stochastic processes and their applications

Stochastic Processes and their Applications 98 (2002) 253–287

www.elsevier.com/locate/spa

A time-reversed representation for the tail probabilities of stationary reflected Brownian motion

Paul Dupuis[a,1], Kavita Ramanan[b,*]

[a]*Lefschetz Center for Dynamical Systems, Brown University, Providence, RI 02912, USA*
[b]*Room 2C-319, Bell Labs, 600 Mountain Avenue, Murray Hill, NJ 07974, USA*

Received 25 February 2001; received in revised form 14 August 2001; accepted 19 August 2001

Abstract

We consider the exponential decay rate of the stationary tail probabilities of reflected Brownian motion X in the N-dimensional orthant $\mathbb{R}_+^N$ having drift b, covariance matrix A, and constraint matrix D. Suppose that the Skorokhod or reflection mapping associated with the matrix D is well-defined and Lipschitz continuous on the space of continuous functions. Under the stability condition $D^{-1}b < 0$, it is known that the exponential decay rate has a variational representation $V(x)$. This representation is difficult to analyze, in part because there is no analytical theory associated with it. In this paper, we obtain a new representation for $V(x)$ in terms of a time-reversed optimal control problem. Specifically, we show that $V(x)$ is equal to the minimum cost incurred to reach the origin when starting at the point x, where the constrained dynamics are described in terms of another constraint matrix $\bar{D}$, and the cost is quadratic in the control as well as the "local time" or constraining term. The equivalence of these representations in fact holds under the milder assumption that the matrices D and $\bar{D}$ satisfy what is known as the completely-$\mathscr{S}$ condition. We then use the time-reversed representation to identify the minimizing large deviation trajectories for a class of RBMs having product form distributions. In particular, we show that the large deviation trajectories associated with product form RBMs that approximate open single-class networks or multi-class feedforward networks do not cycle. © 2001 Published by Elsevier Science B.V.

MSC: primary 60F10; 60J65; secondary 49L99.

* Corresponding author.
E-mail address: kavita@research.bell-labs.com (K. Ramanan).

[1] Research supported in part by the National Science Foundation (NSF-DMS-0072004) and the Army Research Office (ARO-DAAD19-99-1-0223) and (ARO-DAAD19-00-1-0549).

PII: S0304-4149(01)00151-X

Keywords: Reflected Brownian motion; Large deviations; Rate function; Stationary distribution; Tail probabilities; Cyclic trajectories; Skew symmetry; Product form; Skorokhod Map; Skorokhod Problem; Optimal control; Time-reversal; Variational problem; Single-class networks; Feedforward networks; Buffer overflow

1. Introduction

1.1. Background

A (b, A, D) reflected Brownian motion (RBM) X in the N-dimensional orthant $\mathbb{R}_+^N$ behaves like an N-dimensional Brownian motion with constant drift b and covariance matrix A in the interior of the orthant, and is instantaneously constrained at the boundary, being pushed along the direction d_j (the jth column of D) when on the relative interior of the face $\{x \in \mathbb{R}_+^N : x_j = 0\}$. Apart from being of general theoretical interest, the study of RBMs in the orthant has also been motivated by the fact that they arise frequently as diffusion or heavy traffic approximations to queueing networks (see Kushner, 2001; Williams, 1996). For example, heavy traffic approximations of single-class open queueing networks can be characterized by RBMs whose constraint matrices satisfy what we refer to as the Harrison–Reiman condition, which is Condition 2.2 below (Reiman, 1984), and likewise multi-class feedforward networks can be approximated by RBMs whose constraint matrices satisfy Condition 2.3 (Peterson, 1991). Stationary distributions of these approximating RBMs have been the focus of much work because they serve as estimates for the stationary measures of the corresponding queueing networks, and are generally easier to compute. Several authors have obtained explicit analytical expressions for the stationary distributions of RBMs in two dimensions (Harrison et al., 1985; Williams, 1985). However, in higher dimensions the only general result available is for RBMs whose stationary densities have an exponential product form. More precisely, Harrison and Williams showed that a certain skew symmetry condition on A and D was necessary and sufficient for any (b, A, D) RBM with a completely-$\mathscr{S}$ matrix D satisfying $D^{-1}b < 0$ to have a product form stationary distribution (Harrison and Williams, 1987a, Theorem 9.2; Dai and Harrison, 1992, Proposition 9). Some product form RBMs serve as tractable approximations to certain queueing networks with rather complicated non-product form stationary distributions. In general, however, explicit solutions for the stationary distributions of even the approximating RBMs are hard to obtain, and one often has to resort to numerical approximations. Dai and Harrison developed one such approximation algorithm and demonstrated its efficacy in solving certain examples (Dai and Harrison, 1992). However, in general the convergence of their approximations depends crucially on a certain reference density which is hard to choose appropriately without any a priori knowledge of the stationary distribution of the RBM.

In this context, the exponential decay rate of the stationary distribution of an RBM is of interest for two reasons. Firstly, a knowledge of the tail behavior of the distribution can lead to better convergence properties for the numerical algorithms proposed in Dai and Harrison (1992). Secondly, since heavy traffic and large deviation limits are often interchangeable (the proof for open feedforward networks and single-class queueing networks can be found in Majewski, 1998a, 2000), the large deviation behavior of the RBM approximates the large deviation behavior of the associated queueing network. The latter is of interest because it characterizes the probability of crucial rare events such as buffer overflow or large delays in the network.

In this paper, we analyze the function $V(\cdot)$ defined below by the variational problem (1). As shown in Section 2.2, $V(\cdot)$ characterizes the exponential decay rate of the tails of the stationary distribution μ of a (b,A,D) RBM under suitable conditions on the data (b,A,D). Roughly speaking, for Borel subsets $B \subset \mathbb{R}^N_+$ that are sufficiently regular (e.g., B is the closure of its interior),

$$-\frac{1}{n}\log \mu(nB) \overset{n\to\infty}{\to} V(B) \quad \text{where } V(B) \doteq \inf_{x\in B} V(x).$$

Let $\mathscr{AC}([0,\infty):\mathbb{R}^N)$ be the space of absolutely continuous functions on $[0,\infty)$ taking values in $\mathbb{R}^N$. Define

$$V(x) \doteq \inf_{\psi\in\mathscr{AC}([0,\infty):\mathbb{R}^N):\psi(0)=0}\ \inf_{\phi\in\Gamma(\psi):\tau_x<\infty} \int_0^{\tau_x} L(\dot\psi(s))\,\mathrm{d}s, \tag{1}$$

where

$$L(\beta) \doteq \tfrac{1}{2}(\beta-b)'A^{-1}(\beta-b), \tag{2}$$

$$\tau_x \doteq \inf\{t \geqslant 0 : \phi(t)=x\}, \tag{3}$$

and where $\Gamma(\psi)$ is the set of images of ψ under the Skorokhod Map (SM) that is associated with D. The Skorokhod Map (which is also referred to as the reflection map in this context) is a key tool in the construction and analysis of RBM. The definition of the SM and some of its main properties are summarized in Section 2.1.

In two dimensions, the simplicity of the geometry allows the variational problem $V(\cdot)$ and the associated minimizing trajectories to be calculated explicitly (Avram et al., 2000; Kieffer, 1995; O'Connell, 1998; Paschalidis, 1996). In dimensions greater than two, one has to allow for the possibility that minimizing trajectories may cycle, i.e., the trajectories may have a decomposition into an infinite number of disjoint pieces that are scaled versions of each other. As a consequence, relatively few explicit expressions or even finite-dimensional representations for $V(x)$ have been derived in higher dimensions. One major difficulty is that although the variational problem (1) can be viewed as an optimal control problem, it cannot be analyzed using classical control theoretical methods (such as dynamic programming techniques). This is primarily due to the fact that the independent variable x of the function $V(x)$ serves as the terminal condition for the trajectories (see (3)), as opposed to the initial condition. In the classical setting with unconstrained dynamics, this problem is usually overcome by reversing the time axis. However, as explained below, this is no longer as straightforward when the dynamics are constrained. The SM, which describes the constrained dynamics, can be defined on a pathwise basis by solving what is known as the Skorokhod Problem (SP) for each input trajectory ψ. For absolutely continuous inputs ψ, one can reformulate the SP in terms of a constrained ODE of the form $\dot\phi=\pi(\phi,\dot\psi)$, $\phi(0)=\psi(0)$, where

$\pi(x,\beta)$ is defined in terms of D, and where the map $x \to \pi(x,\beta)$ is discontinuous. In spite of this discontinuity, under appropriate conditions one can show that this ODE admits a decent qualitative theory (i.e., existence and uniqueness of solutions, etc.) if one is solving forward in time (Dupuis and Ishii, 1991). However, in strong contrast to the classical theory for ODEs under a Lipschitz continuity condition, there is non-uniqueness if one attempts to solve the constrained ODE backward in time. This makes the development of dynamic programming principles difficult. Numerical approximations of the rate function associated with a class of RBMs have been proposed by Majewski (1998b), but these suffer from a combinatorial explosion with increase in dimension, and do not identify certain qualitative properties of the minimizing paths such as cyclicity.

1.2. Main results

Our main result (Theorem 3.6) is that $V(x)$ is equal to the value function $\bar{V}(x)$ of a related optimal control problem [see (17)] that is more amenable to analysis. Roughly speaking, $\bar{V}(x)$ is the minimum cost to reach the origin for any trajectory ϕ that starts at x, where the dynamics are specified by a SM associated with a "time-reversed" constraint matrix $\bar{D}$ (specified in terms of A and D) and a cost that is a quadratic function of the control $\dot{\psi}$ as well as of the local time $\dot{\eta} = \dot{\phi} - \dot{\psi}$. The proof uses time-reversal arguments and also establishes a one-to-one correspondence between the minimizing trajectories of the two problems.

In the cases we have examined in detail, the minimizing trajectories for the time-reversed problem have a simpler form than those of the original variational problem. For example, we show in Section 4 that for product form RBMs the minimizing trajectory for the time-reversed problem is associated with a constant control or input velocity $\dot{\psi}$, whereas in general the optimal control for the original variational problem (1) is only piecewise constant. Moreover, using this property we show that the large deviation trajectories associated with product form RBMs that arise as diffusion approximations to open single-class networks or multi-class feedforward networks do not exhibit cycling behavior.

The outline of the paper is as follows. In Section 2 we introduce the background theory related to the Skorokhod Problem and large deviations that is necessary to establish that under suitable assumptions the function $V(x)$ with the variational representation (1) characterizes the exponential decay rate of the tails of a stationary (b,A,D) RBM. In Section 3 we introduce the time-reversed optimal control problem and the variational representation for its value function $\bar{V}(x)$. The main result $V=\bar{V}$ is established in Theorem 3.6 using pathwise time-reversal arguments under the assumption that D and $\bar{D}$ are completely-$\mathscr{S}$. In Section 4 we specialize to the case of RBMs with product form stationary distributions. We identify the large deviation minimizing trajectories and show that for the subclass of RBMs whose constraint matrices D are either of Harrison–Reiman type, or satisfy Condition 2.3, these trajectories do not cycle. For the special case of three dimensions, we also show that the minimizing large deviation trajectories of product form RBMs having generalized Harrison–Reiman constraint matrices do not cycle. We illustrate the implications of these results for determining the most likely way in which buffers overflow in queueing networks via a three-dimensional example in Section 5.

1.3. Notation

Some common notation used throughout the paper is as follows. We use capital letters to denote matrices, and lower case letters to denote vectors or scalars. Inequalities involving vectors and matrices will be understood componentwise. Vectors are understood to be column vectors, and the ith component of a vector v is denoted $(v)_i$, or when there is no ambiguity, simply by v_i. Given a matrix D, D_{ij} denotes the entry in the ith row and jth column and d_i denotes the ith column vector of D. Note that by this convention, $D_{ij} = (d_j)_i$. Moreover, we use D' to denote the transpose of the matrix D. I represents the $N \times N$ identity matrix, $\mathbb{R}^N_+$ is used to denote the N-dimensional orthant $\bigcap_{i=1,\dots,N} \{x : x_i \geqslant 0\}$, and $\{e_i,\ 1 = 1, \dots, N\}$ is the standard orthonormal basis in $\mathbb{R}^N$.

Given any subset $A \subset \mathbb{R}^N$, we let A° denote its interior and ∂A its boundary. For $E = \mathbb{R}^N$ or $\mathbb{R}^N_+$ we use $\mathscr{C}([0,\infty) : E)$ (respectively, $\mathscr{C}_+([0,\infty) : E)$) to denote the set of continuous functions f on $[0,\infty)$ taking values in E whose initial value $f(0)$ lies in $\mathbb{R}^N$ (respectively, $\mathbb{R}^N_+$). We define $\mathscr{I}([0,\infty) : E)$, $\mathscr{I}_+([0,\infty) : E)$ to be the analogous sets of (componentwise) non-decreasing continuous functions and likewise let $\mathscr{AC}([0,\infty) : E)$ and $\mathscr{AC}_+([0,\infty) : E)$ be the corresponding sets of absolutely continuous functions. We use $\imath$ to represent the identity mapping $\imath : \mathbb{R}_+ \to \mathbb{R}_+$ so that $\imath(t) = t$ for every $t \geqslant 0$. In general, quantities with a bar on them (like $\bar{\phi}$) are used to refer to the time-reversed variational problem $\bar{V}$ defined by (17), while plain quantities without bars (like ϕ) refer to the original variational problem (1). Starred quantities (like $\bar{\psi}^*$ or ψ^*) are used to represent optimal or minimizing trajectories for the respective variational problems.

2. Large deviations for stationary reflected Brownian motion

In this section we formulate the well-known variational representation (1) for the decay rate of the tails of the stationary distribution μ of a (b, A, D) RBM X. We first show that the decay rate can equivalently be represented as the large deviation rate function for the sequence $\{\tilde{\mu}_n\}$, where $\tilde{\mu}_n$ is the stationary distribution of a $(b, A/n, D)$ RBM, and then use standard results to show that the latter has the representation (1). In Section 2.1 we define the Skorokhod Problem, the associated Skorokhod Map and reflected Brownian motion (RBM). In Section 2.2 we show that the decay rate has the variational representation $V(x)$ given in (1).

2.1. The Skorokhod Map

Let a constraint matrix D be given, and recall that d_j is the jth column of D. For each point x on the boundary of $\mathbb{R}^N_+$, let $I(x) \doteq \{i : x_i = 0\}$. We introduce the set-valued function

$$d(x) \doteq \left\{ \sum_{i \in I(x)} a_i d_i : a_i \geqslant 0, \left\| \sum_{i \in I(x)} a_i d_i \right\| = 1 \right\}$$

that describes the set of directions of constraint allowed at each point $x \in \partial \mathbb{R}_+^N$. The Skorokhod Problem (SP) assigns to every path $\psi \in \mathscr{C}_+([0,\infty):\mathbb{R}^N)$ a path $\phi \in \mathscr{C}_+([0,\infty):\mathbb{R}_+^N)$ that starts at $\phi(0) = \psi(0)$, but is constrained to $\mathbb{R}_+^N$ as follows. If ϕ is in the interior of $\mathbb{R}_+^N$ then the evolution of ϕ mimics that of ψ, in that the increments of the two functions are the same until ϕ hits the boundary of $\mathbb{R}_+^N$. When ϕ is on the boundary a constraining "force" is applied to keep ϕ in the domain, and this force can only be applied in one of the directions in $d(\phi(t))$, and only for t such that $\phi(t)$ is on the boundary. The precise definition is as follows. For $\eta \in \mathscr{C}([0,\infty):\mathbb{R}^N)$ and $t \in [0,\infty)$ we let $|\eta|(t)$ denote the total variation of η on $[0,t]$ with respect to the Euclidean norm on $\mathbb{R}^N$.

Definition 2.1 (*Skorokhod Problem*). Let $\psi \in \mathscr{C}_+([0,\infty):\mathbb{R}^N)$ be given. Then (ϕ,η) solves the SP for ψ (with respect to $\mathbb{R}_+^N$ and the constraint matrix D) if $\phi(0) = \psi(0)$, and if for all $t \in [0,\infty)$
1. $\phi(t) = \psi(t) + \eta(t)$,
2. $\phi(t) \in \mathbb{R}_+^N$,
3. $|\eta|(t) < \infty$,
4. $|\eta|(t) = \int_{[0,t]} 1_{\{\phi(s) \in \partial \mathbb{R}_+^N\}} d|\eta|(s)$,
5. There exists a Borel measurable function $\gamma : [0,\infty) \to \mathbb{R}^N$ such that $d|\eta|$-almost everywhere $\gamma(t) \in d(\phi(t))$, and such that

$$\eta(t) = \int_{[0,t]} \gamma(s) d|\eta|(s).$$

Note that η changes only when ϕ is on the boundary, and only in the directions $d(\phi)$. We let $\Gamma(\psi)$ denote the set of ϕ such that $(\phi, \phi - \psi)$ solve the SP for ψ and refer to the (in general multi-valued) mapping Γ as the Skorokhod Map (SM). Note that if Γ is single-valued (e.g. when Γ is Lipschitz continuous), then with some abuse of notation we will write $\phi = \Gamma(\psi)$ if $(\phi, \phi - \psi)$ solve the SP for ψ.

Remark. In the definition of the SP given above, we have specialized to the case where the domain is $\mathbb{R}_+^N$ and there are N directions of constraint. (For a more general definition, see Dupuis and Ishii, 1991; Dupuis and Ramanan, 1999a.) In this setting the SP and SM are often referred to as the dynamic complementarity problem and the reflection map, respectively (Harrison and Reiman, 1981; Mandelbaum and Van der Heyden, 1987). Moreover, η is often represented as $D\theta$, where $\theta \in \mathscr{I}([0,\infty):\mathbb{R}^N)$ (Harrison and Reiman, 1981). We refer to θ as the local time.

A basic assumption we make on the constraint matrix D is that it satisfies the completely-$\mathscr{S}$ condition stated below. When D is completely-$\mathscr{S}$, its diagonal elements must be positive, and thus without loss of generality we assume throughout the paper that $D_{ii} = 1$ for $i = 1, \ldots, N$. Note that this means $d_i' e_i = 1$ for $i = 1, \ldots, N$.

Definition 2.2 (*Completely-$\mathscr{S}$*). $D \in \mathbb{R}^{N \times N}$ is said to be completely-$\mathscr{S}$ if for each principal matrix $\tilde{D}$ of D there is $y \geqslant 0$ such that $\tilde{D} y > 0$, where the inequalities are interpreted componentwise.

It is well-known that the SM Γ is well-defined on all of $\mathscr{C}_+([0,\infty):\mathbb{R}^N)$ if and only if the associated constraint matrix D is completely-$\mathscr{S}$ (Bernard and El Kharroubi, 1991). We now consider some additional regularity conditions on the constraint matrix D.

Condition 2.1 (*Regular SP*). *$D \in \mathbb{R}^{N \times N}$ is invertible and the associated Skorokhod Map Γ is Lipschitz continuous* (*with respect to the topology of uniform convergence on compact sets*) *and is defined for every $\psi \in \mathscr{C}_+([0,\infty):\mathbb{R}^N)$.*

Observe that D must in particular be completely-$\mathscr{S}$ in order to satisfy Condition 2.1. General assumptions that ensure Condition 2.1 can be found in Dupuis and Ishii (1991) and Dupuis and Ramanan (1999a,b). Below we provide three sufficient conditions under which D satisfies Condition 2.1 (Dupuis and Ramanan, 1999b; Harrison and Reiman, 1981). Note, however, that these are not necessary conditions for Condition 2.1 to hold (Dupuis and Ramanan, 1999b, Section 2.4). Recall that I is the $N \times N$ identity matrix.

Condition 2.2 (*Harrison–Reiman*). *$D \in \mathbb{R}^{N \times N}$ has the form $I - V$, where V is a non-negative off-diagonal matrix with spectral radius less than one.*

Condition 2.3 (*Multiclass feedforward*). *$D \in \mathbb{R}^{N \times N}$ has the form $I - V$, where V is upper triangular ($v_{ij} = 0$ if $i \leqslant j$) or lower triangular ($v_{ij} = 0$ if $i \geqslant j$).*

Condition 2.4 (*Generalized Harrison–Reiman*). *$D \in \mathbb{R}^{N \times N}$ has the form $I - V$, where $|V|$ is an off-diagonal matrix with spectral radius less than one.*

Note that Condition 2.4 includes Conditions 2.2 and 2.3 as special cases. When D has the regularity properties dictated by Condition 2.1, the associated SM serves as an extremely convenient tool for the pathwise construction of reflected Brownian motion. We use $\imath$ to denote the identity map. Since the only RBMs we consider admit a pathwise construction, we will take the following as our definition of RBM.

Definition 2.3 (*Reflected Brownian motion*). Given a drift vector $b \in \mathbb{R}^N$, covariance matrix $A = \sigma\sigma' > 0$, N-dimensional standard Brownian motion W, domain $\mathbb{R}_+^N$ and constraint matrix D satisfying Condition 2.1, let Γ be the associated Skorokhod Map. Then

$$X \doteq \Gamma(b\imath + \sigma W)$$

is a (b, A, D) RBM.

2.2. The large deviation rate function for stationary RBM

We now characterize the large deviation rate function for a sequence of scaled stationary RBMs in terms of the variational problem (1) for $V(x)$. We first recall the definition of a large deviation principle (LDP) (Varadhan, 1984, Definition 2.1).

Definition 2.4 (*Large deviation principle*). A sequence of probability measures $\{\mu_n\}$ defined on a complete separable metric space $(\mathscr{X},\mathscr{B})$ is said to satisfy the LDP with rate function J if for all $\Theta\in\mathscr{B}$,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mu_n(\Theta)\leqslant-\inf_{x\in\bar{\Theta}}J(x),$$

and

$$\liminf_{n\to\infty}\frac{1}{n}\log\mu_n(\Theta)\geqslant-\inf_{x\in\Theta^\circ}J(x),$$

where $J:\mathscr{X}\to\mathbb{R}\cup\{\infty\}$ is a function with compact level sets. A sequence of random variables $\{X_n\}$ defined on some measure space taking values in a complete separable metric space $(\mathscr{X},\mathscr{B})$ is said to satisfy an LDP with rate function $J(\cdot)$ if the corresponding induced measures satisfy a LDP with the same rate function.

Given a constraint matrix D, define

$$\mathscr{C}\doteq-d(0)=\left\{-\sum\alpha_i d_i:\alpha_i\geqslant 0\right\},$$

where as before d_i is the ith column of the constraint matrix D. We now state a stability condition that ensures existence of the invariant distribution for a (b,A,D) RBM.

Condition 2.5 (*Stability*). $b\in\mathscr{C}^\circ$.

Note that when D is invertible, Condition 2.5 is equivalent to the inequality $D^{-1}b<0$. When D satisfies Condition 2.1 it was shown in Budhiraja and Dupuis (1999) that Condition 2.5 is necessary and sufficient for the trajectory $\Gamma(x+b\imath)(t)$ to reach the origin in finite time for any $x\in\mathbb{R}_+^N$, and for the associated (b,A,D) RBM X to be positive recurrent. (Recall that $\imath:\mathbb{R}_+\to\mathbb{R}_+$ is the identity mapping.) For the case when the constraint matrix satisfies the Harrison–Reiman condition (Condition 2.2), the fact that $b\in\mathscr{C}^\circ$ is necessary and sufficient for the positive recurrence of RBM was first established in Harrison and Williams (1987a, Section 6). Note that if the RBM is positive recurrent it has an invariant distribution.

Let μ be the invariant distribution of the (b,A,D) RBM. We are interested in the large deviation principle for the sequence $\{\mu_n\}$ defined by $\mu_n(B)=\mu(nB)$, and will use the fact that $\mu_n=\tilde{\mu}_n$, where $\tilde{\mu}_n$ is the invariant distribution for the $(b,A/n,D)$ RBM. This will allow us to appeal to existing results in the large deviation literature, and in particular to the classical results of Freidlin and Wentzell on large deviations for the invariant distributions of small noise Markov processes.

Theorem 2.5. *Suppose the* (b,A,D) *RBM* X *on the orthant* $\mathbb{R}^N_+$ *is such that* $A=\sigma\sigma'$ *is positive definite and Conditions* 2.1 *and* 2.5 *are satisfied. Let* Γ *be the associated Skorokhod Map. Then* $\{\mu_n\}$ *satisfies the LDP with rate function* $V(x)$ *given by* (1).

Proof. Consider the RBM $X_n \doteq \Gamma(b\iota + \sigma W/\sqrt{n})$. X_n is well-defined by Condition 2.1, and possesses a unique invariant distribution $\tilde{\mu}_n$ under Condition 2.5. Fix any $T\in(0,\infty)$ and $x\in\mathbb{R}^N_+$. We also consider the family of RBMs with starting position x defined by $X^x_n \doteq \Gamma(x + b\iota + \sigma W/\sqrt{n})$. Schilder's theorem (which states the large deviation principle for Brownian motion with drift b and covariance A/n, (Varadhan, 1984, p. 19)), the contraction mapping theorem (Varadhan, 1984, Remark 1), and the continuity of Γ imply that $\{X^x_n\}$ satisfies a large deviation principle on $[0,T]$ with rate function J^x_T, where

$$J^x_T(\phi) \doteq \inf_{\psi\in\mathscr{AC}([0,\infty):\mathbb{R}^N):\psi(0)=x,\phi=\Gamma(\psi)} \frac{1}{2}\int_0^T L(\dot{\psi}(t))\,\mathrm{d}t,$$

and where L is given by (2). Straightforward extensions of standard arguments from Freidlin–Wentzell theory (Freidlin and Wentzell, 1984, Theorem 4.3, Chapter 4) can then be used to establish the representation (1) of the large deviation rate function for the sequence of invariant measures $\{\tilde{\mu}_n\}$. We then use the relation $\mu_n=\tilde{\mu}_n$ to deduce that $\{\mu_n\}$ satisfies the LDP with the same rate function $V(\cdot)$. (Also see Majewski, 1998a,b, for an alternate derivation of the variational representation for the particular case when D satisfies either Condition 2.2 or 2.3.) □

3. An associated control problem

Theorem 2.5 established that under suitable assumptions the exponential decay rate of the stationary distribution of reflected Brownian motion can be characterized as the minimum cost $V(x)$ (expressed as a quadratic function $L(\cdot)$ of the control $\dot{\psi}$) till $\phi(t)=\Gamma(\psi)(t)$ reaches x. In this section we establish that $V(x)$ also has a representation in terms of another "time-reversed" optimal control problem that appears more amenable to analysis. More precisely, we show in Theorem 3.6 that there exists another constraint matrix $\bar{D}$ such that $V(x)$ can be represented as a minimum cost problem involving trajectories that start at x and accumulate cost till they reach the origin. Under the assumption that the constraint matrices D and $\bar{D}$ are completely-$\mathscr{S}$, the representation takes the same form as (1), save that the constraints on the endpoints are reversed, and the running cost is quadratic in the local time as well as the control. The new constraint matrix and costs are chosen in such a way that the least cost incurred to produce a trajectory ϕ (going from 0 to x) in the original variational problem is equal to the least cost incurred to realize the same trajectory traced backward in time (going from x to 0) in the time-reversed optimal control problem. In fact, as established in Lemma 3.1 and explained in detail below, this equivalence actually holds at a more local level. In the absence of boundaries, it is easy to see how the cost structure should be changed in order to achieve this. However this problem is more subtle when boundaries are present.

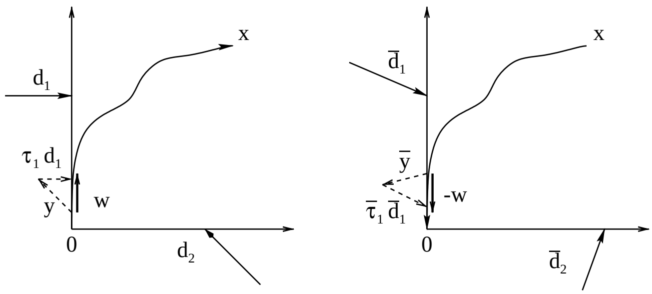

Fig. 1. The time-reversed cost structure.

Our objective is to equate the cost of moving an infinitesimal amount in a certain direction w away from a point in the original problem with the cost of moving an infinitesimal amount away from the same point in the opposite direction $-w$ for the time-reversed problem. When the point is in the interior of the domain, by definition of the SP the controls y and $\bar{y}$ are equal to the velocities w and $-w$ produced in the forward and time-reversed variational problems, respectively. Thus it is clear that for the costs to be equal, the definition of the cost function in the time-reversed problem should be $\bar{L}(y) \doteq L(-y)$ (see (4) below), regardless of the time-reversed constraint matrix $\bar{D}$. Now suppose instead that the point is on the boundary, say in the relative interior of the face $\{z \in \mathbb{R}^N_+ : z_1 = 0\}$, and that the velocity moves along the boundary so that $w_1 = 0$ (see Fig. 1). Then the cost to move along w in the forward problem is the infimum of $L(y)$ over all "controls" y that map to w under the SM associated with D. Fix y and note that from the definition of the SM it must have a representation of the form $y = w - \tau_1 d_1$ for some $\tau_1 \geqslant 0$. Likewise, any velocity $\bar{y}$ used in the time-reversed problem to achieve an infinitesimal displacement in the direction $-w$ must have a representation $-w - \bar{\tau}_1 \bar{d}_1$ for some $\bar{\tau}_1 \geqslant 0$. However, since $w - \tau_1 d_1 \neq w + \bar{\tau}_1 \bar{d}_1$ unless $\tau_1 = \bar{\tau}_1 = 0$, we do not in general have the equality $\bar{L}(\bar{y}) = L(y)$. Now choose $\bar{y}$ to be the specific velocity corresponding to $\bar{\tau}_1 = \tau_1$ in the representation. Lemma 3.1 shows that by adding a "boundary" cost to the interior cost $\bar{L}(\bar{y})$, which is a quadratic function of the local time vector $\bar{\tau}$, and by choosing an appropriate $\bar{D}$, the cost of using $\bar{y}$ in the time-reversed problem can be made equal to $L(y)$. Note that for the example in $\mathbb{R}^2$ illustrated in Fig. 1, just a linear cost $\bar{c}\bar{\tau}$ is adequate, but when the same logic is applied to points on the intersection of multiple faces, it turns out that the cost must in fact be quadratic in the local time as well as the control.

Before we can state the lemma, we need to introduce some notation. Consider a (b, A, D) RBM X and recall the definition of L given in (2). Consider the function $\bar{L} : \mathbb{R}^N \to \mathbb{R}_+$ defined by

$$\bar{L}(y) \doteq L(-y), \tag{4}$$

and a new constraint matrix $\bar{D}$ defined by

$$\bar{D} \doteq -D + 2AM^{-1}, \tag{5}$$

where M is the diagonal matrix with $M_{ii} \doteq A_{ii} = e_i' A e_i$. Let

$$v_i \doteq 2AM^{-1}e_i, \tag{6}$$

and observe that if d_i and $\bar{d}_i$ are the ith columns of the matrices D and $\bar{D}$, respectively, then

$$\bar{d}_i = -d_i + v_i, \tag{7}$$

and $(\bar{d}_i)_i = 1$. We will use the vector $\bar{c} \in \mathbb{R}^N$ defined by

$$\bar{c} \doteq 2M^{-1}b, \tag{8}$$

so that

$$\bar{c}_i \doteq (\bar{c})_i = 2b'M^{-1}e_i.$$

In addition, we define the $N \times N$ matrix $\bar{K}$ by

$$\bar{K}_{ij} \doteq 2e_i'M^{-1}AM^{-1}e_j - e_i'M^{-1}De_j - e_j'M^{-1}De_i. \tag{9}$$

Rewriting $\bar{K}_{ij}$ in the form

$$\bar{K}_{ij} = e_i'M^{-1}[2A - DM - MD']M^{-1}e_j,$$

and using the fact that A is symmetric and $D_{ii} = 1$, it is easy to see that $\bar{K}$ is symmetric ($\bar{K}_{ij} = \bar{K}_{ji}$) and $\bar{K}_{ii} = 0$. Rearranging and expanding the relation (5), we observe that

$$2A_{jk} = A_{jj}(d_k)_j + A_{jj}(\bar{d}_k)_j = A_{jj}D_{jk} + A_{jj}\bar{D}_{jk}.$$

Lemma 3.1. *Given any $J \subset \{1,\ldots,N\}$, for every $w \in \mathbb{R}^N$ that satisfies $w_i = 0$ for $i \in J$, we have*

$$\left[\bar{L}\left(-w - \sum_{i\in J}\tau_i\bar{d}_i\right) + \sum_{i\in J}\bar{c}_i\tau_i - \sum_{i,j\in J}\bar{K}_{ij}\tau_i\tau_j\right] = L\left(w - \sum_{i\in J}\tau_i d_i\right). \tag{10}$$

Proof. Using definition (2) of L, we see that the right-hand side of (10) is equal to

$$\frac{1}{2}(w-b)'A^{-1}(w-b) - \sum_{i\in J} d_i'A^{-1}(w-b)\tau_i + \frac{1}{2}\sum_{i\in J}\sum_{j\in J} d_i'A^{-1}d_j\tau_i\tau_j. \tag{11}$$

On the other hand, definitions (4) and (7) of $\bar{L}$ and $\bar{d}_i$, respectively, imply that

$$\bar{L}\left(-w - \sum_{i\in J}\tau_i\bar{d}_i\right) = L\left(w - \sum_{i\in J}\tau_i(d_i - v_i)\right),$$

and analogous to (11) we can expand the right-hand side of the last display to obtain

$$\frac{1}{2}(w-b)'A^{-1}(w-b) - \sum_{i\in J}\left[(d_i - v_i)'A^{-1}(w-b)\right]\tau_i + \frac{1}{2}\sum_{i\in J}\sum_{j\in J}(d_i - v_i)'A^{-1}(d_j - v_j)\tau_i\tau_j. \tag{12}$$

Subtracting (12) from (11) we see that $L(w - \sum_{i\in J}\tau_i d_i) - \bar{L}(-w - \sum_{i\in J}\tau_i \bar{d}_i)$ is equal to

$$-\sum_{i\in J} v_i'A^{-1}(w-b)\tau_i - \frac{1}{2}\sum_{i,j\in J} v_i'A^{-1}v_j\tau_i\tau_j + \frac{1}{2}\sum_{i,j\in J} v_i'A^{-1}d_j\tau_i\tau_j + \frac{1}{2}\sum_{i,j\in J} d_i'A^{-1}v_j\tau_i\tau_j.$$

Since $w_i = 0$ for $i \in J$, substitution of definition (6) of v_i reduces the above display to

$$\sum_{i\in J} 2e_i'M^{-1}b\tau_i - \sum_{i,j\in J}[2e_i'M^{-1}AM^{-1}e_j - e_i'M^{-1}De_j - e_j'M^{-1}De_i]\tau_i\tau_j,$$

which, on substituting the values for $\bar{c}_i$ and $\bar{K}_{ij}$ from (8) and (9) becomes

$$\sum_{i\in J}\bar{c}_i\tau_i - \sum_{i,j\in J}\bar{K}_{ij}\tau_i\tau_j.$$

Combining the last three displays we obtain (10), thus completing the proof. □

We now establish three more results required to convert the local equivalence proved in Lemma 3.1 to the main result in Theorem 3.6. Fix $x \in \mathbb{R}_+^N$, and consider any function $\phi \in \mathscr{AC}([0,\infty):\mathbb{R}_+^N)$ such that $\phi(0)=0$ and $\tau_x < \infty$, where $\tau_x \doteq \inf\{t \geqslant 0 : \phi(t) = x\}$. Let $S_\phi(x)$ be the set of all trajectories $\psi \in \mathscr{AC}([0,\infty):\mathbb{R}^N)$ with $\psi(0) = 0$ for which there is $\theta \in \mathscr{I}([0,\infty):\mathbb{R}^N)$ such that $\theta(0) = 0$ and

$$\dot{\psi}(s) = \dot{\phi}(s) - \sum_{i=1}^{N}\dot{\theta}_i(s)d_i \quad \text{for } s \in [0,\tau_x], \text{ and } \dot{\theta}_i(s) > 0 \text{ only if } \phi_i(s) = 0.$$

From Definition 2.1, it is easy to see that $S_\phi(x)$ is the set of absolutely continuous trajectories that map to ϕ under the SP. Note that no assumptions are made on the existence or uniqueness of solutions to the SP for arbitrary ψ in this definition. For each such trajectory ϕ we define the minimal cost

$$V_\phi(x) = \inf_{\psi\in S_\phi(x)} \int_0^{\tau_x} L(\dot{\psi}(s))\,\mathrm{d}s.$$

We also define an analogous set $\bar{S}_\phi(x)$, save that the condition $\psi(0) = 0$ is replaced by $\psi(0) = x$, d is replaced by $\bar{d}$ and $\tau_0 \doteq \inf\{t \geqslant 0 : \phi(t) = 0\}$ is used in lieu of τ_x.

For $\phi \in \mathscr{AC}([0,\infty):\mathbb{R}^N_+)$ with $\phi(0)=x$ let

$$\bar{V}_\phi(x) \doteq \inf_{\psi \in \bar{S}_\phi(x)} \int_0^{\tau_0} [\bar{L}(\dot{\psi}(s)) + \bar{c}'\dot{\theta}(s) - \dot{\theta}'(s)\bar{K}\dot{\theta}(s)]\,\mathrm{d}s.$$

Lemma 3.2. *Fix $x \in \mathbb{R}^N_+$, and consider any function $\phi \in \mathscr{AC}([0,\infty):\mathbb{R}^N_+)$ such that $\phi(0)=0$, $\tau_x < \infty$, and $\phi(t) \neq 0$ for $t \in (0,\tau_x)$. Define $\bar{\phi}(t) \doteq \phi(\tau_x - t)$. Then*

$$V_\phi(x) = \bar{V}_{\bar{\phi}}(x).$$

Proof. For $\psi \in S_\phi(x)$ let θ be the corresponding vector of increasing processes. Let $\bar{\phi}(\cdot)$ be the time-reversed trajectory defined by $\phi(\tau_x - \cdot)$, and let

$$\dot{\bar{\theta}}(t) \doteq \dot{\theta}(\tau_x - t)$$

for $t \in [0,\tau_x]$. In addition, let

$$\bar{\tau}_0 \doteq \inf\{t \geqslant 0 : \bar{\phi}(t) = 0\}.$$

Then since $\phi(t) \neq 0$ for $t \in (0,\tau_x)$, we have $\bar{\phi}(0) = x$, $\bar{\phi}(\tau_x) = 0$, $\bar{\tau}_0 = \tau_x < \infty$, and $\dot{\bar{\phi}}(t) = -\dot{\phi}(\tau_x - t)$ for $t \in [0,\bar{\tau}_0]$. Moreover, for every $s \in [0,\tau_x]$ $\dot{\theta}_i(s) > 0$ implies $\phi_i(s) = 0$ if and only if for every $s \in [0,\bar{\tau}_0]$ $\dot{\bar{\theta}}_i(s) > 0$ implies $\bar{\phi}_i(s) = 0$. For $s \in [0,\bar{\tau}_0]$ define

$$\bar{\psi}(s) \doteq \bar{\phi}(s) - \bar{D}\bar{\theta}(s). \tag{13}$$

It is easy to see that $\psi \in S_\phi(x)$ if and only if $\bar{\psi} \in \bar{S}_{\bar{\phi}}(x)$. In fact, on differentiating Eq. (13) and substituting for $\dot{\bar{\theta}}$ in terms of $\dot{\psi}$ and $\dot{\phi}$ one obtains the explicit relation

$$\dot{\bar{\psi}}(s) = \bar{D}D^{-1}\dot{\psi}(\tau_x - s) - [I + \bar{D}D^{-1}]\dot{\phi}(\tau_x - s). \tag{14}$$

Recall from Section 2.1 that for $y \in \mathbb{R}^N$, $I(y) \doteq \{i \in \{1,\ldots,N\} : y_i = 0\}$. Using the definition of $\bar{L}(\cdot)$, the properties stated above and Lemma 3.1, we deduce that

$$\begin{aligned}
&\int_0^{\tau_x} L(\dot{\psi}(s))\,\mathrm{d}s \\
&\quad = \int_0^{\tau_x} L\left(\dot{\phi}(s) - \sum_{i=1}^N \dot{\theta}_i(s) d_i\right) \mathrm{ds} \\
&\quad = \int_0^{\tau_x} L\left(\dot{\phi}(s) - \sum_{i \in I(\phi(s))} \dot{\theta}_i(s) d_i\right) \mathrm{d}s \\
&\quad = \int_0^{\tau_x} \left[\bar{L}\left(-\dot{\phi}(s) - \sum_{i \in I(\phi(s))} \dot{\theta}_i(s) \bar{d}_i\right)\right. \\
&\qquad \left. + \sum_{i \in I(\phi(s))} \bar{c}_i \dot{\theta}_i(s) - \sum_{i,j \in I(\phi(s))} \bar{K}_{ij} \dot{\theta}_i(s) \dot{\theta}_j(s)\right] \mathrm{d}s
\end{aligned}$$

$$= \int_0^{\bar{\tau}_0} \left[\bar{L} \left(\dot{\bar{\phi}}(s) - \sum_{i \in I(\bar{\phi}(s))} \dot{\bar{\theta}}_i(s) \bar{d}_i \right) \right.$$

$$\left. + \sum_{i \in I(\bar{\phi}(s))} \bar{c}_i \dot{\bar{\theta}}_i(s) - \sum_{i,j \in I(\bar{\phi}(s))} \bar{K}_{ij} \dot{\bar{\theta}}_i(s) \dot{\bar{\theta}}_j(s) \right] \mathrm{d}s$$

$$= \int_0^{\bar{\tau}_0} \left[\bar{L} \left(\dot{\bar{\phi}}(s) - \sum_{i=1}^{N} \dot{\bar{\theta}}_i(s) \bar{d}_i \right) + \sum_{i=1}^{N} \bar{c}_i \dot{\bar{\theta}}_i(s) - \sum_{i,j=1}^{N} \bar{K}_{ij} \dot{\bar{\theta}}_i(s) \dot{\bar{\theta}}_j(s) \right] \mathrm{d}s$$

$$= \int_0^{\bar{\tau}_0} \left[\bar{L}(\dot{\bar{\psi}}(s)) + \bar{c}' \dot{\bar{\theta}}(s) - \dot{\bar{\theta}}'(s) \bar{K} \dot{\bar{\theta}}(s) \right] \mathrm{d}s.$$

Since for every $\phi \in \mathscr{AC}([0,\infty):\mathbb{R}^N)$ there exists a one–one correspondence $\psi \leftrightarrow \bar{\psi}$ between the sets $S_\phi(x)$ and $\bar{S}_{\bar{\phi}}(x)$ (given explicitly by (14)), the last display and the definitions of $V_\phi(x)$ and $\bar{V}_\phi(x)$ show that $V_\phi(x) = \bar{V}_{\bar{\phi}}(x)$. □

Let

$$T(x) \doteq \{\phi \in \mathscr{AC}([0,\infty), \mathbb{R}^N_+) : \phi(0) = 0, \tau_x < \infty\},$$

and define the minimal cost

$$U(x) \doteq \inf_{\phi \in T(x)} V_\phi(x). \tag{15}$$

Similarly, let

$$\bar{T}(x) \doteq \{\phi \in \mathscr{AC}([0,\infty), \mathbb{R}^N_+) : \phi(0) = x, \tau_0 < \infty\},$$

and

$$\bar{U}(x) \doteq \inf_{\phi \in \bar{T}(x)} \bar{V}_\phi(x). \tag{16}$$

Theorem 3.3. *Let U and $\bar{U}$ be defined by* (15) *and* (16). *Then $U = \bar{U}$. Moreover, for any $x \in \mathbb{R}^N_+$, $U(x)$ is achieved at a trajectory ϕ that satisfies $\phi(t) \neq 0$ for $t > 0$ if and only if $\bar{U}(x)$ is attained at the corresponding trajectory $\bar{\phi}$ defined by*

$$\bar{\phi}(t) = \phi(\tau_x - t),$$

where τ_x is defined by (3).

Proof. Since $L \geqslant 0$, in the definition of $U(x)$ we can restrict to trajectories ϕ which satisfy $\phi(t) \neq 0$ for $t > 0$. Similarly, since by (10) and the fact that $L \geqslant 0$ the integrand in the definition of $\bar{V}_\phi$ is also non-negative, in the definition of $\bar{U}(x)$ we can restrict to ϕ which satisfy $\phi(t) \neq x$ for $t > 0$. Making use of the bijection $\phi \leftrightarrow \bar{\phi}$ between the sets $\{\phi \in \mathscr{AC}([0,\infty):\mathbb{R}^N) : \phi(0) = 0, \phi(t) \neq 0, t > 0, \phi(\tau_x) = x, \tau_x < \infty\}$

and $\{\phi \in \mathscr{AC}([0,\infty):\mathbb{R}^N):\phi(0)=x,\phi(t)\neq x, t>0, \phi(\tau_0)=0,\tau_0<\infty\}$, and Lemma 3.2 we conclude that

$$
\begin{aligned}
U(x) &= \inf_{\{\phi\in\mathscr{AC}([0,\infty):\mathbb{R}^N):\,\phi(0)=0,\phi(t)\neq 0,t>0,\phi(\tau_x)=x,\tau_x<\infty\}} V_\phi(x)\\
&= \inf_{\{\phi\in\mathscr{AC}([0,\infty):\mathbb{R}^N):\,\phi(0)=0,\phi(t)\neq 0,t>0,\phi(\tau_x)=x,\tau_x<\infty\}} \bar{V}_{\bar{\phi}}(x)\\
&= \inf_{\{\phi\in\mathscr{AC}([0,\infty):\mathbb{R}^N):\,\phi(0)=x,\phi(t)\neq x,t>0,\phi(\tau_0)=0,\tau_0<\infty\}} \bar{V}_\phi(x)\\
&= \bar{U}(x).
\end{aligned}
$$

The second statement follows automatically from the proof above. □

In the theorem just proved the infimization is over trajectories ϕ, rather than over the set of "control" processes ψ. Indeed, ψ enters only indirectly through the definition of V_ϕ. This formulation has the advantage of not requiring any regularity of the corresponding SM, since the question of which ϕ is the image of a given ψ is never raised. However, $U(x)$ is not automatically equal to $V(x)$, since the variational problem (1) for the latter requires infimization over all absolutely continuous ψ.

Thus to complete the derivation of a "reverse time" formulation of the large deviations variational problem, we first show in Lemma 3.4 that when D is completely-$\mathscr{S}$, the image of an absolutely continuous trajectory under the associated SM is also absolutely continuous.

Lemma 3.4. *Suppose $D\in\mathbb{R}^{N\times N}$ is completely-$\mathscr{S}$ and Γ is the associated SM. Then given any $\psi\in\mathscr{AC}([0,\infty):\mathbb{R}^N)$, $\phi\in\Gamma(\psi)$ implies $\phi\in\mathscr{AC}([0,\infty):\mathbb{R}^N)$.*

Proof. For $\psi\in\mathscr{C}([0,\infty):\mathbb{R}^N)$, it was proved in Bernard and El Kharroubi (1991, Lemma 1) that if D is completely-$\mathscr{S}$ and $\phi\in\Gamma(\psi)$ then there exists $K<\infty$ such that for any $0\leqslant t_1\leqslant t_2<\infty$,

$$|\phi(t_2)-\phi(t_1)|\leqslant K\operatorname{Osc}(\psi,[t_1,t_2]),$$

where

$$\operatorname{Osc}(\psi,[t_1,t_2]) \doteq \sup\{|\psi(t)-\psi(s)|: t_1\leqslant s<t\leqslant t_2\}.$$

Fix $T<\infty$. Since ψ is absolutely continuous, given $\varepsilon>0$ we can find $\delta>0$ so that if $\{(a_i,b_i),\ i=1,\ldots,n\}$ is any collection of non-overlapping intervals in $[0,T]$ with $\sum_{i=1}^n (b_i-a_i)<\delta$, then $\sum_{i=1}^n|\psi(b_i)-\psi(a_i)|<\varepsilon$. Given such a collection of intervals, let $a_i\leqslant a_i^*\leqslant b_i^*\leqslant b_i$ be such that $|\psi(b_i^*)-\psi(a_i^*)|=\operatorname{Osc}(\psi,[a_i,b_i])$. It follows that the non-overlapping intervals $\{(a_i^*,b_i^*),\ i=1,\ldots,n\}$ satisfy $\sum_{i=1}^n(b_i^*-a_i^*)<\delta$, and thus

$$\sum_{i=1}^n|\phi(b_i)-\phi(a_i)|\leqslant K\sum_{i=1}^n|\psi(b_i^*)-\psi(a_i^*)|\leqslant K\varepsilon.$$

Since $T<\infty$ is arbitrary, this shows that ϕ is absolutely continuous on $[0,\infty)$. □

We now use this property to show that $U(x) = V(x)$ in Theorem 3.5. In addition, it will also be convenient to know when $\bar{U}$ can be replaced by a variational problem whose form is similar to that of V. Let $\bar{\Gamma}$ denote the SM that is associated with the constraint matrix $\bar{D}$, and assume for the purposes of the following definitions that $\bar{D}$ is completely-$\mathscr{S}$ and invertible. Note that in this case $\bar{\Gamma}$ could be multi-valued. Let

$$\bar{V}(x) = \inf_{\psi \in \mathscr{AC}([0,\infty):\mathbb{R}^N):\psi(0)=x} \ \inf_{\phi \in \bar{\Gamma}(\psi):\tau_0<\infty} \int_0^{\tau_0} [\bar{L}(\dot{\psi}(s)) + \bar{c}'\dot{\theta}(s) - \dot{\theta}'(s)\bar{K}\dot{\theta}(s)]\,\mathrm{d}s, \tag{17}$$

where

$$\tau_0 \doteq \inf\{t \geqslant 0 : \phi(t) = 0\}, \tag{18}$$

and

$$\theta(s) \doteq (\bar{D}^{-1})'(\phi(s) - \psi(s)). \tag{19}$$

Theorem 3.5. *Suppose D is completely-$\mathscr{S}$, and define $V(x)$ and $U(x)$ for $x \in \mathbb{R}^N_+$ by* (1) *and* (15), *respectively. Then $V(x) = U(x)$. Suppose $\bar{D}$ defined by* (5) *is completely-$\mathscr{S}$ and invertible, and define $\bar{V}(x)$ and $\bar{U}(x)$ for $x \in \mathbb{R}^N_+$ by* (17) *and* (16), *respectively. Then $\bar{V}(x) = \bar{U}(x)$.*

Proof. Given any $\phi \in T(x)$, let $\psi \in S_\phi(x)$. Then ψ is a candidate infimizer in the definition of $V(x)$, which shows that $V(x) \leqslant U(x)$. Next let ψ be a candidate minimizer in the definition of $V(x)$, i.e., $\psi \in \mathscr{AC}([0,\infty):\mathbb{R}^N)$, with $\psi(0) = 0$ and $\tau_x < \infty$, for some $\phi \in \Gamma(\psi)$ with $\tau_x \doteq \inf\{t \geqslant 0 : \phi(t) = x\}$. From Lemma 3.4 it follows that ϕ is absolutely continuous whenever ψ is absolutely continuous and D is completely-$\mathscr{S}$. Thus $\phi \in T(x)$ and $\psi \in S_\phi(x)$, and therefore $U(x) \leqslant V(x)$.

Since $\bar{D}$ is invertible, $\bar{V}$ is well-defined. The proof that $\bar{V}(x) = \bar{U}(x)$ when $\bar{D}$ is completely-$\mathscr{S}$ is similar and therefore omitted. □

We can now state our main result. Although it is an immediate corollary of the results of this section, we will state it as a theorem due to its importance.

Theorem 3.6. *Let $V(x)$, $\bar{U}(x)$ and $\bar{V}(x)$ be defined by* (1), (16) *and* (17), *respectively. If D is completely-$\mathscr{S}$ then $V(x) = \bar{U}(x)$. If in addition $\bar{D}$ is also completely-$\mathscr{S}$ and invertible, then $V(x) = \bar{V}(x)$. Moreover, ϕ^* with $\phi^*(t) \neq 0$ for $t > 0$ is an optimal trajectory for $V(x)$ if and only if $\bar{\phi}^*(\cdot) \doteq \phi^*(\tau_x^* - \cdot)$ with $\bar{\phi}^*(t) \neq x$ for $t > 0$ (where τ_x^* is defined as in* (3) *with ϕ replaced by ϕ^*) is an optimal trajectory for $\bar{V}(x)$.*

Proof. Follows directly from Theorems 3.3 and 3.5 and Lemma 3.2. □

The paper (Dupuis and Ramanan, 2000) considers problems of the form (17) for general convex functions and constraint matrices. In all cases, the corresponding optimal controls $\dot{\bar{\psi}}$ are constant in time, though they typically depend on the initial position x. With the particular choice of the convex function $\bar{L}$ defined in (4), the setup of Dupuis and Ramanan (2000) corresponds to the case $\bar{K}=\bar{c}=0$ in (17). In the next section we analyze the control problem (17) for the case when $\bar{K}=0$, $\bar{c}\neq 0$ (which corresponds to product form RBMs) and show once again that the optimal controls are constant in time. In fact, in this special case the optimal controls turn out to also be independent of the initial position x. In the general two-dimensional case, it can be verified (by using (14) and the two-dimensional results of Avram et al., 2000, for instance) that the optimal controls continue to be constant in time, though not necessarily independent of x. An important problem for future work is to determine if this qualitative property holds in the general case with $\bar{K}\neq 0$, $\bar{c}\neq 0$ in arbitrary dimensions.

4. Large deviations of product form RBMs

Here we apply the results of the last section to RBMs that have the so-called product form stationary distributions. We introduce this class of RBMs in Section 4.1 and write down explicitly the known exponential decay rate of the tails of the stationary distribution for such RBMs. Although the decay rate for any RBM in this class is known, to the best of our knowledge the minimizing large deviation path that achieves this value has not been identified. These trajectories are of interest from the perspective of design of queueing networks approximated by RBMs since they provide insight into the manner in which large buffer contents build up. However, direct calculation of these trajectories from the original variational problem (1) for $V(x)$ appears to be non-trivial.

In Section 4.2 we examine the time-reversed variational problem (17) for this specific case. Since dynamic programming principles can be applied to the time-reversed formulation, we obtain a simplification that enables us to identify the minimizing paths. Indeed, we show in Theorem 4.1 that optimal controls for the time-reversed problem [i.e., $\dot{\bar{\psi}}$ in (17)] are constant, while optimal controls for the original variational problem (1) are in general only piecewise constant.

Since the minimizing trajectories $\bar{\phi}^*$ associated with the time-reversed problem in the product form case are images of trajectories with constant velocity under the SM, in Section 4.3 we first study images of this type. In Section 4.3.1 we consider the general case when D is completely-$\mathscr{S}$. In Section 4.3.2 we show that if D is either Harrison–Reiman or satisfies Condition 2.3, then the image is piecewise linear with at most N or 2^N-1 changes of slope, respectively. In Section 4.3.3 we show that in three dimensions, if $D\in\mathbb{R}^{3\times 3}$ is generalized Harrison–Reiman, then the image is piecewise affine with at most 5 changes of slope. Using the fact (proved in Theorem 3.5) that the minimizing large deviation trajectories ϕ^* are time-reversals of $\bar{\phi}^*$, this allows us to conclude in Section 4.3.4 that the minimizing large deviation trajectories of product form RBMs with matrices satisfying any of the above conditions do not cycle.

The results of this section demonstrate that the time-reversed perspective offers a considerable simplification over the original variational problem, at least for the product form case.

4.1. The skew-symmetry condition

Harrison and Williams (1987a,b) and Williams (1987) showed that for any (b, A, D) RBM X with a completely-$\mathscr{S}$ constraint matrix D such that $D^{-1}b < 0$, the following skew symmetry condition

$$2A_{jk} = A_{jj}(d_k)_j + A_{kk}(d_j)_k = A_{jj}D_{jk} + A_{kk}D_{kj}$$

is necessary and sufficient for the stationary density p (with respect to Lebesgue measure) of X to have a separable exponential form. Note that although they assume that D has the additional regularity stated in Condition 2.2, this is not required in the proof of the particular theorem (Harrison and Williams, 1987b, Theorem 9.2; Dai and Harrison, 1992, Proposition 9) More precisely, they showed that

$$p(x) = C\prod_{i=1}^{N} u_i \exp(-u_i x_i)$$

for some constant $C < \infty$, where (translating Harrison and Williams, 1987b, Eq. (7) into our notation)

$$u \doteq -2M^{-1}D^{-1}b. \tag{20}$$

Thus for the class of RBMs with completely-$\mathscr{S}$ D matrices satisfying the skew-symmetry condition and $D^{-1}b < 0$, the exponential decay rate of the tails of the distribution is linear and has the known explicit form

$$x'u = -2x'M^{-1}D^{-1}b.$$

4.2. The time-reversed control problem

In this section we analyze the time-reversed optimal control problem (17) for product-form RBMs. The skew symmetry condition introduced in the last section can be equivalently expressed as

$$2A = MD' + DM. \tag{21}$$

Comparing (5) and (21) we note that this condition is satisfied if and only if the time-reversed constraint matrix $\bar{D}$ satisfies the relationship

$$\bar{D} = MD'M^{-1}. \tag{22}$$

Moreover, from Definition (9) of $\bar{K}$ it is easy to see that (21) holds if and only if the matrix $\bar{K}$ is identically zero. Thus for product form RBMs, the time-reversed control problem (17) reduces to

$$\bar{V}(x) = \inf_{\psi\in\mathscr{AC}([0,\infty):\mathbb{R}^N):\psi(0)=x} \inf_{\phi\in\bar{\Gamma}(\psi):\tau_0<\infty} \int_0^{\tau_0} [\bar{L}(\dot{\psi}(s)) + \bar{c}'\dot{\theta}(s)]\,\mathrm{d}s, \tag{23}$$

where $\bar{\Gamma}$ is the Skorokhod Map associated with $\bar{D}$, and τ_0 and θ are defined as before via (18)–(19).

In Theorem 4.1 below we derive the structure of the optimal control for $\bar{V}(x)$. We first provide some intuition behind the result. Given $w \in \mathbb{R}^N$, let $D_w V$ ($D_w \bar{V}$) denote the directional derivative of V (respectively $\bar{V}$) along w. Whenever V is smooth $D_w V(x) = D_w \bar{V}(x)$ exists for all $w \in \mathbb{R}^N$, and dynamic programming principles (see for example, Kushner and Dupuis, 1992, Chapter 3) suggest that for $x \in (\mathbb{R}_+^N)^\circ$ the Hamilton–Jacobi–Bellman equation

$$\inf_{\beta} [D_\beta \bar{V}(x) + \bar{L}(\beta)] = 0 \tag{24}$$

related to the control problem (17) should be satisfied. Moreover, one would expect that when $\phi(t) = x \in (\mathbb{R}_+^N)^\circ$ the optimal control that should be applied must be equal to the value $\bar{\beta}^*$ at which the infimum in (24) is achieved. For the product form case, if D is sufficiently regular to ensure that the tails of the stationary distribution satisfy an LDP with rate defined by the variational problem (1) and if $\bar{D} = MD'M^{-1}$ is completely-$\mathscr{S}$, then by Theorem 3.6 we have $\bar{V}(x) = V(x) = x'u$. Both of these conditions hold if, for example, D satisfies Condition 2.4. Thus using the definitions of $\bar{L}$ and $\bar{V}$ and differentiating the left-hand side of Eq. (24), simple algebra yields

$$\bar{\beta}^* = -b - Au = -b + 2AM^{-1}D^{-1}b = MD'M^{-1}D^{-1}b, \tag{25}$$

where the second equality uses the definition (20) of u and the last equality follows from the skew symmetry condition (21). Note that since $\bar{V}$ is linear in the product form case, $\bar{\beta}^*$ is independent of x. This leads one to conjecture that the minimizing trajectory in (17) may be given by

$$\bar{\psi}^*(t) = x + \bar{\beta}^* t. \tag{26}$$

The discussion above was heuristic as it neglected the effect of boundaries. In Theorem 4.1 below we prove this conjecture by direct verification. We first define $\bar{\phi}^*$ by

$$\bar{\phi}^*(t) \doteq \bar{\Gamma}(x + \bar{\beta}^* \imath)(t), \tag{27}$$

let $\bar{\tau}_0^*$ be defined by (18) with ϕ replaced by $\bar{\phi}^*$, and for $s \in [0, \bar{\tau}_0^*]$ let

$$\phi^*(s) \doteq \bar{\phi}^*(\bar{\tau}_0^* - s).$$

In the next theorem we prove that $\bar{\beta}^*$ defines the velocity of the unique minimizer in the definition of $\bar{V}(x)$. As a consequence of the explicit correspondence between the minimizing paths for the forward and backward problems $U(x)$ and $\bar{U}(x)$ proved in Lemma 3.3, it then follows that ϕ^* is the unique minimizing path in the definition of $U(x)$, up to the normalization $\phi^*(t) \neq 0$ for $t > 0$. It is therefore also the most likely path from the perspective of the related large deviation problem when D is sufficiently regular to ensure that $V(x)$ characterizes the exponential decay rate for the stationary RBM.

Theorem 4.1. *Suppose Conditions* 2.4 *and* 2.5 *hold, and in addition the skew-symmetry condition* (21) *is satisfied. Let* $\bar{V}$ *and* $\bar{\psi}^*$ *be defined as in* (17) *and* (26) *respectively. Then the infimum in* (17) *is uniquely achieved when* $\psi = \bar{\psi}^*$.

Proof. From the discussion in Section 4.1 and Theorem 2.5, we can conclude that $V(x) = u'x$, where u is defined by (20). Since Conditions 2.4 and 2.5 imply that D and therefore $\bar{D} = MD'M^{-1}$ is completely-$\mathscr{S}$ and invertible (it in fact also satisfies Condition 2.4), by Theorem 3.6 $\bar{V}(x) = V(x) = u'x$. Recall that $\bar{\Gamma}$ is the SM associated with the matrix $\bar{D}$, let $\bar{\phi}^*$ be defined by (27) and suppose $\dot{\bar{\theta}}^* \doteq \bar{D}^{-1}(\dot{\bar{\phi}}^* - \bar{\beta}^*)$. Define

$$Z^*(t) \doteq \int_0^t [\bar{L}(\bar{\beta}^*) + \bar{c}'\dot{\bar{\theta}}^*(s)]\,\mathrm{d}s.$$

Note that since $\bar{d}_i = MD'M^{-1}e_i$,

$$u'\bar{d}_i = -2b'(D')^{-1}M^{-1}MD'M^{-1}e_i = -2b'M^{-1}e_i = -\bar{c}_i.$$

Moreover, the definition of $\bar{\beta}^*$ as the optimizer in (24) implies that $\bar{L}(\bar{\beta}^*) = -u'\bar{\beta}^*$. Using these two properties, the definition of $\bar{\theta}^*$, and the relation $\dot{\bar{\phi}}^*(t) = \bar{\beta}^* + \sum_{i=1}^N \dot{\bar{\theta}}_i^*(t)\bar{d}_i$, we conclude that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} Z^*(t) &= \bar{L}(\bar{\beta}^*) + \bar{c}'\dot{\bar{\theta}}^*(t) \\
&= -u'\bar{\beta}^* + \bar{c}'\dot{\bar{\theta}}^*(t) \\
&= -u'\dot{\bar{\phi}}^*(t) + \sum_{i=1}^N [u'\bar{d}_i + \bar{c}_i]\dot{\bar{\theta}}_i^*(t) \\
&= -u'\dot{\bar{\phi}}^*(t) \\
&= -\frac{\mathrm{d}}{\mathrm{d}t}\bar{V}(\bar{\phi}^*(t)).
\end{aligned}$$

Now since $\bar{D}$ satisfies (22), we infer that

$$\bar{D}^{-1}\bar{\beta}^* = M(D')^{-1}M^{-1}MD'M^{-1}D^{-1}b = D^{-1}b.$$

Hence Condition 2.5 (which is equivalent to $D^{-1}b < 0$) holds if and only if $\bar{D}^{-1}\bar{\beta}^* < 0$. Thus by Budhiraja and Dupuis (1999, Theorem 3.12) $\bar{\tau}_0^* < \infty$, and integrating the equality in the last display gives $Z^*(\bar{\tau}_0^*) = \bar{V}(x)$. This shows that $\bar{\beta}^*$ is an optimal control in the variational problem (23), i.e., it achieves the value $\bar{V}(x)$.

Using the fact that $\bar{L}(\bar{\beta}) > -u'\bar{\beta}$ if $\bar{\beta} \neq \bar{\beta}^*$, it follows that any other trajectory gives strictly greater cost. Suppose that the control $\bar{\beta}(t)$ is not equal to $\bar{\beta}^*$ on a set of positive Lebesgue measure (prior to the first time the controlled trajectory reaches the origin). Let $\bar{\phi}$, $\bar{\theta}$, $\bar{\tau}_0$ and Z be defined as the starred quantities were. Then for a subset of $[0, \bar{\tau}_0]$ of positive Lebesgue measure,

$$\frac{\mathrm{d}}{\mathrm{d}t} Z(t) > -\frac{\mathrm{d}}{\mathrm{d}t}\bar{V}(\bar{\phi}(t)).$$

Thus either $\bar{\tau}_0 = \infty$, in which case $\bar{\beta}$ is not admissible, or else $Z(\bar{\tau}_0) > \bar{V}(x)$, in which case $\bar{\beta}$ is not optimal. □

Remark. Observe that for the product form case the milder assumption that D be completely-$\mathscr{S}$ (and $D^{-1}b < 0$) is sufficient to ensure that $V(x) = \bar{V}(x)$ and that the exponential decay rate of the tails of the stationary distribution is equal to $u'x$. Thus it is natural to ask whether Theorem 4.1 still holds without the additional regularity (i.e. Condition 2.4) imposed on D. The proof uses the regularity in two places. Firstly it is needed to ensure (as in Theorem 2.5) that the exponential decay rate $u'x$ indeed has the variational representation $V(x)$. Secondly it is used to deduce that $\bar{\tau}_0^* < \infty$ (which follows from the fact that $\bar{\Gamma}$ is Lipschitz continuous and $\bar{D}^{-1}b < 0$), which is necessary for $\bar{\psi}^*$ to be a valid control in the definition (23) of $\bar{V}$.

4.3. When do most likely large deviation trajectories cycle?

Theorem 4.1 established that (under Conditions 2.4 and 2.5) for the product form case the optimal control for the time-reversed variational problem $\bar{V}(x)$ is constant. In contrast, the optimal control for the original variational problem typically does not have this property, as is borne out by the relation (14) between the two controls as well as by the example given in Section 5. Since the minimizing trajectories for the time-reversed variational problem are images of constant velocity trajectories under the SM, in this section we first analyze trajectories of this form.

In Section 4.3.1 we consider SMs that are associated with a completely-$\mathscr{S}$ constraint matrix D and show that the image of a piecewise affine trajectory under such a SM is again piecewise affine (see Definition 4.2). In Sections 4.3.2 and 4.3.3 we show that when additional regularity properties are imposed on D, then ϕ is finitely piecewise affine. As a corollary of these results, in Section 4.3.4 (Theorem 4.9) we derive sufficient conditions for when the most likely large deviation path associated with a product-form RBM does not cycle.

4.3.1. Images of affine trajectories under SMs associated with a completely-$\mathscr{S}$ matrix

We start by quoting a lemma of Bernard and Kharroubi (1991, Proposition 3 and Remark 4) that shows that the images of affine trajectories under the SM are piecewise affine (see definition below) whenever the SM is associated with a completely-$\mathscr{S}$ matrix.

Definition 4.2 (*Piecewise affine*). We say $\phi \in \mathscr{C}([0,\infty) : \mathbb{R}^N)$ is piecewise affine if for every $t \in [0,\infty)$ there exists $v_t \in \mathbb{R}^N$ and $\varepsilon_t > 0$ such that for $s \in [0, \varepsilon_t)$,

$$\phi(t+s) = \phi(t) + s v_t.$$

Moreover, ϕ is m-piecewise affine if for some $k \leqslant m$, there exist $0 = t_0 < t_1 < t_2 < \cdots < t_{k-1} < t_k = \infty$ and vectors $v^1, \ldots, v^k \in \mathbb{R}^N$ with $v^i \neq v^{i+1}$ for $i = 1, \ldots, k-1$ such that for $i = 1, \ldots, k$ and $t \in [t_{i-1}, t_i)$,

$$\phi(t) = \phi(t_{i-1}) + (t - t_{i-1}) v^i.$$

Moreover, ϕ is said to be finitely piecewise affine if it is m-piecewise affine for some $m < \infty$.

Remark. Note that the definition of a piecewise affine function ϕ given above allows for the possibility that $v^k = 0$ for some $k \in \mathbb{N}$. Thus the graph $\{x \in \mathbb{R}^N : \phi(t) = x$ for some $t \geqslant 0\}$ of ϕ may have fewer changes of slope than the function ϕ.

Lemma 4.3 (Bernard and Kharroubi, 1991). *Let Γ be the SM associated with a completely-$\mathscr{S}$ matrix $D \in \mathbb{R}^{N\times N}$. If ψ is piecewise affine, then there exists $\phi \in \Gamma(\psi)$ that is also piecewise affine.*

Recall the definition that $I(x) = \{i \in \{1,\dots,N\} : x_i = 0\}$. Let Γ be the SM associated with a completely-$\mathscr{S}$ matrix $D \in \mathbb{R}^{N\times N}$ and suppose $\phi \in \Gamma(x + u\iota)$ is piecewise affine (the existence of such a ϕ is guaranteed by Lemma 4.3). Let $t_0 = 0$. Then by Definition 4.2 and the definition of the SP it is clear that for as long as $t_{i-1} < \infty$, there exists $t_i > t_{i-1}$ and a corresponding set J_i and vector v^i such that

$$I(\phi(s)) = J_i \quad \text{and} \quad \dot{\phi}(s) = v^i \text{ for } s \in (t_{i-1}, t_i), \tag{28}$$

there exists $\alpha_j \geqslant 0$, $j \in J_i$, such that

$$v^i = u + \sum_{j\in J_i} \alpha_j d_j, \tag{29}$$

$$v^i_j = 0 \text{ for } j \in J_i \quad \text{and} \quad I(\phi(t_i)) \supset J_i \text{ if } t_i < \infty. \tag{30}$$

Moreover, ϕ is k-piecewise affine if and only if $t_j = \infty$ for some $j \leqslant k$.

A natural question to ask is under what conditions is ϕ finitely piecewise affine. It is straightforward to check that in two dimensions, as long as D is completely-$\mathscr{S}$, ϕ is always finitely piecewise affine. (It can in fact be shown to be 3-piecewise affine.) However, this is not true in higher dimensions. For example, in Bernard and Kharroubi (1991, p. 160) it was shown that if Γ is the SM associated with the constraint matrix $D \in \mathbb{R}^{3\times 3}$ such that $D_{12} = D_{23} = D_{31} = 3/2$, $D_{13} = D_{21} = D_{32} = 0$, $u = -(1,1,1)$ and $x \in \partial\mathbb{R}^N_+$, then there exists $\phi \in \Gamma(x + u\iota)$ with infinitely many changes of slope which cycles outward along the boundary. It is also possible to show that with the same choice of $u = -(1,1,1)$ and another choice of D (for example $D_{12} = D_{23} = D_{31} = 6/5$ and $D_{13} = D_{21} = D_{32} = -3/5$), there exists a corresponding trajectory $\phi \in \Gamma(x + u\iota)$ that again has infinitely many changes of slope, but this time cycles inward along the boundary and reaches 0 at some finite time and then stays there. In both cases the matrices D are completely-$\mathscr{S}$, but they do not satisfy additional regularity conditions such as the generalized Harrison-Reiman condition (Condition 2.4). In the next two sections we impose additional regularity conditions on D that guarantee that ϕ is finitely piecewise affine.

4.3.2. *Sufficient conditions for finitely piecewise affine paths in $\mathbb{R}^N_+$*

Theorem 4.4 below shows that images of constant velocity trajectories under the SM associated with two classes of matrices have a finite number of changes of slope and hence do not cycle.

Theorem 4.4. *Let Γ be the SM associated with the constraint matrix D and suppose $\psi = x + u\iota$ for some $x \in \mathbb{R}_+^N$ and $u \in \mathbb{R}^N$, and let $\phi \in \Gamma(\psi)$. Then the following properties hold.*

1. *If D satisfies Condition 2.2 then ϕ is either N-piecewise affine or there exists $T < \infty$ such that ϕ is N-piecewise affine on $[0, T]$ and $\dot{\phi}(t) = \phi(t) = 0$ for all $t > T$.*
2. *If D satisfies Condition 2.3 then ϕ is 2^N-piecewise affine.*

Proof. Suppose D satisfies Condition 2.2. Then D is completely-$\mathscr{S}$ and by Lemma 4.3 ϕ is piecewise affine. The key to the proof under this condition is to show that $I(\phi(t))$ is monotonically increasing in t. In other words, once ϕ enters a given face of $\partial\mathbb{R}_+^N$, it will stay in the closure of that face from that time on, and the trajectory can only move from any given face (say F_1) to a face having strictly lower dimension (say F_2), if $I(x) \subset I(y)$ when $x \in F_1, y \in F_2$. Since ϕ is easily seen to be affine within faces, this means that ϕ can undergo at most N changes of slope.

Let $t_0 = 0$, define t_i, $i = 1, \ldots,$ as described after Lemma 4.3, and choose $j \in \{1, \ldots, N\}$ such that $t_j < \infty$. If no such j exists, then $t_1 = \infty$ and it follows that ϕ is of constant velocity. So we assume henceforth that such a j exists and let the vectors v^i, and sets J_i satisfy (28)–(30) for $i = 1, \ldots, j+1$. For simplicity we denote J_j by J, v^j by v and $I(\phi(t_j))$ by K. First note that $J \neq \{1, \ldots, N\}$, because otherwise from (28) and uniqueness of the solution to the SP it follows that $\dot{\phi}(t) = 0$ for all $t \geqslant t_{j-1}$ which implies that $t_j = \infty$ and contradicts the choice of j. We now show that $J_{j+1} = K$.

From the definition of the SP we know there exist $\alpha_i \geqslant 0$, $i \in J$, such that $v = u + \sum_{i \in J} \alpha_i d_i$. Define $\alpha_i = 0$ for $i \in K \setminus J$. Clearly, $v_i < 0$ for $i \in K \setminus J$, and so $v_i = 0$ for $i \in J$ implies $v_i \leqslant 0$ for all $i \in K$. Let D_K be the submatrix of D that consists of all entries D_{ij} with $i \in K$ and $j \in K$. Since D satisfies Condition 2.2, it follows that D_K is invertible, and moreover $D_K^{-1} \geqslant 0$ (see Berman and Plemmons, 1979, or Dupuis and Ramanan, 2000, Lemmas 4.6 and 4.7). Let $\bar{\alpha}_i$, $i \in K$, solve

$$\left(v + \sum_{i \in K} \bar{\alpha}_i d_i \right)_j = 0$$

for $j \in K$. In other words, if $\bar{\alpha}$ and $\bar{v}$ are the column vectors in $\mathbb{R}^{|K|}$ with entries $\bar{\alpha}_i$ and v_i, $i \in K$, then

$$\bar{v} + D_K \bar{\alpha} = 0.$$

Since $\bar{v} \leqslant 0$ and $D_K^{-1} \geqslant 0$, it follows that $\bar{\alpha} \geqslant 0$. We now define $\tilde{\alpha}_i = \alpha_i + \bar{\alpha}_i \geqslant 0$ for $i \in K$. Let $w = u + \sum_{i \in K} \tilde{\alpha}_i d_i$. Then for any $k \in K$,

$$w_k = \left(u + \sum_{i \in K} \tilde{\alpha}_i d_i \right)_k = \left(v + \sum_{i \in K} \bar{\alpha}_i d_i \right)_k = 0.$$

By uniqueness of the SM, if $\phi = \Gamma(x + u\iota)$, then since $\Gamma(\phi(t_j) + u\iota)(t) = \phi(t_j + t)$ it follows that $v^{j+1} = w$ and $J_{j+1} = K$. Since from (30) we know that $K \supset J_j$, $K \neq J_j$, we have shown that ϕ moves to a strictly lower dimensional face at t_j whenever $t_j < \infty$. It will stay in this face until it hits a face of even lower dimension or till time ∞.

Now consider the case when D satisfies Condition 2.3, and assume without loss of generality that D is lower triangular. Then it follows from the definition of the SM that there exists $\theta \in \mathscr{I}([0,\infty):\mathbb{R}^N)$ with $\theta(0)=0$ such that for $i=1,\ldots,N$,

$$\phi_i = \psi_i + \sum_{j=1}^{N} d_{ij}\theta_j = \psi_i + \theta_i + \sum_{j<i} d_{ij}\theta_j, \tag{31}$$

where $\dot{\theta}_i > 0$ only if $\phi_i = 0$. Let Γ_1 denote the one-dimensional Skorokhod Map with domain $\mathbb{R}_+$ and $d_1 = e_1$ at $\{0\}$ (given explicitly by (33)). Then it is easy to see that (31) holds if and only if ϕ satisfies

$$\begin{aligned}
\phi_1 &= \Gamma_1(\psi_1),\\
\phi_2 &= \Gamma_1(\psi_2 + d_{21}\theta_1),\\
&\vdots\\
\phi_i &= \Gamma_1\left(\psi_i + \sum_{j<i} d_{ij}\theta_j\right),\\
&\vdots\\
\phi_N &= \Gamma_1\left(\psi_N + \sum_{j<N} d_{ij}\theta_j\right),
\end{aligned} \tag{32}$$

where $\theta_i = \phi_i - \psi_i - \sum_{j<i-1} d_{ij}\theta_j$.

The one-dimensional Skorokhod Map Γ_1 is known to have the explicit expression (Skorokhod, 1961)

$$\Gamma_1(f)(t) = f(t) - \left[\inf_{s\in[0,t]} f(s)\right] \wedge 0. \tag{33}$$

We first prove some properties of Γ_1. Suppose $f \in \mathscr{C}([0,\infty):\mathbb{R}^1)$ with $f(0)\in\mathbb{R}_+$ is 1-piecewise affine and let $g=\Gamma_1(f)$. Then clearly $f = x + v\iota$ for some $x\in\mathbb{R}_+$ and $v\in\mathbb{R}$. If $v \geqslant 0$, then $g = x + v\iota$ and hence is 1-piecewise affine. On the other hand if $v<0$, then $g(t) = x + vt$ for $t\in[0,-x/v]$, and $g(t)=0$ for $t \geqslant -x/v$. Thus $g=\Gamma_1(f)$ has at most one change of slope, and so is 2-piecewise affine. Now suppose f is m-piecewise affine with slope v^i on the interval (t_{i-1},t_i) for $i=1,\ldots,m$, where $t_m \doteq \infty$, and let $g=\Gamma_1(f)$. Then using the behavior of Γ_1 on affine trajectories and the fact that for any i $\Gamma_1(g(t_{i-1}) + v^i\iota)(t) = g(t_{i-1}+t)$ for $t\in[0,t_i - t_{i-1}]$, it follows that g can have at most one change in slope in each interval (t_{i-1},t_i). Thus g is $2m$-piecewise affine, and moreover the number of points at which either f or g has a change of slope is bounded by $2m-1$. Moreover, these $2m-1$ points also contain the changes of slope of any function obtained as a linear combination of f and g and/or addition by a 1-piecewise affine function.

Applying these properties to the mappings in (32), it follows that since ψ_1 is 1-piecewise affine, the set of points where either ψ_1, ϕ_1, θ_1, or ψ_2 has a change of slope is bounded by 1 (where the property for ψ_2 holds trivially since it has no change of slope). Now suppose there is a set with cardinality $m-1$ that contains the points of change of slopes of all the functions θ_j, $j<i$, and ψ_i. Then by (32) and the properties of Γ_1 listed in the last paragraph, there must exist a set of cardinality $2m-1$ that contains the points of change of slopes of the functions ϕ_i, θ_j, $j<i+1$, and ψ_i. Since we have shown this to be true for $m=2$, proceeding by induction we conclude that the total number of points of change of slope for any of the functions θ_i, $i \leqslant N$, is bounded by 2^N-1. Since each ϕ_i is a linear combination of these functions along with translation by the affine function ψ_i, it follows that ϕ is 2^N-piecewise affine. □

4.3.3. More general conditions for finitely piecewise affine paths in $\mathbb{R}^3_+$

The main result of this section, Theorem 4.8, shows that if Γ is associated with a generalized Harrison–Reiman matrix $D \in \mathbb{R}^{3\times 3}$, then any trajectory ϕ obtained by the application of Γ to a constant velocity input is 6-piecewise affine. The proof uses Lemma 4.6 and Theorem 4.7, which together show that if $D \in \mathbb{R}^{N\times N}$ is a generalized Harrison–Reiman matrix, then ϕ never leaves the boundary $\partial\mathbb{R}^N_+$ after the first time $t_0>0$ that it hits the boundary. Note that Theorem 4.7 is true for arbitrary dimensions, while Theorem 4.8 is valid only in three dimensions. Recall the definition $I(x)=\{i \in \{1,\dots,N\} : x_i=0\}$. Given a finite set A we use $|A|$ to denote its cardinality. We first prove an elementary lemma that shows that functions obtained as images of an affine function under a regular Skorokhod Map cannot have two different slopes on the same face.

Lemma 4.5. *Let Γ be the SM associated with a matrix $D \in \mathbb{R}^{N\times N}$ that satisfies Condition* 2.1, *and let $\phi \doteq \Gamma(x+u\iota)$ for some $x \in \mathbb{R}^N_+$ and $u \in \mathbb{R}^N$. If there exist $0 \leqslant t_1 < t_2 \leqslant t_3 < t_4$ and $J \subseteq \{1,\dots,N\}$ such that $I(\phi(s))=J$ for $s \in (t_1,t_2)\cup(t_3,t_4)$, then there exists $v \in \mathbb{R}^J$ with $v_i=0$ for $i \in J$ such that $\dot{\phi}(s)=v$ for $s \in (t_1,t_2)\cup(t_3,t_4)$.*

Proof. Since ϕ is piecewise affine by Lemma 4.3, it is clear from (28) and (29) that for $i=1,2$ there exist $v^i \in \mathbb{R}^J$ with $v^i_j=0$ for $j \in J$ and

$$v^i = u + \sum_{j\in J} \alpha^i_j d_j,$$

for some $\alpha^i_j \geqslant 0$, $j \in J$, such that $\dot{\phi}(s)=v^1$ for $s\in(t_1,t_2)$ and $\dot{\phi}(s)=v^2$ for $s \in (t_3,t_4)$. By the definition of the SP this implies that for any y with $I(y)=J$ there exists $\varepsilon>0$ such that for $s\in[0,\varepsilon]$ $\Gamma(y+u\iota)(s)=y+v^1 s$ and $\Gamma(y+u\iota)(s)=y+v^2 s$. The uniqueness of Γ then dictates that $v^1=v^2$. □

Lemma 4.6. *Let Γ be the SM associated with a matrix $D \in \mathbb{R}^{N\times N}$ that satisfies Condition* 2.4, *and let $\phi \doteq \Gamma(x+u\iota)$ for some $x \in \mathbb{R}^N_+$ and $u \in \mathbb{R}^N$. Suppose for $s \in (0,\infty)$ that $I(\phi(s)) \neq \emptyset$. Then there exists*

$$i \in I(\phi(s)) \text{ such that } u_i \leqslant 0. \tag{34}$$

Proof. By Lemma 4.3 we know that ϕ is piecewise affine. Thus, using the definition of the SP, for every $t \in [0,\infty)$ there exist $\varepsilon_t > 0$, $J \subset \{1,\dots,N\}$ and $v \in \mathbb{R}^N$ with $v_i = 0$, $i \in J$, such that $I(\phi(s)) = J$, $\dot{\phi}(s) = v$ for $s \in (t, t+\varepsilon_t)$ and $\varepsilon_t = \inf\{s > 0 : I(\phi(t+s)) \neq J\}$, so that if $\varepsilon_t < \infty$ then $I(\phi(t+\varepsilon_t)) \neq J$. Assume $J \neq \emptyset$. Clearly any $t \in (0,\infty)$ can be expressed as the limit of a non-decreasing sequence $s_n \uparrow t$, where each $s_n \in (t_n, t_n + \varepsilon_{t_n})$ for some sequence $t_n \in (0,\infty)$ (for example, choose $t_n \uparrow t$). Moreover, if (34) is satisfied by $s = s_n$ for $n = 1,\dots,$ then the continuity of ϕ, the upper semicontinuity of $I(\cdot)$, the finiteness of $\{1,\dots,N\}$ and the fact that $s_n \to t$ dictate that (34) is also satisfied for $s = t$. Thus property (34) is preserved under limits, and hence to prove the lemma it is enough to establish (34) for $s \in (t, t+\varepsilon_t)$, $t \in [0,\infty)$.

Fix any $t \in [0,\infty)$ and let ε_t, v and $J \neq \emptyset$ be as defined above. By definition of the SP, there exist $\alpha_i \geqslant 0$, $i \in J$, such that

$$v = u + \sum_{i \in J} \alpha_i d_i. \tag{35}$$

Let $D_J \in \mathbb{R}^{|J| \times |J|}$ be the submatrix of D having only entries D_{ij}, $i,j \in J$. Moreover, define the matrices D_J^+ and D_J^- by

$$D_J^+ \doteq D_J \vee 0 - I \quad \text{and} \quad D_J^- = D_J \wedge 0 + I,$$

where 0 and I here represent the $|J| \times |J|$ zero and identity matrix, respectively. Note that $D_J = D_J^+ + D_J^-$. Also, since $0 \leqslant I - D_J^- \leqslant |I - D_J|$ and Condition 2.4 implies $\sigma(|I - D_J|) < 1$, it follows from Berman and Plemmons (1979, Corollary 2.1.5, p. 27) that $\sigma(I - D_J^-) < 1$. Along with the inequality $I - D_J^- \geqslant 0$, this implies that $(D_J^-)^{-1} \geqslant 0$ (Dupuis and Ramanan, 2000, Lemmas 5.6 and 5.7; Berman and Plemmons, 1979). Let v^J be the vector v_i, $i \in J$, and define u^J and α^J analogously. Then (35) and the fact that $v^J = 0$ implies that

$$u^J + D_J \alpha^J = u^J + D_J^+ \alpha^J + D_J^- \alpha^J = 0.$$

Rearranging terms in the above display and applying $(D_J^-)^{-1}$ to each term yields

$$(D_J^-)^{-1} u^J = -(D_J^-)^{-1} D_J^+ \alpha^J - \alpha^J \leqslant 0,$$

where the last inequality follows from the fact that $\alpha^J \geqslant 0$, $D_J^+ \geqslant 0$ and $(D_J^-)^{-1} \geqslant 0$. Once again using the fact $(D_J^-)^{-1} \geqslant 0$, the last display shows that $u^J \not> 0$. Since $J = I(\phi(s))$ for $s \in (t, t+\varepsilon_t)$, this establishes (34). $\square$

Remark. Note that the above proof does not require that the matrix D satisfy Condition 2.4, but only uses the less restrictive condition that the matrix $D^- \doteq D \wedge 0 + I$ be completely-$\mathscr{S}$.

Theorem 4.7. *Let Γ be the SM associated with a matrix $D \in \mathbb{R}^{N \times N}$ that satisfies Condition* 2.4, *and let $\phi \doteq \Gamma(x + u\iota)$ for some $x \in \mathbb{R}_+^N$ and $u \in \mathbb{R}^N$. Then $\phi(s) \in \partial\mathbb{R}_+^N$ for all $s \in (t_0, \infty)$, where*

$$t_0 \doteq \inf\{t > 0 : \phi(t) \in \partial\mathbb{R}_+^N\}.$$

Proof. The case $t_0 = \infty$ follows trivially. So we assume without loss of generality $t_0 < \infty$ and argue by contradiction. Suppose there exists $t_* \in (t_0, \infty)$ such that $\phi(t_*) \in (\mathbb{R}^N_+)^\circ$, and let $s_* \doteq \sup\{t \leqslant t_* : \phi(t) \in \partial\mathbb{R}^N_+\}$. Since ϕ is continuous and $(\partial\mathbb{R}^N_+)^\circ$ is open we know $s_* \in [t_0, t_*)$, and from the definition of the SP we deduce that for $s \in [s_*, t_*]$,

$$\phi(s) = \phi(s_*) + u(s - s_*).$$

The above display and the fact that $\phi(s) \in (\mathbb{R}^N_+)^\circ$ for $s \in (s_*, t_*)$ imply that $u_j > 0$ for every $j \in I(\phi(s_*))$. However, since $s_* > 0$ and $I(\phi(s_*)) \neq \emptyset$, this contradicts Lemma 4.6, and so it follows that $\phi(s) \in \partial\mathbb{R}^N_+$ for all $s \in (t_0, \infty)$. □

Theorem 4.8. *Let Γ be the SM associated with a constraint matrix $D \in \mathbb{R}^{3\times 3}$ that satisfies Condition* 2.4, *and let $\phi \doteq \Gamma(x + u\iota)$ for some $x \in \mathbb{R}^3_+$, $u \in \mathbb{R}^3$. Then ϕ is either* 5*-piecewise affine, or there exists $T < \infty$ such that ϕ is* 5*-piecewise affine on* $[0, T]$ *and $\dot\phi(t) = \phi(t) = 0$ for all $t > T$.*

Proof. Recall that $d_i \in \mathbb{R}^3$ is the ith column of D. We first show that a necessary condition for D to satisfy Condition 2.4 is that

$$|(d_1)_2(d_2)_3(d_3)_1| + |(d_1)_3(d_2)_1(d_3)_2| < 1. \tag{36}$$

Observe that the spectral radius of the matrix $|I - D|$ is the largest root (in absolute value) of the function

$$f(\lambda) = \lambda^3 - [|(d_1)_2(d_2)_1| + |(d_2)_3(d_3)_2| + |(d_3)_1(d_1)_3|]\lambda$$
$$- |(d_1)_2(d_2)_3(d_3)_1| - |(d_1)_3(d_2)_1(d_3)_2|.$$

Since $f(\lambda) \uparrow \infty$ as $\lambda \uparrow \infty$ and f is continuous, a necessary condition for all roots of f to be less than one is that $f(1) > 0$, which in turn implies (36).

Let $t_0 \doteq \inf\{t > 0 : \phi(t) \in \partial\mathbb{R}^3_+\}$. If $t_0 = \infty$ the theorem follows trivially. So assume henceforth that $t_0 < \infty$. Since Condition 2.4 implies D is completely-$\mathscr{S}$ and Γ is single-valued, by Lemma 4.3 ϕ is piecewise affine. Thus as stated after Lemma 4.3 for $i = 1, \ldots,$ as long as $t_{i-1} < \infty$, there exists $t_i \in (t_{i-1}, \infty]$, $J_i \subset \{1, 2, 3\}$ and $v^i \in \mathbb{R}^3$ that satisfy (28)–(30).

Since ϕ has constant velocity on $[0, t_0)$, it is enough to show that ϕ is either 4-piecewise affine on $[t_0, \infty)$ or is 5-piecewise affine on $[t_0, \infty)$ with the final piece having zero velocity and staying at the origin. We now argue by contradiction. Suppose this were not true. Then it must be that

$$t_i < \infty \text{ for } i = 1, \ldots, 4 \quad \text{and} \quad v^5 \neq 0. \tag{37}$$

Also, for $i = 1, \ldots, 5$, since $t_{i-1} \geqslant t_0$ it follows from Lemma 4.6 that $J_i \neq \emptyset$. Recall the definition $\tau_0 \doteq \inf\{t \geqslant 0 : \phi(t) = 0\}$. We conclude that $|J_i| \neq 3$ for $i = 1, \ldots, 4$, because that would imply that $v^i = 0$ and $t_i = \infty$, which would violate (37). Now suppose $|J_i| = 2$ for some $i \in \{1, 2, 3\}$. If $v^i_j \geqslant 0$ for $j \notin J_i$ then ϕ must have a velocity of the form

ae_j, $a \geqslant 0$, and thus $t_i = \infty$. Since this would contradict (37), it must be that $v^i_j < 0$ for $j \notin J_i$, which implies that $\phi(t_i) = 0$. After time $t_i \phi$ will move away from the origin in some direction v^{i+1}. Since $\phi(t) \in \mathbb{R}^3_+$ this requires that $v^{i+1}_j \geqslant 0$ for $j = 1, 2, 3$, which in turn implies $t_{i+1} = \infty$ and hence contradicts (37).

Thus we conclude that if ϕ is not 4-piecewise affine on $[t_0, \infty)$, we must have $|J_i| = 1$ for $i = 1, 2, 3$. From the uniqueness of the SP and the continuity of ϕ it is clear that in order for this to happen, $J_1 \neq J_2 \neq J_3$, and so (by relabeling indices if necessary) we can assume that $J_i = \{i\}$ for $i = 1, 2, 3$. Note that in that case

$$I(\phi(t_1)) = \{1, 2\}, \quad I(\phi(t_2)) = \{2, 3\} \quad \text{and} \quad I(\phi(t_3)) = \{3, 1\}. \tag{38}$$

By Lemma 4.6 this implies that $u \leqslant 0$ and the definition of the SP and (38) dictate that for $i = 1, 2, 3$,

$$v^i = u + (-u_i) d_i \quad \text{and} \quad v^i_{i+1} < 0, \tag{39}$$

and for $i = 2, 3$

$$v^i_{i-1} > 0, \tag{40}$$

where we identify 0 with 3 and 1 with 4. Moreover, for $i = 2, 3$ $u_i < 0$ since (39) and (40) imply that if $u_i = 0$, then $v^i = u$ and $u_{i-1} > 0$, which contradicts the fact that $u \leqslant 0$.

Since for $t \in (t_3, t_4) \phi(t) \notin (\mathbb{R}^3_+)^\circ$ and (by (37)) $\phi(t) \neq 0$, it follows that $J_4 \neq \{1, 2, 3\}$ and $J_4 \neq \emptyset$. Suppose $|J_4| = 2$. Then since $I(\phi(t_3)) = \{3, 1\}$ by (38), ϕ remains in the set $\{ae_2,\ a \geqslant 0\}$, and thus $J_4 = \{3, 1\}$. For (37) to hold and $\phi(t) \in \mathbb{R}^3_+$ we must have $v^4_2 < 0$ (or else $t_4 = \infty$) and (since $v^4_2 < 0$ implies $\phi(t_4) = 0$) $v^5 \geqslant 0$, $v^5 \neq 0$. This implies that after $t_4 \phi$ must travel away from the origin along a ray. Since ϕ cannot return to the interior of $\mathbb{R}^3_+$, it must either move along an edge or the relative interior of a face. In the former case, (38) shows that ϕ will intersect itself transversally. However, this contradicts the uniqueness of the SP. In the latter case, (39) shows that ϕ will have two different velocities on the same face, which contradicts Lemma 4.5. Thus we conclude that $J_4 \neq \{3, 1\}$, and consequently it must be that $|J_4| = 1$. Since $J_3 = \{3\}$, $I(\phi(t_3)) = \{3, 1\}$ and $J_4 \neq J_3$, we know in fact that $J_4 = \{1\}$. Furthermore, once again by Lemma 4.5, it follows that $v^4 = v^1$, and since v^4_3 must be positive, (40) and the strict inequality $u_i < 0$ now hold also for $i = 1$ in addition to $i = 2, 3$. Since $u < 0$, the identity in (39) along with the inequality (40) implies that for $i = 1, 2, 3$, $(d_i)_{i-1} > 0$ and $u_{i-1} + (-u_i)(d_i)_{i-1} > 0$, which in turn requires that

$$|(d_i)_{i-1}| > \frac{|u_{i-1}|}{|u_i|},$$

and therefore that

$$|(d_2)_1 (d_3)_2 (d_1)_3| > 1.$$

However, this contradicts (36), and so ϕ must either be 4-piecewise affine on $[t_0, \infty)$ or 5-piecewise affine on $[t_0, \infty)$ with $\phi(t_4) = 0$ and $v^5 = 0$. This concludes the proof. □

The last theorem established that in three dimensions, ϕ can have at most five changes of slope, and in fact that the graph of ϕ can have at most 4 changes of slope (see the remark after Definition 4.2). The following example shows that this is tight in the sense that there exist x, u and D satisfying Condition 2.4 for which $\phi = \Gamma(x + u\iota)$ has precisely five changes of slope.

A trajectory $\phi \doteq \Gamma(x + u\iota)$ *in* $\mathbb{R}^3_+$ *with five changes of slope.* Let Γ be the SM associated with the constraint matrix

$$D = \begin{bmatrix} 1 & \frac{1}{10} & -\frac{1}{3} \\ -\frac{1}{3} & 1 & \frac{1}{10} \\ -\frac{1}{3} & -\frac{1}{3} & 1 \end{bmatrix},$$

and note that since $|I - D|$ is substochastic, Condition 2.4 is automatically satisfied. Let $u = -(1, 11, 111)$, $x = (1, 21, 211)$ and let $\phi \doteq \Gamma(x + u\iota)$. We show below that ϕ has four changes of slope. The first piece of the trajectory ϕ clearly has velocity u and lies in the interior $(\mathbb{R}^3_+)^\circ$, and the first change of slope occurs at $t_0 = 1$, when ϕ hits $(0, 10, 100)$. It then moves into the relative interior of the face $\{x \in \mathbb{R}^3 : x_1 = 0\}$ with projected velocity $v^1 = (0, -34/3, -334/3)$, and its second change of slope occurs at $t_1 = 1 + 30/34$, when it reaches the point $(0, 0, 30/17)$. It then moves along the face $\{x \in \mathbb{R}^3 : x_2 = 0, x_1 > 0, x_3 > 0\}$ at a velocity $v^2 = (1/10, 0, -344/3)$. Since $v^2_1 > 0$ and $v^2_3 < 0$, it is easy to see that there exists $t_2 > t_1$ such that $\phi(t_2) \in \{x : x_2 = x_3 = 0\}$. The fourth piece of ϕ then acquires a velocity $v^3 = (-114/3, 1/10, 0)$ and moves along the relative interior of the face $\{x \in \mathbb{R}^3 : x_3 = 0\}$. As before, since $v^3_1 < 0$ and $v^3_2 > 0$, there must exist $t_3 > t_2$ such that $I(\phi(t_3)) = \{1, 3\}$. Note that the radial homogeneity of the SM and the fact that $v^1_1 < 0$ shows that at t_3 ϕ cannot move into the relative interior of the face $\{x \in \mathbb{R}^3 : x_1 = 0\}$. Using the definition of the SP it is easy to see that the fifth piece of ϕ moves along the face $\{x \in \mathbb{R}^3 : x_1 = x_3 = 0, x_2 > 0\}$ at a velocity $v^4 = (0, -509/40, 0)$ till it reaches the origin at some $t_4 < \infty$. It then stays there for all $t \geqslant t_4$ since $u \in \mathscr{C}^\circ$ (or equivalently $D^{-1}u < 0$) implies $\dot{\phi}(t) = v^5 = 0$ for all $t > t_4$.

It is interesting to contrast this three-dimensional example, which satisfies Condition 2.4, with three-dimensional examples that satisfy Conditions 2.2 or 2.3. Under Condition 2.2 we are guaranteed that ϕ would have no more than four changes of slope. Under Condition 2.3 we are guaranteed fewer than seven changes of slope, but since Condition 2.4 implies Condition 2.3 it is clear from Theorem 4.8 that in dimension 3 this bound is not tight.

4.3.4. Sufficient conditions for no cycling of large deviation trajectories

As a corollary of the results proved in the last few sections, we now state conditions under which the large deviation minimizing trajectories of a (b, A, D) RBM do not cycle.

Theorem 4.9. *Consider a* (b, A, D) *RBM* X *satisfying Condition* 2.5 *and the skew-symmetry relation* (21). *If* $D \in \mathbb{R}^{N \times N}$ *satisfies either Condition* 2.2 *or* 2.3, *then for every* $x \in \mathbb{R}^N_+$ *the minimizing trajectory* ϕ^* *for the large deviation rate function* $V(x)$ *of* X *is* N*-piecewise affine or* 2^N*-piecewise affine, respectively. Moreover, if* $D \in \mathbb{R}^{3 \times 3}$ *and satisfies Condition* 2.4, *then for any* $x \in \mathbb{R}^3_+$ *the minimizing trajectory* ϕ^* *for the rate function* $V(x)$ *is* 5*-piecewise affine.*

Proof. First note that if D satisfies either Condition 2.2, 2.3 or 2.4, then the product form time-reversed constraint matrix $\bar{D} = MD'M^{-1}$ defined in (22) also satisfies the same condition. The theorem then follows directly as a consequence of Theorems 3.6, 4.1, 4.4 and 4.8. □

Remark. Note that minimizing large deviation trajectories of RBM's in $\mathbb{R}^3$ can actually have five changes of slope. Indeed, if D, u and ϕ are as described in the example at the end of Section 4.3.3, consider the product form $(b, A, \tilde{D})$ RBM defined by $\tilde{D} \doteq D'$, $A \doteq (\tilde{D} + \tilde{D}')/2$ and $b \doteq \tilde{D}(\tilde{D}')^{-1}u$. Then by Theorem 4.8 the minimizing large deviation trajectory for this RBM is equal to the time-reversal of ϕ and hence has five changes of slope.

5. Applications to a 3-D queueing network

In this section we demonstrate how the results of this paper can be used to find the most probable manner in which buffers overflow in a queueing network during periods of high congestion. Specifically we consider the single class three-buffer open queueing network example illustrated in Fig. 2. As mentioned in the introduction, it was shown in Reiman (1984) that under fairly general assumptions on the distributions of the arrival, service and routing processes, the queue length process associated with an open single-class queueing network can be approximated in heavy traffic by a RBM whose constraint matrix satisfies Condition 2.2. In particular, explicit expressions for the drift, covariance matrix and constraint matrix of the RBM were derived in terms of the primitive data associated with the queueing network. Since here we are more interested in the structure of the approximating RBM's, we only give a very rough idea of how the parameters of the RBM are derived from the original queueing network, referring the reader to Reiman (1984) for more details.

Heavy traffic RBM approximation for a 3-D open network. Consider a queueing network with N stations, and suppose that in the heavy traffic limit, the long term average service time and the long term exogenous average interarrival time for station i are denoted by $1/\mu_i$ and $1/\lambda_i$, respectively. Similarly, let s_i and a_i, respectively, denote

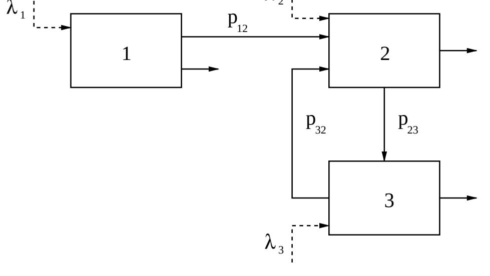

Fig. 2. A three-dimensional queueing network.

the corresponding limits of the long term variances of the interarrival and service times at station i. We let P denote the routing matrix of the network, where P_{ij} can be interpreted as the long term average fraction of customers serviced at station i that get routed to station j (with a fraction $1 - \sum_j P_{ij}$ leaving the network). The fact that it is an open network follows from the fact that the routing matrix is substochastic (i.e. $\sum_j P_{ij} < 1$). When the network is heavily loaded, the (long term) net average input rate nearly equals the net average capacity at each station. Thus in the heavy traffic limit (see Reiman, 1984, Condition (24)), we must have

$$\lambda = (I - P')\mu. \tag{41}$$

It was shown in Reiman (1984) that (under suitable assumptions) the queueing network in heavy traffic can be approximated by a RBM whose constraint matrix is given by $D = I - P'$.

For concreteness, in what follows we consider the particular three-dimensional network illustrated in Fig. 2, whose routing matrix is given by

$$P \doteq \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{4} & 0 \end{bmatrix}.$$

Then the constraint matrix for the approximating RBM is equal to

$$D = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{4} \\ 0 & -\frac{1}{3} & 1 \end{bmatrix},$$

and it is easy to check that D satisfies Condition 2.2. Finally we assume without loss of generality that $\mu_i \doteq 1$ for $i = 1, 2, 3$, and note that by (41) it follows that $\lambda = D\mu = (1, 1/4, 2/3)$.

Time-reversal of a product form approximating RBM. We now assume that the limit interarrival and service terms are independent and identically distributed according to an exponential distribution. This implies that $\lambda_i^2 a_i = \mu_i^2 s_i = 1$, and from Reiman (1984, (27) and (28)) and (41) we conclude that the covariance matrix A_{P} for the approximating RBM satisfies $(A_{\mathrm{P}})_{ii} = 2\mu_i$, $(A_{\mathrm{P}})_{ij} = -[\mu_i p_{ij} + \mu_j p_{ji}]$ for $i \neq j$, $i, j = 1, 2, 3$. Thus for the particular example considered above we have

$$A_{\mathrm{P}} = \begin{bmatrix} 2 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 2 & -\frac{7}{12} \\ 0 & -\frac{7}{12} & 2 \end{bmatrix}.$$

It is easy to verify that in this case (A_{P}, D) satisfy the skew-symmetric condition (21), and thus the approximating RBM is of product form.

Using relation (22) we see that the time-reversed constraint matrix $\bar{D}_{\mathrm{P}}$ in the product form case is equal to

$$\bar{D}_{\mathrm{P}} = M_{\mathrm{P}} D' (M_{\mathrm{P}})^{-1} = D' = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{3} \\ 0 & -\frac{1}{4} & 1 \end{bmatrix},$$

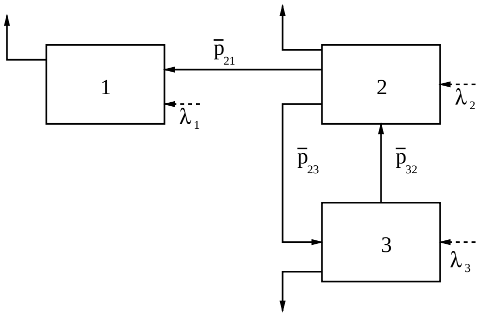

Fig. 3. The time-reversed queueing network (product form case).

where recall that M_P is the diagonal matrix whose diagonal entries coincide with those of A_P. Note that $\bar{D}_P$ satisfies Condition 2.2 and $I-\bar{D}'_P$ is sub-stochastic and nonnegative. Thus $\bar{D}_P$ admits an interpretation as the constraint matrix of the RBM approximation of a "time-reversed" queueing network with routing matrix

$$\bar{P}_P = I - \bar{D}'_P = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & \frac{1}{3} & 0 \end{bmatrix}.$$

The associated "time-reversed" queueing network is illustrated in Fig. 3. Note that this network is simply the original network with the routing arrows reversed. As we will see below, this simple correspondence holds only in the product form case.

Minimizing large deviation trajectories for the product form RBM. We now calculate the minimizing large deviation trajectories for the product-form RBM that approximates the queueing process of Fig. 2. From (Reiman, 1984, expression (24)), in constructing our example we can choose the drift independently of the other system parameters A_P and D. Here we assume that the drift b is given by

$$b \doteq [-44, -44, -44]'.$$

Straightforward calculations show that $D^{-1}b = -[44, 84, 72] < 0$, and hence Condition 2.5 is satisfied. Using (25) we infer that the constant slope of the optimal control in the product form time-reversed variational problem (23) is given by

$$\bar{\beta}^* = \bar{D}_P D^{-1} b = [-2, -60, -51]'.$$

Now fix $x = (10, 10, 10)$, let $\bar{\psi}^* \doteq x + \bar{\beta}^* \iota$ and let $\bar{\phi}^* \doteq \bar{\Gamma}_P(\bar{\psi}^*)$, where $\bar{\Gamma}_P$ is the SM associated with $\bar{D}_P$. Then simple calculations show that

$$\bar{\phi}^*(t) = \begin{cases} (10,10,10) + t(-2,-60,-51) & \text{for } t \in [0, t_1], \\ (29/3, 0, 3/2) + (t-t_1)(-32, 0, -66) & \text{for } t \in [t_1, t_2], \\ (295/33, 0, 0) + (t-t_2)(-44, 0, 0) & \text{for } t \in [t_2, t_3], \\ (0,0,0) & \text{for } t \in [t_3, \infty), \end{cases}$$

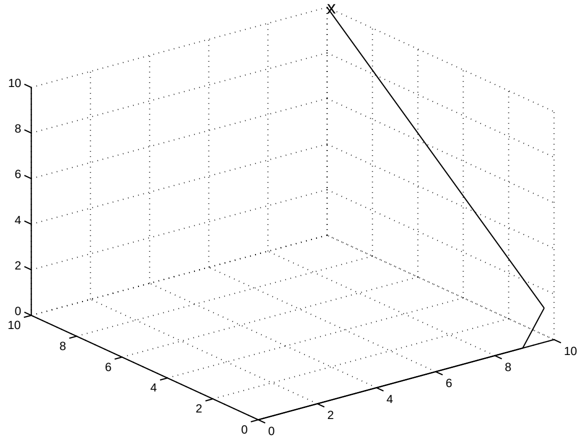

Fig. 4. The minimizing large deviation trajectory ($x = (10, 10, 10)$).

where $t_1 = 1/6$, $t_2 = t_1 + 1/44$ and $t_3 = t_2 + 295/1452$. The minimizing trajectory $\bar{\phi}^*$ is illustrated in Fig. 4. Note that it is 4-piecewise affine with $v^4 = 0$, as predicted by Theorem 4.4. Moreover, by Theorem 3.6 the most likely way in which the RBM reaches a level $(10, 10, 10)$ is given by the path $\phi^*(\cdot) \doteq \bar{\phi}^*(t_3 - \cdot)$. Since the large deviation behavior of the RBM closely approximates that of the queueing network for open single-class networks, this shows that for all the scaled buffer contents to reach a common high level of 10, buffer 1 first increases, and then buffer 3, and finally buffer 2 also joins in and all three buffers build up till they reach the level 10.

Time-reversal of a non-product form approximating RBM. We once again consider the network in Fig. 2, with all parameter values being the same as in the product form case except that the coefficient of variation $\mu_1\sqrt{s_1}$ for the service times at the first station is now $\sqrt{2}$ instead of 1. Once again, from Reiman (1984, (27) and (28)) and relation (41) it follows that

$$A = \begin{bmatrix} 3 & -1 & 0 \\ -1 & \frac{9}{4} & -\frac{7}{12} \\ 0 & -\frac{7}{12} & 2 \end{bmatrix}.$$

Using the expression $\bar{D} = -D + 2AM^{-1}$ in (5), we see that the time-reversed constraint matrix is given by

$$\bar{D} = \begin{bmatrix} 1 & -\frac{8}{9} & 0 \\ -\frac{1}{6} & 1 & -\frac{1}{3} \\ 0 & -\frac{5}{27} & 1 \end{bmatrix}.$$

Since $I - \bar{D}$ is nonnegative and substochastic it clearly satisfies Condition 2.2, and so the time-reversed non-product form RBM can be associated with another single-class

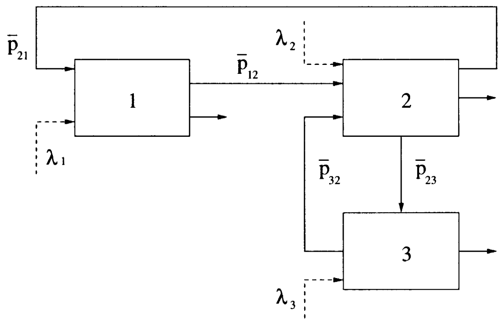

Fig. 5. The time-reversed queueing network (general case).

open queueing network which has routing matrix $\bar{P} \doteq I - \bar{D}'$. By Theorem 3.5 we see that the most likely way in which the RBM approximation to the original queueing network in Fig. 2 reaches a point $x \in \mathbb{R}_+^N$ is obtained as the time-reversal of the solution to the variational problem that corresponds to the network in Fig. 5, where optimality is measured with respect to the quadratic cost function specified in (17).

As illustrated in Fig. 5, note that unlike the product form case, the time-reversed network here is clearly not obtained by just reversing the routes of the original network. Indeed, in general $I-\bar{D}$ may not even be nonnegative and thus may have no interpretation in terms of an open single-class queueing network. (This is the case, for example, if one assumes that $\lambda_1 s_1^2 = 2$, while all other interarrival and service distributions are exponential.)

The presence of the $\bar{L}$ term in the time-reversed variational problem in (17) suggests a connection with a large deviation limit for a time-reversed network. However, for non-product form RBMs one cannot always associate a single class network with the time-reversed variational problem. In any case (to the authors' knowledge) no rigorous connection has been established between the variational problem in (17) and any time-reversal of the original stochastic network.

References

Avram, F., Dai, J.G., Hasenbein, J.J., 2000. Explicit solutions for variational problems in the quadrant. Special issue of Queueing Systems: Stochastic Models with Tractable Steady State Characteristics 37, 261–291.

Bernard, A., El Kharroubi, A., 1991. Régulation de processus dans le premier orthant de $\mathbb{R}^n$. Stochast. Stochast. Rep. 34, 149–167.

Berman, A., Plemmons, R.J., 1979. Nonnegative Matrices in the Mathematical Sciences. Academic Press, New York.

Budhiraja, A., Dupuis, P., 1999. Simple necessary and sufficient conditions for the stability of constrained processes. SIAM J. Appl. Math. 59, 1686–1700.

Dai, J.G., Harrison, J.M., 1992. Reflected Brownian motion in an orthant: numerical methods for steady-state analysis. Ann. Appl. Probab. 2, 65–86.

Dupuis, P., Ishii, H., 1991. On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications. Stochastics 35, 31–62.

Dupuis, P., Ramanan, K., 1999a. Convex duality and the Skorokhod Problem I. Probab. Theor. Related Fields 115, 153–195.
Dupuis, P., Ramanan, K., 1999b. Convex duality and the Skorokhod Problem II. Probab. Theor. Related Fields 115, 197–236.
Dupuis, P., Ramanan, K., 2000. An explicit formula for the solution of certain optimal control problems on domains with corners. Theor. Probab. Math. Statist. 63.
Freidlin, M.I., Wentzell, A.D., 1984. Random Perturbations of Dynamical Systems. Springer, New York.
Harrison, J.M., Landau, H., Shepp, L., 1985. The stationary distribution of reflected Brownian motion in a planar region. Ann. Probab. 13, 744–757.
Harrison, J.M., Reiman, M.I., 1981. Reflected Brownian motion on an orthant. Ann. Probab. 9, 302–308.
Harrison, J.M., Williams, R., 1987a. Brownian models of open queueing networks with homogeneous customer populations. Stochastics 22, 77–115.
Harrison, J.M., Williams, R., 1987b. Multidimensional reflected Brownian motions having exponential stationary distributions. Ann. Probab. 15, 115–137.
Kieffer, G., 1995. The large deviation principle for two-dimensional stable systems. PhD Thesis, University of Massachusetts.
Kushner, H.J., 2001. Heavy Traffic Analysis of Controlled and Uncontrolled Queueing and Communication Networks. Springer, New York, Berlin.
Kushner, H., Dupuis, P., 1992. Numerical methods for stochastic control problems in continuous time. Applications of Mathematics, Vol. 24. Springer, New York.
Majewski, K., 1998a. Heavy traffic approximations of large deviations of feedforward queueing networks. Queueing Systems 28, 125–155.
Majewski, K., 1998b. Large deviations of the steady-state distribution of reflected processes with applications to queueing systems. Queueing Systems 29, 351–381.
Majewski, K., 2000. Pathwise heavy-traffic convergence of single-class queueing networks. Preprint.
Mandelbaum, A., Van der Heyden, L., 1987. Complementarity and reflection. Unpublished work.
O'Connell, N., 1998. Large deviations for queue lengths at a multi-buffered resource. J. Appl. Probab. 35, 240–245.
Paschalidis, I., 1996. Large deviations in high speed communication networks. PhD Thesis. MIT, Cambridge, MA.
Peterson, W.P., 1991. Diffusion approximations for networks of queues with multiple customer types. Math. Oper. Res. 9, 90–118.
Reiman, M.I., 1984. Open queueing networks in heavy traffic. Math. Oper. Res. 9, 441–458.
Skorokhod, A.V., 1961. Stochastic equations for diffusion in a bounded region. Theor. Probab. Appl. 6, 264–274.
Varadhan, S.R.S., 1984. Large deviations and Applications. CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.
Willams, R.J., 1985. Recurrence classification and invariant measure for reflected Brownian motion in a wedge. Ann. Probab. 13, 758–778.
Williams, R.J., 1987. Reflected Brownian motion with skew symmetric data in a polyhedral domain. Probab. Theor. Rel. Fields 75, 459–485.
Williams, R.J., 1996. On the approximation of queueing networks in heavy traffic. In: Kelly, F.P., Zachary, S., Ziedins, I. (Eds.), Stochastic Networks: Theory and Applications, Proceedings of the Royal Society Workshop, Edinburgh, 1995, OUP, pp. 35–56.