \to \theta$ as $t \to \infty$, with probability one. Moreover, it is also well-known how this result can be used to establish the fact that, as $n \to \infty$, $D^{n,*}$ converges uniformly, on compact sets, to the function θt. We can also derive a simple approximation for C_1, under a similar scaling: just let, for each $n \geq 1$, $t \geq 0$,

$$C_1^{n,*}(t) = \frac{1}{n} \sum_{k=1}^{A_1(t)} Z_{n,k}^*,$$

where $Z_{n,k}^* = \inf\{j \in \mathbb{Z} : \frac{j}{n} > Z_k\}$. Under these assumptions, we see that the sequence of stochastic processes $X^{n,*} = C_1^{n,*} - D^{n,*}$ converges uniformly on compact sets to X. Moreover, $X^{n,*}$ can be interpreted as a PRP system, with exponential services. In this case, negative jumps from the $C_1^{n,*}$ process correspond to catastrophes, positive jumps from $C_1^{n,*}$ correspond to batch arrivals, and the jumps from $D^{n,*}$ correspond to times of service completion.

Remark This type of approximation scheme is different from what was used in Fralix [20], where similar ideas were used to derive expressions for the time-dependent moments of the $M/G/1$ workload. In [20], the workload process was approximated with a sequence of PRP systems, with batch arrivals, and deterministic services. We could have used a similar kind of approximation for this case as well in this paper.

4.2.3 The compound Poisson with positive drift case, without a Brownian part

This case is analogous to the previous one, in that we can again use a fluid scaling result to approximate a positive drift θ. Again, let $\{D(t); t \geq 0\}$ denote a Poisson process with rate θ, and define $D^{n,*}$ and $C_1^{n,*}$ in precisely the same way as in the PRP approximation for compound Poisson processes with negative drift. Thus, $X^{n,*} := D^{n,*} + C_1^{n,*}$ converges uniformly on compact sets to a compound Poisson process X with positive, deterministic drift θ, and jumps that occur at a Poisson rate γ, and have distribution $P(Z_1 \in dx)$. Moreover, each $X^{n,*}$ is also a PRP system: here, customers arrive with infinite amounts of work, and are removed by negative jumps in the $C_1^{n,*}$ model. These negative jumps correspond to catastrophes, the positive jumps of the $C_1^{n,*}$ correspond to batch arrivals, and the small jumps from the $D^{n,*}$ process correspond to single arrivals, which arrive more and more frequently as n gets large.

4.2.4 An arbitrary Lévy process

Again, suppose that X is a Lévy process with characteristic exponent given by (8). The Lévy-Itô decomposition says that X can be expressed as the sum of three independent Lévy processes $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$, where $X^{(1)}$ is a Brownian motion with characteristic exponent

$$\Psi_1(\omega) = -i\theta\omega + \frac{\sigma^2\omega^2}{2},$$

$X^{(2)}$ is a compound Poisson process with characteristic exponent

$$\Psi_2(\omega) = \nu(\mathbb{R} \setminus (-1,1)) \int_{|x|\geq 1} (1 - e^{i\omega x}) \frac{1}{\nu(\mathbb{R} \setminus (-1,1))} \nu(dx),$$

and $X^{(3)}$ is a limit of a countable sum of "compensated" compound Poisson processes: $X^{(3)}$ has characteristic exponent

$$\Psi_3(\omega) = \int_{0<|x|<1} (1 - e^{i\omega x} + i\omega x)\nu(dx).$$

To show how our Wiener-Hopf factorization for PRP systems implies the well-known factorization for Lévy processes, we will first have to understand precisely how $X^{(3)}$ is constructed. Recall that, in the proof of the Lévy-Itô decomposition (see pg. 50 of [31], the proof of part (i) of Theorem 2.10), it is shown that $X^{(3)}$ is the L^2 limit of a sequence of compound Poisson processes $X_n^{(3)}$, for $n \geq 1$, where $X_n^{(3)}$ has Lévy exponent

$$\int_{2^{-(n+1)} \leq |x| < 1} (1 - e^{i\omega x} + i\omega x)\nu(dx)$$

in the sense that for each $t \geq 0$,

$$\lim_{n\to\infty} E\left(\sup_{0\leq s\leq t} (X^{(3)} - X_n^{(3)})^2 \right) = 0.$$

Thus, it immediately follows that if X_n represents a Lévy process with Lévy exponent

$$\Psi_n(\omega) \quad = \quad -i\theta\omega + \frac{\sigma^2\omega^2}{2}$$

$$
\begin{aligned}
&+ \quad \nu(\mathbb{R} \setminus (-1,1)) \int_{|x|\geq 1} (1-e^{i\omega x}) \frac{1}{\nu(\mathbb{R} \setminus (-1,1))} \nu(dx) \\
&+ \quad \int_{2^{-(n+1)}<|x|<1} (1-e^{i\omega x} + i\omega x)\nu(dx),
\end{aligned}
$$

then, for each $t \geq 0$,

$$
\lim_{n\to\infty} E\left(\sup_{0\leq s\leq t} (X_n(s) - X(s))^2 \right) = 0.
$$

This implies that, for each t, as $n \to \infty$,

$$
\sup_{0\leq s\leq t} X_n(s) \Rightarrow \sup_{0\leq s\leq t} X(s),
$$

where $\Rightarrow$ denotes convergence in distribution (a simple proof of this follows from first establishing the fact that they converge in L^2). Thus, by the Lévy continuity theorem (see e.g. Theorem 5.3 of Kallenberg [29]), it follows that as $n \to \infty$,

$$
\sup_{0\leq s\leq e_q} X_n(s) \Rightarrow \sup_{0\leq s\leq e_q} X(s).
$$

Showing the corresponding convergence of the infimum follows from the elementary fact that, for each t,

$$
\inf_{0\leq s\leq t} X_n(s) - \inf_{0\leq s\leq t} X(s) \quad = \quad \sup_{0\leq s\leq t} -X(s) - \sup_{0\leq s\leq t} -X_n(s)
$$

and so

$$
\begin{aligned}
\left| \inf_{0\leq s\leq t} X_n(s) - \inf_{0\leq s\leq t} (X(s)) \right| \quad &= \quad \left| \sup_{0\leq s\leq t} -X(s) - \sup_{0\leq s\leq t} -X_n(s) \right| \\
&\leq \quad \sup_{0\leq s\leq t} |X_n(s) - X(s)|.
\end{aligned}
$$

Finally, we have that for each t, $X_n(t)$ converges in L^2 to $X(t)$, which also implies convergence in distribution. These are the crucial observations that we will need for proving the Wiener-Hopf factorization, as we will now show.

4.3 The Wiener-Hopf factorization

We are now ready to see how the Wiener-Hopf factorization for Lévy processes follows as a consequence of the Wiener-Hopf identity for PRP systems, whose arrival rates, service rates, and jump distributions do not, at any time, depend on the level of the process.

4.3.1 The classical case

We begin with establishing the well-known version of the Wiener-Hopf factorization, for Lévy processes.

Theorem 4.1 *Suppose X is a Lévy process, and let e_q be an exponential random variable, independent of X, with rate $q > 0$. Then $\inf_{0\leq s\leq e_q} X(s)$ and $X(e_q) - \inf_{0\leq s\leq e_q} X(s)$ are independent.*

Proof Suppose first that $\tilde{X}$ is a Lévy process that consists of only a Brownian component and a compound Poisson component. Hence, there exists a sequence of PRP systems $\{\tilde{X}_n\}_{n\geq 1}$, such that $\tilde{X}_n$ converges uniformly on compact sets to $\tilde{X}$. The reader should also notice that each $\tilde{X}_n$ process is also a Lévy process. Moreover, Corollary 3.1 showed that the Wiener-Hopf factorization is valid for each PRP system with state-independent jumps, and so from the Lévy continuity theorem, we deduce for each $\omega \in \mathbb{R}$ that

$$\begin{aligned} E[e^{i\omega \tilde{X}(e_q)}] &= \lim_{n\to\infty} E[e^{i\omega \tilde{X}_n(e_q)}] \\ &= \lim_{n\to\infty} E[e^{i\omega \sup_{0\le s\le e_q} \tilde{X}_n(s)}]E[e^{i\omega \inf_{0\le s\le e_q} \tilde{X}_n(s)}] \\ &= E[e^{i\omega \sup_{0\le s\le e_q} \tilde{X}(s)}]E[e^{i\omega \inf_{0\le s\le e_q} \tilde{X}(s)}]. \end{aligned}$$

Thus, the Wiener-Hopf factorization holds for all Lévy processes that consist of only a Brownian motion and a compound Poisson part. Using this result, in conjunction with the proof of the Lévy-Itô decomposition, verifies that the Wiener-Hopf factorization holds for an arbitrary Lévy process. $\diamondsuit$

Finally, as a consequence of what was previously discussed at the end of Section 4, we can use these techniques to show that both $\inf_{0\le s\le e_q} X(s)$ and $X(e_q) - \inf_{0\le s\le e_q} X(s)$ are each infinitely divisible, which is again well-known.

4.3.2 The reflected case

We now show how to use Corollary 3.2 to deduce an analogous factorization for Lévy processes.

Theorem 4.2 *Suppose X represents a Lévy process, and let e_q be an exponential random variable with rate $q > 0$, which is independent of X. Moreover, let $R := \{R(t); t \ge 0\}$ represent the reflection of X, with a reflected barrier at state zero. Then, assuming $X(0) = x \ge 0$,*

$$E_x[e^{i\omega R(e_q)}] = E_0[e^{i\omega R(e_q)}]E_x[e^{i\omega \inf_{0\le u\le e_q} R(u)}].$$

Proof The proof of this result is completely analogous to the proof of Theorem 4.1. First, we establish that it holds for a Lévy process X that consists of only a Brownian and compound Poisson part. The general statement then again follows as before, from the proof of the Lévy-Itô decomposition. $\diamondsuit$

This factorization result, for Lévy processes, does not seem to be explicitly known, however direct computations of $E_x[e^{i\omega R(e_q)}]$ have appeared in the literature in some instances.

For example, it is not hard to see that $\inf_{0\le s\le e_q} R(s)$ is equal in distribution to $(x+\inf_{0\le s\le e_q} X(s)) \vee 0$ (here $X(0) = 0$). Hence, if X is spectrally positive, $-\inf_{0\le s\le e_q} X(s)$ is just an exponential random variable, and so $(x + \inf_{0\le s\le e_q} X(s)) \vee 0$ will have a tractable transform. The original Wiener-Hopf factorization can then be used to compute the transform of $R(e_q) - \inf_{0\le s\le e_q} R(s)$. Thus, we can use this factorization to easily derive the transform for the reflected process, starting in any state $x \ge 0$. We emphasize though that these transforms are already known in the spectrally positive case: see e.g. Bingham [10], Bekker et al. [9], or Chapter 9, Theorem 3.10 of Asmussen [6]; but what is interesting is that the transform expressions given in these references do not immediately suggest that such a factorization is possible.

Having said this, there are references in the literature that make heavy use of factorizing the transform of $R(e_q)$, starting at an arbitrary initial level $x \ge 0$. Indeed, both Theorem 9.1 of [1], which holds when X is a Brownian motion, and Theorem 2.1 of [3], which holds when X is the difference of two Poisson processes (the free process of an $M/M/1$ queue), are special cases of Theorem 4.2. The authors of [1, 3] proved both of these results by taking the known transform of the reflected version, and showing through algebraic manipulations that the transform can be factored into the product of two other transforms, where one represents the reflection, starting in state 0.

These transforms were then used in Abate and Whitt [1, 3] to show how the moments of the reflected process at time t, i.e. $E[R(t)^n \mid R(0) = 0]$, for both Brownian motion and the $M/M/1$ queue, can be expressed in terms of cumulative distribution functions associated with the busy period. They also discuss how to use the factorization to compute analogous forms for $E[R(t) \mid R(0) = x]$, for any arbitrary $x \ge 0$. A similar technique is also used in [3] to compute $P(R(t) = j \mid R(0) = 0)$: there they emphasize that it is not immediately obvious how one can use their factorization to compute $P(R(t) = j \mid R(0) = i)$ for any arbitrary $i, j \geq 0$. Later, in [4], the authors succeeded in performing this computation, by applying another technique. However, now that we have successfully identified both of the random variables present in the factorization, we can use an inversion approach analogous to that found in [3] to derive expressions for $P(R(t) = j \mid R(0) = i)$, which match those found in [4]. Readers wishing to read more about such ideas should consult [19].

We would also like to refer the reader to the interesting Lemma 5.1 and Example 3 of Palmowski and Vlasiou [35]. There they give a proof of Theorem 4.2, which is obtained through the use of duality arguments, along with the well-known Wiener-Hopf factorization for Lévy processes, without reflection. Their result actually pertains to the stationary distribution of a reflected Lévy process that experiences a state-dependent transition at random, Erlang-distributed locations, but it is not difficult to see that the ideas used to derive this stationary distribution can also be used to derive the distribution of the reflected process at an independent exponential time. Indeed, Equation 5.2 of [35] gives the Laplace-Stieltjes transform of $R(e_q)$, given $R(0) = B$ (which is random), and what is interesting about this expression is that it has already been factored. Technically, this equation is only valid for a reflected spectrally positive Lévy process, but the arguments given in Example 3 of [35] also establish the factorization, in the general case.

5 Random Walks

The purpose of this section is to illustrate the simplicity of our approach, when it is used to derive the classical Wiener-Hopf factorization for random walks. Indeed, the general Wiener-Hopf identity that we established within the context of PRP systems also appears within the context of random walks, as we will now show.

Suppose $\{Z_n\}_{n\geq 1}$ represents a sequence of random variables, which are defined as follows. Set $S_0 = s \in \mathbb{Z}$, and let $S_1 = s + Z_1$, where the distribution of Z_1 depends on s. Moreover, set $S_2 = S_1 + Z_2$, and assume that the distribution of Z_2 depends on S_1. In general, for each $n \geq 0$, we set $S_{n+1} = S_n + Z_{n+1}$, where the distribution of Z_{n+1} depends on S_n.

To emphasize the dependence of the distribution of Z_{n+1} on S_n, we will write $Z(s)$ as the distribution of Z_{n+1}, given $S_n = s$.

Notice that the Z terms used in this section are very much analogous to the Z terms used in the PRP section: in both cases, they represent the sizes of jumps, and these jumps are allowed to depend on the current level.

We also let $S^l := \{S_n^l\}_{n\geq 0}$ represent our state-dependent random walk, with reflection at level l. This type of process is the random-walk analogue to the Q_l process that was defined in Section 3, where l was our reflecting barrier. Again, S^l jumps upward in a manner that is exactly the same as S does, but potential downward jumps from an arbitrary level k to any level lower than l are instead made to level l.

We are now ready to state our main result of this section.

Theorem 5.1 *Suppose G_p represents a geometric random variable on $\{0, 1, 2, 3, \ldots\}$, independent of our random walk S. Furthermore, suppose that $S_0 = s \in \mathbb{Z}$. Then for each $l \leq s$, $k \geq 0$,*

$$P(S_{G_p} = k + l \mid \inf_{0\leq n\leq G_p} S_n = l) = P(S_{G_p}^l = k + l).$$

Proof Notice that

$$\mathbf{1}(S_n \geq k + l, \inf_{0\leq u\leq n} S_u = l) = \sum_{j=0}^{k-1} \sum_{t=1}^{n} \sum_{m=k}^{\infty} \mathbf{1}(S_{t-1} = j + l, \inf_{0\leq u\leq t-1} S_u = l, Z_t(k + l) = m - j, \tau_{m+l,k+l}(t) > n),$$

where $\tau_{m+l,k+l}(t) = \inf\{n > t : W_n < k + l\}$, under $W_t = m + l$. We can proceed in precisely the same manner as before: after taking expectations of both sides, we find that

$$P(S_n \geq k + l, \inf_{0\leq u\leq n} S_u = l) = \sum_{j=0}^{k-1} \sum_{t=1}^{n} \sum_{m=k}^{\infty} P(\tau_{m+l,k+l} > n - t) P(Z(k + l) = m - j) P(S_{t-1} = j + l, \inf_{0\leq u\leq t-1} S_u = l),$$

where again, we interpret $\tau_{m+l,k+l}$ to be $\tau_{m+l,k+l}(0)$. Finally, if we now multiply both sides of this equality by $(1-p)p^n$, and then sum over all possible n, we conclude, after some algebra and conditioning, that for each $k \geq 1$,

$$P(S_{G_p} \geq k+l \mid \inf_{0 \leq u \leq G_p} S_u = l) = \sum_{j=0}^{k-1} P(S_{G_p} = j+l \mid \inf_{0 \leq u \leq G_p} S_u = l) \sum_{m=k}^{\infty} \frac{p(1-E[p^{\tau_{m+l,k+l}}])}{1-p} P(Z(k+l) = m-j).$$

Due to the fact that

$$\sum_{k=0}^{\infty} P(S_{G_p} = k+l \mid \inf_{0 \leq u \leq G_p} S_u = l) = 1,$$

we deduce that this system of equations has only one solution that is a probability mass function. Since we can use the same type of techniques to conclude that $P(S^l_{G_p} = k+l)$ satisfies precisely the same system of equations, we can again conclude that

$$P(S_{G_p} = k+l \mid \inf_{0 \leq n \leq G_p} S_n = l) = P(S^l_{G_p} = k+l)$$

and this completes the proof. $\diamondsuit$
The analogous Wiener-Hopf identity for the reflected random walk can also be derived in an analogous manner: we leave the details to the interested reader.

We feel that this argument possesses many interesting qualities. Firstly, just as in the analogous proof for the PRP systems, increasing or decreasing ladder heights are not used in order to establish the Wiener-Hopf identity. Secondly, notice that no use is made of stopping times, or the strong Markov property: only the Markov property is used to derive our system of equations. Finally, Palm measures are not needed in this setting either, simply because we are working in discrete-time.

If we now suppose that the $\{Z_n\}_{n\geq 1}$ sequence is i.i.d., our identity again reduces to the classical Wiener-Hopf factorization, for integer-valued random walks. Moreover, by an obvious scaling argument, the Wiener-Hopf factorization for an arbitrary random walk also follows as a consequence of the Wiener-Hopf factorization for integer-valued random variables.

6 Applications to birth-death processses, and diffusions

Our Wiener-Hopf identity, in its most general form, allows for the jumps of our PRP system to be state-dependent. Hence, if we assume that customers with label m bring with them an exponential amount of work with rate 1, which is processed at rate μ_m, we find that our PRP system is just a continuous-time Markov chain on the integers. Moreover, if customers arrive one-at-a-time to the system, and no catastrophes occur, we may further claim that our PRP system is actually a birth-death process on the integers. Thus, our Wiener-Hopf identity is valid for birth-death processes; this fact will be thoroughly exploited throughout our calculations.

6.1 Birth-death processes

Suppose that $Q := \{Q(t); t \geq 0\}$ represents a birth-death process on the integers, with birth rates $\{\lambda_n\}_{n\in\mathbb{Z}}$ and death rates $\{\mu_n\}_{n\in\mathbb{Z}}$. Moreover, let e_q represent an exponential random variable with rate $q > 0$, independent of Q. Throughout, we will assume that Q is ergodic, and we will let π represent its stationary distribution.

We begin by showing how our Wiener-Hopf identity can be used to derive expressions for the probability mass function of $Q(e_q)$. What makes this approach interesting is that the expressions that we derive are amenable to continuous-mapping arguments: this implies that analogous expressions can also be derived for any diffusion process, assuming that the diffusion is the weak limit of a sequence of properly scaled birth-death processes.

6.1.1 Some initial calculations

By Corollary 4.1.1 of Abate and Whitt [3], we can make use of the reversibility of birth-death processes to say that for each $n \in \mathbb{Z}$,

$$P_0(Q(e_q) = n) = \frac{\pi_n E_n[e^{-q\tau_0}]}{\sum_{k\in\mathbb{Z}} \pi_k E_k[e^{-q\tau_0}]}$$

where P_n is meant to represent the conditional probability that $Q(0) = n$. Hence, we find that there is a very nice relationship between the probability mass function of $Q(e_q)$, and the hitting time distribution of Q. We should also mention here that an analogous expression actually holds in the case where the birth-death process is not ergodic: see [19] for details.

However, suppose that we are interested in computing the pmf of $Q(e_q)$, when $Q(0) = n_0 \neq 0$. While the same method will tell us that

$$P_{n_0}(Q(e_q) = n) = \frac{\pi_n E_n[e^{-q\tau_{n_0}}]}{\sum_{j\in\mathbb{Z}} \pi_j E_j[e^{-q\tau_{n_0}}]}$$

we must be careful: how do we know that $E_n[e^{-q\tau_{n_0}}]$ is tractable? This is a very legitimate question, as there are many instances where $E_n[e^{-q\tau_{n_0}}]$ will be tractable for some choices of n_0, but not for others. In such an instance, we refer to state n_0 as a reference point: the reference point is the point that we would like to appear in the hitting-time Laplace-Stieltjes transforms found in our probability mass function for $Q(e_q)$. Fortunately, the Wiener-Hopf identity allows us to derive computable expressions, since it allows us to use whatever reference point we'd like, for each initial value.

Our next example illustrates how the Wiener-Hopf identity can be useful towards deriving the pmf of $Q(e_q)$, when $Q(t)$ represents the number of customers present in an $M/M/s$ queueing system at time t. The reader will clearly see that this identity plays a large role in the calculations: in essence, it allows us to express $M/M/s$ quantities in terms of simple quantities that are associated with simpler systems, namely the $M/M/1$ queue and the $M/M/\infty$ queue.

6.1.2 The $M/M/s$ queue

Recall that the $M/M/s$ queue is a birth-death process on $\{0, 1, 2, \ldots\}$ with birth rates $\lambda_n = \lambda$, for $n \geq 0$, and death rates $\mu_n = \min\{n, s\}\mu$, for $n \geq 1$. A classical reference on the time-dependent behavior of the $M/M/s$ queue is Saaty [37], which makes use of the classical approach found in Bailey [8].

From what we have already seen in [3], we find that if $Q(0) = s$, then for each $n \geq 0$,

$$P_s(Q(e_q) = n) = \frac{\pi_n E_n[e^{-q\tau_s}]}{\sum_{j\geq 0} \pi_j E_j[e^{-q\tau_s}]}.$$

We now aim to derive an expression for the pmf of $Q(e_q)$ that contains Laplace-Stieltjes transforms of the hitting time τ_s, for *every* feasible initial condition. To see why the appearance of this particular hitting-time transform is desirable, notice that if $k < s$, $E_k[e^{-q\tau_s}]$ is the Laplace-Stieltjes transform of the amount of time it takes our $M/M/s$ process to go from level k to level s, but this is the same as the Laplace-Stieltjes transform of the amount of time it takes us to go from k to s in an $M/M/\infty$ queue, with arrival rate λ and service rate μ. Similarly, for $k > s$, $E_k[e^{-q\tau_s}]$ is just the LST of the amount of time it takes us to go from level k to level s in an $M/M/1$ queue, with arrival rate λ and service rate $s\mu$. Hence, all of the terms in our expression for $P_s(Q(e_q) = k)$ can theoretically be derived from two simpler models, the $M/M/1$ queue and the $M/M/\infty$ queue.

Indeed, for $k > s$, we already have a closed-form expression for $E_k[e^{-q\tau_s}]$: letting $\psi(q) = E_{s+1}[e^{-q\tau_s}]$ be the busy period of an $M/M/1$ queue with arrival rate λ and service rate $s\mu$, we see that

$$E_k[e^{-q\tau_s}] = \psi(q)^{k-s}.$$

Hence, we will focus on the case where $k < s$. Letting $\{Q_{M/M/\infty}(t); t \geq 0\}$ represent the queue-length process of an $M/M/\infty$ queue (including the customers in service), we can follow a classical argument found in Darling and Siegert [17] to find that

$$\begin{aligned} P_k(Q_{M/M/\infty}(e_q) = s) &= P_k(Q_{M/M/\infty}(e_q) = s, \tau_s \leq e_q) \\ &= P_s(Q_{M/M/\infty}(e_q) = s) E_k[e^{-q\tau_s}] \end{aligned}$$

which shows that

$$E_k[e^{-q\tau_s}] = \frac{P_k(Q_{M/M/\infty}(e_q) = s)}{P_s(Q_{M/M/\infty}(e_q) = s)}.$$

Hence, we see that our LST of interest can in fact be computed.

To compute $P_k(Q_{M/M/\infty}(e_q) = s)$, we will need to make use of the following known lemma. This was also observed in Flajolet and Guillemin [18], but we repeat it here for convenience.

Lemma 6.1 *For a positive real number q,*

$$\int_0^\infty q e^{-(qt+\rho(1-e^{-\mu t}))} dt = M\left(1, \frac{q}{\mu} + 1, -\rho\right)$$

where M is Kummer's function, i.e.

$$M(a, b, z) = \sum_{n=0}^{\infty} \frac{(a)_n z^n}{(b)_n n!}$$

with $(a)_0 = 1$, and for $n \geq 1$, $(a)_n = (a)(a+1)\cdots(a+n-1)$.

Proof Applying partial integration gives

$$\int_0^\infty e^{-\rho(1-e^{-\mu t})} q e^{-qt} dt = 1 - \rho\mu \int_0^\infty e^{-(q+\mu)t} e^{-\rho(1-e^{-\mu t})} dt.$$

After further applying partial integration infinitely many more times, we arrive at the result. ♢

Lemma 6.2 *For each $k \leq s$,*

$$P_k(Q_{M/M/\infty}(e_q) = s) = \sum_{j=0}^{k} \sum_{m=0}^{k+s-2j} \binom{k}{j} \binom{k+s-2j}{m} \frac{(\rho)^{s-j}(-1)^m}{(s-j)!} \frac{q}{q+(j+m)\mu} M\left(1, \frac{q}{\mu} + j + m + 1, -\rho\right).$$

Proof This identity can be derived from the known fact that, at a fixed time $t \geq 0$, $Q(t)$ is the convolution of a binomial random variable with parameters $(k, e^{-\mu t})$ and a Poisson random variable with parameter $\rho(1 - e^{-\mu t})$. The result then follows by integrating the pmf of $Q(t)$, and applying Lemma 6.1. ♢

From this lemma, we can now conclude that the Laplace-Stieltjes transform $E_k[e^{-q\tau_s}]$ can be expressed as a ratio of two such sums. In other words, we may state the following lemma.

Lemma 6.3 *For each $k \leq s$, we see that*

$$E_k[e^{-q\tau_s}] = \frac{\sum_{j=0}^{k} \sum_{m=0}^{k+s-2j} \binom{k}{j} \binom{k+s-2j}{m} \frac{(\rho)^{s-j}(-1)^m}{(s-j)!} \frac{q}{q+(j+m)\mu} M\left(1, \frac{q}{\mu} + j + m + 1, -\rho\right)}{\sum_{j=0}^{s} \sum_{m=0}^{2(s-j)} \binom{s}{j} \binom{2(s-j)}{m} \frac{(\rho)^{s-j}(-1)^m}{(s-j)!} \frac{q}{q+(j+m)\mu} M\left(1, \frac{q}{\mu} + j + m + 1, -\rho\right)}.$$

Our next step is to use the Wiener-Hopf identity to compute probabilities of the form $P_k(Q(e_q) = n)$, for arbitrary $k, n \geq 0$. Notice that we already have a nice expression for such a pmf, when $k = s$.
Case 1: $k > s$, $n \leq s$. Notice that

$$\begin{aligned} P_k(Q(e_q) = n) &= P_k(Q(e_q) = n, \tau_s \leq e_q) \\ &= P_k(Q(e_q) = n \mid \tau_s \leq e_q) E_k[e^{-q\tau_s}] \\ &= P_s(Q(e_q) = n) E_k[e^{-q\tau_s}] \end{aligned}$$

and so we conclude that, from our previous calculations, this probability is already tractable.
Case 2: $k > s$, $n > s$. This case is much more interesting. Now, it is possible for our process to go from k to n, without ever reaching level s in $[0, e_q]$. Proceeding in the same manner as in Case 1, we find that

$$\begin{aligned} P_k(Q(e_q) = n) &= P_k(Q(e_q) = n, \tau_s \leq e_q) + P_k(Q(e_q) = n, \tau_s > e_q) \\ &= P_s(Q(e_q) = n) E_k[e^{-q\tau_s}] + \sum_{l=s+1}^{\min\{n,k\}} P_k(Q(e_q) = n \mid \inf_{0 \leq u \leq e_q} Q(u) = l) P_k(\inf_{0 \leq u \leq e_q} Q(u) = l). \end{aligned}$$

However, notice that

$$\begin{aligned} P_k(\inf_{0 \leq u \leq e_q} Q(u) = l) &= P_k(\tau_l \leq e_q) - P_k(\tau_{l-1} \leq e_q) \\ &= E_k[e^{-q\tau_l}] - E_k[e^{-q\tau_{l-1}}] \\ &= \psi(q)^{k-l} - \psi(q)^{k-l+1} \end{aligned}$$

and, furthermore, we find from the Wiener-Hopf identity that, conditional on $\inf_{0 \leq u \leq e_q} Q(u) = l$, Q behaves as an $M/M/1$ queue on $[0, e_q]$ with arrival rate λ and service rate $s\mu$. Hence, we can also say that

$$P_l(Q(e_q) = n \mid \inf_{0 \leq u \leq e_q} Q(u) = l) = \left(1 - \frac{\lambda\psi(q)}{s\mu}\right)\left(\frac{\lambda\psi(q)}{s\mu}\right)^{n-l}.$$

Case 3: $0 \leq k < s$, $n \geq s$. This case is analogous to Case 1: by following the same techniques, we find that

$$P_k(Q(e_q) = n) = P_s(Q(e_q) = n) E_k[e^{-q\tau_s}].$$

Now we can use Lemma 6.3 to express $E_k[e^{-q\tau_s}]$ in terms of Kummer functions.
Case 4: $0 \leq k < s$, $n < s$. As expected, this case is analogous to Case 2, but the expression in this case is more complicated than the other expressions. Here

$$\begin{aligned} P_k(Q(e_q) = n) &= P_k(Q(e_q) = n, \tau_s \leq e_q) + P_k(Q(e_q) = n, \tau_s > e_q) \\ &= P_s(Q(e_q) = n) E_k[e^{-q\tau_s}] + \sum_{l=\max\{k,n\}}^{s-1} P_k(Q(e_q) = n \mid \sup_{0 \leq u \leq e_q} Q(u) = l) P_k(\sup_{0 \leq u \leq e_q} Q(u) = l). \end{aligned}$$

However, we can again observe that

$$P_k(\sup_{0 \leq u \leq e_q} Q(u) = l) = E_k[e^{-q\tau_l}] - E_k[e^{-q\tau_{l+1}}]$$

and conditional on $\sup_{0 \leq u \leq e_q} Q(u) = l$, we observe from our Wiener-Hopf identity that Q behaves as an $M/M/l/l$ queue on $[0, e_q]$, starting at level l. Hence, we see that

$$P_k(Q(u) = n \mid \sup_{0 \leq u \leq e_q} Q(u) = l) = \frac{\frac{\rho^n}{n!} E_n[e^{-q\tau_l}]}{\sum_{j=0}^{l} \frac{\rho^j}{j!} E_j[e^{-q\tau_l}]}.$$

Hence, the entire pmf of $Q(e_q)$ can be expressed in terms of Kummer functions.

There is an important lesson to be learned from our calculations of the pmf of $Q(e_q)$. We have shown that, through a proper choice of initial point and reference point, our probability mass function of $Q(e_q)$ can be expressed in terms of quantities related to three simpler models: the $M/M/1$ queue, the $M/M/l/l$ queue, and the $M/M/\infty$ queue.

6.1.3 The $M/M/s/K$ queue

This technique can also be used to derive the pmf of $Q(e_q)$, when $\{Q(t); t \geq 0\}$ represents the queue-length process of an $M/M/s/K$ queueing system, where s represents the number of servers and K the total system capacity, which is a bound on the number of customers that can be in the system at any given time. By choosing our reference point to be s, we see that this model can be expressed in terms of three simpler models: the $M/M/\infty$ queue, the $M/M/l/l$ queue, and the $M/M/1/(K-s)$ queue.

The relevant hitting-time transforms for the $M/M/1/(K-s)$ queue can be derived from the $M/M/1$ queue, since we can use the pmf of an $M/M/1$ queue at an exponential time to derive the LST of the time it takes us to go from level j_1 to level j_2 in an $M/M/1$ queue, when $j_1 < j_2$. Such a result can then be used to derive all of the corresponding hitting-time transforms for an $M/M/1/(K-s)$ queue. Once this has been observed, the Wiener-Hopf identity can be used to derive the pmf of $Q(e_q)$ for the $M/M/s/K$ queue, in a manner analogous to what was done above for the $M/M/s$ queue.

6.1.4 Time-dependent moments

It is possible to make use of the Wiener-Hopf identity to derive the moments of $Q(e_q)$ as well. To illustrate the main idea, we first suppose that $\{Q(t); t \geq 0\}$ represents an $M/M/1$ queue-length process, with arrival rate λ and service rate μ. It has been shown in Abate and Whitt [2] that, for each $t \geq 0$,

$$E[Q(t) \mid Q(0) = 0] = \frac{\rho}{1-\rho} P(R_\tau \leq t)$$

where τ represents the busy period of an $M/M/1$ queue, and R_τ represents the residual busy period, i.e. for each $t > 0$,

$$P(R_\tau > t) = \frac{1}{E[\tau]} \int_t^\infty P(\tau > x) dx.$$

Letting e_q be an exponential r.v. with rate $q > 0$, independent of Q, gives

$$\begin{aligned} E[Q(e_q) \mid Q(0) = 0] &= \frac{\rho}{1-\rho} E[e^{-qR_\tau}] \\ &= \frac{\rho}{1-\rho} \frac{1 - E[e^{-q\tau}]}{qE[\tau]} \\ &= \frac{\lambda(1 - E[e^{-q\tau}])}{q} \end{aligned}$$

which implies that the first moment of $Q(e_q)$ is tractable, assuming we start in state 0.

The Wiener-Hopf identity can be used in many ways to compute the first moment of $Q(e_q)$, for any initial condition. One possible procedure is the following. Suppose that $Q(0) = n_0 \geq 0$. Then

$$\begin{aligned} E[Q(e_q) \mid Q(0) = n_0] &= E[Q(e_q) \mid \inf_{0 \leq s \leq e_q} Q(s) = 0, Q(0) = n_0] P(\inf_{0 \leq s \leq e_q} Q(s) = 0 \mid Q(0) = n_0) \\ &+ \sum_{k=1}^{n_0} E[Q(e_q) \mid \inf_{0 \leq s \leq e_q} Q(s) = k, Q(0) = n_0] P(\inf_{0 \leq s \leq e_q} Q(s) = k \mid Q(0) = n_0) \\ &= E[Q(e_q) \mid Q(0) = 0] P(\inf_{0 \leq s \leq e_q} Q(s) = 0 \mid Q(0) = n_0) \\ &+ \sum_{k=1}^{n_0} (E[Q(e_q) \mid Q(0) = 0] + k) P(\inf_{0 \leq s \leq e_q} Q(s) = k \mid Q(0) = n_0) \\ &= E[Q(e_q) \mid Q(0) = 0] + \sum_{k=1}^{n_0} k P(\inf_{0 \leq s \leq e_q} Q(s) = k \mid Q(0) = n_0) \end{aligned}$$

$$= \quad \frac{\lambda(1 - E[e^{-q\tau}])}{q} + \sum_{k=1}^{n_0} k\psi(q)^{n_0-k}(1 - \psi(q)).$$

The important step to understand in this derivation is the second equality: if $\inf_{0 \leq s \leq e_q} Q(s) = k$, we can say from the Wiener-Hopf identity that $Q(e_q)$ is equal in distribution to an $M/M/1$ queue on the states $\{k, k+1, k+2, \ldots\}$ with arrival rate λ and service rate μ. This result agrees with the result given in [3], and also in [21]. With a bit of patience, higher moments can also be computed through the use of this approach.

An analogous procedure can be used to compute the moments of $Q(e_q)$, for more complicated processes. Suppose now that $\{Q(t); t \geq 0\}$ represents the queue-length process of an $M/M/s$ queue, with arrival rate λ and service rate μ, and s servers. While the transient moments of the $M/M/s$ queue have been studied in Marcellán and Pérez [33], the point here is to show how to construct the moments from simpler birth-death processes.

The key to computing the moments of $Q(e_q)$ for an arbitrary initial condition is to first compute the moments, while assuming that $Q(0) = s$, since we will want to again use s as a reference point when we apply the Wiener-Hopf identity. Again, since Q is a reversible process, we can say that

$$E[Q(e_q) \mid Q(0) = s] = \pi_0 \sum_{k=0}^{s-1} kE_k[e^{-q\tau_s}]\frac{\rho^k}{k!} + s\pi_0\frac{\rho^s}{s!} + \pi_0\frac{\rho^s}{s!}\sum_{k=s+1}^{\infty} kE_k[e^{-q\tau_s}](\rho/s)^{k-s}.$$

There are a few observations here that are worth noting. First, notice that the term

$$\pi_0 \sum_{k=0}^{s} kE_k[e^{-q\tau_s}]\frac{\rho^k}{k!} \quad = \quad P(Q(\infty) \leq s)E_s[Q_{M/M/s/s}(e_q)]$$

where $Q_{M/M/s/s}$ represents an $M/M/s/s$ loss model with arrival rate λ, service rate μ, and s servers, and this is a known expected value; see Abate and Whitt [5] for details. Second, we see that

$$\begin{aligned}
\pi_0\frac{\rho^s}{s!}\sum_{k=s+1}^{\infty} kE_k[e^{-q\tau_s}](\rho/s)^{k-s} \quad &= \quad \pi_0\frac{\rho^s}{s!}\sum_{k=s+1}^{\infty}(k-s)E_k[e^{-q\tau_s}](\rho/s)^{k-s} \\
&+ \quad \pi_0\frac{\rho^s}{s!}\sum_{k=s+1}^{\infty} sE_k[e^{-q\tau_s}](\rho/s)^{k-s} \\
&= \quad P(Q(\infty) \geq s)E_0[Q_{M/M/1}(e_q)] + sP(Q(\infty) \geq s)P_0(Q(e_q) \geq 1)
\end{aligned}$$

where $Q_{M/M/1}$ represents an $M/M/1$ queue with arrival rate λ and service rate $s\mu$. Thus, we conclude that $E[Q(e_q) \mid Q(0) = s]$ is a quantity that can be computed.

To get $E[Q(e_q) \mid Q(0) = i]$ for an arbitrary $i \geq 0$, we now invoke the Wiener-Hopf identity. Suppose first that $i < s$. Then

$$\begin{aligned}
E[Q(e_q) \mid Q(0) = i] \quad &= \quad \sum_{j=i}^{s-1} E[Q(e_q) \mid \sup_{0 \leq s \leq e_q} Q(s) = j, Q(0) = i]P(\sup_{0 \leq s \leq e_q} Q(s) = j \mid Q(0) = i) \\
&+ \quad E[Q(e_q) \mid Q(0) = s]P(\tau_s \leq e_q)
\end{aligned}$$

and we observe from the Wiener-Hopf identity that, conditional on $\sup_{0 \leq s \leq e_q} Q(s) = j$, $Q(e_q)$ behaves as an $M/M/j/j$ queue on $\{0, 1, 2, \ldots, j\}$, which means that

$$E[Q(e_q) \mid \sup_{0 \leq s \leq e_q} Q(s) = j, Q(0) = i] \quad = \quad E[Q_{M/M/j/j}(e_q) \mid Q_{M/M/j/j}(0) = j].$$

All of the other terms in the sum are, for similar reasons, also tractable. A similar argument can be used to derive $E[Q(e_q) \mid Q(0) = i]$ for $i > s$; we omit the details.

We also point out that a similar argument can be used to derive moment expressions for the $M/M/s$ queue with exponential reneging, i.e. the $M/M/s - M$ queue, which is the model studied in Garnett et al. [24]. Such moments would be decomposed into components from an $M/M/s/s$ queue, and a $M/M/1 - M$ queue, and the $M/M/1 - M$ queue moments have recently been studied in [22].

6.2 Diffusion processes

Our goal is now to show how the results derived above for birth-death processes can be used to establish similar expressions for diffusion processes. We shall do so for the case of regulated Brownian motion

6.2.1 Regulated Brownian motion

Suppose that $\{B(t); t \geq 0\}$ represents a Brownian motion, with drift $\mu = -1$ and volatility $\sigma^2 = 1$. We are interested in understanding the time-dependent behavior of $\{R(t); t \geq 0\}$, where

$$R(t) = B(t) - \inf_{0 \leq u \leq t} B(u)$$

i.e. R is the one-sided reflection of B. Granted, since B is a Lévy process, we can already use the Wiener-Hopf factorization to derive the Laplace-Stieltjes transform of $R(e_q)$. However, we will instead be interested in showing how our Wiener-Hopf identity can also be used to derive the probability density function of $R(e_q)$.

To derive this pdf, we will need to know a bit about the distribution of the hitting times associated with a Brownian motion. Following the classical argument of applying the optional sampling theorem to the Wald martingale, we see that

$$E_x[e^{-q\tau_0}] = e^{-(-1+\sqrt{1+2q})x}.$$

Moreover, R has a unique stationary distribution π, where $\pi(dx) = 2e^{-2x}dx$.

We will now compute the density of $R(e_q)$, given $R(0) = x_0$: we denote this density at the point x as $f_{R(e_q)}(x; x_0)$. Again, we will need to break the calculation up into cases. Considering first the case where $x > x_0$, we may use our Wiener-Hopf identity, along with a weak-convergence argument to find that

$$\begin{aligned} P_{x_0}(R(e_q) > x) &= E_{x_0}[e^{-q\tau_0}] \frac{\int_x^\infty E_y[e^{-q\tau_0}]\pi(dy)}{\int_0^\infty E_y[e^{-q\tau_0}]\pi(dy)} \\ &+ \int_0^{x_0} \frac{\int_x^\infty E_y[e^{-q\tau_0}]\pi(dy)}{\int_z^\infty E_y[e^{-q\tau_0}]\pi(dy)} dP(\inf_{0 \leq u \leq e_q} R(u) \leq z). \end{aligned}$$

For $x \geq 0$, we can use our expressions for both the hitting-time LST and the stationary distribution to show that

$$\begin{aligned} \int_x^\infty E_y[e^{-q\tau_0}]\pi(dy) &= \int_x^\infty e^{-(-1+\sqrt{1+2q})y} 2e^{-2y} dy \\ &= \frac{2}{1+\sqrt{1+2q}} e^{-(1+\sqrt{1+2q})x}. \end{aligned}$$

Also, for $0 < z < x_0$,

$$\begin{aligned} P_{x_0}(\inf_{0 \leq u \leq e_q} R(u) \leq z) &= P_{x_0}(\tau_z \leq e_q) \\ &= E_{x_0}[e^{-q\tau_z}] \\ &= E_{x_0 - z}[e^{-q\tau_0}] \\ &= e^{-(-1+\sqrt{1+2q})(x_0 - z)} \end{aligned}$$

so for positive z, we find that the density of $\inf_{0 \leq u \leq e_q} R(u)$ is just

$$dP(\inf_{0 \leq u \leq e_q} R(u) \leq z) = (-1 + \sqrt{1+2q}) e^{-(-1+\sqrt{1+2q})x_0} e^{(-1+\sqrt{1+2q})z} dz.$$

Plugging everything in, we can now say that

$$\begin{aligned} P_{x_0}(R(e_q) > x) &= e^{-(-1+\sqrt{1+2q})x_0} e^{-(1+\sqrt{1+2q})x} \\ &+ \int_0^{x_0} e^{-(1+\sqrt{1+2q})x} e^{(1+\sqrt{1+2q})z} (-1+\sqrt{1+2q}) e^{-(-1+\sqrt{1+2q})x_0} e^{(-1+\sqrt{1+2q})z} dz \\ &= e^{-(-1+\sqrt{1+2q})x_0} e^{-(1+\sqrt{1+2q})x} \left[1 + \frac{(-1+\sqrt{1+2q})}{2\sqrt{1+2q}} \left[e^{2\sqrt{1+2q}x_0} - 1 \right] \right] \end{aligned}$$

and so after taking derivatives and multiplying by (-1), we find that the transient density of $R(e_q)$, for $x > x_0$, is just

$$\begin{aligned} f_{R(e_q)}(x; x_0) &= (1+\sqrt{1+2q}) e^{-(-1+\sqrt{1+2q})x_0} e^{-(1+\sqrt{1+2q})x} \\ &+ \frac{q}{\sqrt{1+2q}} e^{-(-1+\sqrt{1+2q})x_0} e^{-(1+\sqrt{1+2q})x} \left[e^{2\sqrt{1+2q}x_0} - 1 \right]. \end{aligned}$$

We will now focus on computing $f_{R(e_q)}(x; x_0)$, for $x < x_0$. After applying our weak-convergence results, we see that

$$\begin{aligned} P_{x_0}(R(e_q) > x) &= 1 - E_{x_0 - x}[e^{-q\tau_0}] + E_{x_0}[e^{-q\tau_0}] \frac{\int_x^\infty E_y[e^{-q\tau_0}]\pi(dy)}{\int_0^\infty E_y[e^{-q\tau_0}]\pi(dy)} \\ &+ \int_0^x \frac{\int_x^\infty E_y[e^{-q\tau_0}]\pi(dy)}{\int_z^\infty E_y[e^{-q\tau_0}]\pi(dy)} dP_{x_0}(\inf_{0 \le u \le e_q} R(u) \le z). \end{aligned}$$

Evaluating this quantity, then taking derivatives shows that the transient density of $R(e_q)$ is just

$$\begin{aligned} f_{R(e_q)}(x; x_0) &= (-1+\sqrt{1+2q}) e^{-(-1+\sqrt{1+2q})x_0} e^{-(1-\sqrt{1+2q})x} \\ &+ (1+\sqrt{1+2q}) e^{-(-1+\sqrt{1+2q})x_0} e^{-(1+\sqrt{1+2q})x} \\ &+ (1-\sqrt{1+2q})(-1+\sqrt{1+2q}) e^{-(-1+\sqrt{1+2q})x_0} e^{-(1-\sqrt{1+2q})x} \\ &- \frac{q}{\sqrt{1+2q}} e^{-(-1+\sqrt{1+2q})x_0} e^{-(1+\sqrt{1+2q})x} \\ &= (-1+\sqrt{1+2q})(2-\sqrt{1+2q}) e^{-(-1+\sqrt{1+2q})x_0} e^{-(1-\sqrt{1+2q})x} \\ &+ \frac{\sqrt{1+2q}+1+q}{\sqrt{1+2q}} e^{-(-1+\sqrt{1+2q})x_0} e^{-(1+\sqrt{1+2q})x}. \end{aligned}$$

A Palm measures

Throughout this paper, we assume that all of our random elements reside on a probability space $(\Omega, \mathcal{F}, P)$, where Ω represents a complete, separable metric space, $\mathcal{F}$ the Borel σ-field generated by the open sets of the metric, and P a probability measure on $\mathcal{F}$. These additional restrictions will be needed in order to properly define a collection of Palm measures, which are used to derive our main result. The reader should not be alarmed by such restrictions, as the space $D[0, \infty)$ endowed with the proper choice of Skorohod metric is a complete, separable metric space, and many queueing processes (and stochastic processes in general) can reside on such a space. Moreover, $\mathbb{R}_+$ is used to represent the nonnegative real line, and $\mathcal{B}$ the Borel σ-field generated by the open sets of $\mathbb{R}_+$.

Let $N := \{N(t); t \ge 0\}$ represent a point process on the nonnegative real line, with mean measure μ, where $\mu(A) = E[N(A)] < \infty$ for all bounded $A \in \mathcal{B}$. Under such assumptions, it is known that N induces a μ-a.e. unique probability kernel $\mathcal{P} : \mathbb{R}_+ \times \mathcal{F} \to [0, 1]$, where for each fixed $E \in \mathcal{F}$, $\mathcal{P}_s(E)$ is a Borel measurable function in s, and for each fixed $s \in \mathbb{R}_+$, $\mathcal{P}_s$ is a probability measure on $\mathcal{F}$. The probability distributions of this kernel are referred to as the Palm measures of N, and these are defined to be the measures that satisfy the following condition: for each $B \in \mathcal{B}$, and each $A \in \mathcal{F}$,

$$E[N(B)\mathbf{1}_A] = \int_B \mathcal{P}_s(A)\mu(ds). \tag{8}$$

An important consequence of equation (8) is the Campbell-Mecke formula; see for instance Kallenberg [28]. The proof of this formula follows from applying a monotone class argument to (8).

Theorem A.1 *(Campbell-Mecke formula) For any measurable stochastic process $\{X(t); t \geq 0\}$, we find that*

$$E\left[\int_0^\infty X(s)N(ds)\right] = \int_0^\infty \mathcal{E}_s[X(s)]\mu(ds)$$

where $\mathcal{E}_s$ represents expectation, under the probability measure $\mathcal{P}_s$.

Throughout, we say that a stochastic process is measurable if it is measurable with respect to the σ-field $\mathcal{A}$, which is generated by sets of the form $A \times C$, where $A \in \mathcal{B}$, and $C \in \mathcal{F}$, i.e. if for each $B \in \mathcal{B}$, $\{(t,\omega); X(t,\omega) \in B\} \in \mathcal{A}$.

The Campbell-Mecke formula is a very important, fundamental result in the theory of Palm measures, and is typically the main tool used when applying Palm measures to a given problem. Readers wishing to consult a rigorous treatment of such measures are referred to Chapters 10-12 of [28]: other classical references on point process theory include the series of textbooks by Daley and Vere-Jones [15, 16].

A collection of sub-σ-fields $\{\mathcal{F}_s; s \geq 0\}$ of $\mathcal{F}$ is said to be a filtration, if for each $s < t$, $\mathcal{F}_s \subset \mathcal{F}_t$. We say that a stochastic process $\{X(t); t \geq 0\}$ is adapted to the filtration if, for each $t \geq 0$, $X(t)$ is measurable with respect to $\mathcal{F}_t$. Associated with a filtration is a collection of σ-fields $\{\mathcal{F}_{s-}; s > 0\}$, where $\mathcal{F}_{s-}$ is the smallest σ-field containing all σ-fields $\mathcal{F}_r$, for $r < s$. These are standard concepts within stochastic calculus, and can be found in virtually any textbook on the subject. Some examples of textbooks that focus on point processes, and include such concepts, are Brémaud [12] and Baccelli and Brémaud [7].

We are now ready to quote a result that is used to derive the main result of this paper. Suppose $N := \{N(t); t \geq 0\}$ represents a point process on $[0,\infty)$, and suppose $\{\mathcal{F}_t; t \geq 0\}$ represents a filtration, to which N is adapted. Within this framework, we say that N is an $\mathcal{F}_t$-Poisson process, if (i) N is adapted to the filtration, and (ii) the distribution of $N(a,b]$, conditional on $\mathcal{F}_a$, is Poisson with rate

$$\mu(a,b] = \int_{(a,b]} \lambda(s)ds$$

for some deterministic function $\lambda : [0,\infty) \to [0,\infty)$ (i.e. $N(a,b]$ is independent of $\mathcal{F}_a$). Under these conditions, we can apply the following result, which is a corollary of a time-dependent analogue of Papangelou's lemma for point processes; see [19] for details.

Proposition A.1 *If N is an $\mathcal{F}_t$-Poisson process, then $\mathcal{P}_t = P$ on $\mathcal{F}_{t-}$, for almost all t (w.r.t. Lebesgue measure).*

References

[1] J. Abate and W. Whitt (1987). Transient behavior of regulated Brownian motion, II: non-zero initial conditions. *Advances in Applied Probability* **19**, 599-631.

[2] J. Abate and W. Whitt (1987). Transient behavior of the $M/M/1$ queue: starting at the origin. *Queueing Systems* **2**, 41-65.

[3] J. Abate and W. Whitt (1988). Transient behavior of the $M/M/1$ queue via Laplace transforms. *Advances in Applied Probability* **20**, 145-178.

[4] J. Abate, M. Kijima and W. Whitt (1991). Decompositions of the $M/M/1$ transition function. *Queueing Systems* **9**, 323-336.

[5] J. Abate and W. Whitt (1998). Calculating transient characteristics of the Erlang loss model by numerical transform inversion. *Stochastic Models* **14**, 663-680.

[6] S. Asmussen (2003). *Applied Probability and Queues*. Springer-Verlag, New York.

[7] F. Baccelli and P. Brémaud (2003). *Elements of Queueing Theory*. Springer-Verlag, Berlin.

[8] N. T. J. Bailey (1954). A continuous-time treatment of a simple queue, using generating functions. *Proceedings of the Royal Statistical Society, Series B*, **16**, 288-291.

[9] R. Bekker, O.J. Boxma and J.A.C. Resing (2009). Lévy processes with adaptable exponent. *Advances in Applied Probability* **41**, 177-205.

[10] N. H. Bingham (1975). Fluctuation theory in continuous time. *Advances in Applied Probability* **7**, 705-766.

[11] O. Boxma, D. Perry, W. Stadje and S. Zacks (2006). A Markovian Growth-Collapse model. *Advances in Applied Probability* **38**, 221-243.

[12] P. Brémaud (1981). *Point Processes and Queues*. Springer-Verlag, New York.

[13] P. Brémaud (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues* Springer-Verlag, New York.

[14] H. Chen and D. D. Yao (2001). *Fundamentals of Queueing Networks*. Springer-Verlag, New York.

[15] D. J. Daley and D. Vere-Jones (2003). *An Introduction to the Theory of Point Processes, Vol. 1*. Applied Probability Trust.

[16] D. J. Daley and D. Vere-Jones (2008). *An Introduction to the Theory of Point Processes, Vol. 2*. Applied Probability Trust.

[17] D. A. Darling and A. J. F. Siegert (1953). The first passage problem for a continuous Markov process. *The Annals of Mathematical Statistics* **24**, 624-639.

[18] P. Flajolet and F. Guillemin (2000). The formal theory of birth-death processes, lattice-path combinatorics and continued fractions. *Advances in Applied Probability* **32**, 750-778.

[19] B. H. Fralix (2009). A time-dependent view of the ASTA problem, with applications to preemptive queues and birth-death processes. Submitted.

[20] B. H. Fralix (2009). Another look at the transient moments of the $M/G/1$ workload process. Under revision.

[21] B. H. Fralix and G. Riaño (2010). A new look at transient versions of Little's law, with applications to $M/G/1$ preemptive Last-Come-First-Served queues. *Journal of Applied Probability* **47**, 459-473.

[22] B. H. Fralix (2011). On the time-dependent moments of Markovian queues with reneging. Submitted.

[23] S. W. Fuhrmann and R. B. Cooper (1985). Stochastic decompositions in the $M/G/1$ queue with generalized vacations. *Operations Research* **33**, 1117-1129.

[24] O. Garnett, A. Mandelbaum and M. Reiman (2004). Designing a call center with impatient customers. *Manufacturing and Service Operations Management* **4**, 208-227.

[25] P. Greenwood and J. Pitman (1980). Fluctuation identities for Lévy processes and splitting at the maximum. *Advances in Applied Probability* **12**, 893-902.

[26] F. Guillemin, P. Robert and B. Zwart (2004). AIMD algorithms and exponential functionals. *Annals of Applied Probability* **14**, 90-117.

[27] D. V. Gusak and V. S. Korolyuk (1968). On the first passage time across a given level for processes with independent increments. *Theory of Probability and its Applications* **13**, 448-456.

[28] O. Kallenberg (1983). *Random Measures.* Akademie-Verlag, Berlin.

[29] O. Kallenberg (2002). *Foundations of Modern Probability.* A series of the Applied Probability Trust: Springer-Verlag, New York.

[30] A. Kuznetsov (2010). Analytic proof of the Pecherskii-Rogozin identity and the Wiener-Hopf factorization. *Theory of Probability and its Applications*, to appear.

[31] A. E. Kyprianou (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications.* Springer-Verlag, Berlin.

[32] A. H. Löpker and J. S. H van Leeuwaarden (2008). Transient moments of the TCP window size process. *Journal of Applied Probability* **45**, 163-175.

[33] F. Marcellán and G. Pérez (2003). The moments of the $M/M/s$ queue-length process. *Queueing Systems* **44**, 281-304.

[34] R. Núñez-Queija (2001). Note on the $GI/GI/1$ queue with LCFS-PR observed at arbitrary times. *Probability in the Engineering and Informational Sciences* **15**, 179-187.

[35] Z. Palmowski and M. Vlasiou (2009). A Lévy input model with additional state-dependent services. Submitted.

[36] E. A. Percheskii and B. A. Rogozin (1969). On the joint distribution of random variables associated with fluctuations of a process with independent increments. *Theory of Probability and its Applications* **14**, 410-423.

[37] T. L. Saaty (1960). Time-dependent solution of the many-server Poisson queue. *Operations Research* **8**, 755-772.

[38] F. W. Steutel and K. van Harn (2004). *Infinite Divisibility of Probability Distributions on the Real Line.* Marcel Dekker, Inc., New York.

[39] W. Whitt (2002). *Stochastic-Process Limits: an introduction to stochastic-process limits and their application to queues.* Springer, New York.