Queueing Systems 37, 259–289, 2001

Explicit Solutions for Variational Problems in the Quadrant

F. AVRAM F.Avram@ma.hw.ac.uk
Department of Statistics and Actuarial Science, Heriot Watt University, Edinburgh, Scotland

J.G. DAI dai@isye.gatech.edu
School of Industrial and Systems Engineering and School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA

J.J. HASENBEIN jhas@mail.utexas.edu
Graduate Program in Operations Research and Industrial Engineering, Department of Mechanical Engineering, University of Texas at Austin, Austin, TX 78712-1063, USA

Received 1 May 1999; revised 3 January 2000

Abstract. We study a variational problem (VP) that is related to semimartingale reflecting Brownian motions (SRBMs). Specifically, this VP appears in the large deviations analysis of the stationary distribution of SRBMs in the d-dimensional orthant $\mathbb{R}^d_+$. When $d = 2$, we provide an explicit analytical solution to the VP. This solution gives an appealing characterization of the optimal path to a given point in the quadrant and also provides an explicit expression for the optimal value of the VP. For each boundary of the quadrant, we construct a "cone of boundary influence", which determines the nature of optimal paths in different regions of the quadrant. In addition to providing a complete solution in the 2-dimensional case, our analysis provides several results which may be used in analyzing the VP in higher dimensions and more general state spaces.

Keywords: variational problems, large deviations, Skorohod problems, reflecting Brownian motions, queueing networks

AMS subject classification: 60F10, 60I65, 60K25, 68M20

1. Introduction

Semimartingale reflecting Brownian motions (SRBMs) in the orthant have been proposed as approximate models of open queueing networks (see, e.g, [16]). Such diffusion processes were first introduced in [17]. Since then, there have been two primary lines of active research related to SRBMs. One line has concentrated on proving limit theorems that justify the Brownian model approximations of queueing networks under heavy traffic conditions. (For recent surveys, see [7,37].) The other focus has been to study the fundamental and analytical properties, including recurrence conditions, of SRBMs. (For a survey, see [36].) The topic of this paper is related to the latter category.

The focus of our paper is a variational problem (VP) which arises from the study of SRBMs. The rare event behavior of the stationary distributions of SRBMs can be analyzed with the help of a large deviations principle (LDP). When such a principle holds, the optimal value of the VP describes the tail behavior of the stationary distribution and the corresponding optimal paths characterize *how* certain rare events are most likely to occur. Below, we provide some motivation both for studying the stationary distributions of SRBMs and in particular for examining the rare event behavior.

The stationary distribution of SRBMs has been a primary object of study because it provides estimates of congestion measures in corresponding queueing networks. Unfortunately, even these Brownian approximations are not immediately tractable. In fact, Harrison and Williams [18,19] showed that the stationary density function admits a separable, exponential density if and only if the covariance and reflection matrices satisfy a certain skew-symmetry condition. When this condition is not satisfied, one must generally resort to developing numerical algorithms to estimate the stationary distribution of the SRBM. One such algorithm has been devised by Dai and Harrison [8]. If one knows the tail behavior of the stationary distribution for the SRBMs, such algorithms can be made to be more efficient. Furthermore, in a recent paper of Majewski [28] it was demonstrated that, roughly speaking, one may switch the heavy traffic and large deviations limits in feed-forward networks, indicating that the rare event behavior of an SRBM can give insight into the rare event behavior of an associated heavily loaded queueing network. This provides ample motivation for studying the *large deviations* theory of SRBMs.

The study of LDPs can be roughly divided into two primary topics:

(i) proving that an LDP holds for a class of processes, and

(ii) analyzing the variational problem which arises from the LDP.

Our study is concerned only with the second topic, but we provide some discussion of the first topic, both here and in the body of the paper, since there is a close relationship between the two. As noted above, the VPs studied in this paper are related to the *stationary distribution* of an SRBM. Thus, the LDP corresponds to SRBM on the entire time interval $[0, \infty)$. When considering LDPs on a finite time interval, the large deviations principle is easier to establish. For example, when the reflection matrix is an $\mathcal{M}$-matrix, as the one used in [17], the SRBM can be defined through a reflection mapping which is Lipschitz continuous on $[0, T]$ for each $T > 0$. In this case, the LDP for the SRBM readily follows from the contraction principle of large deviations theory, as demonstrated in [12]. To investigate large deviations theory for the stationary distribution of an SRBM, we must consider SRBMs on the interval $[0, \infty)$, which complicates matters considerably. However, LDP for SRBMs of this type have been established in special cases. In particular, Majewski has established such an LDP for a stationary SRBM when the reflection matrix R is an $\mathcal{M}$-matrix [29] and when the reflection matrix has a special structure arising from feed-forward queueing networks [26]. For a general stationary SRBM, establishing an LDP remains an open problem (see conjecture 4.1), even in two dimensions.

The major thrust of this paper is to investigate the VP which arises for the aforementioned LDPs. Our analysis provides a complete, explicit solution to the VP when the state space is the 2-dimensional orthant $\mathbb{R}^2_+$. In particular, we characterize the optimal paths to a given point $v \in \mathbb{R}^2_+$. It turns out that the optimal path to v is influenced by a boundary if v is contained within a cone associated with that boundary. We identify precisely each of these "cones of boundary influence". When v is not in either of the cones, the optimal path is a direct, linear path. When v is contained in one of these cones, however, the optimal path first travels along a boundary, and then travels directly to v. Furthermore, such a path leaves the boundary and enters the interior at a *unique* entrance angle which can be determined directly from the problem data. For VPs which arise from a large deviations analysis of random walks in the quadrant, Ignatyuk et al. [21] demonstrated that similar behavior is manifested in the optimal paths. Specifically, they are able to identify analogous regions of boundary influence in the solutions to such VPs.

Another work closely related to our study is [29]. In this paper, Majewski examines a general class VPs in high dimensions and among other things, provides a general purpose numerical algorithm to solve these VPs. Our work complements his numerical work for the 2-dimensional case. Also, a crucial assumption in implementing such numerical algorithms is that the optimal path consists of a finite number of linear pieces. An intermediate result in our paper shows that an optimal path, in two dimensions, consists of at most two linear pieces.

It seems clear from the literature (see, e.g., [1]) that explicitly characterizing the solution to VPs which arise from queueing networks or other systems is significantly harder in three dimensions or higher versus the 2-dimensional case. Although this paper primarily focuses on the 2-dimensional problem in $\mathbb{R}^2_+$, it should be noted that several of our results indeed hold in higher dimensions and for general polygonal state spaces. In addition to this more direct connection to higher dimensional problems, we hope that the problem framework which we establish in the SRBM setting will provide motivation for further research into the interesting and challenging open problems beyond the 2-dimensional case.

There is a large body of literature on LDPs for random walks and queueing networks. The book [33] by Shwartz and Weiss contains an excellent list of references. We can only provide a short survey of the latest works which are most closely related to our study. Recent work on LDPs for queueing networks include [13,30,31] on multi-buffer single-server systems, [4] on acyclic networks, [22] on 2-station networks with feedback, and [1,10] on general queueing networks with feedback. The works by Ignatyuk et al. [21] and by Borovkov and Mogulskii [5] investigate random walks that are constrained to an orthant. Knessl and Tier [23–25] used a perturbation approach to study rate functions for some queueing systems.

We now provide a brief outline of the paper. In section 2, we introduce the Skorohod problem and the VP. The main result of this paper (theorem 3.1) is stated in section 3. In section 4 we introduce semimartingale reflecting Brownian motions and the large deviations principle that connects the VP with the SRBM. We examine the VP in depth in section 5 and characterize the optimal escape paths in section 6. Finally, we provide some examples in section 7.

2. The Skorohod and variational problems

Let $d \geqslant 1$ be an integer. Throughout this paper, θ is a constant vector in $\mathbb{R}^d$, Γ is a $d \times d$ symmetric and strictly positive definite matrix, and R is a $d \times d$ matrix. In this section, we first define the Skorohod problem associated with the matrix R, and then define the variational problem (VP) associated with (θ, Γ, R).

2.1. The Skorohod problem

Let $C([0,\infty), \mathbb{R}^d)$ be the set of continuous functions $x : t \in [0,\infty) \to x(t) \in \mathbb{R}^d$. A function $x \in C([0,\infty), \mathbb{R}^d)$ is called a path and is sometimes denoted by $x(\cdot)$. The space $C([0,\infty), \mathbb{R}^d)$ is endowed with a topology in which convergence means uniform convergence in each finite interval.

We now define the Skorohod problem associated with R and state space $\mathbb{R}^d_+$ (sometimes called an R-regulation). Note that all vector inequalities should be interpreted componentwise and all vectors are assumed to be column vectors.

Definition 2.1 (The Skorohod problem). Let x be a path. An R-regulation of x is a pair of paths $(z, y) \in C([0,\infty), \mathbb{R}^d) \times C([0,\infty), \mathbb{R}^d)$ such that

$$z(t) = x(t) + R\,y(t), \qquad t \geqslant 0, \tag{2.1}$$
$$z(t) \geqslant 0, \qquad t \geqslant 0, \tag{2.2}$$
$$y(\cdot) \text{ is nondecreasing}, \qquad y(0) = 0, \tag{2.3}$$
$$\int_0^\infty z_i(s)\,\mathrm{d}y_i(s) = 0, \qquad i = 1, \ldots, d. \tag{2.4}$$

When the R-regulation (y, z) of x is unique for each $x \in C([0,\infty), \mathbb{R}^d)$, the mapping

$$\psi : x \to \psi(x) = z$$

is called the reflection mapping from $C([0,\infty), \mathbb{R}^d)$ to $C([0,\infty), \mathbb{R}^d_+)$. When the R-regulation of x is not unique, we use $\psi(x)$ to denote the set of all z which are components of an R-regulation (y, z) of x. When the triple (x, y, z) is used, it is implicitly assumed that (y, z) is an R-regulation of x.

Bernard and El Kharroubi [3] proved that there exists an R-regulation for *every* x with $x(0) \geqslant 0$ if and only if R is completely-$\mathcal{S}$ as defined in definition 2.2 below. For a $d \times d$ matrix R and a subset $\mathcal{D} \subset \{1, \ldots, d\}$, the principal submatrix associated with $\mathcal{D}$ is the $|\mathcal{D}| \times |\mathcal{D}|$ matrix obtained from R by deleting the rows and columns that are not in $\mathcal{D}$, where $|\mathcal{D}|$ is the cardinality of $\mathcal{D}$.

Definition 2.2. A $d \times d$ matrix R is said to be an *$\mathcal{S}$-matrix* if there exists a $u \geqslant 0$ such that $Ru > 0$. The matrix R is *completely-$\mathcal{S}$* if each principal submatrix of R is an $\mathcal{S}$-matrix.

2.2. The variational problem

In this section we introduce the variational problem (VP) of interest to us. This problem arises in the study of large deviations for semimartingale reflecting Brownian motions (SRBMs) to be defined in section 4, and we will make this connection in section 4.2.

Recall that $\psi(x)$ maps x to *one* unique path, if the corresponding Skorohod problem has a unique solution. If the Skorohod problem is non-unique, then $\psi(x)$ represents a set of paths corresponding to x. Now, in order to establish a general framework for posing VPs, we wish to include cases for which the Skorohod problem is not unique. For $T > 0$ and $v \in \mathbb{R}^d_+$, we will adopt the following convention. We will take

$$\psi(x)(T) = v$$

to signify that there exists a $z \in \psi(x)$ such that $z(T) = v$. Next, for vectors $v \in \mathbb{R}^d$ and $w \in \mathbb{R}^d$ we define the inner product

$$\langle v, w\rangle = v'\Gamma^{-1}w$$

with the associated norm $\|v\| = \sqrt{\langle v, v\rangle}$.

We are now prepared to present the VP that will be the main focus of this paper.

Definition 2.3 (The Variational Problem – VP).

$$I(v) \equiv \inf_{T\geqslant 0} \inf_{x\in\mathcal{H}^d, \psi(x)(T)=v} \frac{1}{2}\int_0^T \left\|\dot{x}(t) - \theta\right\|^2 \mathrm{d}t, \tag{2.5}$$

where $\mathcal{H}^d$ is the space of all absolutely continuous functions $x(\cdot) : [0, \infty) \to \mathbb{R}^d$ which have square integrable derivatives on bounded intervals and have $x(0) = 0$.

Definition 2.4. Let $v \in \mathbb{R}^d_+$. If a given triple of paths (x, y, z) is such that the triple satisfies the Skorohod problem, $z(T) = v$ for some $T \geqslant 0$, and

$$\frac{1}{2}\int_0^T \left\|\dot{x}(t) - \theta\right\|^2 \mathrm{d}t = I(v),$$

then we will call (x, y, z) an *optimal triple*, for VP (2.5), with optimal value $I(v)$. The function x is called an *optimal path* if it is the first member of an optimal triple and z is called an *optimal reflected path* if it is the last member of an optimal triple. Such a triple (x, y, z) is also sometimes referred to as a *solution* to the VP (2.5).

3. 2-dimensional results

In this section, we state our main theorem, which gives an explicit solution to the VP in terms of the problem data, (θ, Γ, R), for the 2-dimensional case. We introduce much of the notation in this section, but defer the proof until section 6. The proof relies on three components:

1. A reduction of the search for optimal paths to the space of piecewise linear functions with at most two segments.
2. An analysis of "locally" optimal paths with a given structure.
3. A quantitative comparison of the VP value for the various types of locally optimal paths to determine the globally optimal path.

It turns out that the solution to the VP in two dimensions can be stated in an appealing way by defining "cones of boundary influence". Both this solution and the proof method yield insights into higher dimensional VPs.

For the majority of this section, we restrict ourselves to the case $d = 2$. We will use the term *face* F_i, to denote one of the axes in $\mathbb{R}^2_+$:

$$F_i = \{v \in \mathbb{R}^2_+ \colon\ v_i = 0\}.$$

We retain the term face because in later sections we will consider faces of the orthant in higher dimensions. To state our main theorem, we need to define the cone C_i associated with a face F_i, $i = 1, 2$. For a face F_i, each cone C_i defines a region of boundary influence on the solutions to the VP. It turns out that the boundary influence depends on two quantities which we will term the "exit velocity" and the "entrance velocity", which will lead to the concept of "reflectivity" of a face. We define and discuss the relationship between these terms presently.

Let p^i be a vector that is orthogonal (under the usual Euclidean inner product) to the ith column of the reflection matrix R, and is normalized with $\|\Gamma p^i\| = 1$. For example, if

$$R = \begin{pmatrix} 1 & r_2 \\ r_1 & 1 \end{pmatrix}, \tag{3.1}$$

then p^1 will be a multiple of $(-r_1, 1)'$ and p^2 will be a multiple of $(1, -r_2)'$.

Definition 3.1. The *exit velocity* a^i associated with face F_i is defined to be

$$a^i = \theta - 2(\theta' p^i)\Gamma p^i. \tag{3.2}$$

We defer the explanation of this term until later in the section.

Definition 3.2. Face F_i is said to be *reflective* if the ith component of a^i is negative, i.e., $a^i_i < 0$.

When F_i is not reflective, C_i is defined to be empty. In this case, the face F_i has no boundary influence on solutions to the VP for any $v \in \mathbb{R}^2_+$.

When F_i is reflective, the characterization of the cone C_i is more involved. We need to define a key notion, the "entrance velocity" associated with face F_i. It is defined to be the "symmetry" of a^i around face F_i. To make this concept precise, let e^i be a directional vector on face F_i, and n^i be a vector that is normal to F_i, pointing to the interior of the state space. We assume that e^i and n^i are normalized so that $\|e^i\| = 1$ and $\|\Gamma n^i\| = 1$. For example, when $\Gamma = I$, we have $e^1 = (0, 1)'$ and $n^1 = (1, 0)'$. One can check that $\langle e^i, \Gamma n^i\rangle = (e^i)'n^i = 0$. Thus, e^i and Γn^i form an orthonormal basis in $\mathbb{R}^2$ under the inner product $\langle\cdot\,,\cdot\rangle$. Therefore, any vector $v \in \mathbb{R}^2$ has the following (unique) representation:

$$v = \langle v, e^i\rangle e^i + \langle v, \Gamma n^i\rangle \Gamma n^i. \tag{3.3}$$

Thus,

$$a^i = \langle a^i, e^i\rangle e^i + \langle a^i, \Gamma n^i\rangle \Gamma n^i.$$

One can then define a symmetry $\tilde{a}^i$ of a^i around face F_i to be

$$\tilde{a}^i = \langle a^i, e^i\rangle e^i - \langle a^i, \Gamma n^i\rangle \Gamma n^i. \tag{3.4}$$

Definition 3.3. We call the symmetry $\tilde{a}^i$ the *entrance velocity* associated with a face F_i.

One can easily check from the definition that

$$\tilde{a}^i_i = -a^i_i, \quad i = 1, 2, \tag{3.5}$$

thus F_i is reflective if and only if $\tilde{a}^i_i > 0$.

When F_i is reflective, C_i is defined to be the cone generated by e^i and $\tilde{a}^i$, namely,

$$C_i = \left\{\alpha e^i + \beta \tilde{a}^i\colon \text{ for } \alpha \geqslant 0, \beta \geqslant 0\right\}.$$

It is possible that $\tilde{a}^i$ points to the outside of $\mathbb{R}^2_+$, even if F_i is reflective. In this case, $C_i \supset \mathbb{R}^2_+$.

The cone C_i identifies precisely the region in which the face F_i has boundary influence. With two cones, C_1 and C_2, defined, we can partition the state space $\mathbb{R}^2_+$ into three regions: $(\mathbb{R}^2_+ \cap C_1) \setminus C_2$, $(\mathbb{R}^2_+ \cap C_2) \setminus C_1$, and one of the two regions, $\mathbb{R}^2_+ \cap C_1 \cap C_2$ or $\mathbb{R}^2_+ \setminus (C_1 \cup C_2)$. Note that one of the latter two regions is always empty, namely, either $\mathbb{R}^2_+ \cap C_1 \cap C_2 = \emptyset$ or $\mathbb{R}^2_+ \setminus (C_1 \cup C_2) = \emptyset$.

Before we state the main theorem of this paper, we introduce some additional notation and terminology. For a $v \in \mathbb{R}^2_+$, let $\tilde{a}^0(v) = (\|\theta\|/\|v\|)v$. The next two expressions will appear in the locally optimal value of the VP for various cases. For $v \in \mathbb{R}^2_+$, let

$$I^0(v) = \langle \tilde{a}^0(v) - \theta, v\rangle, \tag{3.6}$$

$$I^i(v) = \langle \tilde{a}^i - \theta, v\rangle, \quad i = 1, 2. \tag{3.7}$$

Now we wish to define three triples, (x^i, y^i, z^i) for $i = 0, 1, 2$, which start at the origin and terminate at v. In will turn out that one or more of these triples will be a solution to the VP. The first triple, (x^0, y^0, z^0), is a *direct triple* to v, with $x^0(t) = \tilde{a}^0(v)t$ for $t \geqslant 0$. Since x^0 always stays in $\mathbb{R}^2_+$, the corresponding reflected path $z^0 = x^0$ and $y^0(t) = 0$ for $t \geqslant 0$ comprise an R-regulation of x for any R. One can more generally define a direct triple from w to v. The next two triples, (x^i, y^i, z^i), $i = 1, 2$, are *broken triples* through the corresponding face. For a face F_1, we introduce a broken

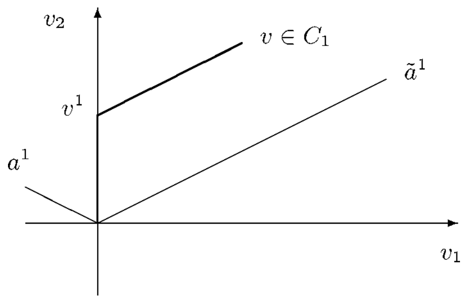

Figure 1. An optimal broken path to $v \in C_1$ through F_1.

triple (x^1, y^1, z^1) from the origin to v through face F_1, which consists of two segments. Each segment of x^1 is linear, and hence we can chose linear y^1 and z^1, within each segment. In the first segment, x^1 has a velocity a^1 such that z^1 stays on the boundary F_1. The segment ends when z^1 reaches $v^1 \in F_1$, where v^1 is uniquely determined by the condition that $v - v^1$ is parallel to $\tilde{a}^1$. The second segment is simply the direct triple traveling in the interior of the state space from v^1 to v, with velocity of x^1 and z^1 being equal to $\tilde{a}^1$. A broken triple through F_2 is defined similarly.

Note that in order for such a broken triple to be well-defined, we must have (i) that F_1 is reflective and (ii) that $v \in C_1$ (see figure 1). In such a case, a^1 is the velocity at which x exits the state space, and $\tilde{a}^1$ is the velocity of x when z enters the interior of the state space. The terms exit velocity and entrance velocity are introduced primarily to define these broken triples, although the interpretations above are not always meaningful when a face is nonreflective.

Now we are prepared to state our main theorem, which completely characterizes the solutions to the VP presented in the previous section, for the 2-dimensional case.

Theorem 3.1. Consider the VP as defined in (2.5) with associated data (θ, Γ, R), with R taking the form (3.1). Let $v \in \mathbb{R}^2_+$ and suppose that R is completely-$\mathcal{S}$ and that the data satisfies the conditions

$$\theta_1 + r_2\theta_2^- < 0, \tag{3.8}$$

$$\theta_2 + r_1\theta_1^- < 0, \tag{3.9}$$

where, for an $a \in \mathbb{R}$, $a^- = \max\{-a, 0\}$. Then

(a) If $v \notin C_1 \cup C_2$, the optimal value is given by $I^0(v)$ and the direct triple (x^0, y^0, z^0) is optimal.

(b) If $v \in C_1 \setminus C_2$, the optimal value is given by $I^1(v)$ and the broken triple (x^1, y^1, z^1) through F_1 is optimal.

(c) If $v \in C_2 \setminus C_1$, the optimal value is given by $I^2(v)$ and the broken triple (x^2, y^2, z^2) through F_2 is optimal.

(d) If $v \in C_1 \cap C_2$, the optimal value is given by $\min\{I^1(v), I^2(v)\}$. When $I^1(v) \leqslant I^2(v)$, the broken triple (x^1, y^1, z^1) through F_1 is optimal. Otherwise, the broken triple (x^2, y^2, z^2) through F_2 is optimal.

Conditions (3.8) and (3.9) are the so-called "recurrence conditions" for the corresponding SRBM (see section 4.1 for more discussion). The proof of theorem 3.1 is deferred until section 6. Several preliminary results needed in the proof, but which are also applicable in higher dimensional problems, are given in section 5.

For $i = 1, 2$, $I^i(v)$ is a linear function of v, whereas $I^0(v)$ is not since $\tilde{a}^0(v)$ depends on v. In fact, in lemma 6.1 we will check that $I^0(v) = \|\theta\| \cdot \|v\| - \langle\theta, v\rangle$. Also it will be verified in theorems 6.1 and 6.3 that, for $v \in C_i$,

$$I^i(v) = \frac{1}{2}\int_0^T \left\|\dot{x}^i(t) - \theta\right\|^2 \mathrm{d}t, \quad i = 0, 1, 2,$$

where T is the first time for z^i to reach v, and we let $C_0 = \mathbb{R}_+^2$. More interestingly, we observe that the optimal value $I^i(v)$, $i = 0, 1, 2$, depends on the velocity of the "last segment" of z^i, which is always given by $\tilde{a}^i$.

In $\mathbb{R}^d$ with $d \geqslant 3$, it is possible to have more complicated types of optimal paths. We would now like to outline three principles which are valid for locally optimal paths in higher dimensions. We do not demonstrate the validity of these propositions in this tract, rather leaving this for a subsequent paper.

(i) *Orthogonality law.* If a locally optimal triple (x, y, z) is such that z traverses a face F, x must have a velocity of the form: $a = \theta + \Gamma p$ while z is on F. In the general setting, p is a vector orthogonal (under the Euclidean inner product) to all the reflection vectors of the face.

(ii) *Norm preservation law.* For a broken triple (x, y, z), which in general may travel along several faces, the norm of the intermediate velocity of x must be equal to the norm of the drift θ.

(iii) *Symmetry law.* For a broken triple (x, y, z) along F, the difference of the velocity of x before and after leaving F must be orthogonal to F with respect to the inner product induced by the covariance matrix.

With these general principles in hand, it is possible to compute locally optimal broken paths of any chosen type (i.e., any prescribed order of traversing the faces). One can then compare the values of each such "locally optimal" path to discover the globally optimal path for a given point. This essentially reduces the resolution of the VP to a numerical task. Unfortunately, to be sure that this numerical task will indeed yield the optimal value for the general VP, one must establish a principle as outlined in step 1 at the beginning of section 3. Such a principle has been established for $\mathbb{R}_+^2$ (see theorem 5.1), but is lacking for more general state spaces.

4. Semimartingale reflecting Brownian motions and large deviations

In this section, we define semimartingale reflecting Brownian motions (SRBMs). Such processes arise in the study of heavy traffic approximations to multiclass queueing networks (see, e.g., [16]). We also discuss conditions under which an SRBM is positive recurrence. Finally, we introduce large deviations principles (LDPs) for an SRBM. The LDPs connect the VP introduced in (2.5) with a corresponding set of SRBMs and provide the motivation for our study of such VPs.

4.1. SRBM

Throughout this section, $\mathcal{B}$ denotes the σ-algebra of Borel subsets of $\mathbb{R}^d_+$. Recall that θ is a constant vector in $\mathbb{R}^d$, Γ is a $d \times d$ symmetric and strictly positive definite matrix, and R is a $d \times d$ matrix. We shall define an SRBM associated with the data $(\mathbb{R}^d_+, \theta, \Gamma, R)$. For this note, a triple $(\Omega, \mathcal{F}, \{\mathcal{F}_t\})$ will be called a *filtered space* if Ω is a set, $\mathcal{F}$ is a σ-field of subsets of Ω, and $\{\mathcal{F}_t\} \equiv \{\mathcal{F}_t,\ t \geqslant 0\}$ is an increasing family of sub-σ-fields of $\mathcal{F}$, i.e., a filtration. If, in addition, $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$, then $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ will be called a filtered probability space.

Definition 4.1 (SRBM). Given a probability measure ν on $(\mathbb{R}^d_+, \mathcal{B})$, a *semimartingale reflecting Brownian motion* (abbreviated as SRBM) associated with the data $(\mathbb{R}^d_+, \theta, \Gamma, R, \nu)$ is an $\{\mathcal{F}_t\}$-adapted, d-dimensional process Z defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}_\nu)$ such that

(i) $\mathbb{P}_\nu$-a.s., Z has continuous paths and $Z(t) \in \mathbb{R}^d_+$ for all $t \geqslant 0$,

(ii) $Z = X + RY$, $\mathbb{P}_\nu$-a.s.,

(iii) under $\mathbb{P}_\nu$,

- (a) X is a d-dimensional Brownian motion with drift vector θ, covariance matrix Γ and $X(0)$ has distribution ν,
- (b) $\{X(t) - X(0) - \theta t, \mathcal{F}_t,\ t \geqslant 0\}$ is a martingale,

(iv) Y is an $\{\mathcal{F}_t\}$-adapted, d-dimensional process such that $\mathbb{P}_\nu$-a.s. for each $j = 1, \dots, d$,

- (a) $Y_j(0) = 0$,
- (b) Y_j is continuous and nondecreasing,
- (c) Y_j can increase only when Z is on the face $F_j \equiv \{x \in \mathbb{R}^d_+\colon\ x_j = 0\}$, i.e., $\int_0^\infty Z_j(s)\,\mathrm{d}Y_j(s) = 0$.

An SRBM associated with the data $(\mathbb{R}^d_+, \theta, \Gamma, R)$ is an $\{\mathcal{F}_t\}$-adapted, d-dimensional process Z together with a family of probability measures $\{\mathbb{P}_x, x \in \mathbb{R}^d_+\}$ defined on some filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\})$ such that, for each $x \in \mathbb{R}^d_+$, (i)–(iv) hold with $P_\nu = P_x$ and ν being the point distribution at x.

A $(\mathbb{R}^d_+, \theta, \Gamma, R, \nu)$-SRBM Z has a fixed initial distribution ν, whereas a $(\mathbb{R}^d_+, \theta, \Gamma, R)$-SRBM has no fixed start point. The latter fits naturally within a Markovian process framework. Condition (iv(c)) is equivalent to the condition that, for each $t > 0$, $Z_j(t) > 0$ implies $Y_j(t-\delta) = Y_j(t+\delta)$ for some $\delta > 0$. Loosely speaking, an SRBM behaves like a Brownian motion with drift vector θ and covariance matrix Γ in the interior of the orthant $\mathbb{R}^d_+$, and it is confined to the orthant by instantaneous "reflection" (or "pushing") at the boundary, where the direction of "reflection" on the jth face F_j is given by the jth column of R. The parameters θ, Γ and R are called the *drift vector*, *covariance matrix* and *reflection matrix* of the SRBM, respectively.

Reiman and Williams [32] showed that a necessary condition for a $(\mathbb{R}^d_+, \theta, \Gamma, R)$-SRBM to exist is that the reflection matrix R is completely-$\mathcal{S}$, as defined in definition 2.2. Taylor and Williams [34] showed that when R is completely-$\mathcal{S}$, for any probability measure ν on $(\mathbb{R}^d_+, \mathcal{B})$, a $(\mathbb{R}^d_+, \theta, \Gamma, R, \nu)$-SRBM Z exists and is unique in distribution.

Let ν be a probability measure on $(\mathbb{R}^d_+, \mathcal{B})$. The measure ν is a stationary distribution for an SRBM Z if for each $A \in \mathcal{B}$,

$$\nu(A) = \int_{\mathbb{R}^d_+} \mathbb{P}_x\big\{Z(t) \in A\big\}\nu(\mathrm{d}x) \quad \text{for each } t \geqslant 0. \tag{4.1}$$

When ν is a stationary distribution, the process Z is stationary under the probability measure $\mathbb{P}_\nu$. Harrison and Williams [18] showed that a stationary distribution, when it exists, is unique and is absolutely continuous with respect to the Lebesgue measure on $(\mathbb{R}^d_+, \mathcal{B})$. We use π to denote the unique stationary distribution when it exists. When π exists, the SRBM Z is said to be positive recurrent. For a $(\mathbb{R}^d_+, \theta, \Gamma, R)$-SRBM Z, when R is an $\mathcal{M}$-matrix as defined in [2], it was proved in [18] that Z is positive recurrent if and only if

$$R^{-1}\theta < 0. \tag{4.2}$$

In the 2-dimensional case, we have a characterization of positive recurrence given by Hobson and Rogers [20] and Williams [35], who have shown that a $(\mathbb{R}^2_+, \theta, \Gamma, R)$-SRBM in the quadrant is positive recurrent if and only if (3.8) and (3.9) hold.

The following theorem by Dupuis and Williams [14] provides a sufficient condition to check the positive recurrence of an SRBM. Let R and θ be given. For $a \in \mathbb{R}^d_+$, set $x^a(t) = a + \theta t$ for $t \geqslant 0$.

Theorem 4.2 (Dupuis and Williams). Suppose that for each $a \in \mathbb{R}^d_+$ and each $z \in \psi(x^a)$, $\lim_{t\to\infty} z(t) = 0$. Then the $(\mathbb{R}^d_+, \theta, \Gamma, R)$-SRBM is positive recurrent for each positive definite matrix Γ.

Using theorem 4.2, Budhiraja and Dupuis [6] provided a slight generalization of the result of Harrison and Williams. For SRBMs in the orthant, no general recurrence condition has yet been established.

4.2. Large deviations

In this section, we provide some motivation for our study of the VPs introduced in (2.5). The primary impetus for our study comes from the theory of large deviations. An excellent reference for this material is Dembo and Zeitouni [9].

Most large deviations analyses can be divided into two principal parts: proving an LDP, and solving the associated VP. For SRBMs in the orthant, considerable progress has been made in the first area by Majewski [29] and we quote his result shortly. The following conjecture, for SRBMs in the d-dimensional orthant, provides motivation for our VP.

Conjecture 4.1 (General Large Deviations Principle). Consider a $(\mathbb{R}_+^d, \theta, \Gamma, R)$-SRBM Z. Suppose that R is a completely-$\mathcal{S}$ matrix and that there exists a probability measure $\mathbb{P}_\pi$ under which Z is stationary. Then for every measurable $A \subset \mathbb{R}_+^d$

$$\limsup_{u\to\infty} \frac{1}{u} \log \mathbb{P}_\pi\big(Z(0)/u \in A\big) \leqslant -\inf_{v\in A^c} I(v) \tag{4.3}$$

and

$$\liminf_{u\to\infty} \frac{1}{u} \log \mathbb{P}_\pi\big(Z(0)/u \in A\big) \geqslant -\inf_{v\in A^o} I(v) \tag{4.4}$$

where A^c and A^o are respectively the closure and interior of A.

Our goal in the next section is to provide some results which simplify the analysis and solution of the VPs which appear in conjecture 4.1. In section 6, we narrow our focus to the 2-dimensional case. For this class of VPs, we are able to provide a complete analytical solution.

Special cases of conjecture 4.1 above have indeed already been established. The most general result of which we are aware was given by Majewski [29]:

Theorem 4.3 (Large Deviations Principle). Consider a $(\mathbb{R}_+^d, \theta, \Gamma, R)$-SRBM Z, where R is an $\mathcal{M}$-matrix and suppose the recurrence condition (4.2) holds. Let $\mathbb{P}_\pi$ be the probability measure under which Z is stationary. For every measurable $A \subset \mathbb{R}_+^d$, (4.3) and (4.4) hold.

For clarification, it should be noted that Majewski states his result for reflection matrices R which he terms K-matrices. This class of matrices is equivalent in our context to what we have chosen to call $\mathcal{M}$-matrices, following Bermon and Plemmons [2]. In this case, the Skorohod problem has a unique solution and the reflection mapping is Lipschitz continuous.

5. Optimal path properties

The variational problem in (2.5) requires a search over a large class of absolutely continuous functions. In this section, we argue that, in $\mathbb{R}^2$, the optimal reflected path can be chosen such that it consists of at most two linear pieces, the first of which travels along one of the boundaries of the positive orthant and another which then traverses the interior. The main result of this section is the following theorem, which will be used to prove theorem 3.1.

Theorem 5.1. Consider the VP as given in (2.5) and let $v \in \mathbb{R}^2_+$. An optimal triple (x, y, z) from 0 to v can always be chosen so that (x, y, z) is a two-segment, piecewise linear path; during the first segment, z stays on one of boundaries of $\mathbb{R}^2_+$, terminating at a point w on the boundary; during the other segment, z is a direct, linear path from w to v. The first segment can be void, that is we may have $w = 0$. In this case, x, and hence z, is a direct, linear path from 0 to v.

When the reflection matrix is an $\mathcal{M}$-matrix, this result can also be inferred from [27], using [29, lemma 14], and lemma 5.2 below. We provide a direct proof in this section, which is based on a series of lemmas that are of independent interest in $\mathbb{R}^d_+$ with arbitrary $d \geqslant 1$.

Our first lemma follows directly from Jensen's inequality. Recall that $\mathcal{H}^d$ is the space of all absolutely continuous functions $x(\cdot) : [0, \infty) \to \mathbb{R}^d$ which have square integrable derivatives on bounded intervals and have $x(0) = 0$.

Lemma 5.1. Let g be a convex function on $\mathbb{R}^d$, and let $x \in \mathcal{H}^d$. Then for $t_1 < t_2$

$$\int_{t_1}^{t_2} g\big(\dot{x}(t)\big)\,\mathrm{d}t \geqslant \int_{t_1}^{t_2} g\left(\frac{x(t_1) - x(t_2)}{t_2 - t_1}\right)\mathrm{d}t. \tag{5.1}$$

In other words, a linear path minimizes this unconstrained variational problem. For our VP, the $g(v)$ we contend with is of the form:

$$\|v - \theta\|^2, \quad v \in \mathbb{R}^d.$$

We now consider the boundary of $\mathbb{R}^d_+$. Note that each face of the boundary can be defined by partitioning the coordinates of $\mathbb{R}^d$ into zero and nonzero components. For a partition $(\mathcal{K}_1, \mathcal{K}_2)$ of $\{1, \ldots, d\}$, we then define a face associated with the partition by letting the coordinates in $\mathcal{K}_1$ be zero and the coordinates of $\mathcal{K}_2$ be nonzero. Note that, for our purposes, the interior of $\mathbb{R}^d_+$ is also considered a face, corresponding to $\mathcal{K}_1 = \emptyset$.

Below, for a partition $(\mathcal{K}_1, \mathcal{K}_2)$ we also let $x_{\mathcal{K}_j}$ be the vector $(x_i,\ i \in \mathcal{K}_j)'$. For the reflection matrix R, we define two submatrices: R_1 is the principal submatrix of R with the rows and columns in $\mathcal{K}_2$ deleted, R_{21} is the submatrix of R with row indices in $\mathcal{K}_2$ and column indices in $\mathcal{K}_1$.

Definition 5.1. We say that a reflected path z is *anchored* to a face corresponding to the partition $(\mathcal{K}_1, \mathcal{K}_2)$ in the interval $[t_1, t_2]$ if

(i) $z_i(t_1) = z_i(t_2) = 0$ for $i \in \mathcal{K}_1$,

(ii) $z_i(t_1) > 0,\ z_i(t_2) > 0$ for $i \in \mathcal{K}_2$,

(iii) $z_i(t) > 0$ for $i \in \mathcal{K}_2$ and for all except finitely many $t \in (t_1, t_2)$.

Lemma 5.2. Consider the VP as given in (2.5) and let (x, y, z) be an optimal triple for this VP. If z is anchored to a face of the boundary of $\mathbb{R}^d_+$ in the interval $[t_1, t_2]$ (for $0 \leqslant t_1 < t_2$), then there exists an optimal triple $(\tilde{x}, \tilde{y}, \tilde{z})$ such that $\dot{\tilde{x}}(t) = a$ for $t \in (t_1, t_2)$ for some constant $a \in \mathbb{R}^d$; $\tilde{z}_i(t) = 0$ for $i \in \mathcal{K}_1$, $t \in [t_1, t_2]$; and $\tilde{z}_i(t) > 0$ for $i \in \mathcal{K}_2$, $t \in [t_1, t_2]$.

Proof. For completeness, we now explicitly write out the variational problem under consideration. For a given $v^1 \in \mathbb{R}^d_+$, $v^2 \in \mathbb{R}^d_+$, $0 \leqslant t_1 < t_2$, consider the minimization problem:

$$\min_x \int_{t_1}^{t_2} \left\| \dot{x}(s) - \theta \right\|^2 \mathrm{d}s \tag{5.2}$$

subject to

$$\begin{aligned}
&z(t) = x(t) + Ry(t), && t_1 \leqslant t \leqslant t_2,\\
&z(t) \geqslant 0, && t_1 \leqslant t \leqslant t_2,\\
&y(\cdot) \text{ is nondecreasing},\\
&\int_{t_1}^{t_2} z_i(s)\,\mathrm{d}y_i(s) = 0, && i = 1, \ldots, d,\\
&z(t_1) = v^1,\\
&z(t_2) = v^2.
\end{aligned}$$

We consider an optimal triple (x, y, z), with z anchored to a face corresponding to some partition $(\mathcal{K}_1, \mathcal{K}_2)$. Without loss of generality, we assume that $y(t_1) = 0$ for each optimal triple (x, y, z). Thus, $z(t_1) = x(t_1) = v^1$.

By the complementarity condition, $\dot{y}_i(t) = 0$ for $i \in \mathcal{K}_2$ on (t_1, t_2). Hence, with our convention that $y(t_1) = 0$, we have $y_i(t) = 0$ for $i \in \mathcal{K}_2$ and $t \in [t_1, t_2]$. Therefore, we have

$$0 = x_{\mathcal{K}_1}(t_2) + R_1 y_{\mathcal{K}_1}(t_2),$$

and it follows that

$$0 = x_{\mathcal{K}_1}(t_2)\left(\frac{t - t_1}{t_2 - t_1}\right) + R_1 y_{\mathcal{K}_1}(t_2)\left(\frac{t - t_1}{t_2 - t_1}\right) \quad \text{for } t_1 \leqslant t \leqslant t_2. \tag{5.3}$$

By condition (iii) of definition 5.1,

$$0 < x_{\mathcal{K}_2}(t_2) + R_{21} y_{\mathcal{K}_1}(t_2),$$

and we then have

$$0 < x_{\mathcal{K}_2}(t_2)\left(\frac{t - t_1}{t_2 - t_1}\right) + R_{21} y_{\mathcal{K}_1}(t_2)\left(\frac{t - t_1}{t_2 - t_1}\right) \quad \text{for } t_1 < t < t_2,$$

which implies

$$0 < x_{\mathcal{K}_2}(t_1)\left(\frac{t_2 - t}{t_2 - t_1}\right) + x_{\mathcal{K}_2}(t_2)\left(\frac{t - t_1}{t_2 - t_1}\right) + R_{21} y_{\mathcal{K}_1}(t_2)\left(\frac{t - t_1}{t_2 - t_1}\right)$$
$$\text{for } t_1 < t < t_2. \tag{5.4}$$

In other words, we can "linearize" the paths x and y as follows:

$$\tilde{x}_i(t) = x_i(t_1) + \left[x_i(t_2) - x_i(t_1)\right]\left(\frac{t - t_1}{t_2 - t_1}\right)$$

and

$$\tilde{y}_i(t) = y_i(t_2)\left(\frac{t - t_1}{t_2 - t_1}\right)$$

for $i = 1, \dots, d$ and $t \in [t_1, t_2]$. Now, if we let

$$\tilde{z}(t) = \tilde{x}(t) + R\tilde{y}(t) \quad \text{for } t \in [t_1, t_2],$$

then (5.3) and (5.4) show that the new reflected path $\tilde{z}$ is anchored to the same face as z on the interval (t_1, t_2). In fact, $\tilde{z}$ is now on the face for the entire interval. Furthermore, it can be checked that $\tilde{z}(t_1) = \tilde{x}(t_1) = x(t_1) = z(t_1)$ and $\tilde{z}(t_2) = z(t_2)$ and hence the new reflected path also has the same endpoints. By lemma 5.1

$$\int_{t_1}^{t_2} \left\|\dot{x}(t) - \theta\right\|^2 \mathrm{d}t \geqslant \int_{t_1}^{t_2} \left\|\frac{x(t_1) - x(t_2)}{t_2 - t_1} - \theta\right\|^2 \mathrm{d}t.$$

Thus, $\tilde{x}$ has equal or lower energy than the original path and hence any optimal path can be reduced to an equivalent optimal path which has a constant derivative while its reflection is on a fixed face of the boundary. □

The reduction of the VP to a class of piecewise linear functions is not particularly surprising. Other authors, including O'Connell [31] and Dupuis and Ishii [11] have achieved similar reductions, but only for special cases, i.e., in $\mathbb{R}^2$ or for a limited class of reflection matrices. Since lemma 5.1 holds for any convex function and in large deviations applications, the kernel which appears in the VP is always convex (for LDPs associated with random vectors), our proof is valid for VPs arising from a wide range of LDPs. We have not specifically addressed the nature of the piecewise linear functions which may solve the VP. In particular, we have not ruled out a piecewise linear function with an infinite number of discontinuities in $\dot{x}$. With the help of the next several lemmas, we can rule out such paths, at least in $\mathbb{R}^2$.

Lemma 5.3 (Scaling lemma). Consider the VP in $\mathbb{R}^d$ as given in (2.5), with target point v in the positive orthant.

(a) For any positive k, $I(kv) = kI(v)$.

(b) If (x, y, z) is an optimal triple for v, then $(\bar{x}, \bar{y}, \bar{z})$ is an optimal triple for kv, where $\bar{x}(t) = kx(t/k)$, $\bar{y}(t) = ky(t/k)$ and $\bar{z}(t) = kz(t/k)$ for $t \geqslant 0$.

Proof. Let (x, y, z) be an optimal triple which solves the Skorohod problem with $z(T) = v$ and

$$\frac{1}{2}\int_0^T \left\|\dot{x}(t) - \theta\right\|^2 \mathrm{d}t = I(v).$$

It is clear that $(\bar{x}, \bar{y}, \bar{z})$ also solves the Skorohod problem with $\bar{z}(kT) = kv$. Furthermore, we have

$$\begin{aligned}\frac{1}{2}\int_0^{kT} \left\|k\frac{\mathrm{d}(x(t/k))}{\mathrm{d}t} - \theta\right\|^2 \mathrm{d}t &= \frac{1}{2}\int_0^{kT} \left\|\dot{x}(t/k) - \theta\right\|^2 \mathrm{d}t \\ &= \frac{1}{2}k\int_0^T \left\|\dot{x}(t) - \theta\right\|^2 \mathrm{d}t = kI(v).\end{aligned}$$

Hence $I(kv) \leqslant kI(v)$. Now since, $k > 0$ is arbitrary, we have $I(v) = I(k^{-1}kv) \leqslant k^{-1}I(kv)$, or $kI(v) \leqslant I(kv)$. Thus, we have $kI(v) = I(kv)$. This proves (a) and the above calculation proves (b). $\square$

A similar scaling lemma for variational problems arising from random walks in an orthant is stated in [21].

Lemma 5.4 (Merge lemma). Let (x^1, y^1, z^1) be an R-regulation triple on $[0, t_1]$ with $z^1(0) = 0$ and $z^1(t_1) = w$. Let (x^2, y^2, z^2) be an R-regulation triple on $[s_2, t_2]$ with $z^2(s_2) = w$ and $z^2(t_2) = v$. Suppose that both x^1 and x^2 are absolutely continuous. Define

$$z(t) = \begin{cases} z^1(t) & \text{for } 0 \leqslant t \leqslant t_1, \\ z^2(t - t_1 + s_2) & \text{for } t_1 \leqslant t \leqslant t_1 + t_2 - s_2, \end{cases}$$

$$x(t) = \begin{cases} x^1(t) & \text{for } 0 \leqslant t \leqslant t_1, \\ x^2(t - t_1 + s_2) - x^2(s_2) + x^1(t_1) & \text{for } t_1 \leqslant t \leqslant t_1 + t_2 - s_2, \end{cases}$$

$$y(t) = \begin{cases} y^1(t) & \text{for } 0 \leqslant t \leqslant t_1, \\ y^2(t - t_1 + s_2) - y^2(s_2) + y^1(t_1) & \text{for } t_1 \leqslant t \leqslant t_1 + t_2 - s_2, \end{cases}$$

and $s = t_1 + t_2 - s_2$. Then (x, y, z) is an R-regulation triple on $[0, s]$ with $z(0) = 0$ and $z(s) = v$ such that x is absolutely continuous on $[0, s]$ and

$$\frac{1}{2}\int_0^s \left\|\dot{x}(t) - \theta\right\|^2 \mathrm{d}t = \frac{1}{2}\int_0^{t_1} \left\|\dot{x}^1(t) - \theta\right\|^2 \mathrm{d}t + \frac{1}{2}\int_{s_2}^{t_2} \left\|\dot{x}^2(t) - \theta\right\|^2 \mathrm{d}t. \tag{5.5}$$

Proof. Since both x^1 and x^2 are absolutely continuous, x is absolutely continuous on $[0, s]$. Also one can check that y is nondecreasing, $z(0) = 0$ and $z(s) = v$. We now check that

$$z(t) = x(t) + Ry(t) \quad \text{for } 0 \leqslant t \leqslant s. \tag{5.6}$$

Clearly, (5.6) is satisfied for $0 \leqslant t \leqslant t_1$. Since (x^1, y^1, z^1) is an R-regulation with $z^1(t_1) = w$, we have $w = z^1(t_1) = x^1(t_1) + Ry^1(t_1)$. For $t_1 \leqslant t \leqslant t_1 + t_2 - s_2$,

$$\begin{aligned}
z(t) &= z^2(t - t_1 + s_2) \\
&= x^2(t - t_1 + s_2) + Ry^2(t - t_1 + s_2) \\
&= z^2(s_2) + x^2(t - t_1 + s_2) - x^2(s_2) + R\big(y^2(t - t_1 + s_2) - y^2(s_2)\big) \\
&= z^1(t_1) + x^2(t - t_1 + s_2) - x^2(s_2) + R\big(y^2(t - t_1 + s_2) - y^2(s_2)\big) \\
&= x^2(t - t_1 + s_2) - x^2(s_2) + x^1(t_1) + R\big(y^2(t - t_1 + s_2) - y^2(s_2) + y^1(t_1)\big) \\
&= x(t) + Ry(t).
\end{aligned}$$

Thus, (5.6) holds for $0 \leqslant t \leqslant s$, from which one can readily show that (x, y, z) is an R-regulation on $[0, s]$. Finally, (5.5) follows from the definition of x. □

In the following, we use E_i to denote the one-dimensional edge $\{v \in \mathbb{R}^d_+ \colon v_j = 0$ for $j \neq i\}$.

Lemma 5.5 (Reduction lemma). Consider the VP in $\mathbb{R}^d$ as given in (2.5). Let $v \in E_i$. Suppose that (x^1, y^1, z^1) is an optimal triple from 0 to v such that $z(t) \in E_i$ for $t \in [s_1, t_1]$ with $s_1 < t_1$ and $z^1(s_1) \neq z^1(t_1)$. Then there exists an optimal triple (x, y, z) on $[0, T]$ from 0 to v such that $z(t) \in E_i$ for $0 \leqslant t \leqslant T$ with $z(T) = v$.

Proof. We start by assuming that $z^1(t_1) = v$. Let (x^1, y^1, z^1) be an optimal triple with $z^1(s_1) = w$ and $z^1(t_1) = v$ such that $z^1(t) \in E_i$ for $t \in [s_1, t_1]$. Let $k = |z^1(s_1)|/|z^1(t_1)|$. Next, by the scaling lemma 5.3 we have $k \leqslant 1$ and by assumption $k \neq 1$, thus $k < 1$. Since $w = kv$, it follows from lemma 5.3 that the triple $(\bar{x}, \bar{y}, \bar{z})$ is an optimal triple from 0 to w, where $\bar{x}(t) = kx^1(t/k)$, $\bar{y}(t) = ky^1(t/k)$ and $\bar{z}(t) = kz^1(t/k)$. Note that $\bar{z}(t) \in E_i$ for $t \in [ks_1, kt_1]$. By lemma 5.4, piecing together the triple $(\bar{x}, \bar{y}, \bar{z})$ on $[0, kt_1]$ with the triple (x^1, y^1, z^1) on $[s_1, t_1]$, we have the triple (x^2, y^2, z^2) on $[0, t_2]$ with $t_2 = kt_1 + t_1 - s_1$. Furthermore, by (5.5), the triple is optimal, and $z^2(t) \in E_i$ for $t \in [s_2, t_2]$ with $s_2 = ks_1$. We now iterate our argument, with a new scaling parameter in each iteration, $k_n = |z^n(s_n)|/|z^n(t_n)| = k^n$. Then we have for each integer $n \geqslant 1$, there exists an optimal triple (x^n, y^n, z^n) on $[s_n, t_n]$ with $z^n(t) \in E_i$ for $t \in [s_n, t_n]$ and $z^n(t_n) = v$, where $s_n = k_n s_{n-1}$ and $t_n = k_n t_{n-1} + (t_{n-1} - s_{n-1})$. Since we had $k < 1$, we have that $k_n \to 0$ and hence s_n and $|z^n(s_n)|$ both converge to zero.

By construction, it can be seen that

$$\big(\dot{x}^n(t), \dot{y}^n(t), \dot{z}^n(t)\big) = (c_1, c_2, c_3) \tag{5.7}$$

on (s_n, t_n) for all n, where each $c_i \in \mathbb{R}^d$ is a constant independent of n. Thus

$$t_n - s_n = \frac{|v - z^n(s_n)|}{|c_3|}.$$

Since $s_n, |z^n(s_n)| \to 0$, this yields $t_n \to |v|/c_3 \equiv T$.

Now we are prepared to construct an optimal triple with the stated properties. We set $(x, y, z) = (c_1 t, c_2 t, c_3 t)$ on $[0, T]$. By (5.7) we have that $(x, y, z) - (x, y, z)(s_1)$ is an R-regulation on $[s_1, t_1]$ and then by lemma 5.4, (x, y, z) is an R-regulation on $[0, t_1 - s_1]$. From the linearity of the Skorohod problem it is therefore also an R-regulation on $[0, T]$. Note that, by construction, we also have $z(T) = v$. It remains only to show that (x, y, z) is optimal. Note that each triple (x^n, y^n, z^n) must have optimal cost. Hence, if we show that

$$\lim_{n\to\infty} \frac{1}{2}\int_0^{t_n} \left\|\dot{x}^n(t) - \theta\right\|^2 \mathrm{d}t = \frac{1}{2}\int_0^{T} \left\|\dot{x}(t) - \theta\right\|^2 \mathrm{d}t$$

then we are done. By repeated application of the scaling lemma 5.3, (5.5), and (5.7), we have

$$\begin{aligned}\frac{1}{2}\int_0^{t_n} \left\|\dot{x}^n(t) - \theta\right\|^2 \mathrm{d}t &= \frac{1}{2}\int_0^{s_n} \left\|\dot{x}^n(t) - \theta\right\|^2 \mathrm{d}t + \frac{1}{2}\int_{s_n}^{t_n} \left\|\dot{x}^n(t) - \theta\right\|^2 \mathrm{d}t \\ &= \frac{1}{2}\left(\prod_{i=1}^{n-1} k_i\right)\int_0^{s_1} \left\|\dot{x}^1 - \theta\right\|^2 \mathrm{d}t + \frac{1}{2}(t_n - s_n)\|c_1 - \theta\|^2.\end{aligned}$$

Since $k_i \to 0$, the first part converges to zero. Since $t_n \to T$ and $s_n \to 0$, the second part, and thus the entire cost, converges to $(T/2)\|c_1 - \theta\|^2$, which is just the cost of the constructed triple (x, y, z). Hence (x, y, z) is an optimal triple.

If $z^1(t_1) \neq v$, we set $k_0 = v/|z^1(t_1)|$ and apply the scaling lemma 5.3 with this k_0. We are then back in the case $z^1(t_1) = v$ and can proceed as before. $\square$

Now, using the a reduction lemma 5.5, we can prove the main theorem of this section, which states that in $\mathbb{R}^2$, an optimal reflected path can be chosen such that it consists of at most two linear pieces.

Proof of theorem 5.1. Throughout the proof we assume that $v \in \mathbb{R}_+^2$ and that all paths are piecewise linear as per lemma 5.2. We divide our argument into three cases.

(1) Let v be in the interior of $\mathbb{R}_+^2$ and let (x, y, z) be an optimal triple to v. Then z either goes directly from the origin to v or else z reaches some point $w \in E_i \setminus \{0\}$ at time $t_2 > 0$. Let us assume that t_2 is the last such time. By lemma 5.2, z must touch another point on $E_i \setminus \{0\}$ at time $t_1 < t_2$, and hence the triple can be chosen so that $z(t) \in E_i$ for $t \in [t_1, t_2]$. But then by lemma 5.5 there exists an optimal triple $(\bar{x}, \bar{y}, \bar{z})$ from 0 to w with $\bar{z}(t) \in E_i$ for $0 \leqslant t \leqslant \bar{t}_2$ and $\bar{z}(\bar{t}_2) = w$. By lemma 5.4, a two-segment triple, formed by merging $(\bar{x}, \bar{y}, \bar{z})$ and (x, y, z), is optimal.

(2) Suppose v is on a face E_1 of $\mathbb{R}^2_+$. Let (x, y, z) be an optimal triple with $z(T) = v$. If z touches another point on $E_1 \setminus \{0\}$, then by lemma 5.5 the triple can be chosen such that it is linear. Otherwise, by lemma 5.2, z has to touch a (last) point $w \in E_2$ and stay on E_2 in some time interval. Again by lemma 5.2, the triple must be linear from w to v. Furthermore, by lemma 5.5, the triple can be chosen to be linear from 0 to w. Hence, in this case, we can again chose an optimal triple with two linear segments.

(3) If $v \in E_2$, then we interchange the roles of E_1 and E_2 in the case 2 argument.

We have thus shown that for any optimal triple, we can choose an equivalent optimal triple that falls into one of the cases in the statement of theorem 5.1. □

A major open problem is to determine whether or not theorem 5.1 can be extended to the case $d > 2$. The problem in higher dimensions is to eliminate consideration of paths which "spiral" around the boundary. In $\mathbb{R}^2$, spiraling cannot occur (without retracing part of a path) and hence the reduction to a piecewise linear path with just two pieces is relatively straightforward. One hopes that in $\mathbb{R}^d$ that one need only consider piecewise linear paths with d pieces, thus reducing the general VP to a finite dimensional optimization problem.

In the case $d = 2$, we have reduced our possible solutions to paths which either go directly from the origin to v through the interior or first travel along one axis and then travel through the interior to v. In particular, we need only search over three *types* of piecewise paths. In the next section, we will argue that this search can be restricted to just three paths.

6. Further constrained variational problems

In this section we provide the analysis of the VP in $\mathbb{R}^2$ which justifies theorem 3.1. We first consider the VP defined in (2.5), adding additional constraints on the allowed path $x(\cdot)$. Recall that the definitions of the direct triple (x^0, y^0, z^0) and the broken triples (x^1, y^1, z^1) and (x^2, y^2, z^2) were given in section 3. When we restrict $x(\cdot)$ such that it only takes values in $\mathbb{R}^2_+$, the corresponding VP has the direct triple (x^0, y^0, z^0) as an optimal triple. We next consider "broken" triples in which the reflected path first moves along a face F_i and then traverses the interior. When F_i is reflective, and $v \in C_i$, the broken triple (x^i, y^i, z^i) is an optimal triple among all two segment broken triples through F_i. Once these principles have been established, the proof of theorem 3.1 will then follow directly.

Before studying these further constrained VPs, we present the following lemma.

Lemma 6.1. Let $I^0(v)$, $I^1(v)$, and $I^2(v)$ be given in (3.6) and (3.7). Then,

(a) $I^0(v) = \|\theta\| \cdot \|v\| - \langle \theta, v\rangle$.

(b) $I^i(v) \leqslant I^0(v)$ for $v \in \mathbb{R}^2_+$ and $i = 1, 2$. Furthermore, $I^i(v) = I^0(v)$ if and only if v is in the same direction as $\tilde{a}^i$.

Proof. Part (a). From the definition of $I^0(v)$ in section 3, we have

$$I^0(v) = \left\langle \frac{v\|\theta\|}{\|v\|} - \theta, v \right\rangle = \frac{\|\theta\|}{\|v\|} \cdot \langle v, v\rangle - \langle \theta, v\rangle = \|\theta\| \cdot \|v\| - \langle \theta, v\rangle.$$

Part (b). First, note that it follows from (3.2) that

$$\langle a^i, a^i\rangle = \langle \theta, \theta\rangle + 4(\theta' p^i)^2 - 4(\theta' p^i)\langle \theta, \Gamma p^i\rangle = \langle \theta, \theta\rangle.$$

Hence, $\|a^i\| = \|\theta\|$. Furthermore, from the definition of $\tilde{a}^i$, we can immediately observe that $\|\tilde{a}^i\| = \|a^i\|$, and thus we have

$$\|a^i\| = \|\tilde{a}^i\| = \|\theta\|. \tag{6.1}$$

Using then (3.6) and (3.7), we conclude

$$I^i(v) = \langle \tilde{a}^i, v\rangle - \langle \theta, v\rangle \leqslant \|\tilde{a}^i\|\|v\| - \langle \theta, v\rangle = I^0(v). \qquad \square$$

6.1. *Interior escape paths*

Now let us consider a point $v \in \mathbb{R}^2_+$ and the VP as defined in (2.5). In this section, we add the additional constraint that $x(\cdot)$ may only take values in $\mathbb{R}^2_+$. We will use $\tilde{I}^0(v)$ to denote the resulting optimal value, namely,

$$\tilde{I}^0(v) = \inf_{T\geqslant 0} \inf_{x\in\mathcal{H}^d, x(T)=v} \frac{1}{2}\int_0^T \|\dot{x}(t) - \theta\|^2 \, \mathrm{d}t, \tag{6.2}$$

where $x(t) \in \mathbb{R}^2_+$ for $0 \leqslant t \leqslant T$.

For the VP (2.5), if we only consider paths which travel in the interior of the orthant, then by theorem 5.2 the optimal path is linear and has constant velocity proportional to the point v that we wish to reach. Hence, we need only determine the optimal speed to minimize the value of the VP. So we set $x(t) = ct\, v$ and the VP in (6.2) reduces to the following:

$$\tilde{I}^0(v) = \inf_{c>0} \frac{1}{2c}\|cv - \theta\|^2. \tag{6.3}$$

This is a one-dimensional minimization problem which has a unique minimum at $c = \|\theta\|/\|v\|$, leading us to the following result.

Theorem 6.1. Among the possible direct triples from the origin to a fixed point v, the direct triple (x^0, y^0, z^0) is optimal. The corresponding minimal cost $\tilde{I}^0(v)$ is $I^0(v)$.

6.2. *Single segment boundary escapes*

We now consider optimal reflected paths which travel along face F_i to reach a point along this face. For $v \in F_i$, we use $\tilde{I}^i(v)$ to denote the resulting optimal value, namely,

$$\tilde{I}^i(v) = \inf_{T\geqslant 0} \inf_{x\in\mathcal{H}^d, z(T)=v} \frac{1}{2}\int_0^T \|\dot{x}(t) - \theta\|^2 \, \mathrm{d}t, \tag{6.4}$$

where $z(\cdot)$ is a reflected path associated with $x(\cdot)$, and $z(t) \in F_i$ for $0 \leqslant t \leqslant T$.

Theorem 6.2. Let $v \in F_i$.

(a) If boundary F_i is reflective, then the broken triple (x^i, y^i, z^i) to v along F_i is optimal for (6.4) with $\tilde{I}^i(v) = I^i(v)$.

(b) If boundary F_i is not reflective, then the direct triple (x^0, y^0, z^0) to v is optimal for (6.4) with $\tilde{I}^i(v) = I^0(v)$.

To prove the theorem we will need the following elementary lemma from calculus.

Lemma 6.2. Let $f(v)$ and $g(v)$ be two differentiable functions on $\mathbb{R}^d$. Let $h(v) = f(v)/g(v)$ be defined for $v \in \mathbb{R}^d$ with $g(v) \neq 0$. Then, for any v satisfying $\nabla h(v) = 0$, there exists a constant k such that

$$\nabla f(v) = k\nabla g(v), \tag{6.5}$$

$$f(v) = kg(v). \tag{6.6}$$

Proof of theorem 6.2. We prove the theorem for $v = (0, 1)' \in F_1$. In this proof, we let $p^1 = (-r_1, 1)'$, without the normalization of section 3. In light of theorem 5.2, the search of an optimal boundary path can be confined to linear paths $x(t) = b\,t$, $t \geqslant 0$, for some $b \in \mathbb{R}^2$. Let (z, y) be an R-regulation associated with $x(\cdot)$ that satisfies $z(t) \in F_1$ for $t \geqslant 0$. Then we have,

$$z_1(t) = b_1 t + y_1(t) + r_2 y_2(t),$$
$$z_2(t) = b_2 t + r_1 y_1(t) + y_2(t).$$

Since it is always cheaper to take a $z(\cdot)$ such that $z_2(t) > 0$ for $0 \leqslant t \leqslant T$, we have $y_2(t) = 0$ for $0 \leqslant t \leqslant T$. Next, because $z_1(t) = 0$ for $0 \leqslant t \leqslant T$, we have $y_1(t) = -b_1 t$ and $b_1 \leqslant 0$. Therefore, $z_2(t) = (b_2 - r_1 b_1)t = (b' p^1)t$. Since $z_2(T) = 1$, $b' p^1$ must be positive and we also must have $T = 1/(b' p^1)$. It follows that

$$\tilde{I}^1(v) = \inf_{b_1 \leqslant 0,\, b' p^1 > 0} \frac{1}{2b' p^1} \|b - \theta\|^2. \tag{6.7}$$

Using lemma 6.2, it can be checked that each critical point b of the unconstrained form of (6.7) must satisfy

$$2\Gamma^{-1}(b - \theta) = 2kp^1, \tag{6.8}$$

$$\|b - \theta\|^2 = 2kbp^1. \tag{6.9}$$

From (6.8), $b - \theta = k\Gamma p^1$, and substituting $k\Gamma p^1$ into the left side of (6.9), we have

$$\left\|k\Gamma p^1\right\|^2 = 2kbp^1 = k\big(\theta + k\Gamma p^1\big)' p^1.$$

Thus, the unconstrained form of (6.7) has two critical points

$$b = \theta + k\Gamma p^1,$$

with $k = 0$ or $k = -2\theta' p^1/\|\Gamma p^1\|^2$. Next, we show that the first critical point $b = \theta$ corresponding to $k = 0$ is not feasible. This is true because, in this case the regulated speed, $\dot{z}_2(t)$, along F_1 would then be given by $\theta_2 - r_1\theta_1$, which, by a detailed calculation, is negative under our stability condition (3.9). The second critical point corresponds to our expression for a^1 given in (3.2). One can check that $(a^1)' p^1 = -\theta' p^1$ which, by (3.9), is positive.

Note that when $|b| \to \infty$, away from the boundary, the function in (6.7) goes to infinity; when b approaches the boundary $b' p^1 = 0$, from the interior of the feasible region, the function in (6.7) goes to infinity. Therefore, the infimum in (6.7) takes place either at a critical point in the interior or at $b_1 = 0$.

If F_1 is not reflective, then by definition, we have $a_1^1 \geqslant 0$. In this case, then there are no critical points in the interior of the feasible region and the minimum must occur at the constraint $b_1 = 0$, which indicates that the direct path along F_1 is optimal and we have $\tilde{I}^1(v) = I^0(v)$.

If F_1 is reflective, then $a_1^1 < 0$, and this quantity is a critical point which is in the interior of the feasible region. The value of the VP (6.4) at the critical point a^1 is

$$\frac{1}{2(a^1)' p^1} \left\| a^1 - \theta \right\|^2 = \frac{1}{2(a^1)' p^1} (k)^2 \left\| \Gamma p^1 \right\|^2 = k > 0$$

and we have

$$k = \left\langle a^1 - \theta, v \right\rangle = \left\langle \tilde{a}^1 - \theta, v \right\rangle = I^1(v),$$

where the first equality follows from (6.8) and the second from the definition of $\tilde{a}^1$. Since $I^1(v) \leqslant I^0(v)$, as demonstrated in lemma 6.1, we then have $\tilde{I}^1(v) = I^1(v)$ and the theorem is proved. □

The theorem demonstrates that if a face is reflective, then a regulated boundary path may be part of an optimal path to a point in the face. In the stochastic setting, we envision such paths as “bouncing paths” which repeatedly bounce against a face and are heavily regulated to keep them within the quadrant.

Furthermore, it should be noted that the stability conditions (3.8) and (3.9) for SRBM in $\mathbb{R}_+^2$ are required for our solution to be valid. In particular the conditions are used in the proof to eliminate a critical point. If one of the stability conditions does not hold, then the VP will have optimal value zero for any points v along some ray in $\mathbb{R}^2$. This means that the VPs given in (4.3)–(4.4), will have value zero for many sets of interest. This result should be expected, since a stationary distribution for the SRBM will not exist in cases where the stability conditions do not hold, hence the LDP of section 4.2 does not hold.

It is a subject of ongoing research to determine explicit stability conditions in higher dimensions and analyze their relation to the corresponding VP. It is possible that heretofore unknown stability conditions will manifest themselves in solving the VP in higher dimensions.

6.3. Two-segment boundary escapes

Recall that, for a point $v \in \mathbb{R}^2_+$, a broken triple (x, y, z) to v through face F_i consists of two segments: during the first segment, z travels from the origin along F_i up to a point $w \in F_i$, and during the second segment, z is a non-regulated path traveling from w to v. When $w = 0$, the broken triple is actually a direct triple.

In this section, we consider the VP (2.5) when constrained to all broken triples through a face. For a point $v \in \mathbb{R}^2_+$, our goal is to determine an optimal broken triple among all such triples. The corresponding optimal value is denoted by $\tilde{I}^i(v)$. The optimal broken triple is an extension of the optimal single segment triple previously considered for a point v on a boundary face. Thus we employ the same symbol $\tilde{I}^i(\cdot)$ to denote the optimal value.

Theorem 6.3. Let $v \in \mathbb{R}^2_+$.

(a) If $v \in C_i$, then (x^i, y^i, z^i) is an optimal broken triple to v through F_i. The optimal broken path z^i has unique breakpoint $w \in F_i$ with $v - w = c\tilde{a}^i$ and $c = -\langle v, \Gamma n^i\rangle / \langle a^i, \Gamma n^i\rangle$. Furthermore, the optimal cost $\tilde{I}^i(v)$ is given by $I^i(v)$.

(b) If $v \notin C_i$, then the direct triple (x^0, y^0, z^0) is optimal among all broken triples to v through F_i with $\tilde{I}^i(v) = I^0(v)$.

Proof. When F_i is not reflective, by part (b) of theorem 6.2, lemma 5.2 and theorem 6.1, the direct triple (x^0, y^0, z^0) is optimal among all broken triples.

Now we assume that F_i is reflective. The optimal total cost for a broken path through F_i to v with a breakpoint at $w = te_i$, $t \geqslant 0$, is

$$\tilde{I}(t, v) = \tilde{I}^i(w) + \tilde{I}^0(v - w).$$

It follows from theorems 6.1 and 6.2 that

$$\tilde{I}(t, v) = \left\langle \tilde{a}^i - \theta, te_i \right\rangle + \|\theta\| \cdot \|v - te_i\| - \langle \theta, v - te_i\rangle. \tag{6.10}$$

Note that $\tilde{I}(t, v) \geqslant tI^i(e_i)$, which goes to infinity as $t \to \infty$. Hence, the minimum of the function $\tilde{I}(\cdot, v)$ must either occur at a critical point in $(0, \infty)$ or at the boundary $t = 0$. So, to minimize this function with respect to t, for $t > 0$, we take derivatives to obtain:

$$\begin{aligned}\frac{\partial}{\partial t}\tilde{I}(t, v) &= \frac{\partial}{\partial t}\Big[t\left\langle \tilde{a}^i - \theta, e_i\right\rangle + \|\theta\| \cdot \|v - te_i\| + t\langle \theta, e_i\rangle - \langle \theta, v\rangle\Big] \\ &= \left\langle \tilde{a}^i - \theta, e_i\right\rangle - \|\theta\| \frac{\langle v - te_i, e_i\rangle}{\|v - te_i\|} + \langle \theta, e_i\rangle = \left\langle \tilde{a}^i, e_i\right\rangle - \|\theta\| \frac{\langle v - te_i, e_i\rangle}{\|v - te_i\|}.\end{aligned}$$

Setting this equal to zero and rearranging yields:

$$\frac{\langle \tilde{a}^i, e_i\rangle}{\|\theta\|} = \frac{\langle v - te_i, e_i\rangle}{\|v - te_i\|}. \tag{6.11}$$

Since $\|\tilde{a}\| = \|\theta\|$ by (6.1), the breakpoint $w = te_i$ then must satisfy

$$\frac{\langle \tilde{a}^i, e_i\rangle}{\|\tilde{a}^i\|} = \frac{\langle v - w, e_i\rangle}{\|v - w\|}.$$

Thus, $v - w$ must be in the same direction of a^i or $\tilde{a}^i$. The first case is not possible for a reflective face F_i. Thus,

$$v - w = c\tilde{a}^i,$$

for some $c > 0$.

To prove part (a), we note that when C_i is nonempty, F_i is reflective. In this case, one can check that the critical point $w = t^* e^i \in F_i$ exists and is unique. To find c, we use the expansion (3.3) on v, (3.4), and the fact that $v = t^* e^i + c\tilde{a}^i$, to obtain

$$\langle v, \Gamma n^i\rangle = -c\langle a^i, \Gamma n^i\rangle$$

or $c = -\langle v, \Gamma n^i\rangle / \langle a^i, \Gamma n^i\rangle$. With breakpoint w, by (6.10) the total cost is

$$\begin{aligned}\tilde{I}(t^*, v) &= \langle \tilde{a}^i - \theta, v - c\tilde{a}^i\rangle + c\big(\|\theta\|\,\|\tilde{a}^i\| - \langle \theta, \tilde{a}^i\rangle\big)\\ &= \langle \tilde{a}^i - \theta, v\rangle + c\big(\|\theta\|\,\|\tilde{a}^i\| - \langle \tilde{a}^i, \tilde{a}^i\rangle\big).\end{aligned}$$

Note that, again using (6.1)

$$\|\theta\|\,\|\tilde{a}^i\| - \langle \tilde{a}^i, \tilde{a}^i\rangle = \|\tilde{a}^i\|^2 - \langle \tilde{a}^i, \tilde{a}^i\rangle = 0.$$

Therefore, we have

$$\tilde{I}(t^*, v) = \langle \tilde{a}^i - \theta, v\rangle = I^i(v).$$

Since $I^i(v) \leqslant I^0(v) = \tilde{I}(0, v)$, as demonstrated in section 3, we have $\tilde{I}^i(v) = I^i(v)$ and part (a) is proved.

When $v \notin C_i$ and F_i is reflective, then no such critical point $w = te_i$, with $t > 0$, exists. Hence, in this case, the optimum occurs at the boundary point $t = 0$, i.e., the optimal triple to v through F_i is simply a direct triple, with corresponding cost $\tilde{I}^i(v) = \tilde{I}(0, v) = I^0(v)$. If face F_i is not reflective, then the optimal triple to v through F_i is also a direct triple, as discussed at the beginning of the proof. This establishes part (b), and hence theorem 6.3. □

6.4. Proof of theorem 3.1

Now we prove the main theorem of the paper. By theorem 5.1 we may conclude that any optimal triple can be reduced to an equivalent direct triple or broken triple through one of the faces. Now, let $v \in \mathbb{R}^2_+$. The remainder of the theorem follows directly from the results we have established in this section. We briefly outline the connection for each case of theorem 3.1:

(a) The fact that the optimal value is $I^0(v)$ follows directly from part (b) of theorem 6.3.

(b) The result follows from theorem 6.3, parts (a) and (b), and lemma 6.1, part (b).

(c) Analogous to (b) above.

(d) In this case the result follows from part (a) of theorem 6.3 and lemma 6.1, part (b).

In each case, there is a broken or direct triple which attains the corresponding minimum value.

7. Examples

In this section we apply the main theorem of the paper in some illustrative examples. In addition to illuminating the results, we expect that this section will provide a connection to previous results obtained for the stationary distribution of SRBMs.

7.1. An SRBM from a tandem network

We next provide an example of the solution to the VP for SRBM data arising from diffusion approximations of 2-station tandem queueing networks [15]. We consider a $(\mathbb{R}^2_+, \theta, \Gamma, R)$-SRBM with the following data. We let $\Gamma = I$, $\theta = (\theta_1, \theta_2)'$, and

$$R = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

It is easy to verify that R is an $\mathcal{M}$-matrix, which implies that the corresponding reflection mapping, and hence the associated SRBM, is well defined. In this case, the recurrence conditions are given by (4.2), which reduce to $\theta_1 < 0$ and $\theta_1 + \theta_2 < 0$. From (3.2) and (3.4), some simple calculations yield:

$$a^1 = \begin{pmatrix} -\theta_2 \\ -\theta_1 \end{pmatrix}, \qquad \tilde{a}^1 = \begin{pmatrix} \theta_2 \\ -\theta_1 \end{pmatrix},$$
$$a^2 = \begin{pmatrix} -\theta_1 \\ \theta_2 \end{pmatrix}, \qquad \tilde{a}^2 = \begin{pmatrix} -\theta_1 \\ -\theta_2 \end{pmatrix}.$$

Furthermore, we will have

$$\begin{aligned} I^0(v) &= \sqrt{\left(\theta_1^2 + \theta_2^2\right)\left(v_1^2 + v_2^2\right)} - (\theta_1 v_1 + \theta_2 v_2), \\ I^1(v) &= -\theta_1(v_1 + v_2) + \theta_2(v_1 - v_2), \\ I^2(v) &= -2(\theta_1 v_1 + \theta_2 v_2). \end{aligned}$$

Since θ_1 has a fixed sign by the recurrence conditions, let us examine more closely the cases $\theta_2 > 0$, $\theta_2 < 0$, and $\theta_2 = 0$.

In the case that $\theta_2 > 0$, we note that $\tilde{a}^1_1 > 0$ and $\tilde{a}^2_2 < 0$. Hence, face F_1 is reflective and face F_2 is nonreflective, with $C_2 = \emptyset$. Within C_1, the broken triple through F_1 is optimal and the optimal value is given by $I^1(v)$ above. Within $\mathbb{R}^2_+ \setminus C_1$ (which is non-empty) the direct triple is optimal and the value of the VP is given by $I^0(v)$. Figure 2 illustrates this case.

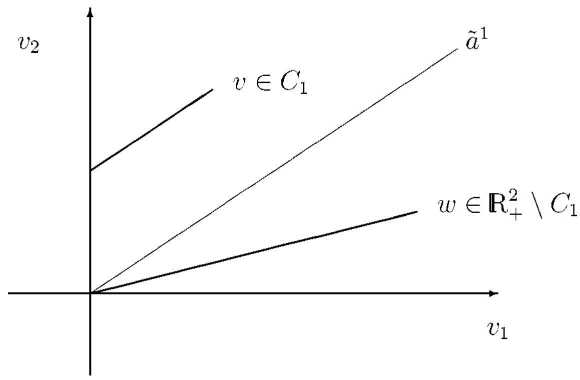

Figure 2. An optimal broken path to $v \in C_1$ and an optimal direct path to $w \in \mathbb{R}^2_+ \setminus C_1$ for $\theta_2 > 0$.

If $\theta_2 < 0$, then $\tilde{a}^1_1 < 0$ and $\tilde{a}^2_2 > 0$. Thus, F_2 is now reflective and F_1 nonreflective, and we have that both C_2 and $\mathbb{R}^2_+ \setminus C_2$ are non-empty.

In the final case, $\theta_2 = 0$, we see that $\tilde{a}^1$ is a multiple of $(0,1)'$ and $\tilde{a}^2$ a multiple of $(1,0)'$. So both faces of the quadrant are nonreflective and thus $C_1 = C_2 = \emptyset$. Furthermore, the direct triple is optimal for all $v \in \mathbb{R}^2_+$ and the optimal value $I^0(v)$ simplifies to

$$I^0(v) = -\theta_1\Big(\sqrt{v_1^2 + v_2^2} + v_1\Big).$$

For this case, Harrison [15] explicitly obtained the density function for the stationary distribution of the SRBM, which is given by $cr^{-1/2}\cos(\phi/2)\exp(-|\theta_1|(v_1 + r))$ with $v = (r\cos(\phi), r\sin(\phi))'$ and some constant $c > 0$. It is reassuring to note that, as expected, the exponent obtained by Harrison exactly matches the VP value above.

7.2. *Skew-symmetric case*

In [18,19] it was demonstrated that the stationary density function for a $(\mathbb{R}^d_+, \theta, \Gamma, R)$-SRBM admits a separable, exponential form if and only if the data satisfies the following *skew-symmetry condition*:

$$2\Gamma = RD^{-1}\Lambda + \Lambda D^{-1}R', \tag{7.1}$$

where $D = \mathrm{diag}(R)$ and $\Lambda = \mathrm{diag}(\Gamma)$. Let us then assume that the skew-symmetry condition holds. In this case, a $(\mathbb{R}^d_+, \theta, \Gamma, R)$-SRBM has a stationary distribution if and only if (4.2) holds. Furthermore, the stationary density is given by $c\exp(-\eta' v)$, where c is a normalizing constant and

$$\eta = -2\Lambda^{-1}DR^{-1}\theta. \tag{7.2}$$

In this subsection, we discuss the solution to the VP in two dimensions in the case that the data satisfies (7.1) and (4.2). This provides a check of the large deviations analysis versus an explicit calculation of the stationary density.

Consider a $(\mathbb{R}^2_+, \theta, \Gamma, R)$-SRBM. Without loss of we assume the following:

$$R = \begin{pmatrix} 1 & r_2 \\ r_1 & 1 \end{pmatrix}, \qquad \Gamma = \begin{pmatrix} \gamma_1 & \gamma_3 \\ \gamma_3 & \gamma_2 \end{pmatrix},$$

where $\gamma_1 > 0$ and $\gamma_2 > 0$. Note that $D = D^{-1}$ is just the identity matrix and the skew-symmetry condition (7.1) in fact reduces to just one equation:

$$2\gamma_3 = r_1\gamma_1 + r_2\gamma_2. \tag{7.3}$$

With the skew-symmetry condition, we have a number of interesting simplifications to the expressions derived for the VP, which are summarized in our next theorem.

Theorem 7.1. Consider a $(\mathbb{R}^2_+, \theta, \Gamma, R)$-SRBM whose data satisfies (7.1) and (4.2). For this SRBM the following hold:

(a) $\tilde{a}^1 = \tilde{a}^2$.

(b) At least one of the two faces is reflective.

(c) If F_1 is reflective, and F_2 is not, $C_2 = \emptyset$ and the cone C_1 covers the entire state space, namely, $C_1 \supset \mathbb{R}^2_+$.

(d) If F_2 is reflective, and F_1 is not, $C_1 = \emptyset$ and the cone C_2 covers the entire state space.

(e) If both faces are reflective, the cones C_1 and C_2 partition the state space $\mathbb{R}^2_+$ and possess a common boundary which has direction $\tilde{a}^1 = \tilde{a}^2$.

(f) For any $v \in \mathbb{R}^+_2$, the optimal value $I(v) = \eta' v$, with η given in (7.2).

Figure 3 illustrates the three general possibilities that can occur for VP solutions in the skew-symmetry case, as outlined in (c)–(e).

Now, let $\tilde{a} = \tilde{a}^1 = \tilde{a}^2$. A consequence of our theorem is that, for any $v \in \mathbb{R}^2_+$, an optimal triple is always a broken triple, except when both faces are reflective and v is in the direction of $\tilde{a}$. In this exceptional case, the direct path to v is optimal. The theorem also verifies that the optimal value $I(v)$ is given by the exponent in [18], as expected.

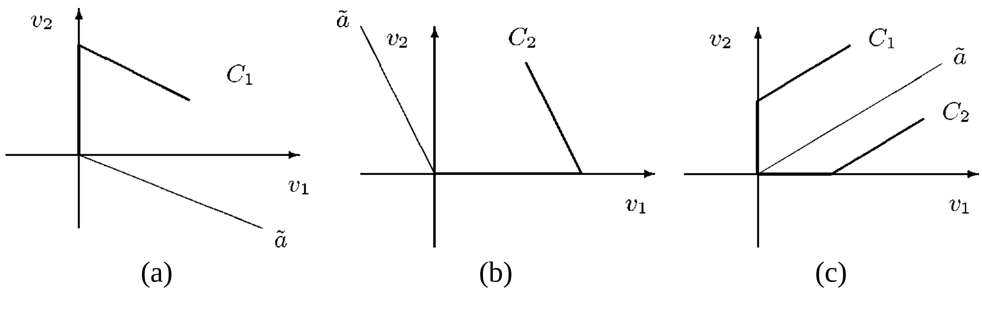

Figure 3. Optimal paths for the skew-symmetric case: (a) F_2 is nonreflective, (b) F_1 is nonreflective, (c) both F_1 and F_2 are reflective.

Proof of theorem 7.1. We first prove (a). Using (3.2), we have

$$a^1 = \theta - 2\frac{(-r_1\theta_1 + \theta_2)}{(r_1^2\gamma_1 - 2r_1\gamma_3 + \gamma_2)}(\gamma_3 - \gamma_1 r_1, \gamma_2 - \gamma_3 r_1)',$$
$$a^2 = \theta - 2\frac{(\theta_1 - r_2\theta_2)}{(\gamma_1 - 2r_2\gamma_3 + r_2^2\gamma_2)}(\gamma_1 - \gamma_3 r_2, \gamma_3 - \gamma_2 r_2)'.$$

By (7.3), we have

$$a_1^1 = -\frac{\gamma_2(-\theta_1 + r_2\theta_2) + r_1\gamma_1(r_1\theta_1 - \theta_2)}{(1 - r_1r_2)\gamma_2}, \tag{7.4}$$
$$a_2^2 = -\frac{r_2\gamma_2(-\theta_1 + r_2\theta_2) + \gamma_1(r_1\theta_1 - \theta_2)}{(1 - r_1r_2)\gamma_1}. \tag{7.5}$$

To prove $\tilde{a}^1 = \tilde{a}^2$, it is sufficient to show that

$$\Gamma^{-1}(\tilde{a}^1 - \theta) = \Gamma^{-1}(\tilde{a}^2 - \theta). \tag{7.6}$$

By the definition of $\tilde{a}^i$, we have

$$\tilde{a}^i = a^i - 2\langle a^i, \Gamma n^i\rangle \Gamma n^i = \theta - 2(\theta' p^i)\Gamma p^i - 2\langle a^i, \Gamma n^i\rangle \Gamma n^i.$$

Thus, (7.6) reduces to

$$-2(\theta' p^1)p^1 - 2\langle a^1, \Gamma n^1\rangle n^1 = -2(\theta' p^2)p^2 - 2\langle a^2, \Gamma n^2\rangle n^2,$$

or

$$-2\frac{(\theta_2 - r_1\theta_1)}{(1 - r_1r_2)\gamma_2}\begin{pmatrix} -r_1 \\ 1 \end{pmatrix} - \frac{2a_1^1}{\gamma_1}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = -2\frac{(\theta_1 - r_2\theta_2)}{(1 - r_1r_2)\gamma_1}\begin{pmatrix} 1 \\ -r_2 \end{pmatrix} - \frac{2a_2^2}{\gamma_2}\begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{7.7}$$

Using the expressions for a_1^1 and a_2^2 in (7.4)–(7.5), one can easily check that (7.7) indeed holds, thus proving $\tilde{a}^1 = \tilde{a}^2$.

To prove (b), we first note from (7.1) that $v'Rv > 0$ for any $v \in \mathbb{R}^2$. Thus, $1 - r_1r_2 = (r_1, -1)R(r_1, -1)' > 0$. Therefore, with (3.5) and part (a), we have that $\tilde{a} = \tilde{a}^1 = \tilde{a}^2 = (-a_1^1, -a_2^2)'$. Furthermore, from (4.2) we also have that $r_2\theta_2 - \theta_1 > 0$ and $r_1\theta_1 - \theta_2 > 0$. Collecting all of these facts we have the following conclusions. If $r_1 > 0$, then (7.4) implies $\tilde{a}_1 > 0$; if $r_2 > 0$, then (7.5) implies $\tilde{a}_2 > 0$. If both $r_1 \leqslant 0$ and $r_2 \leqslant 0$, then R^{-1} has nonnegative entries. This fact, along with the linear equalities (7.4)–(7.5), imply that at least one of the $-a_i^i$ is positive. Since at least one component of $\tilde{a}$ is positive in every case, we have now established (b).

Parts (c)–(e) follow directly from parts (a), (b) and the definitions of C_i and reflectivity.

It remains to prove (f). From (c)–(e), we know that, for any $v \in \mathbb{R}_+^2$, $I(v)$ is equal to either $\langle \tilde{a}^1 - \theta, v\rangle$ or $\langle \tilde{a}^2 - \theta, v\rangle$. Since $\tilde{a} = \tilde{a}^1 = \tilde{a}^2$, it is sufficient to prove

$\langle \tilde{a} - \theta, v \rangle = \eta' v$ for any $v \in \mathbb{R}^2_+$, or equivalently, $\Gamma^{-1}(\tilde{a} - \theta) = \eta$. Since $\Gamma^{-1}(\tilde{a} - \theta)$ is equal to either side of (7.7), we have

$$\Gamma^{-1}(\tilde{a} - \theta) = \left(-2\frac{(\theta_1 - r_2\theta_2)}{(1 - r_1 r_2)\gamma_1}, -2\frac{(\theta_2 - r_1\theta_1)}{(1 - r_1 r_2)\gamma_2} \right)'$$

which is equal to η as desired. This concludes the proof of the theorem. □

8. Extensions and further research

We now comment on extending these ideas to more general polyhedral state spaces. It is our hope that our study will provide a framework and road map especially for further research on higher dimensional problems.

It is clear that much of the analysis in section 5 will hold in higher dimensions and for general polyhedral state spaces. However, even for more general regions in two dimensions, we may have to choose an optimal escape path from more than three possible types. In other words, the VP can still be reduced to finite choice problem, but it may be more difficult to easily characterize the solutions as we did for the orthant in section 6.

We encounter more serious difficulties when passing to the VP in three or more dimensions. A primary challenge to extending the results is to investigate if an analog to theorem 5.1 holds in higher dimensions. From a computational standpoint, it would be most desirable if we could eliminate paths which “spiral” around the boundary before traversing the interior of the orthant. An example of such a principle is a conjecture of Majewski [29] which states that at most d pieces are needed to solve the VP in $\mathbb{R}^d_+$. This would limit our search in three dimensions to 10 types of paths, significantly reducing computational effort, even if it is no longer practical to completely characterize the solutions to the VP, as we have for the quadrant. It is an interesting open problem to characterize under what conditions, the VP may be so reduced to finite choice problem. Further work is also needed in proving LDPs like conjecture 4.1 for SRBMs and even more general regulated processes.

Acknowledgements

We are grateful to C. Knessl, Kurt Majewski, Neil O’Connell, and Ruth Williams for generously sharing their ideas and preprints. Part of this research was performed while Jim Dai was spending his sabbatical visit to Stanford from December 1998 to June 1999. He would like to thank the financial support of The Georgia Tech Foundation, the Department of Engineering-Economic Systems/Operations Research and the Graduate School of Business of Stanford. Partial support from NSF grants DMI-9457336 and DMI-9813345 is also acknowledged. A good portion of this research was performed while John Hasenbein was visiting the Centro de Investigación en Matemáticas (CIMAT) in Guanajuato, Mexico. He would like to thank CIMAT and the National

Science Foundation, who jointly provided an International Research Fellowship (NSF Grant INT-9971484), which made this visit possible.

References

[1] R. Atar and P. Dupuis, Large deviations and queueing networks: Methods for rate function identification, Technical Report LCDS 98-19, Lefschetz Center for Dynamical Systems, Brown University (1998).

[2] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences* (Academic Press, New York, 1979).

[3] A. Bernard and A. El Kharroubi, Régulations déterministes et stochastiques dans le premier "orthant" de $\mathbb{R}^n$, Stochastics Stochastics Rep. 34 (1991) 149–167.

[4] D. Bertsimas, I. Paschalidis and J.N. Tsitsiklis, On the large deviations behaviour of acyclic networks of $G/G/1$ queues, Ann. Appl. Probab. 8 (1998) 1027–1069.

[5] A.A. Borovkov and A.A. Mogulskii, Large deviations for stationary Markov chains in a quarter plane, in: *Probability Theory and Mathematical Statistics*, Tokyo, 1995 (World Sci. Publishing, River Edge, NJ, 1996) pp. 12–19.

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

[7] H. Chen and A. Mandelbaum, Hierarchical modeling of stochastic networks II: Strong approximations, in: *Stochastic Modeling and Analysis of Manufacturing Systems*, ed. D.D. Yao (Springer, New York, 1994) chapter 3, pp. 107–132.

[8] J.G. Dai and J.M. Harrison, Reflected Brownian motion in an orthant: Numerical methods for steady-state analysis, Ann. Appl. Probab. 2 (1992) 65–86.

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Applications of Mathematics, Vol. 38 (Springer, New York, 1998).

[10] P. Dupuis and R.S. Ellis, The large deviation principle for a general class of queueing systems, Trans. Amer. Math. Soc. 347 (1996) 2689–2751.

[11] P. Dupuis and H. Ishii, Large deviation behaviour of two competing queues, in: *Proc. of the 22nd Annual Conf. on Information Sciences and Systems*, March 1988.

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

[13] P. Dupuis and K. Ramanan, A Skrokhod problem formulation and large deviation analysis of a processor sharing model, Queueing Systems 28 (1998) 109–124.

[14] P. Dupuis and R.J. Williams, Lyapunov functions for semimartingale reflecting Brownian motions, Ann. Probab. 22 (1994) 680–702.

[15] J.M. Harrison, The diffusion approximation for tandem queues in heavy traffic, Adv. in Appl. Probab. 10 (1978) 886–905.

[16] J.M. Harrison and V. Nguyen, Brownian models of multiclass queueing networks: Current status and open problems, Queueing Systems 13 (1993) 5–40.

[17] J.M. Harrison and M.I. Reiman, Reflected Brownian motion on an orthant, Ann. Probab. 9 (1981) 302–308.

[18] J.M. Harrison and R.J. Williams, Brownian models of open queueing networks with homogeneous customer populations, Stochastics 22 (1987) 77–115.

[19] J.M. Harrison and R.J. Williams, Multidimensional reflected Brownian motions having exponential stationary distributions, Ann. Probab. 15 (1987) 115–137.

[20] D.G. Hobson and L. Rogers, Recurrence and transience of reflecting Brownian motion in the quadrant, Math. Proc. Cambridge Phil. Soc. 113 (1994) 387–399.

[21] I. Ignatyuk, V. Malyshev and V. Scherbakov, Boundary effects in large deviations problems, Russian Math. Surveys 49 (1994) 41–99.
[22] G. Kieffer, The large deviation principle for two-dimensional stable systems, Ph.D. thesis, University of Massachusetts (1995).
[23] C. Knessl, Diffusion approximation to two parallel queues with processor sharing, IEEE Trans. Automat. Control 36 (1991) 1356–1367.
[24] C. Knessl, On the diffusion approximation to a fork-join queuing model, SIAM J. Appl. Math. 51 (1991) 160–171.
[25] C. Knessl and C. Tier, A diffusion model for two tandem queues with general renewal input, Comm. Statist. Stochastic Models 15 (1999) 299–343.
[26] K. Majewski, Large deviations of stationary reflected Brownian motions, in: *Stochastic Networks: Theory and Applications*, eds. F.P. Kelly, S. Zachary and I. Ziedins, Royal Statistical Society (Oxford Univ. Press, Oxford, 1996).
[27] K. Majewski, Solving variational problems associated with large deviations of reflected Brownian motions, Working paper (December, 1996).
[28] K. Majewski, Heavy traffic approximations of large deviations of feedforward queueing networks, Queueing Systems 28 (1998) 125–155.
[29] K. Majewski, Large deviations of the steady-state distribution of reflected processes with applications to queueing systems, Queueing Systems 29 (1998) 351–381.
[30] N. O'Connell, Large deviations for departures from a shared buffer, J. Appl. Probab. 34 (1997) 753–766.
[31] N. O'Connell, Large deviations for queue lengths at a multi-buffered resource, J. Appl. Probab. 35 (1998) 240–245.
[32] M.I. Reiman and R.J. Williams, A boundary property of semimartingale reflecting Brownian motions, Probab. Theory Related Fields 77 (1988) 87–97 and 80 (1989) 633.
[33] A. Shwartz and A. Weiss, *Large Deviations for Performance Analysis* (Chapman & Hall, New York, 1995).
[34] L.M. Taylor and R.J. Williams, Existence and uniqueness of semimartingale reflecting Brownian motions in an orthant, Probab. Theory Related Fields 96 (1993) 283–317.
[35] R.J. Williams, Recurrence classification and invariant measure for reflected Brownian motion in a wedge, Ann. Probab. 13 (1985) 758–778.
[36] R.J. Williams, Semimartingale reflecting Brownian motions in the orthant, in: *Stochastic Networks*, eds. F.P. Kelly and R.J. Williams,The IMA Volumes in Mathematics and its Applications, Vol. 71 (Springer, New York, 1995) pp. 125–137.
[37] R.J. Williams, On the approximation of queueing networks in heavy traffic, in: *Stochastic Networks: Theory and Applications*, eds. F.P. Kelly, S. Zachary and I. Ziedins, Royal Statistical Society (Oxford Univ. Press, Oxford, 1996).