*The Annals of Applied Probability*
2015, Vol. 25, No. 6, 3209–3250
DOI: 10.1214/14-AAP1072


# STEADY-STATE SIMULATION OF REFLECTED BROWNIAN MOTION AND RELATED STOCHASTIC NETWORKS[1]

BY JOSE BLANCHET AND XINYUN CHEN

*Columbia University and Stony Brook University*

This paper develops the first class of algorithms that enable unbiased estimation of steady-state expectations for multidimensional reflected Brownian motion. In order to explain our ideas, we first consider the case of compound Poisson (possibly Markov modulated) input. In this case, we analyze the complexity of our procedure as the dimension of the network increases and show that, under certain assumptions, the algorithm has polynomial-expected termination time. Our methodology includes procedures that are of interest beyond steady-state simulation and reflected processes. For instance, we use wavelets to construct a piecewise linear function that can be guaranteed to be within $\varepsilon$ distance (deterministic) in the uniform norm to Brownian motion in any compact time interval.

**1. Introduction.** This paper studies simulation methodology that allows estimation, without any bias, of steady-state expectations of multidimensional reflected processes. Our algorithms are presented with companion rates of convergence. Multidimensional reflected processes, as we shall explain, are very important for the analysis of stochastic queueing networks. However, in order to motivate the models that we study, let us quickly review a formulation introduced by Kella (1996).

Consider a network of $d$ queueing stations indexed by $\{1, 2, \ldots, d\}$. Suppose that jobs arrive to the network according to a Poisson process with rate $\lambda$, denoted by $(N(t) : t \geq 0)$. Specifically, the $k$th arrival brings a vector of job requirements $\mathbf{W}(k) = (W_1(k), \ldots, W_d(k))^T$ which are nonnegative random variables (r.v.'s), and they add to the workload at each station right at the moment of arrival. So if the $k$th arrival occurs at time $t$, the workload of the $i$th station (for $i \in \{1, \ldots, d\}$) increases by $W_i(k)$ units right at time $t$. We assume that $\mathbf{W} = (\mathbf{W}(k) : k \geq 1)$ is a sequence of i.i.d. (independent and identically distributed) nonnegative r.v.'s. For fixed $k$, the coordinates of $\mathbf{W}(k)$ are not necessarily independent; however, $\mathbf{W}$ is assumed to be independent of $N(\cdot)$.

Throughout the paper we shall use boldface to write vector quantities, which are encoded as columns. For instance, we write $\mathbf{y} = (y_1, \ldots, y_d)^T$.

Received January 2012; revised September 2014.
[1]Supported in part by the Grants NSF-CMMI-0846816 and NSF-CMMI-1069064.
*MSC2010 subject classifications.* 60J65, 65C05.
*Key words and phrases.* Reflected Brownian motion, steady-state simulation, dominated coupling from the past, wavelet representation.

The total amount of external work that arrives to the $i$th station up to (and including) time $t$ is denoted by

$$J_i(t) = \sum_{k=1}^{N(t)} W_i(k).$$

Now, assume that the workload at the $i$th station is processed as a fluid by the server at a rate $r_i$, continuously in time. This means that if the workload in the $i$th station remains strictly positive during the time interval $[t, t+dt]$, then the output from station $i$ during this time interval equals $r_i\,dt$. In addition, suppose that a proportion $Q_{i,j} \geq 0$ of the fluid processed by the $i$th station is circulated to the $j$th server. We have that $\sum_{j=1}^{d} Q_{i,j} \leq 1$, $Q_{i,i} = 0$, and we define $Q_{i,0} = 1 - \sum_{j=1}^{d} Q_{i,j}$. The proportion $Q_{i,0}$ corresponds to the fluid that goes out of the network from station $i$.

The dynamics stated in the previous paragraph are expressed formally by a differential equation as follows. Let $Y_i(t)$ denote the workload content of the $i$th station at time $t$. Then for given $Y_i(0)$, we have

$$\begin{aligned} dY_i(t) &= dJ_i(t) - r_i I\big(Y_i(t) > 0\big)\,dt + \sum_{j:j\neq i} Q_{j,i} r_j I\big(Y_j(t) > 0\big)\,dt \\ &= dJ_i(t) - r_i\,dt + \sum_{j:j\neq i} Q_{j,i} r_j\,dt \\ &\quad + r_i I\big(Y_i(t) = 0\big)\,dt - \sum_{j:j\neq i} Q_{j,i} r_j I\big(Y_j(t) = 0\big)\,dt \end{aligned} \tag{1}$$

for $i \in \{1, \ldots, d\}$. It is well known that the resulting vector-valued workload process, $\mathbf{Y}(t) = (Y_1(t), \ldots, Y_d(t))^T$, is Markovian. The differential equation (1) admits a unique piecewise linear solution that is right-continuous and has left limits (RCLL). This can be established by elementary methods, and we shall comment on far-reaching extensions shortly.

The equations given in (1) take a neat form in matrix notation. This notation is convenient when examing stability issues and other topics which are related to the steady-state simulation problem we investigate. In particular, let $\mathbf{r} = (r_1, \ldots, r_d)^T$ be the column vector corresponding to the service rates, write $R = (I - Q)^T$ and define

$$\mathbf{X}(t) = \mathbf{J}(t) - R\mathbf{r}t,$$

where $\mathbf{J}(t)$ is a column vector with its $i$th coordinate equal to $J_i(t)$. Then equation (1) can be expressed as

$$\mathbf{Y}(t) = \mathbf{Y}(0) + \mathbf{X}(t) + R\mathbf{L}(t), \tag{2}$$

where $\mathbf{L}(t)$ is a column vector with its $i$th coordinate equal to

$$L_i(t) = \int_0^t r_i I\big(Y_i(s) = 0\big)\,ds.$$

As mentioned earlier, $\mathbf{Y} = (\mathbf{Y}(t) : t \geq 0)$ is a Markov process. Let us assume that $Q^n \to 0$ as $n \to \infty$. This assumption is synonymous with the assumption that the network is open. In detail, for each $i$ such that $\lambda_i > 0$, there exists a path $(i_1, i_2, \ldots, i_k)$ satisfying that $\lambda_i Q_{i,i_1} Q_{i_1,i_2} \cdots Q_{i_{k-1},i_k} > 0$ with $i_k = 0$ and $k \leq d$. In addition, under this assumption the matrix $R^{-1}$ exists and has nonnegative coordinates. To ensure stability, we assume that $R^{-1} E\mathbf{X}(1) < 0$—inequalities involving vectors are understood coordinate-wise throughout the paper. It follows from Theorem 2.4 of Kella and Ramasubramanian (2012) that $\mathbf{Y}(t)$ converges in distribution to $\mathbf{Y}(\infty)$ as $t \to \infty$, where $\mathbf{Y}(\infty)$ is an r.v. with the (unique) stationary distribution of $\mathbf{Y}(\cdot)$.

The first contribution of this paper is that we develop an exact sampling algorithm (i.e., simulation without bias) for $\mathbf{Y}(\infty)$. This algorithm is developed in Section 2 of this paper under the assumption that $\mathbf{W}(k)$ has a finite moment-generating function. In addition, we analyze the order of computational complexity (measured in terms of expected random numbers generated) of our algorithm as $d$ increases, and we show that it is polynomially bounded.

Moreover, we extend our exact sampling algorithm to the case in which there is an independent Markov chain driving the arrival rates, the service rates, and the distribution of job sizes at the time of arrivals. This extension is discussed in Section 3.

The workload process $(\mathbf{Y}(t) : t \geq 0)$ is a particular case of a reflected (or constrained) stochastic network. Although the models introduced in the previous paragraphs are interesting in their own right, our main interest is the steady-state simulation techniques for reflected Brownian motion. These techniques are obtained by abstracting the construction formulated in (2). This abstraction is presented in terms of a Skorokhod problem, which we describe as follows. Let $\mathbf{X} = (\mathbf{X}(t) : t \geq 0)$ with $\mathbf{X}(0) \geq 0$, and $R$ be an $M$-matrix $R$ so that the inverse $R^{-1}$ exists and has nonnegative coordinates. To solve the Skorokhod problem requires finding a pair of processes $(\mathbf{Y}, \mathbf{L})$ satisfying equation (2), subject to:

(i) $\mathbf{Y}(t) \geq 0$ for each $t$,
(ii) $L_i(\cdot)$ nondecreasing for each $i \in \{1, \ldots, d\}$ and $L_i(0) = 0$,
(iii) $\int_0^t Y_i(s)\, dL_i(s) = 0$ for each $t$.

Eventually we shall take the input process $\mathbf{X}(\cdot)$ as a Brownian motion with constant drift $\mathbf{v} = E\mathbf{X}(1)$ and nondegenerate covariance matrix $\Sigma$. There then exists a strong solution (i.e., path-by-path and not only in law) to the stochastic differential equation (SDE) (2) subject to the Skorokhod problem constraints (i) to (iii), and the initial condition $\mathbf{Y}(0)$. This was proved by Harrison and Reiman (1981), who introduced the notion of reflected Brownian motion (RBM). When $R$ is an $M$-matrix, $R^{-1}\boldsymbol{\mu} < 0$ is a necessary and sufficient condition for the stability of an RBM; see Harrison and Williams (1987). Our algorithm for the RBM is motivated by the fact that in great generality (i.e., only requiring the existence of variances of service

times and inter-arrival times), the so-called generalized Jackson networks (which are single-server queues connected with Markovian routing) converge weakly to a reflected Brownian motion in a heavy traffic asymptotic environment as in Reiman (1984). Moreover, recent papers from Gamarnik and Zeevi (2006) and Budhiraja and Lee (2009) have shown that convergence occurs also at the level of steady-state distributions. Therefore, reflected Brownian motion (RBM) plays a central role in queueing theory.

The second contribution of this paper is the development of an algorithm that allows estimation with no bias of $E[g(\mathbf{Y}(\infty))]$ for positive and continuous functions $g(\cdot)$. Moreover, given $\varepsilon > 0$, we provide a simulation algorithm that outputs a random variable $\mathbf{Y}_\varepsilon(\infty)$ that can be guaranteed to be within $\varepsilon$ distance (say in the Euclidian norm) from an unbiased sample $\mathbf{Y}(\infty)$ from the steady-state distribution of RBM. This contribution is developed in Section 4 of this paper. We show that the number of Gaussian random variables generated to produce $\mathbf{Y}_\varepsilon(\infty)$ is of order $O(\varepsilon^{-a_C-2}\log(1/\varepsilon))$ as $\varepsilon \searrow 0$, where $a_C$ is a constant only depending on the covariance matrix of the Brownian motion; see Section 4.4. In the special case when the $d$-dimensional Brownian motion has nonnegative correlations, the number of random variables generated is of order $O(\varepsilon^{-d-2}\log(1/\varepsilon))$.

Our methods allow estimation without bias of $E[g(\mathbf{Y}(t_1), \mathbf{Y}(t_2), \ldots, \mathbf{Y}(t_m))]$ for a positive function $g(\cdot)$ continuous almost everywhere and for any $0 < t_1 < t_2 < \cdots < t_m$. Simulation of RBM has been studied in the literature. In the one-dimensional setting it is not difficult to sample RBM exactly; this follows, for instance, from the methods in Devroye (2009). The paper of Asmussen, Glynn and Pitman (1995) also studies the one-dimensional case and provides an enhanced Euler-type scheme with an improved convergence rate. The work of Burdzy and Chen (2008) provides approximations of reflected Brownian motion with orthogonal reflection (the case in which $R = I$).

With regard to steady-state computations, the work of Dai and Harrison (1992) provides numerical methods for approximating the steady-state expectation by numerically evaluating the density of $\mathbf{Y}(\infty)$. In contrast to our methods, Dai and Harrison's procedure is based on projections in mean-squared norm with respect to a suitable reference measure. Since such an algorithm is nonrandomized, it is therefore, in some sense, preferable to simulation approaches, which are necessarily randomized. However, the theoretical justification of Dai and Harrison's algorithm relies on a conjecture that is believed to be true but has not been rigorously established; see Dai and Dieker (2011). In addition, no rate of convergence is known for this procedure, even assuming that the conjecture is true.

Finally, we briefly discuss some features of our procedure and our strategy at a high level. There are two sources of bias that arise in the setting of steady-state simulation of RBM. First, discretization error in the simulation of the process $\mathbf{Y}$ is inevitable due to the continuous nature of Brownian motion, especially when the reflection matrix $R$ is not the identity. This issue is present even in finite time

horizon. The second issue is, naturally, that we are concerned with steady-state expectations which inherently involve, in principle, an infinite time horizon.

In order to concentrate on removing the bias issues arising from the infinite horizon, we first consider the reflected compound Poisson case where we can simulate the solution of the Skorokhod problem in any finite interval exactly and without any bias. Our strategy is based on the dominated coupling from the past (DCFTP). This technique was proposed by Kendall (2004), following the introduction of coupling from the past by Propp and Wilson (1996). The idea behind DCFTP is to construct suitable upper- and lower-bound processes that can be simulated in stationarity and backward in time. We take the lower bound to be the process identically equal to zero. We use results from Harrison and Williams (1987) (for the RBM) and Kella (1996) (for the reflected compound Poisson process), to construct an upper bound process based on the solution of the Skorokhod problem with reflection matrix $R = I$. It turns out that simulation of the stationary upper-bound process backward involves sampling the infinite horizon maximum (coordinatewise) from $t$ to infinity of a $d$-dimensional compound Poisson Process with negative drift. We use sequential acceptance/rejection techniques (based on a exponential tilting distributions used in rare-event simulation) to simulate from an infinite horizon maximum process.

Then we turn to RBM. A problem that arises, in addition to the discretization error given the continuous nature of Brownian motion, is the fact that in dimensions higher than one (as in our setting) RBM never reaches the origin. Nevertheless, it will be arbitrarily close to the origin, and we shall certainly leverage off this property to obtain simulation that is guaranteed to be $\varepsilon$-close to a genuine steady-state sample. Now in order to deal with the discretization error we use wavelet-based techniques. We take advantage of a well-known wavelet construction of Brownian motion; see Steele (2001).

Instead of simply simulating Brownian motion using the wavelets, which is the standard practice, we simulate the wavelet coefficients jointly with suitably defined random times. Consequently, we are able to guarantee with probability one that our wavelet approximation is $\varepsilon$-close in the uniform metric to Brownian motion in any compact time interval (note that $\varepsilon$ is deterministic and defined by the user; see Section 4.2).

Finally, we use the following fact. Let process $\mathbf{Y}$ be the solution to the Skorokhod problem. Then the process $\mathbf{Y}$, as a function of the input process $\mathbf{X}$, is Lipschitz continuous with a computable Lipschitz constant, under the uniform topology. These observations combined with an additional randomization, in the spirit of Beskos, Peluchetti and Roberts (2012), allow estimation with no bias of the steady-state expectation.

We strongly believe that the use of tolerance-enforced coupling based on wavelet constructions, as we illustrate here, can be extended more broadly in the numerical analysis of the Skorokhod and related problems.

We perform some numerical experiments to validate our algorithms. Our results are reported in Section 5. Further numerical experiments are pursued in a companion paper, in which we also discuss further implementation issues and some adaptations, which are specially important in the case of RBM.

The rest of the paper is organized as follows: in Section 2, we consider the problem of exact simulation from the steady-state distribution of the reflected compound Poisson process discussed earlier; we then show how our procedure is adapted without major complications to Markov-modulated input in Section 3; in Section 4, we continue explaining the main strategy to be used for the reflected Brownian motion case; finally, the numerical experiments are given in Section 5.

**2. Exact simulation of reflected compound Poisson processes.** The model that we consider has been explained at the beginning of the Introduction. We summarize the assumptions that we shall impose next.

*Assumptions*:

(A1) the matrix $R$ is an $M$-matrix;

(A2) $R^{-1}E\mathbf{X}(1) < 0$ (recall that inequalities apply coordinate-wise for vectors);

(A3) there exists $\boldsymbol{\theta} > 0$, $\boldsymbol{\theta} \in \mathbb{R}^d$ such that

$$E\big[\exp(\boldsymbol{\theta}^T \mathbf{W}(k))\big] < \infty.$$

We have commented on (A1) and (A2) in the Introduction. Assumption (A3) is important in order to do exponential tilting when we simulate a stationary version of the upper-bound process.

In addition to (A1) to (A3), we shall assume that one can simulate from exponential tilting distributions associated to the marginal distribution of $\mathbf{W}(k)$. That is, we can simulate from $P_{\theta_i}(\cdot)$ such that

$$P_{\theta_i}\big(W_1(k) \in dy_1, \ldots, W_d(k) \in dy_d\big) = \frac{\exp(\theta_i y_i)}{E\exp(\theta_i W_i(k))} P\big(W_1(k) \in dy_1, \ldots, W_d(k) \in dy_d\big),$$

where $\theta_i \in \mathbb{R}$ and $E\exp(\theta_i W_i(k)) < \infty$. We will determine the value of $\theta_i$ through assumption (A3b), as given below.

Let us briefly explain our program, which is based on DCFTP. First, we will construct a *stationary* dominating process $(\mathbf{Y}^+(s) : -\infty < s \le 0)$ that is *coupled* with our target process, that is, a stationary version of the process $(\mathbf{Y}(s) : -\infty < s \le 0)$ satisfying the Skorokhod problem (2). Under coupling, the dominating process satisfies

$$R^{-1}\mathbf{Y}(s) \le R^{-1}\mathbf{Y}^+(s), \tag{3}$$

for each $s \le 0$. We then simulate the process $\mathbf{Y}^+(\cdot)$ backward up to a time $-\tau \le 0$ such that $\mathbf{Y}^+(-\tau) = 0$. Following the tradition of the CFTP literature, we call a time $-\tau$ such that $\mathbf{Y}^+(-\tau) = 0$ a coalescence time. Since $\mathbf{Y}(s) \ge 0$, inequality (3)

yields $\mathbf{Y}(-\tau) = 0$. The next and final step in our strategy is to evolve the solution $\mathbf{Y}(s)$ of the Skorokhod problem (2) forward from $s = -\tau$ to $s = 0$ with $\mathbf{Y}(-\tau) = 0$, *using the same input that drives the construction of* $(\mathbf{Y}^+(s) : -\tau \le s \le 0)$ so that $\mathbf{Y}$ and $\mathbf{Y}^+$ are coupled. The output is therefore $\mathbf{Y}(0)$, which is stationary. The precise algorithm will be summarized in Section 2.2.

So, a crucial part of the whole plan is the construction of $\mathbf{Y}^+(\cdot)$ together with a coupling that guarantees inequality (3). In addition, the coupling must be such that one can use the driving randomness that defines $\mathbf{Y}^+(\cdot)$ directly as an input to the Skorokhod problem (2) that is then used to evolve $\mathbf{Y}^+(\cdot)$. We shall first start by constructing a time reversed stationary version of a suitable dominating process $\mathbf{Y}^+$.

2.1. *Construction of the dominating process.* In order to construct the dominating process $\mathbf{Y}^+(\cdot)$, we first need the following result attributed to Kella (1996) (Lemma 3.1).

LEMMA 1. *There exists* $\mathbf{z}$ *such that* $E\mathbf{X}(1) < \mathbf{z}$ *and* $R^{-1}\mathbf{z} < 0$. *Moreover, if*

$$\mathbf{Z}(t) = \mathbf{X}(t) - \mathbf{z}t,$$

*and* $\mathbf{Y}^+(\cdot)$ *is the solution to the Skorokhod problem*

$$\begin{aligned} &d\mathbf{Y}^+(t) = d\mathbf{Z}(t) + d\mathbf{L}^+(t), \qquad \mathbf{Y}^+(0) = \mathbf{y}_0, \\ &\mathbf{Y}^+(t) \ge 0, \qquad Y_j^+(t)\, dL_j^+(t) = 0, \qquad L_j^+(0) = 0, \qquad dL_j^+(t) \ge 0, \end{aligned} \tag{4}$$

*then* $0 \le R^{-1}\mathbf{Y}(t) \le R^{-1}\mathbf{Y}^+(t)$ *for all* $t \ge 0$ *where* $\mathbf{Y}(\cdot)$ *solves the Skorokhod problem*

$$\begin{aligned} &d\mathbf{Y}(t) = d\mathbf{X}(t) + R\, d\mathbf{L}(t), \qquad \mathbf{Y}(0) = \mathbf{y}_0, \\ &\mathbf{Y}(t) \ge 0, \qquad Y_j(t)\, dL_j(t) = 0, \qquad L_j(0) = 0, \qquad dL_j(t) \ge 0. \end{aligned}$$

We note that computing $\mathbf{z}$ from the previous lemma is not difficult. One can simply pick $\mathbf{z} = E\mathbf{X}(1) + \delta\mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1)^T$ and with $\delta$ chosen so that $0 < \delta R^{-1}\mathbf{1} < -R^{-1}E\mathbf{X}(1)$. In what follows we shall assume that $\mathbf{z}$ has been selected in this form, and we shall assume without loss of generality that $E[\mathbf{Z}(1)] < 0$.

The Skorokhod problem corresponding to the dominating process can be solved explicitly. It is not difficult to verify [see, e.g., Harrison and Reiman (1981)] that if $\mathbf{Y}^+(0) = 0$, the solution of the Skorokhod problem (4) is given by

$$\mathbf{Y}^+(t) = \mathbf{Z}(t) - \min_{0 \le u \le t} \mathbf{Z}(u) = \max_{0 \le u \le t} \big(\mathbf{Z}(t) - \mathbf{Z}(u)\big), \tag{5}$$

where the running maximum is obtained coordinate-by-coordinate.

In order to construct a stationary version of $\mathbf{Y}^+(\cdot)$ backward in time, we first extend $\mathbf{Z}(\cdot)$ to a two-sided compound Poisson process with $\mathbf{Z}(0) = 0$. We define a

time-reversal of $\mathbf{Z}(\cdot)$ as $\mathbf{Z}^{\leftarrow}(t) = -\mathbf{Z}(-t)$. It is easy to check that $\mathbf{Z}^{\leftarrow}(\cdot)$ has stationary and independent increments that are identically distributed as those of $\mathbf{Z}(\cdot)$.

For any given $T \leq 0$, we define a process $\mathbf{Z}_T^{\leftarrow}$ via $\mathbf{Z}_T^{\leftarrow}(t) = \mathbf{Z}^{\leftarrow}(T + t)$ for $0 \leq t \leq |T|$. And for any given $\mathbf{y} \geq 0$ we define $\mathbf{Y}_T^+(t, \mathbf{y})$ for $0 \leq t \leq |T|$ to be the solution to the Skorokhod problem with input process $\mathbf{Z}_T^{\leftarrow}$, initial condition $\mathbf{Y}_T^+(0, \mathbf{y}) = \mathbf{y}$ and reflection matrix $R = I$. In detail, $\mathbf{Y}_T^+(\cdot, \mathbf{y})$ solves

$$
\begin{aligned}
&d\mathbf{Y}_T^+(t, \mathbf{y}) = d\mathbf{Z}_T^{\leftarrow}(t) + d\mathbf{L}_T^+(t, \mathbf{y}), \qquad \mathbf{Y}_T^+(0, \mathbf{y}) = \mathbf{y}, \\
&\mathbf{Y}_T^+(t, \mathbf{y}) \geq 0, \qquad Y_{T,j}^+(t, \mathbf{y})\, dL_{T,j}^+(t, \mathbf{y}) = 0, \\
&L_{T,j}^+(0, \mathbf{y}) = 0, \qquad dL_{T,j}^+(t, \mathbf{y}) \geq 0.
\end{aligned} \tag{6}
$$

According to (5), if $\mathbf{y} = 0$,

$$
\mathbf{Y}_T^+(t, 0) = \max_{0 \leq u \leq t} \big(\mathbf{Z}_T^{\leftarrow}(t) - \mathbf{Z}_T^{\leftarrow}(u)\big). \tag{7}
$$

Since $E[\mathbf{Z}(1)] < 0$, the process $\mathbf{Y}^+$ satisfying the Skorokhod problem (4) with orthogonal reflection ($R = I$) possesses a unique stationary distribution. So, we can construct a stationary version of $(\mathbf{Y}^+(s) : -\infty < s \leq 0)$ as

$$
\mathbf{Y}_*^+(s) = \lim_{T \to -\infty} \mathbf{Y}_T^+(-T - s, 0). \tag{8}
$$

The following representation of $\mathbf{Y}_*^+(\cdot)$ is known in the queueing literature; still we include a short proof to make the presentation self-contained.

PROPOSITION 1. *Given any* $t \geq 0$,

$$
\mathbf{Y}_*^+(-t) = -\mathbf{Z}(t) + \max_{t \leq u < \infty} \mathbf{Z}(u). \tag{9}
$$

PROOF. Expression (7) together with the definition of $\mathbf{Z}_T^{\leftarrow}(\cdot)$ yields

$$
\begin{aligned}
\mathbf{Y}_T^+(-T + s, 0) &= \max_{0 \leq u \leq -T+s} \big(\mathbf{Z}^{\leftarrow}(s) - \mathbf{Z}^{\leftarrow}(T + u)\big) = \max_{T \leq r \leq s} \big(\mathbf{Z}^{\leftarrow}(s) - \mathbf{Z}^{\leftarrow}(r)\big) \\
&= \max_{T \leq r \leq s} \big(-\mathbf{Z}(-s) + \mathbf{Z}(-r)\big) = -\mathbf{Z}(-s) + \max_{T \leq r \leq s} \mathbf{Z}(-r).
\end{aligned}
$$

Let $-s = t \geq 0$ and $-r = u \geq 0$, and we obtain $\mathbf{Y}_T^+(-T - t, 0) = -\mathbf{Z}(t) + \max_{t \leq u \leq -T} \mathbf{Z}(u)$. Now send $-T \to \infty$ and arrive at (9), thereby obtaining the result. □

2.2. *The structure of the main simulation procedure.* We now are ready to explain our main algorithm to simulate unbiased samples from the steady-state distribution of $\mathbf{Y}$. For this purpose, let us first define

$$
\mathbf{M}(t) = \max_{t \leq u < \infty} \mathbf{Z}(u),
$$

for $t \geq 0$ so that $\mathbf{Y}_*^+(-t) = \mathbf{M}(t) - \mathbf{Z}(t)$. Since $E[\mathbf{Z}(1)] < 0$, it follows that $\mathbf{M}(0) < \infty$, and hence $(\mathbf{M}(t) : t \geq 0)$ is a stochastic process with finite value. We assume that we can simulate $\mathbf{M}(\cdot)$ jointly with $\mathbf{Z}(\cdot)$ until the coalescence time $\tau$, and we shall explain how to perform such simulation procedures in Section 2.3.

ALGORITHM 1 [Exact sampling of $\mathbf{Y}(\infty)$]. *Step* 1: Simulate $(\mathbf{M}(t), \mathbf{Z}(t))$ jointly until time $\tau \geq 0$ such that $\mathbf{Z}(\tau) = \mathbf{M}(\tau)$.

*Step* 2: Set $\mathbf{X}_{-\tau}^{\leftarrow}(t) = \mathbf{Z}(\tau) - \mathbf{Z}(\tau - t) + \mathbf{z} \times t$, and compute $\mathbf{Y}_{-\tau}(t, 0)$ for $0 \leq t \leq \tau$ that solves the Skorokhod problem with input process $\mathbf{X}_{-\tau}^{\leftarrow}(t)$ and initial value $\mathbf{Y}_{-\tau}(0, 0) = 0$. In detail, $\mathbf{Y}_{-\tau}(t, 0)$ solves

$$\begin{aligned} d\mathbf{Y}_{-\tau}(t, 0) &= d\mathbf{X}_{-\tau}^{\leftarrow}(t) + R\, d\mathbf{L}_{-\tau}(t, 0), \\ \mathbf{Y}_{-\tau}(t, 0) &\geq 0, \qquad Y_{-\tau, j}(t, 0)\, dL_{-\tau, j}(t, 0) = 0, \\ L_{-\tau, j}(0, 0) &= 0, \qquad dL_{-\tau, j}(t, 0) \geq 0, \end{aligned}$$

for $\tau$ units of time.

*Step* 3: Output $\mathbf{Y}_{-\tau}(\tau, 0)$ which has the distribution of $\mathbf{Y}(\infty)$.

In step 2, The constant $\mathbf{z}$ is chosen according to Lemma 1 such that $\mathbf{Z}(t) = \mathbf{X}(t) - \mathbf{z}t$. The time is $-\tau$ precisely the coalescence time as in a DCFTP algorithm. The following proposition summarizes the validity of this algorithm.

PROPOSITION 2. *The previous algorithm terminates with probability one, and its output is an unbiased sample from the distribution of* $\mathbf{Y}(\infty)$.

PROOF. The argument is similar to the classic Lyones construction. Let us start by first noting that

$$\mathbf{Y}_+^*(0) = \mathbf{M}(0) = 0 \vee \bigl(-U_1 \boldsymbol{\mu} + \mathbf{W}(1) + \mathbf{M}'\bigr).$$

Here $U_1$ is the arrival time of the first job and follows an exponential distribution. $\mathbf{M}' = \max_{0 \leq t < \infty} \mathbf{Z}(t + U_1) - \mathbf{Z}(U_1) < \infty$ is equal in distribution to $\mathbf{M}(0)$. Then $P(\mathbf{Y}_+^*(0) = 0) = P(U_1 \geq \max_i (W_i(1) + M_i')/\mu_i) > 0$ since $U_1$ has infinite support and is independent of both $\mathbf{W}(1)$ and $\mathbf{M}'$. Therefore, $\mathbf{Y}^+(\infty)$ has an atom at zero. This implies that $\tau < \infty$ with probability one. Actually, we will show later that $E[\exp(\delta\tau)] < \infty$ for some $\delta > 0$ in Theorem 1. Let $T < 0$, and note that, thanks to Lemma 1, for $t \in (0, |T|]$

$$R^{-1}\mathbf{Y}_T(t, 0) \leq R^{-1}\mathbf{Y}_T^+(t, 0). \tag{10}$$

In addition, by monotonicity of the solution to the Skorokhod problem in terms of its initial condition [see Kella and Whitt (1996)], we also have [using the definition of $\mathbf{Y}_T^+(t, \mathbf{y})$ from (6) and $\mathbf{Y}_*^+(T)$ from (8)] that

$$\mathbf{Y}_T^+(t, 0) \leq \mathbf{Y}_T^+\bigl(t, \mathbf{Y}_*^+(T)\bigr) = \mathbf{Y}_*^+(T + t). \tag{11}$$

So $\mathbf{Y}_*^+(T+t)=0$ implies $\mathbf{Y}_T^+(t,0)=0$. One step further, as $R^{-1}$ has nonnegative coordinates, equations (10) and (11) imply that $\mathbf{Y}_T(t,0)=0$. Consequently, if $-T>\tau\geq 0$,

$$\mathbf{Y}_T\big(|T|-\tau,0\big)=0,$$

which in particular yields that $\mathbf{Y}_T(-T,0)=\mathbf{Y}_{-\tau}(\tau,0)$. We then obtain that

$$\lim_{T\to-\infty}\mathbf{Y}_T(-T,0)=\mathbf{Y}_{-\tau}(\tau,0),$$

thereby concluding that $\mathbf{Y}_\tau(-\tau,0)$ follows the distribution $\mathbf{Y}(\infty)$ as claimed. □

Step 2 in Algorithm 1.1 is straightforward to implement because the process $\mathbf{X}_{-\tau}^{\leftarrow}(\cdot)$ is piecewise linear, and the solution to the Skorokhod problem, namely $\mathbf{Y}_{-\tau}(\cdot,0)$, is also piecewise linear. The gradients are simply obtained by solving a sequence of linear system of equations which are dictated by evolving the ordinary differential equations given in (1). Therefore, the most interesting part is the simulation of the stochastic object $(\mathbf{M}(t):0\leq t\leq\tau)$ in step 1, as we will discuss in Section 2.3.

2.3. *Simulation of the stationary dominating process*. As customary, we use the notation $E_0(\cdot)$ or $P_0(\cdot)$ to indicate the conditioning $\mathbf{Z}(0)=0$. We define $\phi_i(\theta)=E_0[\exp(\theta Z_i(1))]$ to be the moment-generating function of $Z_i(1)$, and let $\psi_i(\theta)=\log(\phi_i(\theta))$. In order to simplify the explanation of the simulation procedure to sample $(\mathbf{M}(t):t\geq 0)$, we introduce the following assumption:

*Assumption*: (A3b) Suppose that in every dimension $i$ there exists $\theta_i^*\in(0,\infty)$ such that

$$\psi_i\big(\theta_i^*\big)=\log E_0\exp\big(\theta_i^* Z_i(1)\big)=0.$$

This assumption is a strengthening of assumption (A3), and it is known as Cramer's condition in the large deviations literature. As we shall explain at the end of Section 2.3, it is possible to dispense this assumption and only work under assumption (A3). For the moment, we continue under assumption (A3b).

We wish to simulate $(\mathbf{Z}(t):0\leq t\leq\tau)$ where $\tau$ is a time such that

$$\mathbf{Z}(\tau)=\mathbf{M}(\tau)=\max_{s\geq\tau}\mathbf{Z}(s)\quad\text{and hence}\quad\forall 0\leq t\leq\tau,\qquad\mathbf{M}(t)=\max_{t\leq s\leq\tau}\mathbf{Z}(s).$$

Recall that $-\tau$ is precisely the coalescence time since $\mathbf{Y}_*^+(-\tau)=0$. We also keep in mind that our formulation at the beginning of the Introduction implies that

$$\mathbf{Z}(t)=\mathbf{J}(t)-R\mathbf{r}t-\mathbf{z}t=\sum_{k=1}^{N(t)}\mathbf{W}(k)-R\mathbf{r}t-\mathbf{z}t,$$

where $\mathbf{z}$ is selected according to Lemma 1. Define

$$\boldsymbol{\mu}=R\mathbf{r}+\mathbf{z},$$

and let $\mu_i > 0$ be the $i$th coordinate of $\boldsymbol{\mu}$. In addition, we assume that we can choose a constant $m > 0$ large enough such that

$$\sum_{i=1}^{d} \exp\bigl(-\theta_i^* m\bigr) < 1. \tag{12}$$

Define

$$T_m = \inf\bigl\{t \geq 0 : Z_i(t) \geq m, \text{ for some } i\bigr\}. \tag{13}$$

Now we are ready to propose the following procedure to simulate $\tau$:

ALGORITHM 1.1 (Simulating the coalescence time). The output of this algorithm is $(\mathbf{Z}(t) : 0 \leq t \leq \tau)$, and the coalescence time $\tau$. Choose the constance $m$ according to (12):

(1) Set $\tau = 0$, $\mathbf{Z}(0) = 0$.

(2) Generate an inter-arrival time $U$ distributed $\mathrm{Exp}(\lambda)$, and sample $\mathbf{W} = (W_1, \ldots, W_d)$ independent of $U$.

(3) Let $\mathbf{Z}(\tau + t) = \mathbf{Z}(\tau) - t\boldsymbol{\mu}$ for $0 \leq t < U$ and $\mathbf{Z}(\tau + U) = \mathbf{Z}(\tau) + \mathbf{W} - U\boldsymbol{\mu}$.

(4) If there exists an index $i$, such that $W_i - U\mu_i \geq -m$, then return to step 2 and reset $\tau \longleftarrow \tau + U$. Otherwise, sample a Bernoulli $I$ with parameter $p = P_0(T_m < \infty)$.

(5) If $I = 1$, simulate a new *conditional path* $(\mathbf{C}(t) : 0 \leq t \leq T_m)$ following the conditional distribution of $(\mathbf{Z}(t) : 0 \leq t \leq T_m)$ given that $T_m < \infty$ and $\mathbf{Z}(0) = 0$. Let $\mathbf{Z}(\tau + t) = \mathbf{Z}(\tau) + \mathbf{C}(t)$ for $0 \leq t \leq T_m$, and reset $\tau \longleftarrow \tau + T_m$. Return to step 2.

(6) Else, if $I = 0$, stop and return $\tau$ along with the feed-in path $(\mathbf{Z}(t) : 0 \leq t \leq \tau)$.

We shall now explain how to execute the key steps in the previous algorithm, namely, steps 4 and 5.

2.3.1. *Simulating a path conditional on reaching a positive level in finite time.* The procedure that we shall explain now is an extension of the one-dimensional procedure given in Blanchet and Sigman (2011); see also the related one-dimensional procedure by Ensor and Glynn (2000). The strategy is to use acceptance/rejection. The proposed distribution is based on importance sampling by means of exponential tilting. In order to describe our strategy, we need to introduce some notation.

We think of the probability measure $P_0(\cdot)$ as defined on the canonical space of right-continuous with left-limits $\mathbb{R}^d$-valued functions, namely, the ambient space of $(\mathbf{Z}(t) : t \geq 0)$ which we denote by $\Omega = D_{[0,\infty)}(\mathbb{R}^d)$. We endow the probability space with the Borel $\sigma$-field generated by the Skorokhod $J_1$ topology; see Billingsley (1999). Our goal is to simulate from the conditional law of

$(\mathbf{Z}(t) : 0 \le t \le T_m)$ given that $T_m < \infty$ and $\mathbf{Z}(0) = 0$, which we shall denote by $P_0^*$ in the rest of this part.

Now let us introduce our proposed distribution, $P_0'(\cdot)$, defined on the space $\Omega' = D_{[0,\infty)}(\mathbb{R}^d) \times \{1, 2, \ldots, d\}$. We endow the probability space with the product $\sigma$-field induced by the Borel $\sigma$-field generated by the Skorokhod $J_1$ topology and all the subsets of $\{1, 2, \ldots, d\}$. So, a typical element $\omega'$ sampled under $P_0'(\cdot)$ is of the form $\omega' = ((\mathbf{Z}(t) : t \ge 0), \text{Index})$, where $\text{Index} \in \{1, 2, \ldots, d\}$. The distribution of $\omega'$ induced by $P_0'(\cdot)$ is described as follows. First, set

$$P_0'(\text{Index} = i) = w_i := \frac{\exp(-\theta_i^* m)}{\sum_{j=1}^d \exp(-\theta_j^* m)}. \tag{14}$$

Now, given $\text{Index} = i$, for every set $A \in \sigma(\mathbf{Z}(s) : 0 \le s \le t)$,

$$P_0'(A | \text{Index} = i) = E_0\big[\exp\big(\theta_i^* Z_i(t)\big) I_A\big].$$

So, in particular, the Radon–Nikodym derivative (i.e., the likelihood ratio) between the distribution of $\omega = (\mathbf{Z}(s) : 0 \le s \le t)$ under $P_0'(\cdot)$ and $P_0(\cdot)$ is given by

$$\frac{dP_0'}{dP_0}(\omega) = \sum_{i=1}^d w_i \exp\big(\theta_i^* Z_i(t)\big).$$

*The distribution of* $(\mathbf{Z}(s) : s \ge 0)$ *under* $P_0'(\cdot)$ *is precisely the proposed distribution that we shall use to apply acceptance/rejection.* It is straightforward to simulate under $P_0'(\cdot)$. First, sample Index according to the distribution (14). Then, conditional on $\text{Index} = i$, the process $\mathbf{Z}(\cdot)$ also follows a compound Poisson process. Given $\text{Index} = i$, under $P_0'(\cdot)$, it follows that $\mathbf{J}(t)$ can be represented as

$$\mathbf{J}(t) = \sum_{k=1}^{\hat{N}(t)} \mathbf{W}'(k), \tag{15}$$

where $\hat{N}(\cdot)$ is a Poisson process with rate $\lambda E[\exp(\theta_i^* W_i)]$. In addition, the distribution of $\mathbf{W}'$ is obtained by exponential titling such that for all $A \in \sigma(\mathbf{W})$,

$$P'(\mathbf{W}' \in A) = E\big[\exp\big(\theta_i^* W_i\big) I_A\big]. \tag{16}$$

In sum, conditional on $\text{Index} = i$, we simply let

$$\mathbf{Z}(t) = \sum_{k=1}^{\hat{N}(t)} \mathbf{W}'(k) - \boldsymbol{\mu} t. \tag{17}$$

Now, note that we can write

$$\begin{aligned} E_0'\big(Z_{\text{Index}}(t)\big) &= \sum_{i=1}^d E_0\big(Z_i(t) \exp\big(\theta_i^* Z_i(t)\big)\big) P'(\text{Index} = i) \\ &= \sum_{i=1}^d \frac{d\phi_i(\theta_i^*)}{d\theta} w_i > 0, \end{aligned}$$

where the last inequality follows by convexity of $\psi_k(\cdot)$ and by definition of $\theta_k^*$. So, we have that $Z_{\text{Index}}(t) \nearrow \infty$ as $t \nearrow \infty$ with probability one under $P_0'(\cdot)$ by the law of large numbers. Consequently $T_m < \infty$ a.s. under $P_0'(\cdot)$.

Recall that $P_0^*(\cdot)$ is the conditional law of $(\mathbf{Z}(t) : 0 \le t \le T_m)$ given that $T_m < \infty$ and $\mathbf{Z}(0) = 0$. In order to assure that we can indeed apply acceptance/rejection theory to simulate from $P_0^*(\cdot)$, we need to show that the likelihood ratio $dP_0/dP_0'$ is bounded:

$$\begin{aligned}
&\frac{dP_0^*}{dP_0'}\big(\mathbf{Z}(t) : 0 \le t \le T_m\big) \\
&\quad = \frac{1}{P_0(T_m < \infty)} \times \frac{dP_0}{dP_0'}\big(\mathbf{Z}(t) : 0 \le t \le T_m\big) \qquad (18) \\
&\quad = \frac{1}{P_0(T_m < \infty)} \times \frac{1}{\sum_{i=1}^d w_i \exp(\theta_i^* Z_i(T_m))}.
\end{aligned}$$

Upon $T_m$, there is an index $L$ ($L$ may be different from Index) such that $\exp(\theta_L^* Z_L(T_m)) \ge \exp(\theta_L^* m)$, therefore

$$\frac{1}{\sum_{i=1}^d w_i \exp(\theta_i^* Z_i(T_m))} \le \frac{1}{w_L \exp(\theta_L^* m)} = \sum_{i=1}^d \exp\big(-\theta_i^* m\big) < 1, \qquad (19)$$

where the last inequality follows by (12). Consequently, plugging (19) into (18) we obtain that

$$\frac{dP_0^*}{dP_0'}\big(\mathbf{Z}(t) : 0 \le t \le T_m\big) \le \frac{1}{P_0(T_m < \infty)}. \qquad (20)$$

We now are ready to summarize our acceptance/rejection procedure and the proof of its validity.

ALGORITHM 1.1.1 (Simulation of paths conditional on $T_m < \infty$).

*Step* 1: Sample $(\mathbf{Z}(t) : 0 \le t \le T_m)$ according to $P_0'(\cdot)$ as indicated via equations (14), (15) and (17).

*Step* 2: Given $(\mathbf{Z}(t) : 0 \le t \le T_m)$, simulate a Bernoulli $I$ with probability

$$\frac{1}{\sum_{i=1}^d w_i \exp(\theta_i^* Z_i(T_m))}.$$

[Note that the previous quantity is less than unity due to (19).]

*Step* 3: If $I = 1$, output $(\mathbf{Z}(t) : 0 \le t \le T_m)$ and Stop, otherwise go to step 1.

PROPOSITION 3. *The probability that $I = 1$ at any given call of step* 3 *in Algorithm* 1.1.1 *is $P_0(T_m < \infty)$. Moreover, the output of Algorithm* 1.1.1 *follows the distribution $P_0^*$.*

PROOF. The result follows directly from the theory of acceptance/rejection; see Asmussen and Glynn (2007), pages 39–42. According to it, since the two probability measures $P_0^*$ and $P_0'$ satisfy

$$\frac{dP_0^*}{dP_0'} \le c = \frac{1}{P_0(T_m < \infty)},$$

as indicated by (18) and (20), one can sample exactly from $P_0^*$ by the so-called acceptance/rejection procedure:

(1) Generate i.i.d. samples $\{\omega_i\}$ from $P_0'$ and i.i.d. random numbers $U_i \sim U[0,1]$ independent of $\{\omega_i\}$.

(2) Define $N = \inf\{n \ge 1 : U_n \le c^{-1}\frac{dP_0^*}{dP_0'}(\omega_i)\}$.

(3) Output $\omega_N$.

The output $w_N$ follows exactly the law $P_0^*$, and $N$ is a geometric random variable with mean $c$; in other words, the probability of accepting a proposal is $c$. In our specific case, we have $c = 1/P_0(T_m < \infty)$, and according to (18) the likelihood ration divided by constant $c$ is

$$c^{-1}\frac{dP_0^*}{dP_0'}(\omega) = \frac{1}{\sum_{i=1}^d w_i \exp(\theta_i^* Z_i(T_m))}.$$

Therefore, Algorithm 1.1.1 has acceptance probability $P(I=1) = P_0(T_m < \infty)$, and it generates a path exactly from $P_0^*$ upon acceptance. □

As the previous result shows, the output of the previous procedure follows exactly the distribution of $(\mathbf{Z}(t) : 0 \le t \le T_m)$ given that $T_m < \infty$ and $\mathbf{Z}(0) = 0$. Moreover, the Bernoulli random variable $I$ has probability $P_0(T_m < \infty)$ of success. So this procedure actually allows both steps 4 and 5 in Algorithm 1.1 to be executed simultaneously. In detail, one simulates a path following the law of $P_0'$ until $T_m$, and then, if the proposed path is accepted, it can be concluded that $T_m$ is finite and the proposed path is exactly a sample path following the law of $P_0^*$; otherwise one can conclude that $T = \infty$.

REMARK. As mentioned earlier, assumption (A3b) is a strengthening of assumption (A3). We can carry out our ideas under assumption (A3) as follows. First, instead of $(\mathbf{M}(t) : t \ge 0)$, we consider the following process $\mathbf{Z}_\mathbf{a}(\cdot)$ and $\mathbf{M}_\mathbf{a}(\cdot)$ defined by

$$\mathbf{Z}_\mathbf{a}(t) := \mathbf{Z}(t) + \mathbf{a}t, \qquad \mathbf{M}_\mathbf{a}(t) = \max_{s \ge t}\big(\mathbf{Z}_\mathbf{a}(s)\big).$$

We shall explain how to choose the nonnegative vector $\mathbf{a} = (a_1, a_2, \ldots, a_d)^T$ in a moment. Note that we can simulate $(\mathbf{M}(t) : t \ge 0)$ jointly with $(\mathbf{Z}(t) : t \ge 0)$ if we are able to simulate $(\mathbf{M}_\mathbf{a}(t) : t \ge 0)$ jointly with $(\mathbf{Z}_\mathbf{a}(t) : t \ge 0)$. Now note that $\psi_i(\cdot)$

is strictly convex and that $\dot{\psi}_i(0) < 0$, so there exists $a_i > 0$ large enough to force the existence of $\theta_i^* > 0$ such that $E\exp(\theta_i^* Z_i(1) + a_i\theta_i^*) = 1$, but at the same time small enough to keep $E(Z_i(1) + a_i) < 0$; again, this follows by strict convexity of $\psi_i(\cdot)$ at the origin. So, if assumption (A3b) does not hold, but assumption (A3) holds, one can then execute Algorithm 1.1 based on the process $\mathbf{Z}_{\mathbf{a}}(\cdot)$.

2.4. *Computational complexity.* In this section we provide a complexity analysis of our algorithm. We first make some direct observations assuming the dimension of the network remains fixed. In particular, we note that the expected number of random variables simulated has a finite moment-generating function in a neighborhood of the origin.

THEOREM 1. *Suppose that* (A1) *to* (A3) *are in force. Let* $\tau$ *be the coalescence time, and* $N$ *be the number of random variables generated to terminate the overall procedure to sample* $\mathbf{Y}(\infty)$. *Then there exists* $\delta > 0$ *such that*

$$E\exp(\delta\tau + \delta N) < \infty.$$

PROOF. This follows directly from classical results about random walks; see Gut (2009). In particular it follows that $E_0'(\exp(\delta T_m)) < \infty$. The rest of the proof follows from elementary properties of compound geometric random variables arising from the acceptance/rejection procedure. □

We are more interested, however, in complexity properties as the network increases. We shall impose some regularity conditions that allow us to consider a sequence of systems indexed by the number of dimensions $d$. We shall grow the size of the network in a meaningful way; in particular, we need to make sure that the network remains stable as the dimension $d$ increases. Additional regularity will also be imposed.

*Assumptions*:

There exists two constants $0 < \delta < 1 < H < \infty$ independent of $d$ satisfying the following conditions:

(C1) $R^{-1}E[\mathbf{X}(1)] < -2\delta R^{-1}\mathbf{1}$ in each network.

(C2) Let $\theta_i^*$ for $i = 1, \ldots, d$ be the tilting parameters as defined in assumption (A3b), then

$$E\exp\big[(\delta + \theta_i^*)W_i\big] \le H < \infty$$

and

$$H > \delta + \theta_i^* \qquad \text{for all } 1 \le i \le d.$$

(C3) The arrival rate $\lambda \in (\delta, H)$.

REMARK. Assumption (C1) implies that $\boldsymbol{\mu} = R\mathbf{r} + \mathbf{z} > \delta\mathbf{1}$, where $\mathbf{z}$ is defined according to Lemma 1. In detail, we choose $\mathbf{z} = E[\mathbf{X}(1)] + \delta\mathbf{1}$ and therefore, $R\mathbf{r} + \mathbf{z} = E[\mathbf{J}(1)] + \delta\mathbf{1} > \delta\mathbf{1}$.

Note that $x \le \exp(ax)/(ae)$ for any $a > 0$ and $x \ge 0$. Plugging in $a = \theta_i^* + \delta$, we have $E[W_i] \le E[\exp((\theta_i^* + \delta)W_i)]/(e(\delta + \theta_i^*)) < H/(e\delta)$ and therefore

$$\boldsymbol{\mu} = \lambda E[\mathbf{W}] + \delta\mathbf{1} < \bigl(H^2/(e\delta) + \delta\bigr)\mathbf{1} = H'\mathbf{1},$$

where $H' = H^2/(e\delta) + \delta$. Similarly, we also have that $E[W_i^2] \le E[4\exp((\theta_i^* + \delta)W_i)]/(e^2(\theta_i^* + \delta)^2) \le 4H/(e^2\delta^2)$, and then we can compute

$$\begin{aligned} E\bigl[Z_i(1)^2\bigr] &= E\left[\left(\sum_{k=1}^{N(1)} W_i(k) - \mu_i\right)^2\right] \le 2E\left[\mu_i^2 + \left(\sum_{k=1}^{N(1)} W_i(k)\right)^2\right] \\ &\le 2\mu_i^2 + 2\bigl(\lambda + \lambda^2\bigr)\frac{4H}{e^2\delta^2} \le 2H'^2 + \frac{8(H^2 + H^3)}{e^2\delta^2} := H''. \end{aligned}$$

In sum, we can conclude that

$$\max_{1\le i\le d} E_0\bigl[Z_i(1)^2\bigr] \le H''.$$

In the complexity analysis, we shall only use the fact that $H$, $H'$ and $H''$ are constants independent of $d$. As a result, for the simplicity of notation, we shall write $H$ for $H$, $H'$ and $H''$ in the rest of this section and assume, without loss of generality, that

$$\boldsymbol{\mu} \le H\mathbf{1} \quad \text{and} \quad \max_{1\le i\le d} E_0\bigl[Z_i(1)^2\bigr] \le H.$$

As discussed in Section 2.3.1, in Algorithm 1.1, we actually do steps 4 and 5 simultaneously. Therefore, we can rewrite Algorithm 1.1 as follows:

ALGORITHM 1.1′ (Simulate the coalescence time).

(1) Set $\tau = 0$, $\mathbf{Z}(0) = 0$, $N = 0$.

(2) Simulate a sample from $\mathbf{W} - U\boldsymbol{\mu}$. Here $U$ is exponentially distributed with mean $1/\lambda$ and independent of $\mathbf{W}$. Record the value of $\mathbf{Z}(t)$ for $\tau \le t \le \tau + U$. Reset $N \leftarrow N + 1$, $\mathbf{Z}(\tau + U) \leftarrow \mathbf{Z}(\tau) + \mathbf{W} - U\boldsymbol{\mu}$, $\tau \leftarrow \tau + U$.

(3) If there exists some index $i$, such that $W_i - Ur_i \ge -m$, return to step 2.

(4) Otherwise, simulate a random walk $\{\mathbf{C}(n)\}$ such that $\mathbf{C}(0) = 0$ and $\mathbf{C}(n) = \mathbf{C}(n-1) + \mathbf{W}'(n) - U'(n)\boldsymbol{\mu}$, where $\mathbf{W}'(n) - U'(n)\boldsymbol{\mu}$ are independent and identically distributed as $\mathbf{W}' - U'\boldsymbol{\mu}$ under the tilted measure $P'$ defined in Section 2.3.1 through (15) to (17). Perform the simulation until $N_m = \inf\{n \ge 0 : C_i(n) > m \text{ for some } i\}$.

(5) Reset $N \leftarrow N + N_m$. Compute $p = 1/\sum_{k=1}^d w_k \exp(\theta_k^* C_k(N_m))$, and sample a Bernoulli $I$ with probability $p$. If $I = 1$, $\mathbf{Z}(\tau + \sum_{k=1}^{N_m} U'(k)) = \mathbf{Z}(\tau) + \mathbf{C}(N_m)$ and $\tau = \tau + \sum_{k=1}^{N_m} U'(k)$. Return to step 2.

(6) If $I = 0$, stop and output $\tau$ with $(\mathbf{Z}(t) : 0 \le t \le \tau)$.

In this algorithm, the total number of random variables required to generate is $d \cdot N$. Use $N(d)$ instead of $N$ to emphasize the dependence on the number of dimensions $d$. The following result shows that our algorithm has polynomial complexity with respect to $d$:

THEOREM 2. *Under assumptions* (C1) *to* (C3),

$$E\big[N(d)\big] = O\big(d^{\gamma}\big) \qquad \text{as } d \to \infty,$$

*for some $\gamma$ depending on $\delta$ and $H$.*

Denote the number of Bernoulli's generated in step 5 by $N_b$ and the number of random variables generated before executing step 4 in a single iteration by $N_a$. By Wald's identity, we can conclude

$$E\big[N(d)\big] = E[N_b]\big(E[N_a] + E[N_m]\big).$$

The following proposition gives an estimate for $E[N_m]$.

PROPOSITION 4. *Under assumptions* (C1) *to* (C3),

$$E[N_m] = O(\log d),$$

*and the coefficient in the bound depends only on $\delta$ and $H$.*

PROOF. First, let us consider the cases in which $W_i$ are uniformly bounded from above by some constant $B$.

Recall that $\phi_i(\theta) = E_0[\exp(\theta Z_i(1))]$. Given Index $= i$, one can check that $E_0'[C_i(1)] = \dot{\phi}_i(\theta_i^*)/(\lambda E[\exp(\theta_i^* W_i)]) \ge \dot{\phi}_i(\theta_i^*)/(\lambda H)$. $N_m$ is a stopping time and $C_i(N_m) < m + B$. By the optional sampling theorem, we have

$$E[N_m] = \sum_{i=1}^{d} \omega_i \frac{E_0'[C_i(N_m)]}{E_0'[C_i(1)]} \le \sum_{i=1}^{d} \omega_i \frac{\lambda H(m+B)}{\dot{\phi}_i(\theta_i^*)}.$$

For each $1 \le i \le d$, we are going to estimate a lower bound for $\dot{\phi}(\theta_i^*)$. Using Taylor's expansion around 0, we have

$$\phi_i\big(\theta_i^*\big) = \phi_i(0) + \theta_i^* \dot{\phi}_i(0) + \frac{(\theta_i^*)^2}{2} \ddot{\phi}_i\big(u_1 \theta_i^*\big),$$

for some $u_1 \in [0, 1]$. As $\phi_i(\theta_i^*) = \phi_i(0) = 1$, we have

$$\theta_i^* \dot{\phi}_i(0) + \frac{(\theta_i^*)^2}{2} \ddot{\phi}_i\big(u_1 \theta_i^*\big) = 0.$$

As $\theta_i^* > 0$,

$$\dot{\phi}_i(0) + \frac{\theta_i^*}{2}\ddot{\phi}_i(u_1\theta_i^*) = 0. \tag{21}$$

Under assumption (C1), $\dot{\phi}_i(0) = E_0[Z_i(1)] < -\delta$. Under assumption (C2), we have that

$$\begin{aligned} E_0\big[\exp\big((\delta+\theta_i^*)Z_i(1)\big)\big] &\le \exp\big(\lambda\log\big(E\big[\exp\big((\delta+\theta_i^*)W_i\big)\big]\big)\big) \\ &\le H^{\lambda} \le H^H \triangleq H_1 < \infty. \end{aligned}$$

As a result,

$$\begin{aligned} \ddot{\phi}_i(u_1\theta_i^*) &= E\big[Z_i(1)^2\exp\big(u_1\theta_i^*Z_i(1)\big)\big] \\ &\le E\big[Z_i(1)^2 I\big(Z_i(1)\le 0\big)\big] + E\big[Z_i(1)^2\exp\big(\theta_i^*Z_i(1)\big)I\big(Z_i(0)>0\big)\big] \\ &\le E\big[Z_i(1)^2\big] + E\big[Z_i(1)^2\exp\big(\theta_i^*Z_i(1)\big)I\big(Z_i(0)>0\big)\big] \\ &\le E\big[Z_i(1)^2\big] + E\big[Z_i(1)^2\exp\big(-\delta Z_i(1)\big)\cdot\exp\big((\delta+\theta_i^*)Z_i(1)\big)\big]. \end{aligned}$$

Besides, one can check that for any $x > 0$, $x^2\exp(-\delta x) \le 4e^{-2}/\delta^2$. Therefore,

$$\begin{aligned} \ddot{\phi}_i(u\theta_i^*) &\le E\big[Z_i(1)^2\big] + \frac{4}{\delta^2}e^{-2}E\big[\exp\big((\delta+\theta_i^*)Z_i(1)\big)\big] \\ &\le H + \frac{4}{\delta^2}e^{-2}H_1. \end{aligned}$$

Plug this result into equation (21) and use that $\dot{\phi}_i(0) < -\delta$ to complete the inequality

$$\theta_i^* \ge \frac{2\delta}{H + 4e^{-2}H_1/\delta^2}. \tag{22}$$

On the other hand, by a Taylor expansion of $\phi_i(\cdot)$ around $\theta_i^*$, we can conclude that

$$\dot{\phi}_i(\theta_i^*) = \frac{\theta_i^*}{2}\ddot{\phi}(u_2\theta_i^*), \tag{23}$$

for some $u_2 \in [0, 1]$. Note that

$$\begin{aligned} \ddot{\phi}_i(u_2\theta_i^*) &= E_0\big[Z_i(1)^2\exp\big(u_2\theta_i^*Z_i(1)\big)\big] \ge E_0\big[Z_i(1)^2\exp\big(u_2\theta_i^*Z_i(1)\big)I(U>1)\big] \\ &\ge E\big[\mu_i^2\exp\big(-\theta_i^*\mu_i\big)I(U>1)\big] \ge \mu_i^2\exp(-H\mu_i)\exp(-\lambda) \\ &\ge \delta^2\exp(-H^2-H). \end{aligned}$$

Thus (22) together with (23) imply

$$\dot{\phi}_i(\theta_i^*) \ge \frac{1}{2}\theta_i^*\delta^2 e^{-H^2-H} \ge \frac{\delta^3 e^{-H^2-H}}{H + 4e^{-2}H_1/\delta^2}. \tag{24}$$

Note that for lower bound (24) to hold, we do not require $W_i$ to be bounded.

Therefore,

$$E[N_m] \le \sum_{i=1}^{d} \omega_i \frac{\lambda H(m+B)}{\dot{\phi}_i(\theta_i^*)} \le \frac{\lambda H(m+B)(H+4e^{-2}H_1/\delta^2)}{\delta^3 e^{-H^2-H}},$$

as $\omega_i > 0$ and $\sum_i \omega_i = 1$.

By (22), we have that $\theta_i^*$ are all uniformly bounded away from 0, so we can choose $m = O(\log d/\min_i \theta_i^*) = O(\log d)$ to satisfy equation (12). Now we can conclude that $E[N_m] = O(\log d)$ as $B$, $H$ and $\delta$ are all constants independent of $d$.

Now, let us consider the more general cases when the $W_i$'s are not bounded from above. Recall that $\mathbf{W}'$ is derived from $\mathbf{W}$ by exponential tilting; see (16). For any $B > 0$, define $\tilde{\mathbf{W}}'$ by $\tilde{W}_i' = W_i' I(W_i' \le B)$ as the truncation of $\mathbf{W}'$, and define the random walk $\tilde{C}_i(n) = \tilde{C}_i(n-1) + \tilde{W}_i'(n) - U'(n)\mu_i$. Let $\tilde{N}_m = \inf\{n : \tilde{C}_i(n) > m \text{ for some } i\}$. Since $\tilde{C}_i(n) \le C_i(n)$, we have $\tilde{N}_m \le N_m$. Our goal is to show that one can choose a proper value for $B$ such that $E[\tilde{N}_m] = O(\log d)$ and hence so is $E[N_m]$.

Since $\tilde{W}_i'$ is bounded from above by $B$, by the optimal stopping theorem, we have

$$E[\tilde{N}_m] \le \sum_{i=1}^{d} \omega_i \frac{m+B}{E[\tilde{C}_i(1)]}.$$

By definition,

$$E\big[\tilde{C}_i(1)\big] = E\big[\big(W_i I(W_i \le B) - U\mu_i\big)\exp\big(\theta_i^*\big(W_i I(W_i \le B) - U\mu_i\big)\big)\big].$$

Since $U\mu_i \ge 0$, we have

$$\begin{aligned} &E\big[\big(W_i I(W_i \le B) - U\mu_i\big)\exp\big(\theta_i^*\big(W_i I(W_i \le B) - U\mu_i\big)\big)\big] \\ &\quad \ge E\big[(W_i - U\mu_i)\exp\big(\theta_i^*(W_i - U\mu_i)\big)\big] - E\big[W_i \exp\big(\theta_i^* W_i\big) I(W_i > B)\big]. \end{aligned}$$

By assumption (C2), $\delta$ and $H > 0$ are constants independent of $d$ such that

$$E\big[\exp\big((\delta + \theta_i^*)W_i\big)\big] \le H < \infty.$$

As a consequence,

$$\begin{aligned} E\big[W_i \exp\big(\theta_i^* W_i\big) I(W_i > B)\big] &\le E\big[W_i \exp(-\delta W_i) I(W_i > B)\exp\big((\delta + \theta_i^*)W_i\big)\big] \\ &\le \max_{w > B}\big\{w\exp(-\delta w)\big\} E\big[\exp\big((\delta + \theta_i^*)W_i\big)\big] \\ &\le B\exp(-\delta B)H \end{aligned}$$

for all $B > 1/\delta$. Recall that by (24),

$$\begin{aligned} E\big[(W_i - U\mu_i)\exp\big(\theta_i^*(W_i - U\mu_i)\big)\big] = E\big[C_i(1)\big] &\ge \dot{\phi}_i\big(\theta_i^*\big)/(\lambda H) \\ &\ge \frac{\delta^3 e^{-H^2-H}}{\lambda H(H + 4e^{-2}H_1/\delta^2)}, \end{aligned}$$

where $H_1 = H^H$. Therefore, we can take $B = O(-\frac{1}{\delta}\log(\frac{\delta^3 e^{-H^2-H}}{2\lambda H^2(H+4\delta e^{-2}H_1/\delta^2)}))$ independent of $d$ such that

$$B\exp(-\delta B)H < \frac{\delta^3 e^{-H^2-H}}{2\lambda H(H+4e^{-2}H_1/\delta^2)} \quad \text{and hence}$$

$$E\big[\tilde{C}_i(1)\big] \geq \frac{\delta^3 e^{-H^2-H}}{2\lambda H(H+4e^{-2}H_1/\delta^2)}.$$

In the end, since $m = O(\log(d))$, we have

$$E[N_m] \leq E[\tilde{N}_m] \leq \frac{2\lambda H(m+B)(2H+8e^{-2}H_1/\delta^2)}{\delta^3 e^{-H^2-H}} = O(\log d). \qquad \square$$

Now we give the proof of the main result in this subsection.

PROOF OF THEOREM 2. Recall that

$$E[N] = E[N_b]\big(E[N_a] + E[N_m]\big).$$

Since $N_b$ is the number of trials required to obtain $I = 0$, $E[N_b] = 1/P(I=0)$. As discussed in Section 2.3.1, $P(I=0) \geq 1 - \sum_{i=1}^d \exp(-\theta_i^* m)$ and hence

$$E[N_b] \leq \frac{1}{1-\sum_{i=1}^d \exp(-\theta_i^* m)} \leq \frac{1}{1-1/d}$$

if we take $m = 2\log d/\min_i \theta_i^*$.

Similarly, we have $E[N_a] = 1/P(U > (m+W_i)/\mu_i, \forall i)$. For any $K > 0$,

$$P\bigg(U > \frac{m+W_i}{\mu_i}, \forall i\bigg) \geq P\bigg(U > \frac{m+K}{\min_i \mu_i}; W_i \leq K \text{ for all } i\bigg).$$

Under assumption (C2), we have

$$P(W_i \leq K \text{ for all } i) \geq 1 - \sum_{i=1}^d P(W_i > K) \geq 1 - dH\exp(-K\delta).$$

Under assumption (C3), we have

$$P\bigg(U > \frac{m+K}{\min_i \mu_i}\bigg) \geq \exp\bigg(-\frac{H(m+K)}{\min_i \mu_i}\bigg).$$

As $U$ and $\mathbf{W}$ are independent,

$$P\bigg(U > \frac{m+W_i}{\mu_i}, \forall i\bigg) \geq \exp\bigg(-\frac{H(m+K)}{\min_i \mu_i}\bigg)\big(1 - dH\exp(-K\delta)\big).$$

Choosing $K = (2\log d + \log H)/\delta$ and plugging in $m = 2\log d/\min_i \theta_i^*$, we get

$$E[N_a] \leq \frac{1}{1-1/d} d^{(2H/(\min_i \mu_i \min_i \theta_i^*)+2H/(\delta \min_i \mu_i))} H^{H/(\delta \min_i \mu_i)}.$$

By Proposition 4 we have $E[N_m] = O(\log d)$. In summary, we have

$$E[N] = E[N_b]\big(E[N_a] + E[N_m]\big) = O\left(\left(\frac{1}{1-1/d}\right)^2 \log d\, d^{2H/(\min_i \mu_i \min_i \theta_i^*)}\right)$$
$$= O\big(d^{1+2H/(\min_i \mu_i \min_i \theta_i^*)}\big).$$

As discussed in the proof of Proposition 4, $\theta_i^* \geq \delta/(H + 4e^{-2}H_1/\delta^2)$ and $\mu_i \geq \delta$ are uniformly bounded away from 0, therefore,

$$E[N] = O\big(d^{1+2H(H+4e^{-2}H_1/\delta)/\delta^2}\big). \qquad \square$$

**3. Extension to Markov-modulated processes.** We shall briefly explain how our development in Section 2, specifically Algorithm 1, can be implemented beyond input with stationary and independent increments. As an example, we shall concentrate on Markov-modulated stochastic fluid networks. Our extension to Markov-modulated networks is first explained in the one-dimensional case, and later we will indicate how to treat the multidimensional setting.

Let $(\hat{I}(t) : t \geq 0)$ be an irreducible continuous-time Markov chain taking values on the set $\{1, \ldots, n\}$. We assume that, conditional on $\hat{I}(\cdot)$, the number of arrivals, $\hat{N}(\cdot)$, follows a time-inhomogeneous Poisson process with rate $\lambda_{\hat{I}(\cdot)}$. We further assume that $\int_0^t \lambda_{\hat{I}(s)}\, ds > 0$ with positive probability. The process $\hat{N}(\cdot)$ is said to be a Markov-modulated Poisson process with intensity $\lambda_{\hat{I}(\cdot)}$. Define $\hat{A}_k$ to be the time of the $k$th arrival, for $k \geq 1$; that is, $\hat{A}_k = \inf\{t \geq 0 : \hat{N}(t) = k\}$.

We assume that the $k$th arrival brings a job requirement equal to $\hat{W}(k)$. We also assume that the $\hat{W}(k)$'s are conditionally independent given the process $\hat{I}(\cdot)$. Moreover, we assume that the moment-generating function $\phi_i(\cdot)$ defined via

$$\phi_i(\theta) = E\big(\exp(\theta \hat{X}(k)) | \hat{I}(\hat{A}_k) = i\big),$$

is finite in a neighborhood of the origin. In simple words, the job requirement of the $k$th arrival might depend upon the environment, $\hat{I}(\cdot)$, at the time of arrival. But, conditional on the environment, the job sizes are independent. Finally, we assume that the service rate at time $t$ is equal to $\mu_{\hat{I}(t)} \geq 0$.

Let $\hat{X}(t) = \sum_{k=1}^{\hat{N}(t)} \hat{W}(k) - \int_0^t \mu_{\hat{I}(s)}\, ds$. Then the workload process, $(Y(t) : t \geq 0)$, can be expressed as

$$Y(t) = \hat{X}(t) - \inf_{0 \leq s \leq t} \hat{X}(s),$$

assuming that $Y(0) = 0$. In order for the process $Y(\cdot)$ to be stable, in the sense of having a stationary distribution, we assume that $\sum_i \pi_i(\lambda_i E[\hat{W} | \hat{I} = i] - \mu_i) < 0$, where $\pi_i$ is the stationary distribution of the Markov chain $\hat{I}$. Following the same argument as in Section 2, we can construct a stationary version of the process $Y(\cdot)$ by a time reversal argument.

Since $\hat{I}(\cdot)$ is irreducible, one can define its associated *stationary* time-reversed Markov chain $I(\cdot)$ with transition rate matrix $\mathcal{A}$; for the existence and detailed description of such reversed chain, see Chapter 2.5 of Asmussen (2003). Let us write $N(\cdot)$ to denote a Markov-modulated Poisson process with intensity $\lambda_{I(\cdot)}$, and let $A_k = \inf\{t \geq 0 : N(t) = k\}$. We consider a sequence $(W(k) : k \geq 1)$ of conditionally independent random variables representing the service requirements (backward in time) such that $\phi_i(\theta) = E(\exp(\theta W(k)) | I(A_k) = i)$.

We then can define $Z(t) = \sum_{k=1}^{N(t)} W(k) - \int_0^t \mu_{I(s)}\, ds$. Following the same arguments as in Section 2, we can run a stationary version $Y^*$ of $Y$ backward via the process

$$Y^*(-t) = \sup_{s \geq t} \big(Z(s) - Z(t)\big).$$

Therefore, $Y^*(-t)$ can be simulated exactly as long as a convenient change of measure can be constructed for the process $(I(\cdot), Z(\cdot))$, so that a suitable adaptation of Algorithm 1.1.1 can be applied. Once the adaptation of Algorithm 1.1.1 is in place, the adaptation of Algorithms 1.1 and 1 is straightforward.

In order to define such change of measure, let us define the matrix $\mathcal{M}(\theta, t) \in \mathbb{R}^{n \times n}$, for $t \geq 0$, via

$$\mathcal{M}_{ij}(\theta, t) = E_i\big[\exp\big(\theta Z(t)\big); I(t) = j\big],$$

where the notation $E_i(\cdot)$ means that $I(0) = i$. Note that $\mathcal{M}(\cdot, t)$ is well defined in a neighborhood of the origin. In what follows we assume that $\theta$ is such that all coordinates of $\mathcal{M}(\theta, t)$ are finite.

It is known [see, e.g., Chapters 11.2 and 13.8 of Asmussen (2003) and the references therein] that $\mathcal{M}(\theta, t) = \exp(tG(\theta))$ where the matrix $G$ is defined by

$$G_{ij}(\theta) = \begin{cases} \mathcal{A}_{ij}, & \text{if } i \neq j, \\ \mathcal{A}_{ii} - \mu_i \theta + \lambda_i \phi_i(\theta), & \text{if } i = j. \end{cases}$$

Besides, $G(\theta)$ has a unique eigenvalue $\beta(\theta)$ corresponding to a strictly positive eigenvector $(u(i, \theta) : 1 \leq i \leq n)$. The eigenvalue $\beta(\theta)$ has the following properties which follow from Propositions 2.4 and 2.10 in Chapter 11.2 of Asmussen (2003):

LEMMA 2.

(1) *$\beta(\theta)$ is convex in $\theta$ and $\dot{\beta}(\theta)$ is well defined.*
(2) $\lim_{t \to \infty} Z(t)/t = \dot{\beta}(0) = \lim_{t \to \infty} \hat{X}(t)/t < 0$.
(3) *$(M(t, \theta) : t \geq 0)$ defined via*

$$M(t, \theta) = \frac{u(I(t), \theta)}{u(I(0), \theta)} \exp\big(\theta Z(t) - t\beta(\theta)\big)$$

*is a martingale.*

As explained in Chapter 13.8 of Asmussen (2003), the martingale $M(\cdot)$ induces a change of measure for the process $(I(\cdot), Z(\cdot))$ as we shall explain. Let $P$ be the probability law of $(I(\cdot), Z(\cdot))$, and define a new probability measure $\tilde{P}$ for $(I(s), Z(s) : s \le t)$ as $d\tilde{P} = M(t, \theta)\, dP$.

We now describe the law of $(I(\cdot), Z(\cdot))$ under $\tilde{P}$. The process $I(\cdot)$ is a continuous time Markov chain with rate matrix $\widetilde{\mathcal{A}}_{ij} = \mathcal{A}_{ij} u(j, \theta)/u(i, \theta)$ for $i \neq j$ (and $\widetilde{\mathcal{A}}_{ii} = -\sum_{j \neq i} \widetilde{\mathcal{A}}_{ij}$). In addition,

$$Z(t) \stackrel{d}{=} \sum_{k=1}^{\tilde{N}(t)} \tilde{W}(k) - \int_0^t \mu_{I(s)}\, ds,$$

where $\tilde{N}$ is a Markov-modulated Poisson process with rate at time $t$ equal to $\phi_{I(t)}(\theta)\lambda(I(t))$, and the $\tilde{W}(k)$'s are conditionally independent given $I(\cdot)$ with moment generating function $\widetilde{\phi}_i(\cdot)$ defined via

$$\widetilde{\phi}_i(\eta; \theta) = \widetilde{E}\big(\exp\big(\eta \tilde{W}(k)\big) | A_k = i\big) = \phi_i(\eta + \theta)/\phi_i(\eta),$$

which is finite in a neighborhood of the origin. In addition, $Z(t)/t \to \dot{\beta}(\theta)$ under $\tilde{P}$.

Because of the stability condition of the system, we have that $\dot{\beta}(0) < 0$. Then, following the same argument as in the remark given at the end of Section 2.3, we may assume the existence of the Cramer root $\theta^* > 0$ such that $\beta(\theta^*) = 0$ and $\dot{\beta}(\theta^*) > 0$. The change of measure that allows adaption of Algorithm 1.1.1 is given by selecting $\theta^* > 0$ as indicated. Now, select $m > 0$ such that

$$K := \exp\big(-\theta^* m\big) \max_{i,j} \frac{u(i, \theta^*)}{u(j, \theta^*)} \le 1. \tag{25}$$

We will use the notation $P_{0,i}(\cdot)$ to denote the law $P(\cdot)$ conditional on $Z(0) = 0$ and $I(0) = i$. Let us write $P^*_{0,i}(\cdot)$ to denote the law of $(Z(t) : 0 \le t \le T_m)$ [under $P_{0,i}(\cdot)$] conditional on $T_m < \infty$. Further, we write $\tilde{P}_{0,i}(\cdot)$ to denote the law of $\tilde{P}(\cdot)$, selecting $\theta = \theta^*$, conditional on $Z(0) = 0$ and $I(0) = i$. Then we have that $\tilde{P}_{0,i}(T_m < \infty) = 1$ [by Lemma 2 since $\dot{\beta}(\theta^*) > 0$], and therefore [by (25)], we have

$$\begin{aligned}
&\frac{dP^*_{0,i}}{d\tilde{P}_{0,i}}\big(\big(I(t), Z(t)\big) : 0 \le t \le T_m\big) \\
&\quad = \frac{u(i, \theta^*)}{u(I(T_m), \theta^*)} \times \frac{\exp(-\theta^* Z(T_m)) I(T_m < \infty)}{P_{0,i}(T_m < \infty)} \\
&\quad \le \frac{K}{P_{0,i}(T_m < \infty)} \le \frac{1}{P_{0,i}(T_m < \infty)}.
\end{aligned}$$

It is clear from this identity, which is completely analogous to identities (18) and (20), which are the basis for Algorithm 1.1.1, that the corresponding adaptation to our current setting follows.

For the $d$-dimensional case ($d > 1$), we first assume the existence of the Cramer root $\theta_j^* > 0$ for each dimension $j \in \{1, \ldots, d\}$. In this setting we also must compute the corresponding positive eigenvector $(u_j(i, \theta_j^*) : 1 \le i \le n)$ for each $j \in \{1, \ldots, d\}$. The desired change of measure that allows the adaptation of Algorithm 1.1.1 is just a mixture of changes of measures such as those described above induced by $M(\cdot, \theta_j^*)$ in each direction, just as discussed in Section 2.3.1, with weight $w_j = \exp(-\theta_j^* m)/\sum_{k=1}^m \exp(-\theta_k^* m)$. The corresponding likelihood ratio is then

$$\frac{dP_{0,i}^*}{d\tilde{P}_{0,i}}\big((I(t), Z(t)) : 0 \le t \le T_m\big) = \frac{1}{\sum_{j=1}^d w_j \exp(\theta_j^* Z_j(T_m)) u_j(I(T_m), \theta_j^*)/u_j(i, \theta_j^*)},$$

and $m$ must be selected so that

$$\sum_{j=1}^d \exp(-\theta_j^* m) \sup_{j,i,k} \frac{u_j(i, \theta_j^*)}{u_j(k, \theta_j^*)} \le 1.$$

**4. Algorithm for reflected Brownian motion.** In this section, we revise our algorithm and explain how we can apply it to the case of reflected Brownian motion. Consider a multidimensional Brownian motion

$$\mathbf{X}(t) = \mathbf{v}t + A\mathbf{B}(t),$$

where $\mathbf{v} \in \mathbb{R}^d$ is the drift vector, and $A \cdot A^T \triangleq \Sigma \in \mathbb{R}^{d \times d}$ is the positive definite covariance matrix. Our target process $\mathbf{Y}(t)$ is the solution to the following Skorokhod problem with input process $\mathbf{X}(\cdot)$ and initial value $\mathbf{Y}(0) = \mathbf{y}_0$:

$$d\mathbf{Y}(t) = d\mathbf{X}(t) + R\, d\mathbf{L}(t), \qquad \mathbf{Y}(0) = \mathbf{y}_0,$$
$$\mathbf{Y}(t) \ge 0, \qquad Y_j(t)\, dL_j(t) \ge 0, \qquad L_j(0) = 0, \qquad dL_j(t) \ge 0.$$

We assume that the reflection matrix $R$ is an $M$-matrix of the form $R = I - Q^T$, where $Q$ has nonnegative coordinates and a spectral radius equal to $\alpha < 1$ so that $R^{-1}$ has only nonnegative elements; see page 304 of Harrison and Reiman (1981). We also assume the stability condition $R^{-1}\mathbf{v} < \mathbf{0}$ for the existence of the steady-state distribution. As discussed in the Harrison and Reiman (1981), there is a unique solution pair $(\mathbf{Y}, \mathbf{L})$ to the Skorokhod problem associated with $\mathbf{X}$, and the process $\mathbf{Y}$ is called a reflected Brownian Motion (RBM). We wish to sample $\mathbf{Y}(\infty)$ (at least approximately, with a pre-defined controlled error).

The stochastic dominance result for reflected Brownian motions that is analogous to Lemma 1 was first developed in the proof of Lemma 12 in Harrison and Williams (1987). In detail, we can construct a dominating process $\mathbf{Y}^+(\cdot)$ as follows. First, we can choose $\mathbf{z} \in \mathbb{R}^d$ such that $\mathbf{v} < \mathbf{z}$ and $R^{-1}\mathbf{z} < \mathbf{0}$. Define a process

$$\mathbf{Z}(t) = \mathbf{X}(t) - \mathbf{z}t := A\mathbf{B}(t) - \boldsymbol{\mu}t, \tag{26}$$

where $\boldsymbol{\mu} = \mathbf{v} - \mathbf{z}$, and let $\mathbf{Y}^+(\cdot)$ be the RBM corresponding to the Skorokhod problem (4), which has orthogonal reflection. Then $R^{-1}\mathbf{Y}(t) \le R^{-1}\mathbf{Y}^+(t)$. As a result, we can assume without loss of generality that the input Brownian motion has strictly negative drift coordinatewise. In sum, the following assumption is in force throughout this section:

ASSUMPTION (D). The input process $\mathbf{Z}(\cdot)$ satisfies (26) with $\mu_i > \delta_0 > 0$ for all $1 \le i \le d$, and we assume that $A$ is nondegenerate so that $A^T A$ is positive definite.

Since $\mathbf{Z}(\cdot)$ has strictly negative drift, following the same argument given for Proposition 1, we can construct a stationary version of the dominating process as

$$\mathbf{Y}^+(-t) = -\mathbf{Z}(t) + \max_{u \ge t} \mathbf{Z}(u) \triangleq \mathbf{Z}(t) - \mathbf{M}(t) \qquad \text{for all } t \ge 0. \tag{27}$$

In order to apply the same strategy as in Algorithm 1 to the RBM, we need to address two problems. First, the input process $\mathbf{Z}$ requires a continuous path description while the computer can only encode and generate discrete objects. Second, the dominating process is a reflected Brownian motion with orthogonal reflection. Therefore the hitting time $\tau$ to the origin is almost surely infinity [see Varadhan and Williams (1985)], which means that Algorithm 1 will not terminate in finite time, in this case. To solve the first problem, we take advantage of a wavelet representation of Brownian motion and use it to simulate a piecewise linear approximation with uniformly small (deterministic) error. To solve the second problem, we define an approximated coalescent time $\tau_\varepsilon$ as the first passage time to a small ball around the origin so that $E[\tau_\varepsilon] < \infty$ and the error caused by replacing $\tau$ with $\tau_\varepsilon$ is bounded by $\varepsilon$. In sum, we concede to an algorithm that is not exact but one that could give any user-defined $\varepsilon$ precision. Nevertheless, at the end of Section 4.1 we will show that we can actually use this $\varepsilon$-biased algorithm to estimate without any bias the steady-state expectation of continuous functions of RBM by introducing an extra randomization step.

Section 4 is organized as follows. In Section 4.1, we will describe the main strategy of our algorithm. In Section 4.2, we use a wavelet representation to simulate a piecewise linear approximation of Brownian motion. In Section 4.3, we will discuss the details in simulating jointly $\tau_\varepsilon$ and the stationary dominating process based on the techniques we have already used for the compound Poisson cases. In the end, in Section 4.4, we will give an estimate of the computational complexity of our algorithm.

4.1. *The structure of the main simulation procedure.* The main strategy of the algorithm is almost the same as Algorithm 1, except for two modifications due to the two issues discussed above: first, instead of simulating the input process $\mathbf{Z}$ exactly, we simulate a piecewise linear approximation $\mathbf{Z}^\varepsilon$ such that $|Z_i^\varepsilon(t) -$

$Z_i(t)| < \varepsilon$ for all indices $i$ and $t \geq 0$; second, instead of sampling the coalescence time $\tau$ such that $\mathbf{M}(\tau) = \mathbf{Z}(\tau)$, we simulate an approximation coalescence time, $\tau_\varepsilon$, such that $\mathbf{M}(\tau_\varepsilon) \leq \mathbf{Z}(\tau_\varepsilon) + \boldsymbol{\varepsilon}$.

With this notation, we now give the structure of our algorithm. The details will be given later in Sections 4.2 and 4.3:

ALGORITHM 2 [Sampling with controlled error of $\mathbf{Y}(\infty)$].

*Step* 1: Let $\tau_\varepsilon \geq 0$ be any time for which $\mathbf{M}(\tau_\varepsilon) \leq \mathbf{Z}(\tau_\varepsilon) + \boldsymbol{\varepsilon}$, and simulate, jointly with $\tau_\varepsilon$, $\mathbf{Z}^{\leftarrow}_{-\tau_\varepsilon}(t) = -\mathbf{Z}^{\varepsilon}(\tau_\varepsilon - t)$ for $0 \leq t \leq \tau_\varepsilon$.

*Step* 2: Define $\mathbf{X}^{\leftarrow}_{-\tau_\varepsilon}(t) = \mathbf{Z}^{\varepsilon}(\tau_\varepsilon) - \mathbf{Z}^{\varepsilon}(\tau_\varepsilon - t) + \mathbf{z}t$, and compute $\mathbf{Y}^{\varepsilon}_{-\tau_\varepsilon}(\tau_\varepsilon, 0)$ which is obtained by evolving the solution $\mathbf{Y}^{\varepsilon}_{-\tau_\varepsilon}(\cdot, 0)$ to the Skorokhod problem

$$
\begin{aligned}
d\mathbf{Y}^{\varepsilon}_{-\tau_\varepsilon}(t,0) &= d\mathbf{X}^{\leftarrow}_{-\tau_\varepsilon}(t) + R\, d\mathbf{L}_{-\tau}(t,0),\\
\mathbf{Y}^{\varepsilon}_{-\tau_\varepsilon}(t,0) &\geq 0, \qquad Y^{\varepsilon}_{-\tau_\varepsilon,j}(t,0)\, dL_{-\tau_\varepsilon,j}(t,0) \geq 0,\\
L_{-\tau_\varepsilon,j}(0,0) &= 0, \qquad dL_{-\tau_\varepsilon,j}(t,0) \geq 0,
\end{aligned}
$$

for $\tau_\varepsilon$ units of time.

*Step* 3: Output $\mathbf{Y}^{\varepsilon}_{-\tau_\varepsilon}(\tau_\varepsilon, 0)$.

First, we show that there exists a stationary version $\{\mathbf{Y}^*(t) : t \leq 0\}$ that is coupled with the dominating stationary process $\{\mathbf{Y}^+(t) : t \leq 0\}$ as given by (27).

LEMMA 3. *There exists a stationary version* $\{\mathbf{Y}^*(t) : t \leq 0\}$ *of* $\mathbf{Y}$ *such that* $R^{-1}\mathbf{Y}^*(t) \leq R^{-1}\mathbf{Y}^+(t)$ *for all* $t \leq 0$.

PROOF. The proof follows the same argument as that of Proposition 2. □

The following proposition shows that the error of the above algorithm has a small and deterministic bound.

PROPOSITION 5. *Suppose* $\mathbf{X} \in \mathbb{R}^d$. *Let* $r = \max_{i,j} R^{-1}_{ij} / \min_{i,j}\{R^{-1}_{ij} : R^{-1}_{ij} > 0\}$. *Then there exists a stationary version* $\mathbf{Y}^*$ *of* $\mathbf{Y}$ *such that in each index* $i$,

$$
\left|Y^*_i(0) - Y^{\varepsilon}_{\tau_\varepsilon,i}(\tau_\varepsilon, 0)\right| \leq \left(\frac{1}{1-\alpha} + dr\right)\varepsilon.
$$

*Here* $0 \leq \alpha < 1$ *is the spectral radius of the matrix* $Q$.

PROOF. Consider three processes on $[-\tau_\varepsilon, 0]$. The first is the coupled stationary process $\mathbf{Y}^*(\cdot)$ as constructed in Lemma 3, which is the solution to the Skorokhod problem with initial value $\mathbf{Y}^*(-\tau_\varepsilon)$ at time $-\tau_\varepsilon$ and input process $\tilde{\mathbf{X}}(\cdot) = \mathbf{X}(\tau_\varepsilon) - \mathbf{X}(-\cdot)$ on $[-\tau_\varepsilon, 0]$; the second is a process $\tilde{\mathbf{Y}}(\cdot)$, which is the solution to the Skorokhod problem with initial value 0 at time $-\tau_\varepsilon$ and input process $\tilde{\mathbf{X}}(\cdot)$; the third is the process $\mathbf{Y}^{\varepsilon}_{-\tau_\varepsilon}(t, 0)$ as we described in the algorithm, which is the solution to the Skorokhod problem with initial value 0 at time $-\tau_\varepsilon$ and input process $\mathbf{X}^{\leftarrow}_{-\tau_\varepsilon}(t)$ as defined in step 2 of Algorithm 2.

By definition, we know that for each index $i$, $|Y_i^+(-\tau_\varepsilon)| < \varepsilon$. Since $R^{-1}\mathbf{Y}(\tau_\varepsilon) \le R^{-1}\mathbf{Y}^+(\tau_\varepsilon)$, the coupled process $Y_i^*(-\tau_\varepsilon) < dr\,\varepsilon$. Note that $\mathbf{Y}^*(\cdot)$ has the same input data as $\tilde{\mathbf{Y}}(\cdot)$ except for their initial values. According to the comparison theorem of Ramasubramanian (2000), the difference between these two processes is uniformly bounded by the difference of their initial values coordinate-wise. Therefore, we can conclude $|Y_i^*(0) - \tilde{Y}_i(0)| < dr\,\varepsilon$.

On the other hand, $\tilde{\mathbf{Y}}(\cdot)$ and $\mathbf{Y}^\varepsilon_{-\tau_\varepsilon}(\cdot, 0)$ have common initial value 0 and input processes whose difference is uniformly bounded by $\varepsilon$. It was proved in Harrison and Reiman (1981) that the Skorokhod mapping is Lipschitz continuous under the uniform metric $d_T(Y^1(\cdot), Y^2(\cdot)) \triangleq \max_{1\le i\le d} \sup_{0\le t\le T} |Y_i^1(t) - Y_i^2(t)|$ for all $0 < T < \infty$, and the Lipschitz constant is equal to $1/(1-\alpha)$, where $0 \le \alpha < 1$ is the spectral radius of $Q$. Therefore, we have that $|\tilde{Y}_i(0) - Y^\varepsilon_{-\tau_\varepsilon,i}(\tau_\varepsilon, 0)| < \varepsilon/(1-\alpha)$.

Simply applying the triangle inequality, we obtain that

$$\left|Y_i^*(0) - Y^\varepsilon_{\tau_\varepsilon,i}(\tau_\varepsilon, 0)\right| \le \left(\frac{1}{1-\alpha} + dr\right)\varepsilon. \qquad \square$$

We conclude this subsection by explaining how to remove the $\varepsilon$-bias induced by Algorithm 2. Let $T$ be any positive random variable with positive density $\{f(t) : t \ge 0\}$ independent of $\mathbf{Y}^*(0)$. Let $g : \mathbb{R}^d \to \mathbb{R}$ be any positive Lipschitz continuous function such that there exists constant $K > 0$ and for all $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^d$, $|g(\mathbf{x}) - g(\mathbf{y})| \le K \max_{i=1} |x_i - y_i|$. As illustrated in Beskos, Peluchetti and Roberts (2012),

$$\begin{aligned} E\left[g\left(\mathbf{Y}^*(0)\right)\right] &= E\left[\int_0^{g(\mathbf{Y}^*(0))} dt\right] = E\left[\int_0^{g(\mathbf{Y}^*(0))} \frac{f(t)}{f(t)}\,dt\right] \\ &= E\left[\frac{1(g(\mathbf{Y}^*(0)) > T)}{f(T)}\right]. \end{aligned}$$

Since $|Y_i^*(0) - Y^\varepsilon_{\tau_\varepsilon,i}(\tau_\varepsilon, 0)| \le (1 + dr)\varepsilon$, we can sample $T$ first, and then select $\varepsilon > 0$ small enough, output $1(g(\mathbf{Y}^\varepsilon_{\tau_\varepsilon}(\tau_\varepsilon, 0)) > T)/f(T)$ as an unbiased estimator of $E[g(\mathbf{Y}^*(0))]$ without the need for computing $\mathbf{Y}^*(0)$ exactly. It is important to have $(\mathbf{Y}^\varepsilon_{\tau_\varepsilon}(\tau_\varepsilon, 0) : \varepsilon > 0)$ coupled as $\varepsilon \to 0$, and this can be achieved thanks to the wavelet construction that we will discuss next.

4.2. *Wavelet representation of Brownian motion.* In this part, we give an algorithm to generate piecewise linear approximations to a Brownian motion path-by-path, with uniform precision on any finite time interval. The main idea is to use a wavelet representation for Brownian motion.

By the Cholesky decomposition, any multidimensional Brownian motion can be expressed as a linear combination of independent one-dimensional Brownian motions. Our goal is to give a piecewise linear approximation to a $d$-dimensional Brownian motion $\mathbf{Z}$ with uniform precision $\varepsilon$ on $[0, 1]$. Suppose that we can write

$\mathbf{Z} = A\mathbf{B}$, where $A$ is the Cholesky decomposition of the covariance matrix, and the $B_i$'s are independent standard Brownian motions. If we are able to give a piecewise linear approximation $\tilde{B}_i$ to each $B_i$ on $[0,1]$ with precision $\varepsilon/(d \cdot a)$ where $a = \max_{i,j} |A_{ij}|$, then $A\tilde{\mathbf{B}}$ is a piecewise linear approximation to $\mathbf{Z}$ with uniform error $\varepsilon$. Therefore, in the rest of this part, we only need to work with a standard one-dimensional Brownian motion.

Now let us introduce the precise statement of a wavelet representation of Brownian motion; see Steele (2001), pages 34–39. First we need to define step function $H(\cdot)$ on $[0,1]$ by

$$H(t) = \begin{cases} 1, & \text{for } 0 \le t < \frac{1}{2}, \\ -1, & \text{for } \frac{1}{2} \le t \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Then define a family of functions

$$H_k(t) = 2^{j/2} H\big(2^j t - l\big)$$

for $k = 2^j + l$ where $j > 0$ and $0 \le l \le 2^j$. Set $H_0(t) = 1$. The following wavelet representation theorem can be seen in Steele (2001):

THEOREM 3. *If $\{W^k : 0 \le k < \infty\}$ is a sequence of independent standard normal random variables, then the series defined by*

$$B_t = \sum_{k=0}^{\infty} \left( W^k \int_0^t H_k(s)\, ds \right)$$

*converges uniformly on $[0,1]$ with probability one. Moreover, the process $\{B_t\}$ defined by the limit is a standard Brownian motion on $[0,1]$.*

Choose $\eta_k = 4 \cdot \sqrt{\log k}$, and note that $P(|W^k| > \eta_k) = O(k^{-4})$, so $\sum_{k=0}^{\infty} P(|W^k| > \eta_k) < \infty$. Therefore, $P(|W^k| > \eta_k, \text{i.o.}) = 0$. The simulation strategy will be to sample $\{W^k\}$ jointly with the finite set $\{k : |W^k| \ge \eta_k\}$.

Note that if we take $j = \lceil \log_2 k \rceil$, as shown in Steele (2001),

$$\sum_{k=1}^{\infty} \left( W^k \int_0^t H_k(s)\, ds \right) \le \sum_{j=0}^{\infty} \Big( 2^{-j/2} \cdot \max_{2^j \le k \le 2^{j+1}-1} |W^k| \Big).$$

Since $\sum_{j=0} 2^{-j/2} \sqrt{j+1} < \infty$, for any $\varepsilon > 0$ there exists $K_0 > 0$, such that

$$\sum_{j = \lceil \log K_0 \rceil} 2^{-j/2} \sqrt{j+1} < \varepsilon. \tag{28}$$

As a result, define

$$K = \max\{k : |W^k| > \eta_k\} \vee K_0 < \infty, \tag{29}$$

then $\sum_{k=K+1}^{\infty} |W^k| \int_0^t H_k(s)\,ds \le \varepsilon$. If we can simulate $\{(W^k)_{k=1}^K, K\}$ jointly,

$$B^{\varepsilon}(t) = \sum_{k=0}^{K} W^k \int_0^t H_k(s)\,ds \tag{30}$$

will be a piecewise linear approximation to a standard Brownian motion within precision $\varepsilon$ in $C[0, 1]$.

Now we show how to simulate $K$ jointly with $\{W^k : 1 \le k \le K\}$. The algorithm is as below with $\rho = 4$ as we have chosen $\eta_k = 4 \cdot \sqrt{\log k}$:

ALGORITHM 2w (Simulate $K$ jointly with $\{W^k\}$).

*Step* 0: Initialize $G = K_0$ and $S$ to be an empty array.

*Step* 1: Set $U = 1$, $D = 0$. Simulate $V \sim \text{Uniform}(0, 1)$.

*Step* 2: While $U > V > D$, set $G \leftarrow G + 1$ and $U \leftarrow P(|W^G| \le \rho\sqrt{\log G}) \times U$ and $D \leftarrow (1 - G^{1-\rho^2/2}) \times U$.

*Step* 3: If $V \ge U$, add $G$ to the end of $S$, that is, $S = [S, G]$, and return to step 1.

*Step* 4: If $V \le D$, $K = \max(S, K_0)$.

*Step* 5: For every $k \in S$, generate $W^k$ according to the conditional distribution of $Z$ given $\{|W| > \rho\sqrt{\log k}\}$; for other $1 \le k \le K$, generate $W^k$ according to the conditional distribution of $W$ given $\{|W| \le \rho\sqrt{\log k}\}$.

In this algorithm, we keep an array $S$, which is used to record the indices such that $|W^k| > \rho\sqrt{\log k}$, and a number $G$ which is the next index to be added into $S$. Precisely speaking, given that the last element in array $S$ is $N$, say, $\max(S) = N$, $G = \inf\{k \ge N + 1 : |W^k| > \rho\sqrt{\log k}\}$. The key part of the algorithm is to simulate a Bernoulli with success parameter $P(G < \infty)$ and to sample $G$ given $G < \infty$.

For this purpose, we keep updating two constants $U$ and $D$ such that $U > P(G = \infty) > D$ and $(U - D) \to 0$ as the number of iterations grows. To illustrate this point, denote the value of $U$ and $D$ in the $m$th iteration by $U_m$ and $D_m$, respectively. Then for all $m > 0$,

$$P(G = \infty) = \prod_{k=N+1}^{\infty} P\bigl(|W^k| \le \rho\sqrt{\log k}\bigr) < \prod_{k=N+1}^{N+m} P\bigl(|W^k| \le \rho\sqrt{\log k}\bigr) = U_m.$$

On the other hand, for all $\rho > \sqrt{2}$ and $N$ large enough,

$$\begin{aligned}\prod_{k=N+m+1}^{\infty} P\bigl(|W^k| \le \rho\sqrt{\log k}\bigr) &> 1 - \sum_{k=N+m+1}^{\infty} P\bigl(|W^k| > \rho\sqrt{\log k}\bigr) \\ &\ge 1 - (N + m + 1)^{1-\rho^2/2},\end{aligned}$$

and hence we conclude that $D_m = (1 - (N + m + 1)^{1-\rho^2/2})U_m < P(G = \infty)$. Because $(1 - (N + m + 1)^{1-\rho^2/2}) \to 1$ as $m \to \infty$, the algorithm proceeds to

steps 3 or 4 after a finite number of iterations, and we can decide whether $G < \infty$ or not.

Now we show that we can actually sample $G$ simultaneously as the Bernoulli with success probability $P(G < \infty)$ is generated. If $V < D$, we conclude that $V < P(G = \infty)$ and hence $G = \infty$ and $K = \max(S)$. Otherwise, we have $G < \infty$. In this case, suppose step 2 ends in the $(m+1)$th iteration and $V > U$. Since $U_m = P(|W^k| \le \rho\sqrt{\log k} \text{ for } k = K+1, \ldots, K+m)$, $U_{m+1} \le V < U_m$ implies nothing but that $K + m + 1 = \inf\{k \ge K+1 : |W^k| > \rho\sqrt{\log k}\}$. Therefore, by definition, $G = K + m + 1$ and should be added into array $S$. Once $S$ and $K$ are generated, $\{W^k : 1 \le k \le K\}$ can be generated jointly with $S$ and $K$ according to step 5.

Also we note that $B^\varepsilon(t)$ has the following nice property:

PROPOSITION 6.

$$B^\varepsilon(1) = B(1).$$

PROOF. The equality follows from the fact that $\int_0^1 H_n(s)\,ds = 0$ for any $n \ge 1$ and $m \ge 1$. □

As a consequence of this property, for any compact time interval $[0, T]$ (without loss of generality, assume $T$ is an integer), in order to give an approximation for $B(t)$ on $[0, T]$ with guaranteed $\varepsilon$ precision uniformly in $[0, T]$, we only need to run the above algorithm $T$ times to get $T$ i.i.d. sample paths $\{B^{\varepsilon,(i)}(t) : t \in [0,1]\}$ for $i = 1, 2, \ldots, T$, and define recursively

$$B^\varepsilon(t) = \sum_{i=1}^{\lfloor t \rfloor} B^{\varepsilon,(i)}(1) + B^\varepsilon_{\lfloor t \rfloor}\big(t - \lfloor t \rfloor\big).$$

4.3. *A conceptual framework for the joint simulation of $\tau_\varepsilon$ and $\mathbf{Z}^\varepsilon$.* Our goal now is to develop an algorithm for simulating $\tau_\varepsilon$ and $(\mathbf{Z}^\varepsilon(t) : 0 \le t \le \tau_\varepsilon)$ jointly. In detail, we want to simulate $\mathbf{Z}^\varepsilon(t)$ forward in time and stop at a random time $\tau_\varepsilon$ such that for any time $s > \tau_\varepsilon$, $Z_i(s) \le Z_i(\tau_\varepsilon) + \varepsilon$ for $1 \le i \le d$.

Because of the special structure of the wavelet representation used in simulating the process $\mathbf{Z}^\varepsilon(\cdot)$, the time $T_m \triangleq \inf\{t \ge 0 : Z_i^\varepsilon(t) > m \text{ for some } 1 \le i \le d\}$ is no longer a stopping time with respect to the filtration generated by $\mathbf{Z}(\cdot)$. As a consequence, we cannot directly carry out importance sampling as in Algorithm 1.1.1. To remedy this problem, we decompose the process $\mathbf{Z}^\varepsilon(t)$ into two parts: a random walk $\{\mathbf{Z}^\varepsilon(n) : n \ge 0\}$ with Gaussian increment and a series of independent Brownian bridges $\{\bar{\mathbf{B}}_n(s) \triangleq \mathbf{Z}^\varepsilon(n+s) - \mathbf{Z}^\varepsilon(n) : s \in [0,1], n \ge 0\}$. Our strategy is to first carry out the importance sampling as in Algorithm 1.1.1 to the random walk $\{\mathbf{Z}^\varepsilon(n) : n \ge 0\}$ to find its upper bound, and next develop a new scheme to control

the upper bounds attained in the intervals $\{(n, n+1) : n \geq 0\}$ for the i.i.d. Brownian bridges $\{\bar{\mathbf{B}}_n(s) : s \in [0,1], n \geq 0\}$.

The whole procedure is based on the wavelet representation of Brownian motion. Let $\{W_n^k(i) : n, k \in \mathbb{N}, i = 1, 2, \ldots, d\}$ be a sequence of i.i.d. standard normal random variables. According to the expression given in Theorem 3, for any $t = n + s$, $s \in [0,1]$,

$$Z_i(t) = Z_i(n) + s\big(Z_i(n+1) - Z_i(n)\big) + \sum_{j=1}^{d} A_{ij}\left(\sum_{k=1}^{\infty} W_n^k(j) \int_0^s H_k(u)\,du\right). \tag{31}$$

Let us put (31) in matrix form,

$$\mathbf{Z}(t) = \mathbf{Z}(n) + s\big(\mathbf{Z}(n+1) - \mathbf{Z}(n)\big) + A \sum_{k=1}^{\infty} \mathbf{W}_n^k \cdot \int_0^s H_k(u)\,du.$$

For all $n \geq 0$ and $s \in [0,1]$, $\bar{\mathbf{B}}_n(s) = A \sum_{k=1}^{\infty} \mathbf{W}_n^k \cdot \int_0^s H_k(u)\,du$. Then the sequence $\{\bar{\mathbf{B}}_n(\cdot) : n \geq 0\}$ is i.i.d. Note that $(Z_i(n+1) - Z_i(n))$ is independent of $\{W_n^k(i) : k \geq 1\}$. We can split the simulation into two independent parts:

(1) Simulate the discrete-time random walk $\{\mathbf{Z}(n) : n \geq 0\}$ with i.i.d. Gaussian increments and $\mathbf{Z}(0) = 0$. That is, $Z_i(0) = 0$ and $Z_i(n+1) = Z_i(n) + \sum_{j=1}^{d} A_{ij} W_{n+1}^0(j) - \mu_i$, where $\{W_n^0(j) : n \geq 0\}$ are i.i.d. standard normals.

(2) For each $n$, simulate $\bar{\mathbf{B}}_n(s)$ to do bridging between $\mathbf{Z}(n)$ and $\mathbf{Z}(n+1)$.

Now, any time $t_0 > 0$ is an approximate coalescence time $\tau_\varepsilon$ if there exists some positive constant $\zeta > 0$ such that the following two conditions hold for all $n \geq t_0$: Condition (1), $\mathbf{Z}(n) \leq \mathbf{Z}(t_0) - \zeta(n - \lceil t_0 \rceil)\mathbf{1} + \boldsymbol{\varepsilon}$, and condition (2), $\max\{\bar{\mathbf{B}}_n(s) : s \in [0,1]\} \leq \zeta(n - \lceil t_0 \rceil)\mathbf{1}$. Based on these observations, we develop an algorithm to simulate the approximate coalescence time $\tau_\varepsilon$ jointly with $\{\mathbf{Z}^\varepsilon(t) : 0 \leq t \leq \tau_\varepsilon\}$.

By Assumption (D), $\mu_i > \delta_0$ for some $\delta_0 > 0$. Let $\zeta = \delta_0/2$, and define $\mathbf{S}(n) = \mathbf{Z}(n) + n\zeta\mathbf{1}$ such that $\{\mathbf{S}(n) : n \geq 0\}$ is a random walk with strictly negative drift. Therefore, condition (1) can be checked by carrying out the importance sampling procedure as in Algorithm 1.1.1 for the random walk $\{\mathbf{S}(n) : n \geq 0\}$. More precisely, since $S_i(n)$ has Gaussian increments, we can compute explicitly that $\theta_i^* = 2(\mu_i - \zeta)/\sigma_i$ and choose $m > 0$ satisfying (12) in order to carry out the importance sampling procedure for the random walk $\{\mathbf{S}(n) : n \geq 0\}$. Suppose we use the importance sampling procedure and find $t_0$ such that $\mathbf{S}(n) \leq \mathbf{S}(t_0)$ for all $n \geq t_0$, and hence condition (1) is satisfied for $t_0$.

About condition (2), recall that $\bar{\mathbf{B}}_n(\cdot)$'s are i.i.d. linear combinations of Brownian bridges, and let $M$ be a random time, finite almost surely, such that

$$M \geq \max\Big\{n \geq t_0 : \max_{0 \leq s \leq 1} \big(\bar{B}_{n,i}(s) - \zeta(n - t_0)\big) > 0 \text{ for some } i\Big\}. \tag{32}$$

Observe that for $t_0$ to be an approximate coalescence time, conditions (1) and (2) must hold simultaneously. If for time $t_0$, for example, condition (1) is satisfied while condition (2) is not, we need to continue the testing procedure and simulation of the process for $t > t_0$. Then, however, the random walk $\{\mathbf{S}(n) : n \geq \lceil t_0 \rceil\}$ should be conditioned on that $\mathbf{S}(n) \leq \mathbf{S}(t_0)$ for the fact that condition (1) holds for $t_0$ reveals "additional information" on the random walk for $n \geq t_0$. Therefore, such "additional information" or "conditioning event" must be incorporated and tracked when conditions (1) and (2) are sequentially tested. All of these conditioning events are described and accounted for in Section 4.3.2, which also includes the overall procedure to sample $\tau_\varepsilon$ jointly with $\mathbf{Z}^\varepsilon$.

Now, let us first provide a precise description of $M$ and explain the simulation algorithm for $M$ in Section 4.3.1.

4.3.1. *Simulating $M$ and $\{\bar{\mathbf{B}}_n^\varepsilon(\cdot) : 1 \leq n \leq M\}$.* Recall that $\bar{\mathbf{B}}_n(t) = A\sum_{k=1}^\infty \mathbf{W}_n^k \cdot \int_0^t H_k(u)\,du$, where $\{W_n^k(i) : n \geq 0, k \geq 1, 1 \leq i \leq d\}$ are i.i.d. standard normals. Note that

$$\sum_{n=0}\sum_{k=1} P\big(|W_n^k(i)| \geq 4\sqrt{\log(n+1)} + 4\sqrt{\log k}\big) \leq \sum_{n=0}\sum_{k=1} \frac{1}{((n+1)k)^4} < \infty.$$

By the Borel–Cantelli lemma, we can conclude that for each $i \in \{1, \ldots, d\}$ there exists $M^i < \infty$ such that for all $(n+1)k > M^i$, $|W_n^k(i)| \leq 4\sqrt{\log(n+1)} + 4\sqrt{\log k}$. Clearly, $\sqrt{\log t} = o(t)$ as $t \to \infty$, so we can select a $m_0$ large enough such that for any $n > m_0$,

$$(n+1)\zeta - ad\left(4\sqrt{\log(n+1)} - \sum_{j=1}^\infty 2^{-j}\sqrt{j}\right) \geq 0.$$

Note that $M^i$ can be simulated jointly with $(W_n^k(i) : n \geq 0, k \geq 1, 1 \leq i \leq d, (n+1)k \leq M^i)$ by adapting Algorithm 2w in Section 4.2 and $M^i$'s are independent of each other. Then, for any $n > \max_{i=1}^d M^i \vee m_0$,

$$\begin{aligned}\bar{\mathbf{B}}_n(t) &= A\sum_{k=1}^\infty \mathbf{W}_n^k \cdot \int_0^t H_k(u)\,du \\ &\leq ad\left(4\sqrt{\log(n+1)} + \sum_{j=1}^\infty 2^{-j/2}\sqrt{j}\right) \leq (n+1)\zeta,\end{aligned}$$

where, $j = \lceil \log_2 k \rceil$. Therefore, we can choose $M = \max_i M^i \vee m_0$.

Now we introduce a variation of Algorithm 2w that will be used in the procedure to simulate $M$ and $\{\bar{B}_n^\varepsilon(\cdot) : 1 \leq n \leq M\}$ jointly. In the following algorithm, a sequence of "conditioning events" of the form $|W^k| \leq \beta_k$, for some given constants $\{\beta^k : \beta^k > 4\sqrt{\log k}\}$, is in force. Let $\Phi(a) = P(|W| < a)$ for all $a > 0$, where $W$ is a standard normal. The random number $K$ to be simulated is defined as in (29).

ALGORITHM 2w′ (Simulate $K$ jointly with $\{W^k : 1 \le k \le K\}$ conditional on $|W^k| \le \beta^k$ for all $k \ge 1$).

*Step* 0: Initialize $G = K_0$ as defined in (28) and $S$ to be an empty array.

*Step* 1: Set $U = 1$, $D = 0$. Simulate $V \sim \text{Uniform}(0, 1)$.

*Step* 2: While $U > V > D$, set $G \leftarrow G + 1$ and $U \leftarrow \frac{\Phi(4\sqrt{\log G})}{\Phi(\beta^k)} \times U$ and $D \leftarrow (1 - G^{-7}) \times U$.

*Step* 3: If $V \ge U$, add $G$ to the end of $S$, that is, $S = [S, G]$, and return to step 1.

*Step* 4: If $V \le D$, $K = \max(S, K_0)$.

*Step* 5: For every $k \in S$, generate $W^k$ according to the conditional distribution of $Z$ given $\{4\sqrt{\log k} < |W| \le \beta^k\}$; for other $1 \le k \le K$, generate $W^k$ according to the conditional distribution of $W$ given $\{|W| \le 4\sqrt{\log k}\}$.

The main difference between Algorithm 2w′ and the original Algorithm 2w is that $U$ and $V$ are now computed from the conditional probability; however, the relations $U > V > D$ and $U - D \to 0$ still hold, and hence Algorithm 2w′ is valid. Based on this, we can now give the main procedure to simulate $M$ and $\{\bar{B}^\varepsilon_n(\cdot) : 1 \le n \le M\}$ jointly:

ALGORITHM 2m (Simulating of $M$ and $\{\bar{\mathbf{B}}^\varepsilon_n(\cdot) : 1 \le n \le M\}$ jointly).

(1) For each index $i$, simulate $M^i$ and $(W^k_n(i) : n \ge 0, k \ge 1, nk < M)$. Compute $M = \max_i M^i \vee m_0$. (As discussed earlier, $M^i$'s are simulated by adapting Algorithm 2w.)

(2) For each $0 \le n \le M$ and each index $i$, $\{W^k_n(i) : k < M^i/n\}$ are already given in step 1. For $k \ge M^i/n$, use Algorithm 2w′ to simulate $K^i_n$ jointly with $\{W^k_n(i) : M^i/n \le k \le K\}$ conditional on $|W^k_n(i)| \le 4(\sqrt{\log(n+1)} + \sqrt{\log k}) \triangleq \beta^k > 4\sqrt{\log k}$.

(3) For any $0 \le n \le M$, compute and output

$$\bar{B}^\varepsilon_{n,i}(t) = \sum_{i=1}^{d} A_{ij} \left( \sum_{k=1}^{K^i_n} W^k_n(i) \int_0^t H_k(u)\, du \right). \tag{33}$$

In step 1 of Algorithm 2m, we can use a similar procedure as in Algorithm 2w′ to impose conditioning events of form $|W^k_n(i)| \le \beta^k_n(i)$ while simulating $M_i$'s jointly with $W^k_n(i)$'s. In this way, we derive an algorithm that is able to simulate $M$ jointly with $\{\bar{\mathbf{B}}^\varepsilon_n(\cdot) : 1 \le n \le M\}$ conditional on $|W^k_n(i)| \le \beta^k_n(i)$ for all $n \ge 0$, $k \ge 1$ and $1 \le i \le d$ for any given sequence of $\{\beta^k_n(i)\}$ such that $\beta^k_n(i) > 4(\sqrt{\log(n+1)} + \sqrt{\log k})$.

ALGORITHM 2m′ (Simulating of $M$ and $\{\bar{\mathbf{B}}^\varepsilon_n(\cdot) : 1 \le n \le M\}$ jointly conditional on $|W^k_n(i)| \le \beta^k_n(i)$ for all $n \ge 0$, $k \ge 1$ and $1 \le i \le d$).

(1) For each index $i$, simulate $M_i$ and $(W_n^k(i) : n \geq 0, k \geq 1, nk < M)$ conditional on $|W_n^k(i)| \leq \beta_n^k(i)$ using a similar procedure as in Algorithm 2w′. Compute $M = \max_i M^i \vee m_0$.

(2) For each $0 \leq n \leq M$ and each index $i$, $\{W_n^k(i) : k < M^i/n\}$ are already given in step 1. For $k \geq M^i/n$, use Algorithm 2w′ to simulate $K_n^i$ jointly with $\{W_n^k(i) : M^i/n \leq k \leq K\}$ conditional on $|W_n^k(i)| \leq 4(\sqrt{\log(n+1)} + \sqrt{\log k})$. [Note that $\beta_n^k(i) > 4(\sqrt{\log(n+1)} + \sqrt{\log k}) > 4\sqrt{\log k}$, and hence this step is well defined.]

(3) For any $0 \leq n \leq M$, compute and output

$$\bar{B}_{n,i}^{\varepsilon}(t) = \sum_{i=1}^{d} A_{ij} \left( \sum_{k=1}^{K_n^i} W_n^k(i) \int_0^t H_k(u)\, du \right).$$

Algorithm 2m′ will be used in the next section in order to keep track of "conditioning events" corresponding to condition (2).

4.3.2. *Keeping track of the conditioning events.* As we have discussed just prior to the beginning of Section 4.3.1, we need to keep track of several conditioning events introduced by conditions (1) and (2). First, let us explain how to deal with the conditioning event corresponding to condition (1). These conditioning events involve only the random walk $\mathbf{S}(\cdot)$. Now we split $\mathbf{S}(\cdot)$ according to the sequences of $\{\Gamma_l : l \geq 1\}$ and $\{\Delta_l : l \geq 1\}$ of random times defined as follows:

(1) Set $\Delta_1 = \min\{n : S_i(n) \leq -2m \text{ for every } i\}$.
(2) Define $\Gamma_l = \min\{n \geq \Delta_l : S_i(n) > S_i(\Delta_l) + m \text{ for some } i\}$.
(3) Put $\Delta_{l+1} = \min\{n \geq \Gamma_l I(\Gamma_l < \infty) \vee \Delta_l : S_i(n) < S_i(\Delta_l) - 2m \text{ for every } i\}$.

Figure 1 illustrates a sample path of the random walk with the sequence of random times $\{\Gamma_l : l \geq 1\}$ and $\{\Delta_l : l \geq 1\}$ in one dimension. The message is that the joint simulation of $\{\mathbf{S}(n) : n \geq 0\}$ with $\{\Gamma_l : l \geq 1\}$ and $\{\Delta_l : l \geq 1\}$ allows us to keep track of the process $\{\max_{m \geq n} \mathbf{S}(m) : n \geq 0\}$, which includes the "additional information" introduced by condition (1). The main steps in the simulation of $\{\mathbf{S}(n) : n \geq 0\}$ jointly with $\{\Gamma_l : l \geq 1\}$ and $\{\Delta_l : l \geq 1\}$ are explained in Lemma 2 through Lemma 4 in Blanchet and Sigman (2011). The approach of Blanchet and Sigman (2011), which works in one dimension, could be modified for multidimensional cases using the change-of-measure as described in Section 2.3.1.

Regarding the verification of condition (2) involving $M$ and the Brownian bridges, as per the discussion in Section 4.3.1, we just need to keep track of certain deterministic $\beta_n^k(i)$ for each $|W_n^k(i)|$, in order to condition on the events of the form $|W_n^k(i)| \leq \beta_n^k(i)$. These events are related to the sequential construction of the random variable $M$ when testing condition (2) as described in Section 4.3.1. Now, we can write down the integrated version of our algorithm for sampling $\tau_\varepsilon$ and $\{\mathbf{Z}^\varepsilon(t) : 0 \leq t \leq \tau_\varepsilon\}$ jointly.

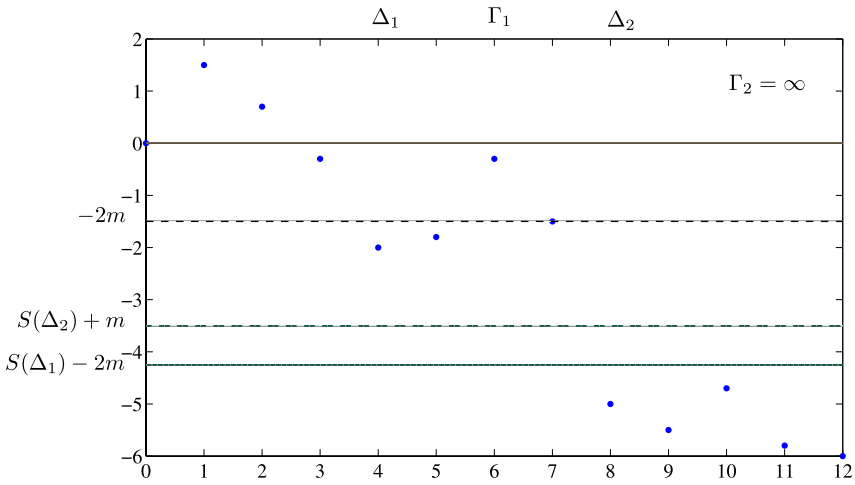


FIG. 1. *Illustration for the random times* $\{\Delta_n\}$ *and* $\{\Gamma_n\}$.

ALGORITHM 2.1 (Simulating $\tau_\varepsilon$ and $\{\mathbf{Z}^\varepsilon(t) : 0 \le t \le \tau_\varepsilon\}$).

The output of this algorithm is $\{\mathbf{Z}^\varepsilon(t) : 0 \le t \le \tau_\varepsilon\}$, and the approximation coalescence time $\tau_\varepsilon$.

(1) Set $\beta_n^k(i) = \infty$ for all $n \ge 1$, $k \ge 1$ and $1 \le i \le d$. Set $L = 0$ and $\tau_\varepsilon = 0$.

(2) Simulate $\mathbf{S}(n)$ until $\Delta_l$, where $l = \min\{j : \Gamma_j = \infty, \Delta_j > \tau_\varepsilon\}$. Compute $\mathbf{Z}^\varepsilon(n) = \mathbf{S}(n) - n\boldsymbol{\zeta}$.

(3) For each $n \in [\tau_\varepsilon, \Delta_l] \cap \mathbb{Z}_+$ and each index $1 \le i \le d$, compute the i.i.d. bridges $\{\bar{\mathbf{B}}_n^\varepsilon(\cdot)\}$ using (33), in which $K_n^i$ is jointly simulated with $(W_n^k(i) : 1 \le k \le K_n^i)$ conditional on that $|W_n^k(i)| \le \beta_n^k(i)$ for all $k \ge 1$ using Algorithm 2w′. Given $\bar{\mathbf{B}}_n^\varepsilon(\cdot)$ and $\mathbf{S}(n)$ for $n \in [\tau_\varepsilon, \Delta_l] \cap \mathbb{Z}_+$, the process $\mathbf{Z}^\varepsilon(t)$ for $t \in [\tau_\varepsilon, \Delta_l]$ can be directly computed. If there exists some $t \ge \Gamma_{l-1}$ such that for all $t \le s \le \Delta_l$, $Z_i^\varepsilon(t) \ge Z_i^\varepsilon(s) - 2\varepsilon$ and $Z_i^\varepsilon(t) \ge Z_i^\varepsilon(\Delta_l) + m - 2\varepsilon$, set $\tau_\varepsilon \leftarrow t$, and go to step 4. Otherwise, set $\tau_\varepsilon \leftarrow \Delta_l$ and return to step 2.

(4) Use Algorithm 2m′ to simulate $M$ jointly with $(\bar{\mathbf{B}}_{\tau_\varepsilon+n}^\varepsilon(\cdot) : 0 \le n \le M)$ conditional on $|W_{\tau_\varepsilon+n}^k(i)| \le \beta_{\tau_\varepsilon+n}^k(i)$ for all $n \ge 0$, $k \ge 1$ and $1 \le i \le d$. Update $\beta_{\tau_\varepsilon+n}^k(i) \leftarrow 4\sqrt{\log(n+1)} + 4\sqrt{\log k}$ for all $n \cdot k \ge M^i$. Keep simulating $\mathbf{S}(n)$ until $n = \Delta_l + M$, and compute $\{\mathbf{Z}^\varepsilon(t) : t \in [\Delta_l, \Delta_l + M]\}$. If there exist some $t$ and $i$ such that $Z_i^\varepsilon(t) > Z_i^\varepsilon(\tau_\varepsilon) + \varepsilon$, set $\tau_\varepsilon \leftarrow t$ and return to step 2.

(5) Otherwise, stop and output $\tau_\varepsilon$ as the approximation coalescence time along with $(\mathbf{Z}^\varepsilon(t) : 0 \le t \le \tau_\varepsilon)$.

4.4. *Computational complexity.* In this part, we will discuss the complexity of our algorithm when $d$ and the other parameters $\boldsymbol{\mu}$ and $A$ are fixed but send the precision parameter $\varepsilon$ to 0. Denote the total number of random variables needed by $N(\varepsilon)$ when the precision parameter for the algorithm is $\varepsilon$.

According to Assumption (D), the input process $\mathbf{Z}(t)$ equals $-\boldsymbol{\mu}t + A\mathbf{B}(t)$ with $\mu_i > \delta_0 > 0$. Let $\max_{i,j} |A_{ij}| = a$. The following result shows that our algorithm's running time is polynomial in $1/\varepsilon$:

THEOREM 4. *Under Assumption (D),*

$$E\big[N(\varepsilon)\big] = O\left(\varepsilon^{-a_C-2}\log\left(\frac{1}{\varepsilon}\right)\right) \qquad \text{as } \varepsilon \to 0,$$

*where $a_C$ is a computable constant depending only on $A$.*

The random variables we need to simulate in the algorithm can be divided into two parts: first, the random variables used to construct the discrete random walk $\mathbf{Z}(n)$ for $n \le T$ and second, the conditional normals used to bridging between $\mathbf{Z}(n-1)$ and $\mathbf{Z}(n)$.

Since $1(|W| > \eta)$ and $1(|W| \le \beta)$ are negatively correlated, it follows that

$$P\big(|W| > \eta \big| |W| \le \beta\big) \le P\big(|W| > \eta\big).$$

Therefore, the expected number of conditional Gaussian random variables used for Brownian bridges between $\mathbf{Z}(n-1)$ and $\mathbf{Z}(n)$ is smaller than the expected number that we would obtain if we use standard Gaussian random variables instead in steps 3 and 4 in Algorithm 2.1. Let $K = \max\{k : |W_k| > \eta_k\} \vee K_0$ as defined in (29). As discussed above, the expected number of truncated Gaussian random variables needed for each bridge $\bar{B}^{\varepsilon}_{n,i}(\cdot)$ is bounded by $E[K]$.

Therefore,

$$E\big[N(\varepsilon)\big] \le \big(dE[K]+1\big)\big(E[T]+1\big).$$

To prove Theorem 2, we first need to study $E[K]$ and $E[T]$.

PROPOSITION 7.

$$E[K] = O\left(\varepsilon^{-2}\log\left(\frac{1}{\varepsilon}\right)\right).$$

PROOF. Recall that $\eta_k = 4\sqrt{\log k}$, and let $p_k = P(|W^k| > \eta_k)$. Then $p_k = O(k^{-4})$. Therefore

$$\begin{aligned} E[K] &= \sum_{n=1}^{\infty} P(K > n) \le K_0 + \sum_{n=K_0+1}^{\infty}\sum_{k=n}^{\infty} p_k \\ &= K_0 + \sum_{k=K_0+1}^{\infty} k \cdot p_k \le K_0 + O\left(\sum_{k=1}^{\infty} k^{-3}\right). \end{aligned}$$

The second term of the left-hand side is finite and independent of $\varepsilon$ and $K_0$.

On the other side,

$$\sum_{j=\log_2 K_0} 2^{-j/2}\sqrt{j+1} \le \frac{2}{\log 2}(\sqrt{K_0})^{-1}\left(\sqrt{\log_2 K_0}+\frac{2}{\log 2}\right).$$

Therefore, we can choose $K_0 = O(\varepsilon^{-2}\log(\frac{1}{\varepsilon}))$ such that $\sum_{j=\log_2 K_0} 2^{-j/2} \times \sqrt{j+1} < \varepsilon$.

In order to get the approximation within error at most $\varepsilon$ for the $d$-dimensional process, according to the Cholesky decomposition as discussed in Section 4.2, we should replace $\varepsilon$ by $\frac{\varepsilon}{da}$. Therefore,

$$E[K] = O\left(\left(\frac{\varepsilon}{da}\right)^{-2}\log\left(\frac{da}{\varepsilon}\right)\right) = O\left(\varepsilon^{-2}\log\left(\frac{1}{\varepsilon}\right)\right). \qquad \square$$

What remains is to estimate $E[T]$. Let $T_a$ be the time before the algorithm executes step 4 in a single iteration. Using the same notation as in Algorithm 2.1 and a similar argument as in Section 2.4, we have

$$E[T] = \frac{E[T_a] + E[T_m | T_m < \infty] + E[M]}{P(T_m < \infty)p},$$

where

$$p = P\left(\max_i Z_i^{\varepsilon}(t) < m+\varepsilon, \forall 0 \le t \le M | \mathbf{Z}(0) = 0; \mathbf{S}(n) < m\right).$$

As $\mathbf{Z}^{\varepsilon}(t) = \mathbf{S}(n) - n\zeta\mathbf{1} + A\bar{\mathbf{B}}_n(t-n)$ and the Brownian bridge $\bar{\mathbf{B}}_n(\cdot)$ is independent of $\mathbf{S}(\cdot)$, it follows that

$$p \ge P\left(\max_i \max_{t\ge 0} Z_i(t) < m | \mathbf{Z}(0) = 0\right).$$

Since $\mathbf{S}(1)$ is a multidimensional Gaussian random vector with strictly negative drift, assumptions (C1) to (C3) are satisfied. Applying Proposition 4, we can get upper bounds for $E[T_m | T_m < \infty]$, $1/P(T_m < \infty)$ and $1/P(\max_i \max_t Z_i(t) < m | \mathbf{Z}(0) = 0)$, which depend only on $d, a$ and $\delta$ and thus are independent of $\varepsilon$. Besides, the bound for $E[M]$ can be estimated by the same method as in Proposition 7 in terms of $\zeta = \delta/2$; hence such a bound is also independent of $\varepsilon$. Therefore, we only need to estimate $E[T_a]$.

PROPOSITION 8. *$E[T_a] = O(\varepsilon^{-a_C})$ as $\varepsilon \to 0$. Here $a_C$ only depends on the matrix $A$. Moreover, in the special cases where $A_{ij} \ge 0$, $a_C = d$.*

PROOF. Recall that $\mathbf{Z}(t) = -\boldsymbol{\mu}t + A\mathbf{B}(t)$ and $\mu_i > \delta = 2\zeta > 0$ as given in Assumption (D). We divide the path of $\mathbf{Z}(t)$ into segments with length $2(m+\varepsilon)/\zeta$,

$$\left\{\left(Z\left(k\cdot\frac{2(m+\varepsilon)}{\zeta}+s\right) : 0 \le s \le \frac{2(m+\varepsilon)}{\zeta}\right) : k \ge 0\right\}.$$

Let

$$N_b = \min\left\{k : A\mathbf{B}\left(k \cdot \frac{2(m+\varepsilon)}{\zeta} + s\right) - A\mathbf{B}\left(k \cdot \frac{2(m+\varepsilon)}{\zeta}\right) \leq \boldsymbol{\varepsilon} \right.$$
$$\left. \text{for all } 0 \leq s \leq \frac{2(m+\varepsilon)}{\zeta}\right\}.$$

By independence and stationarity of the increments of Brownian motion, $N_b$ is a geometric random variable with parameter

$$p = P\left(A\mathbf{B}(s) \leq \boldsymbol{\varepsilon} \text{ for all } 0 \leq s \leq \frac{2(m+\varepsilon)}{\zeta}\right).$$

On the other hand, since $-\mu_i < -2\zeta$, we have:

(1) $Z_i(N_b \cdot \frac{2(m+\varepsilon)}{\zeta} + s) \leq Z_i(N_b \cdot \frac{2(m+\varepsilon)}{\zeta}) + \varepsilon$, for all $0 \leq s \leq \frac{2(m+\varepsilon)}{\zeta}$.

(2) $Z_i((N_b+1) \cdot \frac{2(m+\varepsilon)}{\zeta}) \leq Z_i(N_b \cdot \frac{2(m+\varepsilon)}{\zeta}) - m$.

Therefore, Algorithm 2.1 should execute step 4 after at most $\frac{2(m+\varepsilon)}{\zeta}(N_b+1)$ units of time in a single iteration,

$$E[T_a] \leq \frac{2(m+\varepsilon)}{\zeta} E[N_b + 1] = \frac{2(m+\varepsilon)}{\zeta}\left(1 + \frac{1}{p}\right).$$

From this inequality, it is now sufficient to show that $p = O(\varepsilon^{a_C})$.

Note that the set $C = \{\mathbf{y} \in \mathbb{R}^d : A\mathbf{y} \leq \boldsymbol{\varepsilon}\}$ forms a cone with vertex $A^{-1}\boldsymbol{\varepsilon}$ in $\mathbb{R}^d$ since $A$ is of full rank under Assumption (D). Define $\tau_C = \inf\{t \geq 0 : \mathbf{B}(t) \notin C\}$ given $\mathbf{B}(0) = 0$, then

$$p = P\left(\tau_C > \frac{2(m+\varepsilon)}{\zeta}\right).$$

If $d = 2$, it is proved by Burkholder (1977) that $a_C = \frac{\pi}{\theta}$ where $\theta \in [0, \pi)$ is the angle formed by the column vectors of $A^{-1}$. Therefore, we can compute explicitly that

$$\theta = \arccos\left(-\frac{A_{11}A_{21} + A_{12}A_{22}}{\sqrt{(A_{11}^2 + A_{12}^2)(A_{21}^2 + A_{22}^2)}}\right),$$

which only depends on $A$.

On the other hand, if $d \geq 3$, applying the results on exit times for Brownian motions given by Corollary 1.3 in DeBlassie (1987),

$$P\left(\tau_C > \frac{2(m+\varepsilon)}{\zeta}\right) \sim u \cdot \|A^{-1}\boldsymbol{\varepsilon}\|^{a_C}$$

as $\varepsilon \to 0$. Here $\|\cdot\|$ represent the Euclidian norm, and $u$ is some constant independent of $\varepsilon$. The rate $a_C$ is determined by the principal eigenvalue of the Laplace–Beltrami operator on $(\mathbf{S}^{d-1} \cap C)$, where $\mathbf{S}^{d-1}$ is a unit sphere centered at the vertex

of $C$, namely $A^{-1}\varepsilon$. The principal eigenvalue only depends on the geometric features of $C$, and it is independent of $\varepsilon$; hence so is $a_C$. Since $A$ is given, we have

$$P\left(\tau_C > \frac{2(m+\varepsilon)}{\zeta}\right) = O\left(\varepsilon^{a_C}\right) \qquad \text{as } \varepsilon \to 0.$$

Computing $a_C$ for $d \geq 3$ is not straightforward in general. However, when $A_{ij} \geq 0$, we can estimate $a_C$ from first principles. Indeed, if $A_{ij} \geq 0$ and we let $a = \max A_{ij}$, we have that

$$C = \{\mathbf{y} \in \mathbb{R}^d : A\mathbf{y} \leq \boldsymbol{\varepsilon}\} \subset \left\{\mathbf{y} \in \mathbb{R}^d : y_i \leq \frac{\varepsilon}{ad}\right\}.$$

As the coordinates of $\mathbf{B}(t)$ are independent,

$$p \geq P\left(\max_{0\leq t\leq 2(m+\varepsilon)/\zeta} B(t) \leq \frac{\varepsilon}{ad}\right)^d,$$

where $B(\cdot)$ is a standard Brownian motion on real line.

Applying the reflection principle, we have

$$\begin{aligned} &P\left(\max_{0\leq t\leq 2(m+\varepsilon)/\zeta} B(t) \leq \frac{\varepsilon}{ad}\right) \\ &\quad = \int_{-\varepsilon/(ad)}^{\varepsilon/(ad)} \frac{1}{\sqrt{2\pi(2(m+\varepsilon)/\zeta)}} \exp\left(-\frac{x^2}{2(2(m+\varepsilon)/\zeta)}\right) \\ &\quad = O(\varepsilon). \end{aligned}$$

As a result, $p = O(\varepsilon^d)$ when the correlations are all nonnegative. □

Given these propositions, we can now prove the main result in this part.

PROOF OF THEOREM 4. As we have discussed,

$$E[N(\varepsilon)] \leq (dE[K]+1)(E[\tau_\varepsilon]+1).$$

First, by Proposition 7, $E[K] = O(\varepsilon^{-2}\log(\frac{1}{\varepsilon}))$. Besides, as discussed above,

$$E[T] \leq \frac{E[T_a] + E[T_m | T_m < \infty] + E[M]}{P(T_m < \infty)P(\max_i \max_{t\geq 0} Z_i(t) < m | \mathbf{Z}(0) = 0)}.$$

According to Proposition 8, $E[T_a] = O(\varepsilon^{-a_C})$, and $a_C$ is a constant when $A$ is fixed. In the end, as we have discussed, $E[T_m | T_m < \infty]$, $P(T_m < \infty)$, $P(\max_i \max_t Z_i(t) < m | \mathbf{Z}(0) = 0)$ and $E[M]$ are independent of $\varepsilon$. Therefore,

$$E[T] = O\left(\varepsilon^{-a_C}\right).$$

In sum, we have

$$E[N(\varepsilon)] = O\left(\varepsilon^{-a_C-2}\log\left(\frac{1}{\varepsilon}\right)\right).$$

□

TABLE 1
*Unbiased estimates of $E[Y_i(\infty)]$ and $E[Y_i^2(\infty)]$ for a network with ten stations in tandem*

| | $E[Y_i(\infty)]$ | | $E[Y_i^2(\infty)]$ | |
|---|---|---|---|---|
| **Station** | **Simulation result** | **True value** | **Simulation result** | **True value** |
| 1 | $1.7919 \pm 0.0521$ | 1.8182 | $10.2755 \pm 0.5289$ | 10.2479 |
| 2 | $0.1761 \pm 0.0068$ | 0.1818 | $0.1511 \pm 0.0170$ | 0.1642 |
| 3 | $0.2171 \pm 0.0083$ | 0.2222 | $0.2242 \pm 0.0224$ | 0.2382 |
| 4 | $0.2706 \pm 0.0102$ | 0.2778 | $0.3462 \pm 0.0339$ | 0.3610 |
| 5 | $0.3516 \pm 0.0131$ | 0.3571 | $0.5717 \pm 0.0590$ | 0.5778 |
| 6 | $0.4737 \pm 0.0171$ | 0.4762 | $0.9840 \pm 0.0871$ | 0.9921 |
| 7 | $0.6632 \pm 0.0233$ | 0.6667 | $1.8472 \pm 0.1513$ | 1.8715 |
| 8 | $1.0033 \pm 0.0345$ | 1.0000 | $4.1004 \pm 0.3377$ | 4.0300 |
| 9 | $1.6497 \pm 0.0542$ | 1.6667 | $10.3734 \pm 0.7823$ | 10.6065 |
| 10 | $3.3200 \pm 0.1040$ | 3.3333 | $39.2015 \pm 2.9950$ | 39.3631 |

**5. Numerical results.** We first implemented Algorithm 1 in order to generate exact samples from the steady-state distribution of stochastic fluid networks, and then we implemented Algorithm 2. Our implementations were performed in Matlab. In all the experiments we simulated 10,000 independent replications, and we displayed our estimates with a margin of error obtained using a 95% confidence interval based on the central limit theorem.

For the case of stochastic fluid networks, we considered a 10-station system in tandem. So, $Q_{i,i+1} = 1$ for $i = 1, 2, \ldots, 9$ and $Q_{10,j} = 0$ for all $j = 1, \ldots, 10$. We assume the arrival rate $\lambda = 1$ and the job sizes are exponentially distributed with unit mean. The service rates $(\mu_1, \ldots, \mu_{10})^T$ are given by $(1.55, 1.5, 1.45, 1.4, 1.35, 1.3, 1.25, 1.2, 1.15, 1.1)$. We are interested in computing the steady-state mean and the second moment of the workload at each station (i.e., $E[Y_i(\infty)]$ and $E[Y_i(\infty)^2]$ for $i = 1, 2, \ldots, 10$). For a network of this type, it turns out that the true values of the quantities we are interested in can be computed from the corresponding Laplace transforms as given in Debicki, Dieker and Rolski (2007).

Both the simulation results and the true values are reported in Table 1. The procedure took a few minutes (less than 5) on a desktop, which is quite a reasonable time.

We then implemented a two-dimensional RBM example. Let us denote the RBM by $\mathbf{Y}(t)$. The parameters to specify $\mathbf{Y}$ are as follows: drift vector $\mu = (-1, -1)$, covariance matrix $\Sigma = [1, 0; 0, 1]$ and reflection matrix $R = [1, -0.2; -0.2, 1]$. For this so-call symmetric RBM, one could compute in close that $E[Y_1(\infty)] = E[Y_2(\infty)] = 5/12 \simeq 0.4167$; see, for instance, Dai and Harrison (1992). The output of our simulation algorithm is reported in Table 2.

Our implementations here are given with the objective of verifying empirically the validity of the algorithms proposed. We stress that a direct implementation of

TABLE 2
*Estimates of $E[Y_i(\infty)]$ for a 2-dimensional RBM with precision $\varepsilon = 0.01$*

| | **Simulation result** | **True value** |
|---|---|---|
| $E[Y_1(\infty)]$ | $0.4164 \pm 0.0137$ | 0.4167 |
| $E[Y_2(\infty)]$ | $0.4201 \pm 0.0131$ | 0.4167 |

Algorithm 2, although capable of ultimately producing unbiased estimations of the expectations of RBM, might not be practical. The simulations took substantially more time to be produced than those reported for the stochastic fluid models. This can be explained by the dependence on $\varepsilon$ in Theorem 4. The bottleneck in the algorithm is finding a time at which both stations are close to $\varepsilon$. An efficient algorithm based on suitably trading a strongly controlled bias with variance can be used to produce faster running times; we expect to report this algorithm in the future.

**Acknowledgments.** The authors thank Offer Kella for pointing out Lemma 1 and thank Amy Biemiller for her editorial assistance. The authors thank the Editor and referees for their useful comments and suggestions.

## REFERENCES

ASMUSSEN, S. (2003). *Applied Probability and Queues*: *Stochastic Modelling and Applied Probability*, 2nd ed. *Applications of Mathematics* (*New York*) **51**. Springer, New York. MR1978607

ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation*: *Algorithms and Analysis*. *Stochastic Modelling and Applied Probability* **57**. Springer, New York. MR2331321

ASMUSSEN, S., GLYNN, P. and PITMAN, J. (1995). Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.* **5** 875–896. MR1384357

BESKOS, A., PELUCHETTI, S. and ROBERTS, G. (2012). $\varepsilon$-strong simulation of the Brownian path. *Bernoulli* **18** 1223–1248. MR2995793

BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

BLANCHET, J. H. and SIGMAN, K. (2011). On exact sampling of stochastic perpetuities. *J. Appl. Probab.* **48A** 165–182. MR2865624

BUDHIRAJA, A. and LEE, C. (2009). Stationary distribution convergence for generalized Jackson networks in heavy traffic. *Math. Oper. Res.* **34** 45–56. MR2542988

BURDZY, K. and CHEN, Z.-Q. (2008). Discrete approximations to reflected Brownian motion. *Ann. Probab.* **36** 698–727. MR2393994

BURKHOLDER, D. L. (1977). Exit times of Brownian motion, harmonic majorization, and Hardy spaces. *Adv. Math.* **26** 182–205. MR0474525

DAI, J. G. and DIEKER, A. B. (2011). Nonnegativity of solutions to the basic adjoint relationship for some diffusion processes. *Queueing Syst.* **68** 295–303. MR2834200

DAI, J. G. and HARRISON, J. M. (1992). Reflected Brownian motion in an orthant: Numerical methods for steady-state analysis. *Ann. Appl. Probab.* **2** 65–86. MR1143393

DEBLASSIE, R. D. (1987). Exit times from cones in $\mathbf{R}^n$ of Brownian motion. *Probab. Theory Related Fields* **74** 1–29. MR0863716

DEVROYE, L. (2009). On exact simulation algorithms for some distributions related to Jacobi theta functions. *Statist. Probab. Lett.* **79** 2251–2259. MR2591982

DEBICKI, K., DIEKER, A. B. and ROLSKI, T. (2007). Quasi-product forms for Lévy-driven fluid networks. *Math. Oper. Res.* **32** 629–647. MR2348239

ENSOR, K. B. and GLYNN, P. W. (2000). Simulating the maximum of a random walk. *J. Statist. Plann. Inference* **85** 127–135. MR1759244

GAMARNIK, D. and ZEEVI, A. (2006). Validity of heavy traffic steady-state approximation in generalized Jackson networks. *Ann. Appl. Probab.* **16** 56–90. MR2209336

GUT, A. (2009). *Stopped Random Walks*: *Limit Theorems and Applications*, 2nd ed. Springer, New York. MR2489436

HARRISON, J. M. and REIMAN, M. I. (1981). Reflected Brownian motion on an orthant. *Ann. Probab.* **9** 302–308. MR0606992

HARRISON, J. M. and WILLIAMS, R. J. (1987). Brownian models of open queueing networks with homogeneous customer populations. *Stochastics* **22** 77–115. MR0912049

KELLA, O. (1996). Stability and nonproduct form of stochastic fluid networks with Lévy inputs. *Ann. Appl. Probab.* **6** 186–199. MR1389836

KELLA, O. and RAMASUBRAMANIAN, S. (2012). Asymptotic irrelevance of initial conditions for Skorohod reflection mapping on the nonnegative orthant. *Math. Oper. Res.* **37** 301–312. MR2931282

KELLA, O. and WHITT, W. (1996). Stability and structural properties of stochastic storage networks. *J. Appl. Probab.* **33** 1169–1180. MR1416235

KENDALL, W. S. (2004). Geometric ergodicity and perfect simulation. *Electron. Commun. Probab.* **9** 140–151 (electronic). MR2108860

PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252. MR1611693

RAMASUBRAMANIAN, S. (2000). A subsidy-surplus model and the Skorokhod problem in an orthant. *Math. Oper. Res.* **25** 509–538. MR1855180

REIMAN, M. I. (1984). Open queueing networks in heavy traffic. *Math. Oper. Res.* **9** 441–458. MR0757317

STEELE, J. M. (2001). *Stochastic Calculus and Financial Applications*. *Applications of Mathematics* (*New York*) **45**. Springer, New York. MR1783083

VARADHAN, S. R. S. and WILLIAMS, R. J. (1985). Brownian motion in a wedge with oblique reflection. *Comm. Pure Appl. Math.* **38** 405–443. MR0792398

INDUSTRIAL ENGINEERING
AND OPERATIONS RESEARCH
COLUMBIA UNIVERSITY
340 S. W. MUDD BUILDING
500 W. 120 STREET
NEW YORK, NEW YORK 10027
USA
E-MAIL: jose.blanchet@columbia.edu

DEPARTMENT OF APPLIED MATHEMATICS
AND STATISTICS
STONY BROOK UNIVERSITY
MATH TOWER B148
STONY BROOK, NEW YORK 11794-3600
USA
E-MAIL: xinyun.chen@stonybrook.edu