

Large Deviations for the Empirical Measures of  
Reflecting Brownian Motion and Related  
Constrained Processes in  $\mathbb{R}_+$

Amarjit Budhiraja\*  
University of North Carolina  
Chapel Hill, NC 27599-3260

Paul Dupuis †  
Lefschetz Center for Dynamical Systems  
Division of Applied Mathematics  
Brown University  
Providence, RI 02912

May 27, 2001

---

\*Research supported in part by the IBM junior faculty development award, University of North Carolina at Chapel Hill.

†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).

## Abstract

We consider the large deviations properties of the empirical measure for one dimensional constrained processes, such as reflecting Brownian motion, the M/M/1 queue, and discrete time analogues. Because these processes do not satisfy the strong stability assumptions that are usually assumed when studying the empirical measure, there is significant probability (from the perspective of large deviations) that the empirical measure charges the point at infinity. We prove the large deviation principle and identify the rate function for the empirical measure for these processes. No assumption of any kind is made with regard to the stability of the underlying process.

**Key Words:** Markov process, constrained process, large deviations, empirical measure, stability, reflecting Brownian motion.

**AMS Classification:** 60F10, 60J25, 93E20.

# 1 Introduction

Let  $\{X_n, n \in \mathbb{N}_0\}$  be a Markov process on a Polish space  $S$ , with transition kernel  $p(x, dy)$ . The empirical measure (or normalized occupation measure) for this process is defined by

$$L^n(A) \doteq \frac{1}{n} \sum_{i=0}^{n-1} \delta_{X_i}(A),$$

where  $\delta_x$  is the probability measure that places mass 1 at  $x$ , and  $A$  is any Borel subset of  $S$ . One of the cornerstones in the general theory of large deviations, due to Donsker and Varadhan in [5], was the development of a large deviation principle for the occupation measures for a wide class of Markov chains taking values in a compact state space. This work also studied the empirical measure large deviation principle (LDP) for continuous time Markov processes, where  $L^n$  is replaced by

$$L^T(A) \doteq \frac{1}{T} \int_0^T \delta_{X(t)}(A) dt$$

and  $X(\cdot)$  is a suitable continuous time Markov process. In subsequent work [6], the results of [5] were extended to Markov processes with an arbitrary Polish state space. These results significantly extended Sanov's theorem, which treats the independent and identically distributed (iid) case and was at that time the state-of-the-art. This work has found many applications since that time, and the general topic has developed into one of the most fertile areas of research in large deviations.

Three main assumptions appear in the empirical measure LDP results proved in [5, 6], and also in most subsequent papers on the same subject. The first is a sort of mixing or transitivity condition, and is the key to the proof of the large deviation lower bound. The second condition is a Feller property on the transition kernel, which is used in the proof of the upper bound. The third condition, and the one of prime interest in the present work, is a strong assumption on the stability of the underlying Markov process. For example, suppose that the underlying process is the solution to the  $\mathbb{R}^n$ -valued stochastic differential equation

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t),$$

where the dimensions of the Wiener process  $W$  and  $b$  and  $\sigma$  are compatible, and  $b$  and  $\sigma$  are Lipschitz continuous. Then satisfaction of the stability assumption would require the existence of a Lyapunov function  $V : \mathbb{R}^n \rightarrow \mathbb{R}$  with the property that

$$\langle b(x), V_x(x) \rangle + \frac{1}{2} \text{tr} [\sigma(x)\sigma'(x)V_{xx}(x)] \rightarrow -\infty \text{ as } \|x\| \rightarrow \infty.$$

Here  $V_x$  and  $V_{xx}$  denote the gradient and Hessian of  $V$ , respectively,  $\text{tr}$  denotes trace, and the prime denotes transpose.

A condition of this sort implies a strong restoring force towards bounded sets, and in fact a force that grows without bound as  $\|x\| \rightarrow \infty$ . It is required for a simple reason, and that is to keep the probability that the occupation measure “charges” points at  $\infty$  small from a large deviation perspective. Under such conditions, the probability that a sample path wanders out to infinity (i.e., eventually escapes every compact set) on the time interval  $[0, T]$  ( $\{0, 1, \dots, n-1\}$  in discrete time) is super-exponentially small. It is a reasonable condition in many circumstances. For example, in the setting of the linear system

$$dX(t) = AX(t)dt + BdW(t),$$

it simply requires that the matrix  $A$  correspond to a stable (deterministic) linear system. Nonetheless, the condition is not satisfied by some very basic processes. Examples include Brownian motion, reflecting Brownian motion, and the Markov process that corresponds to the  $M/M/1$  queue.

In the present paper we consider a class of one dimensional reflected (or constrained) processes which includes the last two examples, and obtain the large deviation principle for the empirical measures without any stability assumptions at all. The restriction to  $\mathbb{R}_+$  allows us to focus on one main issue: how to deal with the possibility that some of the mass of the occupation measure is placed on the point  $\infty$  (asymptotically and from the perspective of large deviations). In a more general setting there are many ways the underlying process can wander out to infinity, and hence the analysis becomes more complex. We defer this general setup to later work.

We begin our study with a discrete time constrained random walk on  $\mathbb{R}_+$ . This Markov chain is introduced in Section 2. Once the empirical measure LDP for this family of models is obtained, one can obtain the LDP for the continuous time Markov processes described by a reflected Brownian motion and a  $M/M/1$  queue via the standard technique of approximating by suitable super exponentially close processes (cf. [5, Section 3]). This is discussed in greater detail in Section 6.

The removal of the strong stability assumption fundamentally changes the nature of both the large deviation result and the proofs. In particular, given that the empirical measure will put mass on infinity, one must have detailed information on how this happens. We will adopt the weak convergence method of [7]. This approach is natural for the problem at hand, and indeed the combination of appropriately constructed test functions and weak convergence supplies us with exactly the sort of information we need (see, e.g., Lemma 3.9).

As stated earlier in the introduction, the fundamental results on the empirical measure LDP for Markov processes were obtained in [5, 6]. Subsequently, a

large amount of work on this topic has been done by various authors in refining these basic results. We refer the reader to [3, 4] for a detailed history of the problem. Most of the available work studies Markov processes that satisfy the strong stability assumptions. One notable exception is the work by Ney and Nummelin [10, 11], where large deviation probabilities for additive functionals of Markov chain are considered, essentially assuming only the irreducibility of the underlying Markov chain. However, the goals there are quite different from ours in that the authors obtain *local* large deviation results. Other authors, such as [1, 8, 2] study the large deviation lower bounds under weaker hypothesis than those in [5, 6]. The proof of the lower bound in these papers does not require any stability assumption on the underlying Markov chain. However, in the absence of strong stability, their lower bound is not (in general) the best possible. To illustrate the basic issue we restrict our attention to the discrete time model introduced in Section 2. Let  $\mathcal{P}(\mathbb{R}_+)$  be the space of probability measures on  $\mathbb{R}_+$ . Let  $F : \mathcal{P}(\mathbb{R}_+) \mapsto \mathbb{R}$  be a continuous and bounded map defined as

$$F(\nu) \doteq \int_{\mathbb{R}_+} f(x)\nu(dx), \quad \nu \in \mathcal{P}(\mathbb{R}_+).$$

Here  $f$  is a real-valued continuous and bounded function on  $\mathbb{R}_+$  such that  $f(x)$  converges, to say  $f^\infty$ , as  $x \rightarrow \infty$ . For such a function the lower bound in the cited papers will imply

$$\begin{aligned} & \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ \exp \left( -n \int_{\mathbb{R}_+} f(x) dL^n(x) \right) \right] \\ & \geq - \inf_{\nu \in \mathcal{P}(\mathbb{R}_+)} \left\{ I_1(\nu) + \int_{\mathbb{R}_+} f(x)\nu(dx) \right\}, \end{aligned} \tag{1.1}$$

where  $I_1$  is defined in Section 2 (see (2.2)). In contrast, for the class of one dimensional processes studied in this paper we will show that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \left[ \exp \left( -n \int_{\mathbb{R}_+} f(x) dL^n(x) \right) \right] \\ & = - \inf_{\nu \in \mathcal{P}(\mathbb{R}_+), \rho \in [0,1]} \left\{ \rho I_1(\nu) + (1 - \rho)J + \rho \int_{\mathbb{R}_+} f(x)\nu(dx) + (1 - \rho)f_\infty \right\}, \end{aligned} \tag{1.2}$$

where  $J$  is the part of the rate function that accounts for the possibility that mass might wander off to infinity. Clearly, the expression on the right side of (1.2) provides a sharper lower bound than the expression on the right side of (1.1). This paper considers a much more general form of the function  $F$ , and in fact we prove the full Laplace principle (and hence the large deviation principle) for the empirical measures  $\{L^n, n \in \mathbb{N}\}$  when considered as elements of  $\mathcal{P}(\bar{\mathbb{R}}_+)$ , where  $\mathcal{P}(\bar{\mathbb{R}}_+)$  is the space of probability measures on the one point compactification of  $\mathbb{R}_+$ .

It is important to note that even though we consider our underlying Markov chain to evolve in a compact Polish space ( $\bar{\mathbb{R}}_+$ ), the usual techniques of proving empirical measure LDP for Markov chains with compact state spaces do not apply. One reason is that if one extends the transition probability function of the Markov chain in (2.1) in the natural fashion by setting  $p(\infty, dy) \doteq \delta_{\{\infty\}}(dy)$ , i.e., making the point at  $\infty$  an absorbing state, then the resulting Markov chain does not satisfy the typical transitivity conditions that are needed for the proof of the lower bound. The proof of the upper bound for compact state space Markov chains, in essence, only uses the Feller property of the Markov chain. It is easy to see that the extended transition probability function introduced above is Feller with respect to the natural topology on  $\bar{\mathbb{R}}_+$  and so one can use the methodology of [5] to obtain an upper bound. This upper bound will be governed by the function

$$I^*(\nu) = \inf_q \left\{ \int_{\bar{\mathbb{R}}_+} R(q(x, \cdot) \parallel p(x, \cdot)) \nu(dx) \right\}, \quad \nu \in \mathcal{P}(\bar{\mathbb{R}}_+),$$

where infimum is taken over all transition probability kernels  $q$  on  $\bar{\mathbb{R}}_+ \times \bar{\mathbb{R}}_+$  with respect to which  $\nu$  is invariant and  $R(\cdot \parallel \cdot)$  denotes the relative entropy function (see Section 2 for precise definitions). Suppose that  $\nu$  puts some mass on  $\infty$  and that  $I^*(\nu) < \infty$ . Let  $q(x, dy)$  achieve the infimum in the definition of  $I^*(\nu)$ . Then the definition of relative entropy implies  $q(\infty, dy) = \delta_{\{\infty\}}(dy)$ , and hence there is no contribution from  $\infty$ :

$$I^*(\nu) = \left\{ \int_{\mathbb{R}_+} R(q(x, \cdot) \parallel p(x, \cdot)) \nu(dx) \right\}.$$

The rate function  $I$  we obtain satisfies  $I(\nu) \geq I^*(\nu)$ , and  $I(\nu) > I^*(\nu)$  if  $I(\nu) < \infty$  and  $\nu(\{\infty\}) > 0$ . Thus the point at  $\infty$  makes a contribution to the rate function, and in fact a careful analysis of the manner in which mass tends to infinity is needed to properly account for this contribution.

We now give a brief outline of the paper. In Section 2, we present the basic discrete time model and state the empirical measure large deviation result (Theorem 2.6) for this model. Sections 3, 4 and 5 are devoted to the proof of Theorem 2.6. Section 3 deals with the Laplace principle upper bound. In Section 4 we present some useful properties of the rate function. Section 5 proves the Laplace principle lower bound. Finally, in Section 6 we indicate how to obtain the empirical measure large deviations for the  $M/M/1$  queue and reflected Brownian motion via super exponentially close approximations by discrete time Markov chains of the form studied in Sections 3-5. We also state a conjecture on the form of the rate function for the LDP of the empirical measure of a one-dimensional Brownian motion.

The notation used in this paper is as follows. We denote the set of positive integers by  $\mathbb{N}$ , the set of nonnegative integers by  $\mathbb{N}_0$ , the set of reals by  $\mathbb{R}$  and

the set of nonnegative reals by  $\mathbb{R}_+$ . The space of real continuous functions on  $\mathbb{R}_+$  with compact support will be denoted by  $C_c(\mathbb{R}_+)$  and the space of real, bounded twice continuously differentiable functions on  $\mathbb{R}_+$  will be denoted by  $C_b^2(\mathbb{R}_+)$ . For a complete separable metric space  $S$ ,  $BM(S)$  denotes the class of real-valued bounded measurable functions on  $S$ ,  $C_b(S)$ , the subclass of  $BM(S)$  of continuous functions on  $S$ ,  $\mathcal{B}(S)$ , the Borel  $\sigma$ -field on  $S$  and  $\mathcal{P}(S)$ , the space of probability measures on  $(S, \mathcal{B}(S))$  endowed with the weak convergence topology. We abbreviate the statement “the sequence  $\nu_n$  in  $\mathcal{P}(S)$  converges weakly to  $\nu$ ” by  $\nu_n \Rightarrow \nu$ .  $\mathcal{M}(S)$  denotes the space of sub-probability measures, i.e., positive finite measures on  $(S, \mathcal{B}(S))$  with total mass not exceeding 1. This space is also given the topology corresponding to the weak convergence of measures. For  $f \in BM(S)$ ,  $\nu \in \mathcal{P}(S)$  and  $g \in C_b(S)$  we denote  $\int_S f(x)d\nu(x)$  by  $\langle f, \nu \rangle$  and  $\sup_{x \in S} |g(x)|$  by  $\|g\|_\infty$ . The probability measure on  $S$  which is concentrated at the single point  $x \in S$  is denoted by  $\delta_x$ , and the indicator function of a set  $A$  by  $1_A$ . A function  $I$  mapping  $S$  into  $[0, \infty]$  is called a rate function if for all  $M \in [0, \infty)$  the level set  $\{x \in S : I(x) \leq M\}$  is compact. The infimum over an empty set, by convention, is taken to be  $\infty$ . As another convention,  $0 \times \infty$  is taken to be 0. A family of random variables with values in a Polish space is said to be tight if the corresponding family of probability laws is tight.

## 2 The Discrete Time Model

Let  $\{W_n, n \in \mathbb{N}_0\}$  be a sequence of iid  $\mathbb{R}$ -valued random variables on the space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $\theta \in \mathcal{P}(\mathbb{R})$  denote the common probability law, and for  $x \in \mathbb{R}$  let  $x^+$  denote the maximum of  $x$  and 0. We recursively define the  $\mathbb{R}_+$ -valued constrained random walk  $\{X_n, n \in \mathbb{N}_0\}$  by setting

$$X_{n+1} \doteq (X_n + W_n)^+, \quad n \in \mathbb{N}_0. \quad (2.1)$$

If  $X_0$  has probability law  $\delta_x$ , then we denote expectation by  $\mathbb{E}_x$ . The transition probability function of the Markov chain  $X_n$  is denoted by  $p(x, dy)$ . The following conditions on  $\theta$  will be used in various places.

**Assumption 2.1** For all  $\alpha \in \mathbb{R}$

$$\int_{\mathbb{R}} \exp(\alpha y) \theta(dy) < \infty.$$

Let  $p^{(k)}(x, dy)$  denote the  $k$ -step transition probability function of the Markov chain  $X_n$ .

**Assumption 2.2** *There exist  $l_0, n_0 \in \mathbb{N}$  such that for all  $x_1, x_2 \in \mathbb{R}_+$*

$$\sum_{i=l_0}^{\infty} \frac{1}{2^i} p^{(i)}(x_1, dy) \ll \sum_{j=n_0}^{\infty} \frac{1}{2^j} p^{(j)}(x_2, dy).$$

**Remark 2.3** Because  $p(x, dy)$  is the transition kernel for a constrained random walk driven by iid noise, one can easily pose conditions on  $\theta$  which guarantee Assumption 2.2. For example, suppose  $\theta$  is absolutely continuous with respect to  $\lambda$ , where  $\lambda$  denotes Lebesgue measure on  $\mathbb{R}$ . A sufficient condition is that there exists  $\delta > 0$  such that  $\frac{d\theta}{d\lambda}(y) > \delta$  for  $y \in (-\delta, \delta)$ . If  $\theta$  is supported on the integers, as would be the case in a discrete time approximation to a queueing model, then one cannot expect a condition such as Assumption 2.2 to hold for all  $x_1$  and  $x_2$ . However, Assumption 2.2 will hold when  $x_1$  and  $x_2$  are restricted to be integers, and this suffices if one wishes to study the large deviations of the empirical measures when only integer-valued initial conditions are considered. This is discussed further in Section 6.

**Definition 2.4 (Relative Entropy)** *For each  $\nu \in \mathcal{P}(\mathbb{R})$ , the relative entropy  $R(\cdot \| \nu)$  is a map from  $\mathcal{P}(\mathbb{R})$  to  $\bar{\mathbb{R}}_+$  defined as follows. If  $\gamma \in \mathcal{P}(\mathbb{R})$  is absolutely continuous with respect to  $\nu$  and if  $\log \frac{d\gamma}{d\nu}$  is integrable with respect to  $\gamma$ , then*

$$R(\gamma \| \nu) \doteq \int_{\mathbb{R}} \left( \log \frac{d\gamma}{d\nu} \right) d\gamma.$$

*In all other cases,  $R(\gamma \| \nu)$  is defined to be  $\infty$ .*

For  $n \in \mathbb{N}$ , the empirical measure  $L^n$  corresponding to the Markov chain  $\{X_i, i = 1, \dots, n\}$  is the  $\mathcal{P}(\mathbb{R}_+)$ -valued random variable defined by

$$L^n(A) \doteq \frac{1}{n} \sum_{j=0}^{n-1} \delta_{X_j}(A), \quad A \in \mathcal{B}(\mathbb{R}_+).$$

The empirical measure is also called the (normalized) occupation measure. Let  $\bar{\mathbb{R}}_+ \doteq \mathbb{R}_+ \cup \{\infty\}$ , the one point compactification of  $\mathbb{R}_+$ . Then  $\bar{\mathbb{R}}_+$  and  $\mathcal{P}(\bar{\mathbb{R}}_+)$  are compact Polish spaces. With an abuse of notation, a probability measure  $\nu \in \mathcal{P}(\mathbb{R}_+)$  will be denoted by  $\nu$  even when considered as an element of  $\mathcal{P}(\bar{\mathbb{R}}_+)$ . In this paper we are interested in the large deviation properties for the family  $\{L^n, n \in \mathbb{N}\}$  of random variables with values in the compact Polish space  $\mathcal{P}(\bar{\mathbb{R}}_+)$ . To introduce the rate function that will govern the large deviation probabilities for this family, we need the following notation and definition.

**Definition 2.5 (Stochastic Kernel)** *Let  $(\mathcal{V}, \mathcal{A})$  be a measurable space and  $S$  a Polish space. Let  $\tau(dy | x)$  be a family of probability measures on  $(S, \mathcal{B}(S))$*



parametrized by  $x \in \mathcal{V}$ . We call  $\tau(dy|x)$  a stochastic kernel on  $S$  given  $\mathcal{V}$  if for every Borel subset  $E$  of  $S$  the map  $x \in \mathcal{V} \mapsto \tau(E|x)$  is measurable. The class of all such stochastic kernels is denoted by  $\mathcal{S}(S|\mathcal{V})$ .

If  $p^* \in \mathcal{S}(\mathbb{R}_+|\mathbb{R}_+)$ , then  $p^*$  is a probability transition function and we will write  $p^*(dy|x)$  as  $p^*(x, dy)$ . We say that a probability measure  $\nu \in \mathcal{P}(\mathbb{R}_+)$  is invariant with respect to a stochastic kernel  $p^* \in \mathcal{S}(\mathbb{R}_+|\mathbb{R}_+)$ , if  $\nu(A) = \int_{\mathbb{R}_+} p^*(x, A)\nu(dx)$  for all  $A \in \mathcal{B}(\mathbb{R}_+)$ . We will also refer to  $\nu$  as a  $p^*$ -invariant probability measure.

We can now define the rate function associated with the family  $\{L^n, n \in \mathbb{N}\}$ . Given  $q^* \in \mathcal{S}(\mathbb{R}|\mathbb{R}_+)$ , we associate the stochastic kernel  $p^* \in \mathcal{S}(\mathbb{R}_+|\mathbb{R}_+)$  which is consistent with  $q^*$  under the constraint mechanism by setting

$$p^*(x, A) \doteq \int_{\mathbb{R}} 1_{\{(x+z)^+ \in A\}} q^*(dz|x), \quad A \in \mathcal{B}(\mathbb{R}_+), \quad x \in \mathbb{R}_+.$$

For  $\nu \in \mathcal{P}(\mathbb{R}_+)$ , let

$$I_1(\nu) \doteq \inf_{\{q^* \in \mathcal{A}_1(\nu)\}} \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot))\nu(dx), \quad (2.2)$$

where  $\mathcal{A}_1(\nu)$  is the collection of all  $q^*$  for which  $\nu$  is  $p^*$ -invariant. Also define by  $\mathcal{P}_{\text{tr}}(\mathbb{R})$  the class of all  $\sigma \in \mathcal{P}(\mathbb{R})$  for which  $\int_{\mathbb{R}} |z|\sigma(dz) < \infty$  and  $\int_{\mathbb{R}} z\sigma(dz) \geq 0$ . Here the subscript tr stands for ‘‘transient.’’ Strictly speaking, this class includes measures that produce a null recurrent process, but transient is more suggestive of what is intended. Let

$$J \doteq \inf_{\sigma \in \mathcal{P}_{\text{tr}}(\mathbb{R})} R(\sigma \parallel \theta). \quad (2.3)$$

Under Assumption 2.1 it follows that  $J < \infty$ . For  $\nu \in \mathcal{P}(\bar{\mathbb{R}}_+)$ , let  $\hat{\nu} \in \mathcal{P}(\mathbb{R}_+)$  be defined as follows. If  $\nu(\mathbb{R}_+) \neq 0$ , then for  $A \in \mathcal{B}(\mathbb{R}_+)$

$$\hat{\nu}(A) \doteq \frac{\nu(A \cap \mathbb{R}_+)}{\nu(\mathbb{R}_+)}.$$

Otherwise,  $\hat{\nu}$  can be taken to be an arbitrary element of  $\mathcal{P}(\mathbb{R}_+)$ . We define the rate function as follows. For  $\nu \in \mathcal{P}(\bar{\mathbb{R}}_+)$

$$I(\nu) \doteq \nu(\mathbb{R}_+)I_1(\hat{\nu}) + (1 - \nu(\mathbb{R}_+))J. \quad (2.4)$$

Our main result is the following.

**Theorem 2.6** *Suppose that Assumptions 2.1 and 2.2 hold. Then for all  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$  and all  $x \in \mathbb{R}_+$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_x [\exp(-nF(L^n))] = - \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\}.$$

Furthermore,  $I(\cdot)$  is a rate function on  $\mathcal{P}(\bar{\mathbb{R}}_+)$ .

**Proof:** The fact that the left hand side in the last display is at most the expression in the right hand side (Laplace principle upper bound) is proved in Theorem 3.1. The reverse inequality (Laplace principle lower bound) is proved in Theorem 5.1. The proof that  $I(\cdot)$  is a rate function on  $\mathcal{P}(\bar{\mathbb{R}}_+)$  is given in Theorem 4.1 (c). ■

**Remark 2.7** Theorem 2.6 can be summarized by the statement that the family  $\{L^n, n \in \mathbb{N}\}$  satisfies the Laplace principle on  $\mathcal{P}(\bar{\mathbb{R}}_+)$  with the rate function  $I(\cdot)$ . From Theorem 1.2.3 of [7] it follows that the family  $\{L^n, n \in \mathbb{N}\}$  satisfies the large deviation principle on  $\mathcal{P}(\bar{\mathbb{R}}_+)$  with the rate function  $I(\cdot)$ .

**Remark 2.8** The convergence in Theorem 2.6 is in fact uniform for  $x$  in compact subsets of  $\mathbb{R}_+$ . See [7, Section 8.4]

We now present an important corollary of Theorem 2.6. Denote by  $\mathcal{S}_0$  the subclass of  $C_b(\mathbb{R}_+)$  consisting of functions  $f$  for which  $f(x)$  converges as  $x \rightarrow \infty$ , with the limit denoted by  $f^\infty$ . Such an  $f$  can be extended to a function  $\bar{f}$  on  $\bar{\mathbb{R}}_+$  by defining  $\bar{f}(\infty) \doteq f^\infty$ . It is easy to see that there is a one to one correspondence between  $\mathcal{S}_0$  and  $C_b(\bar{\mathbb{R}}_+)$ , given by  $f \in \mathcal{S}_0 \mapsto \bar{f} \in C_b(\bar{\mathbb{R}}_+)$  and  $g \in C_b(\bar{\mathbb{R}}_+) \mapsto g|_{\mathbb{R}_+} \ni \mathcal{S}_0$ .

**Corollary 2.9** Let  $F : \mathcal{P}(\mathbb{R}_+) \mapsto \mathbb{R}$  be given by

$$F(\nu) \doteq G(\langle \nu, f_1 \rangle, \dots, \langle \nu, f_k \rangle),$$

where  $G \in C_b(\mathbb{R}^k)$  and  $f_i \in \mathcal{S}_0$ ,  $i = 1, \dots, k$ . Then for all  $x \in \mathbb{R}_+$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_x [\exp(-nF(L^n))] = - \inf_{\nu \in \mathcal{P}(\mathbb{R}_+), \rho \in [0,1]} \{\rho I_1(\nu) + (1-\rho)J + F^*(\rho, \nu)\},$$

where  $F^* \in C_b([0,1] \times \mathcal{P}(\mathbb{R}_+))$  is defined by

$$F^*(\rho, \nu) \doteq G(\rho \langle f_1, \nu \rangle + (1-\rho)f_1^\infty, \dots, \rho \langle f_k, \nu \rangle + (1-\rho)f_k^\infty).$$

### 3 Laplace Principle Upper Bound

The main result of this section is the following.

**Theorem 3.1** Suppose that Assumption 2.1 holds, and define  $I$  by (2.4). Then for all  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$  and all  $x \in \mathbb{R}_+$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_x [\exp(-nF(L^n))] \leq - \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\}. \quad (3.1)$$

Throughout this paper there will be many constructions and results that are analogous to those used in [7, Chapter 8] to study empirical measures under the classical strong stability assumption. While there are some differences in these constructions, in an effort to streamline the presentation we will emphasize those parts of the analysis that are new.

Our first step in the proof will be to give a variational representation for the prelimit expression on the left side of (3.1). This representation will take the form of the value function for a controlled Markov chain with an appropriate cost function. The representation will also be used in the proof of the lower bound in Section 5. We begin with the construction of a controlled Markov chain.

For  $n \in \mathbb{N}$  and  $j = 0, \dots, n$ , let  $\nu_j^n(dy | x, \gamma)$  be a stochastic kernel in  $\mathcal{S}(\mathbb{R} | \mathbb{R}_+ \times \mathcal{M}(\mathbb{R}_+))$ . For  $n \in \mathbb{N}$  and  $x \in \mathbb{R}_+$ , define a controlled sequence of  $\mathbb{R}_+ \times \mathcal{M}(\mathbb{R}_+) \times \mathbb{R}$ -valued random variables  $\{(\bar{X}_j^n, \bar{L}_j^n, \bar{W}_j^n), j = 0, \dots, n\}$  on some probability space  $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}_x)$  as follows. We set  $\bar{X}_0^n \doteq x$  and  $\bar{L}_0^n \doteq 0$ , and then for  $k = 0, \dots, n-1$  recursively define

$$\begin{aligned} \bar{\mathbb{P}}_x(\bar{W}_k^n \in dy | (\bar{X}_j^n, \bar{L}_j^n), j = 0, \dots, k) &\doteq \nu_k^n(dy | \bar{X}_k^n, \bar{L}_k^n) \\ \bar{X}_{k+1}^n &\doteq (\bar{X}_k^n + \bar{W}_k^n)^+ \\ \bar{L}_{k+1}^n &\doteq \bar{L}_k^n + \frac{1}{n} \delta_{\bar{X}_k^n}. \end{aligned} \quad (3.2)$$

We denote  $\bar{L}_n^n$  by  $\bar{L}^n$ . We now give the variational representation for

$$W^n(x) \doteq -\frac{1}{n} \log \mathbb{E}_x[\exp(-nF(L^n))] \quad (3.3)$$

in terms of the controlled sequences introduced above.

**Lemma 3.2** *Fix  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$  and let  $W^n(x)$  be defined via (3.3). Then, for all  $n \in \mathbb{N}$  and  $x \in \mathbb{R}_+$*

$$W^n(x) = \inf_{\{\nu_j^n\}} \bar{\mathbb{E}}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) + F(\bar{L}^n) \right], \quad (3.4)$$

where the infimum is taken over all possible sequences of stochastic kernels  $\{\nu_j^n; j = 0, \dots, n\}$  in  $\mathcal{S}(\mathbb{R} | \mathbb{R}_+ \times \mathcal{M}(\mathbb{R}_+))$ .

The proof of this lemma is similar to the proof of Theorem 4.2.2 of [7]. The main difference between the two results is that the latter gives a representation which involves the transition probability function,  $p(x, dy)$ , of the Markov chain rather than the probability law,  $\theta$ , of the noise sequence. In the representation, the original empirical measures  $L^n$  are replaced by  $\bar{L}^n$ , which are the empirical

measures for the process generated by the stochastic kernels  $\nu_j^n$ . We pay a cost of relative entropy for “twisting” the distribution from  $\theta$  to  $\nu_j^n$ , plus a cost of  $\bar{\mathbb{E}}_x F(\bar{L}^n)$  that depends on where the controlled empirical measure ends up. The representation exhibits  $W^n(x)$  as the minimum expected total cost.

Let  $\epsilon > 0$  be arbitrary. In view of the preceding lemma, for each  $n \in \mathbb{N}$  we can find a sequence of “ $\epsilon$ -optimal” stochastic kernels  $\{\bar{\nu}_j^n, j = 0, \dots, n\}$ , such that

$$W^n(x) + \epsilon \geq \bar{\mathbb{E}}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\bar{\nu}_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) + F(\bar{L}^n) \right]. \quad (3.5)$$

Since  $F$  is bounded, both  $|F(\bar{L}^n)|$  and  $|W^n(x)|$  are bounded above by  $\|F\|_\infty$ . Thus

$$\Delta \doteq \sup_n \bar{\mathbb{E}}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\bar{\nu}_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) \right] < \infty. \quad (3.6)$$

Define  $\bar{\nu}^n \in \mathcal{P}(\bar{\mathbb{R}}_+ \times \mathbb{R}^2)$  by

$$\bar{\nu}^n(dx \, dy \, dz) \doteq \frac{1}{n} \sum_{j=0}^{n-1} \delta_{\bar{X}_j^n}(dx) \bar{\nu}_j^n(dy | \bar{X}_j^n, \bar{L}_j^n) \delta_{\bar{W}_j^n}(dz). \quad (3.7)$$

**Lemma 3.3** *Suppose Assumption 2.1 holds. Then  $\{\bar{\nu}^n, n \in \mathbb{N}\}$  is a tight family of  $\mathcal{P}(\bar{\mathbb{R}}_+ \times \mathbb{R}^2)$ -valued random variables.*

The result is a consequence of (3.6) and follows along the lines of the proof of Proposition 8.2.5 of [7]. We remark that unlike in the above cited Proposition, we do not need to make any stability assumption on the underlying Markov chain. This is one of the key advantages of working with the representation in Lemma 3.2 given in terms of the probability law of the noise sequence  $(\theta)$  rather than the transition probability function of the Markov chain  $(p(x, dy))$ . The main idea is to show a tightness property of the average of the measures  $\bar{\nu}_j^n$ , which follows from the bound on the relative entropy and Assumption 2.1. Tightness of the third marginal of  $\bar{\nu}^n$  follows from this and the fact that  $\bar{\nu}_j^n$  selects the distribution of  $\bar{W}_j^n$ , while tightness of the first marginal is automatic since the space  $\bar{\mathbb{R}}_+$  is compact.

Take a convergent subsequence of  $\bar{\nu}^n$  and denote the limit point by  $\bar{\nu}$ . To simplify the notation, we retain  $n$  to denote this convergent subsequence, so that

$$\bar{\nu}^n \Rightarrow \bar{\nu} \quad \text{as } n \rightarrow \infty. \quad (3.8)$$

*For the rest of this section, we will assume that Assumption 2.1 is satisfied, and that (3.8) holds.*

For  $\gamma \in \mathcal{P}(\bar{\mathbb{R}}_+ \times \mathbb{R}^2)$  and  $1 \leq i < j \leq 3$ , denote by  $(\gamma)_i$  and  $(\gamma)_{i,j}$  the  $i$ -th marginal and the  $(i, j)$ -th marginal of  $\gamma$ , respectively. For example,  $(\gamma)_{2,3}$  is the element of  $\mathcal{P}(\mathbb{R} \times \mathbb{R})$  defined as

$$(\gamma)_{2,3}(A \times B) \doteq \gamma(\bar{\mathbb{R}}_+ \times A \times B), \quad A, B \in \mathcal{B}(\mathbb{R}).$$

**Lemma 3.4** *Let  $\bar{\nu}$  be as in (3.8). Then*

$$(\bar{\nu})_{1,2} = (\bar{\nu})_{1,3}, \quad a.s.$$

**Proof:** It suffices to show that for all  $g \in C_b(\bar{\mathbb{R}}_+)$  and  $h \in C_b(\mathbb{R})$ ,

$$\left| \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} g(x)h(y)(\bar{\nu}^n)_{1,2}(dx \, dy) - \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} g(x)h(y)(\bar{\nu}^n)_{1,3}(dx \, dy) \right| \quad (3.9)$$

converges to 0, as  $n \rightarrow \infty$ , in probability. Observe that

$$\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} g(x)h(y)(\bar{\nu}^n)_{1,2}(dx \, dy) = \frac{1}{n} \sum_{j=0}^{n-1} g(\bar{X}_j^n) \int_{\mathbb{R}} h(y) \bar{\nu}_j^n(dy \mid \bar{X}_j^n, \bar{L}_j^n)$$

and

$$\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} g(x)h(y)(\bar{\nu}^n)_{1,3}(dx \, dy) = \frac{1}{n} \sum_{j=0}^{n-1} g(\bar{X}_j^n) h(\bar{W}_j^n).$$

Thus the expression in (3.9) can be rewritten as

$$\Lambda \doteq \left| \frac{1}{n} \sum_{j=0}^{n-1} g(\bar{X}_j^n) \left( h(\bar{W}_j^n) - \int_{\mathbb{R}} h(y) \bar{\nu}_j^n(dy \mid \bar{X}_j^n, \bar{L}_j^n) \right) \right|.$$

We use the fact that the summands form a martingale difference. In other words, for  $0 \leq j < k \leq n-1$  and an arbitrary real bounded and measurable function  $\psi$ ,

$$\bar{\mathbb{E}}_x \left[ \psi(\bar{X}_j^n, \bar{L}_j^n, \bar{W}_j^n, \bar{X}_k^n, \bar{L}_k^n) \left( h(\bar{W}_k^n) - \int_{\mathbb{R}} h(y) \bar{\nu}_k^n(dy \mid \bar{X}_k^n, \bar{L}_k^n) \right) \right] = 0.$$

It follows that  $\bar{\mathbb{E}}_x[\Lambda^2]$  is  $O(1/n)$ . Thus the expression in (3.9) converges to 0 in probability, as  $n \rightarrow \infty$ . This proves the lemma. ■

**Lemma 3.5** *Let  $\{\bar{\nu}^n\}$  and  $\bar{\nu}$  be as in (3.8). Then*

$$\limsup_{C \rightarrow \infty} \sup_n \bar{\mathbb{E}}_x \int_{|z| > C} |z| (\bar{\nu}^n)_3(dz) = 0, \quad (3.10)$$

$$\bar{\mathbb{E}}_x \int_{\mathbb{R}} |z| (\bar{\nu})_3(dz) < \infty. \quad (3.11)$$

The proof of the lemma is similar to proof of Proposition 5.3.2 of [7]. As with the tightness, the key ingredient in the proof is the uniform bound on the expected relative entropy stated in (3.6).

For  $c \in (0, \infty)$ , define a real continuously differentiable function  $F_c$  on  $\mathbb{R}$  by

$$F_c(x) \doteq \begin{cases} \frac{1}{2} & \text{if } x \in (-\infty, c] \\ \frac{x^2}{2c^2} - \frac{x}{c} + 1 & \text{if } x \in (c, c^2 + c] \\ x - \frac{c^2}{2} - c + \frac{1}{2} & \text{if } x \in (c^2 + c, \infty). \end{cases} \quad (3.12)$$

We now present an elementary lemma concerning the function  $F_c$ .

**Lemma 3.6** *For all  $x, y \in \mathbb{R}$*

$$F_c(y) - F_c(x) = (y - x)F'_c(x) + R_c(x, y),$$

where  $F'_c$  denotes the derivative of  $F_c$  and the remainder  $R_c(x, y)$  satisfies the inequality

$$R_c(x, y) \leq \frac{|y - x|}{c} + |y - x|1_{|y-x| \geq c}. \quad (3.13)$$

**Proof:** We will only prove the result for the case when  $x \leq y$ . The result for  $y \leq x$  follows in a similar fashion. Note that the derivative  $F'_c$  is given as

$$F'_c(x) = \begin{cases} 0 & \text{if } x \in (-\infty, c] \\ \frac{x}{c^2} - \frac{1}{c} & \text{if } x \in (c, c^2 + c] \\ 1 & \text{if } x \in (c^2 + c, \infty). \end{cases}$$

Since  $F'_c(x)$  is a continuous function, we have

$$\begin{aligned} F_c(y) - F_c(x) &= \int_x^y F'_c(u) du \\ &= \int_x^y (F'_c(u) - F'_c(x)) du + (y - x)F'_c(x). \end{aligned}$$

Now define

$$R_c(x, y) \doteq \int_x^y (F'_c(u) - F'_c(x)) du.$$

Since  $F'_c$  is an increasing function bounded above by 1,

$$R_c(x, y) \leq \left( \int_x^y (F'_c(u) - F'_c(x)) du \right) 1_{|y-x| \leq c} + |y - x|1_{|y-x| \geq c}. \quad (3.14)$$

Now let  $x, y$  be such that  $|y - x| \leq c$ . Then

$$\begin{aligned} \int_x^y (F'_c(u) - F'_c(x)) du &\leq (y - x)(F'_c(y) - F'_c(x)) \\ &\leq \frac{(y - x)^2}{c^2} \\ &\leq \frac{|y - x|}{c}, \end{aligned}$$

where the first inequality follows on noting that  $F'_c(u) - F'_c(x) < F'_c(y) - F'_c(x)$  for all  $u \leq y$ , the second inequality is a consequence of the fact that for all  $-\infty < x < y < \infty$ ,  $F'_c(y) - F'_c(x) \leq (y - x)/c^2$  and the final inequality is obtained on using that  $|y - x| \leq c$ . Using the last inequality in (3.14), we have (3.13). This proves the lemma. ■

**Lemma 3.7** For  $c \in (0, \infty)$ , let  $F_c$  be given via (3.12). For  $n \in \mathbb{N}$ , let  $\bar{\nu}^n$  be defined via (3.7). Then

$$\begin{aligned} \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} y F'_c(x) (\bar{\nu}^n)_{1,3}(dx dy) &\geq -3 \int_{\mathbb{R}} |y| 1_{|y| \geq c} (\bar{\nu}^n)_3(dy) \\ &\quad - \frac{1}{c} \int_{\mathbb{R}} |y| (\bar{\nu}^n)_3(dy) - \frac{F_c(\bar{X}_0^n)}{n}. \end{aligned} \tag{3.15}$$

**Proof:** For  $j \in \{0, \dots, n-1\}$ , define  $\xi_j^n \doteq \bar{X}_{j+1}^n - \bar{X}_j^n$ . We begin by observing that

$$\begin{aligned} &\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} y F'_c(x) (\bar{\nu}^n)_{1,3}(dx dy) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \bar{W}_i^n F'_c(\bar{X}_i^n) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \bar{W}_i^n F'_c(\bar{X}_i^n) 1_{|\bar{W}_i^n| \geq c} + \frac{1}{n} \sum_{i=0}^{n-1} \bar{W}_i^n F'_c(\bar{X}_i^n) 1_{|\bar{W}_i^n| < c} \\ &\geq -\frac{1}{n} \sum_{i=0}^{n-1} |\bar{W}_i^n| 1_{|\bar{W}_i^n| \geq c} + \frac{1}{n} \sum_{i=0}^{n-1} \bar{W}_i^n F'_c(\bar{X}_i^n) 1_{|\bar{W}_i^n| < c} \\ &\geq -\frac{1}{n} \sum_{i=0}^{n-1} |\bar{W}_i^n| 1_{|\bar{W}_i^n| \geq c} + \frac{1}{n} \sum_{i=0}^{n-1} \xi_i^n F'_c(\bar{X}_i^n) 1_{|\bar{W}_i^n| < c} \\ &\geq -2 \frac{1}{n} \sum_{i=0}^{n-1} |\bar{W}_i^n| 1_{|\bar{W}_i^n| \geq c} + \frac{1}{n} \sum_{i=0}^{n-1} \xi_i^n F'_c(\bar{X}_i^n). \end{aligned} \tag{3.16}$$

The first inequality follows on recalling that  $F'_c$  is non-negative and bounded above by 1, the second equality is a consequence of the fact that  $F'_c(x)$  equals 0 for  $x \leq c$  and on the set  $\{\bar{X}_i^n > c\} \cap \{|\bar{W}_i^n| \leq c\}$ ,  $\xi_i^n$  equals  $\bar{W}_i^n$  and finally the third inequality is a consequence of the fact that  $|\xi_i^n| \leq |\bar{W}_i^n|$  and  $F'_c(x)$  takes values in  $[0, 1]$ . Next observe that, from Lemma 3.6

$$\begin{aligned} 0 &\leq \frac{1}{n} F_c(\bar{X}_n^n) \\ &= \frac{1}{n} \sum_{i=0}^{n-1} (F_c(\bar{X}_{i+1}^n) - F_c(\bar{X}_i^n)) + \frac{F_c(\bar{X}_0^n)}{n} \\ &= \frac{1}{n} \sum_{i=0}^{n-1} \xi_i^n F'_c(\bar{X}_i^n) + \frac{1}{n} \sum_{i=0}^{n-1} R_c(\bar{X}_i^n, \bar{X}_{i+1}^n) + \frac{F_c(\bar{X}_0^n)}{n}. \end{aligned}$$

Also, note that

$$R_c(\bar{X}_i^n, \bar{X}_{i+1}^n) \leq \frac{|\xi_i^n|}{c} + |\xi_i^n| 1_{|\xi_i^n| \geq c}.$$

Combining the last two observations we have

$$\begin{aligned} \frac{1}{n} \sum_{i=0}^{n-1} \xi_i^n F'_c(\bar{X}_i^n) &\geq -\frac{1}{n} \sum_{i=0}^{n-1} |\xi_i^n| 1_{|\xi_i^n| \geq c} - \frac{1}{c} \frac{1}{n} \sum_{i=0}^{n-1} |\xi_i^n| - \frac{F_c(\bar{X}_0^n)}{n} \\ &\geq -\frac{1}{n} \sum_{i=0}^{n-1} |W_i^n| 1_{|W_i^n| \geq c} - \frac{1}{c} \frac{1}{n} \sum_{i=0}^{n-1} |W_i^n| - \frac{F_c(\bar{X}_0^n)}{n}. \end{aligned} \tag{3.17}$$

Finally, substituting (3.17) in (3.16) gives

$$\begin{aligned} &\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} y F'_c(x) (\bar{\nu}^n)_{1,3}(dx dy) \\ &\geq -3 \frac{1}{n} \sum_{i=0}^{n-1} |\bar{W}_i^n| 1_{|\bar{W}_i^n| \geq c} - \frac{1}{c} \frac{1}{n} \sum_{i=0}^{n-1} |\bar{W}_i^n| - \frac{F_c(\bar{X}_0^n)}{n} \\ &= -3 \int_{\mathbb{R}} |y| 1_{|y| \geq c} (\bar{\nu}^n)_3(dy) - \frac{1}{c} \int_{\mathbb{R}} |y| (\bar{\nu}^n)_3(dy) - \frac{F_c(\bar{X}_0^n)}{n}. \end{aligned}$$

This proves the lemma. ■

**Lemma 3.8** *Let  $\bar{\nu}$  be as in (3.8). For  $c \in (0, \infty)$ , let  $F_c$  be defined via (3.12). Let  $\{c_k\}_{k \in \mathbb{N}}$  be a sequence in  $(0, \infty)$  such that  $c_k \rightarrow \infty$  as  $k \rightarrow \infty$ , and for all  $k \in \mathbb{N}$*

$$\bar{\mathbb{E}}_x[(\bar{\nu})_3\{c_k\}] = 0.$$

Then

$$\liminf_{k \rightarrow \infty} \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z F'_{c_k}(x) (\bar{\nu})_{1,3}(dx dz) \geq 0, \quad a.s. \tag{3.18}$$



**Proof:** Note that from Lemma 3.7

$$\begin{aligned} & \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z F'_{c_k}(x)(\bar{\nu}^n)_{1,3}(dx dz) \\ & \geq -3 \int_{\mathbb{R}} |z| 1_{|z| \geq c_k}(\bar{\nu}^n)_3(dz) - \frac{1}{c_k} \int_{\mathbb{R}} |z|(\bar{\nu}^n)_3(dz) - \frac{F_c(\bar{X}_0^n)}{n}, \end{aligned} \quad (3.19)$$

for all  $k \in \mathbb{N}$ , a.s. We would like to show that the analogous inequality holds with  $\bar{\nu}^n$  replaced by  $\bar{\nu}$ . By using the Skorohod representation theorem, we can assume without loss of generality that  $\bar{\nu}^n$  converges to  $\bar{\nu}$  almost surely. The term  $F_c(\bar{X}_0^n)/n$  disappears in the limit  $n \rightarrow \infty$ , since  $\bar{X}_0^n$  takes a fixed deterministic value for all  $n$ . We will now show that if  $\bar{\nu}^n \rightarrow \bar{\nu}$  almost surely then all other terms in the above inequality converge to the corresponding terms with  $\bar{\nu}^n$  replaced by  $\bar{\nu}$ .

We begin by considering  $\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z F'_{c_k}(x)(\bar{\nu}^n)_{1,3}(dxdz)$ . Note that for all  $g \in C_b(\mathbb{R})$ ,

$$\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} g(z) F'_{c_k}(x)(\bar{\nu}^n)_{1,3}(dxdz) \rightarrow \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} g(z) F'_{c_k}(x)(\bar{\nu})_{1,3}(dxdz)$$

a.s. We can approximate the identity by a bounded continuous function  $g(z)$  which equals  $z$  when  $|z| \leq C$  for a large constant  $C$ . The uniform integrability expressed in Lemma 3.5 then justifies the replacement of  $g(z)$  in the last display by  $z$ , giving

$$\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z F'_{c_k}(x)(\bar{\nu}^n)_{1,3}(dxdz) \rightarrow \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z F'_{c_k}(x)(\bar{\nu})_{1,3}(dxdz) \quad (3.20)$$

a.s. See, e.g., the proof of Lemma 5.3.6 in [7]. In exactly the same way, we have that

$$\int_{\mathbb{R}} |z| (\bar{\nu}^n)_3(dz) \rightarrow \int_{\mathbb{R}} |z| (\bar{\nu})_3(dz), \quad (3.21)$$

a.s. as  $n \rightarrow \infty$ . Finally we consider  $\int_{|z| \geq c_k} |z|(\bar{\nu}^n)_3(dz)$ . The condition assumed of the sequence  $c_k$  implies that the function  $z \mapsto |z| 1_{|z| \geq c_k}$  is continuous w.p.1 under  $\bar{\nu}$ , a.s. When combined with the uniform integrability stated in Lemma 3.5, it follows that

$$\int_{|z| \geq c_k} |z|(\bar{\nu}^n)_3(dz) \rightarrow \int_{|z| \geq c_k} |z|(\bar{\nu})_3(dz), \quad (3.22)$$

a.s. Thus taking the limit as  $n \rightarrow \infty$  in (3.19), we have from (3.20), (3.21) and (3.22) that

$$\begin{aligned} \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z F'_{c_k}(x)(\bar{\nu})_{1,3}(dx dz) & \geq -3 \int_{\mathbb{R}} |z| 1_{|z| \geq c_k}(\bar{\nu})_3(dz) \\ & \quad - \frac{1}{c_k} \int_{\mathbb{R}} |z|(\bar{\nu})_3(dz) \end{aligned}$$

Finally the proof is completed via an application of Lemma 3.5, (3.11) on letting  $k \rightarrow \infty$  in the last display. ■

Let  $\bar{\nu}$  be as in (3.4) and let  $\bar{q}$  be a stochastic kernel in  $\mathcal{S}(\mathbb{R} | \bar{\mathbb{R}}_+)$  such that

$$(\bar{\nu})_{1,2}(dx dy) = (\bar{\nu})_1(dx) \otimes \bar{q}(dy|x), \quad (3.23)$$

i.e. for  $A \in \mathcal{B}(\bar{\mathbb{R}}_+)$ ,  $B \in \mathcal{B}(\mathbb{R})$ ,

$$(\bar{\nu})_{1,2}(A \times B) = \int_A \bar{q}(B|x)(\bar{\nu})_1(dx).$$

As an immediate consequence of Lemma 3.8, we have the following result.

**Lemma 3.9** For  $\bar{P}_x$  - a.e.  $\bar{\omega}$  in the set  $\{(\bar{\nu})_1(\mathbb{R}_+) < 1\}$ ,

$$\int_{\mathbb{R}} |y| \bar{q}(dy|\infty) < \infty \quad (3.24)$$

and

$$\int_{\mathbb{R}} y \bar{q}(dy|\infty) \geq 0. \quad (3.25)$$

**Proof:** The inequality in (3.24) follows from the fact that

$$\begin{aligned} (1 - (\bar{\nu})_1(\mathbb{R}_+)) \int_{\mathbb{R}} |y| \bar{q}(dy|\infty) &\leq \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} |y| (\bar{\nu})_{1,2}(dx dy) \\ &= \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} |z| (\bar{\nu})_{1,3}(dx dz) \\ &= \int_{\mathbb{R}} |z| (\bar{\nu})_3(dz) \\ &< \infty, \end{aligned}$$

a.s., where the last step follows from Lemma 3.5, (3.11). In order to see (3.25), note that if the sequence  $\{c_k\}$  is chosen as in Lemma 3.8, then (3.18) holds. Furthermore, note that for all  $(x, z) \in \bar{\mathbb{R}}_+ \times \mathbb{R}$ ,  $|zF'_{c_k}(x)| \leq |z|$  and  $zF'_{c_k}(x) \rightarrow z1_{\{\infty\}}(x)$ , as  $k \rightarrow \infty$ . Thus observing that  $\int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} |z| (\bar{\nu})_{1,3}(dx dz) < \infty$ , we have via an application of dominated convergence theorem that

$$\begin{aligned} \lim_{k \rightarrow \infty} \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z F'_{c_k}(x) (\bar{\nu})_{1,3}(dx dz) &= \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} z 1_{\{\infty\}}(x) (\bar{\nu})_{1,3}(dx dz) \\ &= \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} y 1_{\{\infty\}}(x) (\bar{\nu})_{1,2}(dx dy) \\ &= (\bar{\nu})_1\{\infty\} \int_{\mathbb{R}} y \bar{q}(dy|\infty), \quad (3.26) \end{aligned}$$

where the second equality follows from Lemma 3.4. Combining (3.18) and (3.26), we have the result. ■

Let  $\bar{\nu}$  and  $\bar{q}$  be related as in (3.23). Define a stochastic kernel  $\bar{p} \in \mathcal{S}(\mathbb{R}_+ | \mathbb{R}_+)$  via

$$\bar{p}(x, A) \doteq \int_{\mathbb{R}} 1_{\{(x+z)^+ \in A\}} \bar{q}(dz|x).$$

Recall that  $(\bar{\nu})_1$  can be decomposed as

$$(\bar{\nu})_1(\cdot) = (\bar{\nu})_1(\mathbb{R}_+) \widehat{(\bar{\nu})}_1(\cdot) + (1 - (\bar{\nu})_1(\mathbb{R}_+)) \delta_\infty(\cdot).$$

The following lemma characterizes  $\widehat{(\bar{\nu})}_1$  as a  $\bar{p}$ -invariant probability measure.

**Lemma 3.10** *Let  $\bar{\nu}$ ,  $\bar{q}$  and  $\bar{p}$  be as above. For  $\bar{\mathbb{P}}_x$  - a.e.  $\bar{\omega}$  in the set  $\{(\bar{\nu})_1(\mathbb{R}_+) > 0\}$ ,  $\widehat{(\bar{\nu})}_1$  is  $\bar{p}$ -invariant, i.e.*

$$\widehat{(\bar{\nu})}_1(A) = \int_{\mathbb{R}} \bar{p}(x, A) \widehat{(\bar{\nu})}_1(dx), \quad \forall A \in \mathcal{B}(\mathbb{R}_+).$$

**Proof:** It suffices to show that for all  $g \in C_c(\mathbb{R}_+)$

$$\int_{\bar{\mathbb{R}}_+} g(x) (\bar{\nu})_1(dx) = (\bar{\nu})_1(\mathbb{R}_+) \int_{\mathbb{R}_+} \left( \int_{\mathbb{R}_+} g(y) \bar{p}(x, dy) \right) \widehat{(\bar{\nu})}_1(dx),$$

a.s. We begin by observing that

$$\begin{aligned} & (\bar{\nu})_1(\mathbb{R}_+) \int_{\mathbb{R}_+} \left( \int_{\mathbb{R}_+} g(y) \bar{p}(x, dy) \right) \widehat{(\bar{\nu})}_1(dx) \\ &= (\bar{\nu})_1(\mathbb{R}_+) \int_{\mathbb{R}_+} \left( \int_{\mathbb{R}} g((x+y)^+) \bar{q}(dy|x) \right) \widehat{(\bar{\nu})}_1(dx) \\ &= (\bar{\nu})_1(\mathbb{R}_+) \int_{\mathbb{R}_+ \times \mathbb{R}} g((x+y)^+) \widehat{(\bar{\nu})}_{1,2}(dx dy) \\ &= \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} H_g(x, y) (\bar{\nu})_{1,2}(dx dy) \\ &= \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} H_g(x, z) (\bar{\nu})_{1,3}(dx dz) \end{aligned} \tag{3.27}$$

a.s., where  $\widehat{(\bar{\nu})}_{1,2} \in \mathcal{P}(\mathbb{R}_+ \times \mathbb{R})$  is defined as

$$\widehat{(\bar{\nu})}_{1,2}(A \times B) \doteq \frac{(\bar{\nu})_{1,2}(A \times B)}{(\bar{\nu})_{1,2}(\mathbb{R}_+ \times \mathbb{R})}, \quad A \in \mathcal{B}(\mathbb{R}_+), \quad B \in \mathcal{B}(\mathbb{R}),$$

and  $H_g \in C_b(\bar{\mathbb{R}}_+ \times \mathbb{R})$  is defined as

$$H_g(x, z) \doteq \begin{cases} g((x+z)^+), & \text{if } (x, z) \in \mathbb{R}_+ \times \mathbb{R} \\ 0, & \text{otherwise.} \end{cases}$$

The next to last equality in (3.27) follows from the compact support property of  $g$  and the last equality is a consequence of Lemma 3.4. Thus in order to complete the proof of the lemma, it suffices to show that

$$\int_{\bar{\mathbb{R}}_+} g(x)(\bar{\nu}^n)_1(dx) - \int_{\bar{\mathbb{R}}_+ \times \mathbb{R}} H_g(x, z)(\bar{\nu}^n)_{1,3}(dx dz) \rightarrow 0 \quad (3.28)$$

in probability as  $n \rightarrow \infty$ . Note that the expression in (3.28) equals

$$\begin{aligned} & \frac{1}{n} \sum_{j=0}^{n-1} g(\bar{X}_j^n) - \frac{1}{n} \sum_{j=0}^{n-1} g((\bar{X}_j^n + \bar{W}_j^n)^+) \\ &= \frac{1}{n} \sum_{j=0}^{n-1} g(\bar{X}_j^n) - \frac{1}{n} \sum_{j=0}^{n-1} g(\bar{X}_{j+1}^n) \\ &= \frac{1}{n} [g(\bar{X}_0^n) - g(\bar{X}_n^n)]. \end{aligned}$$

From the boundedness of  $g$ , it follows that the above expression is  $O(1/n)$ . This proves (3.28) and hence the lemma. ■

Finally, we ready to prove Theorem 3.1.

**Proof of Theorem 3.1:** Recall from (3.5) that

$$\begin{aligned} W^n(x) + \epsilon &\geq \bar{\mathbb{E}}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\bar{\nu}_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) + F(\bar{L}^n) \right] \\ &= \bar{\mathbb{E}}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\delta_{\bar{X}_j^n} \otimes \bar{\nu}_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \delta_{\bar{X}_j^n} \otimes \theta(\cdot)) + F(\bar{L}^n) \right] \\ &\geq \bar{\mathbb{E}}_x \left[ R \left( \frac{1}{n} \sum_{j=0}^{n-1} \delta_{\bar{X}_j^n} \otimes \bar{\nu}_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \frac{1}{n} \sum_{j=0}^{n-1} \delta_{\bar{X}_j^n} \otimes \theta(\cdot) \right) + F(\bar{L}^n) \right] \\ &= \bar{\mathbb{E}}_x [R((\bar{\nu}^n)_{12} \| (\bar{\nu}^n)_1 \otimes \theta) + F((\bar{\nu}^n)_1)], \end{aligned} \quad (3.29)$$

where the second line in the above display exploits a property of relative entropy with respect to the decomposition of measures according to their marginals ([7, C.3.3]), and the second inequality follows from Jensen's inequality. Next, recalling (3.8), we can assume without loss of generality that  $\bar{\nu}^n \Rightarrow \bar{\nu}$  a.s. as  $n \rightarrow \infty$ . Thus from (3.29) we have that

$$\begin{aligned} \liminf_{n \rightarrow \infty} W^n(x) + \epsilon &\geq \bar{\mathbb{E}}_x \left[ \liminf_{n \rightarrow \infty} (R((\bar{\nu}^n)_{12} \| (\bar{\nu}^n)_1 \otimes \theta) + F((\bar{\nu}^n)_1)) \right] \\ &\geq \bar{\mathbb{E}}_x [R((\bar{\nu})_{12} \| (\bar{\nu})_1 \otimes \theta) + F((\bar{\nu})_1)], \end{aligned} \quad (3.30)$$

where the first inequality follows from Fatou's lemma and the last inequality is a consequence of lower semi-continuity of  $R(\cdot \| \cdot)$  and the continuity of  $F$ .

Next note that the definitions of  $I_1$  and  $J$  and Lemmas 3.9 and 3.10 imply

$$\begin{aligned}
R((\bar{\nu})_{12} \| (\bar{\nu})_1 \otimes \theta) &= \int_{\bar{\mathbb{R}}_+} R(\bar{q}(\cdot|x) \| \theta(\cdot))(\bar{\nu})_1(dx) \\
&= (\bar{\nu})_1(\mathbb{R}_+) \int_{\mathbb{R}_+} R(\bar{q}(\cdot|x) \| \theta(\cdot))(\widehat{(\bar{\nu})_1})(dx) \\
&\quad + (1 - (\bar{\nu})_1(\mathbb{R}_+))R(\bar{q}(\cdot|\infty) \| \theta(\cdot)) \\
&\geq (\bar{\nu})_1(\mathbb{R}_+)I_1(\widehat{(\bar{\nu})_1}) + (1 - (\bar{\nu})_1(\mathbb{R}_+))J \\
&= I((\bar{\nu})_1).
\end{aligned}$$

When combined with (3.30) we have

$$\begin{aligned}
\liminf_{n \rightarrow \infty} W^n(x) + \epsilon &\geq \bar{\mathbb{E}}_x \{I((\bar{\nu})_1) + F((\bar{\nu})_1)\} \\
&\geq \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\}.
\end{aligned}$$

Since  $\epsilon > 0$  is arbitrary, the result follows. ■

## 4 Properties of the Rate Function

In this section we will prove the following result.

**Theorem 4.1** *Let the function  $I : \mathcal{P}(\bar{\mathbb{R}}_+) \mapsto [0, \infty]$  be given via (2.4). Then the following conclusions hold.*

- (a)  $I$  is a convex function.
- (b) Suppose that for some  $\alpha \in (0, \infty)$ ,

$$\int_{\mathbb{R}} e^{\alpha|z|} \theta(dz) < \infty.$$

Let  $\pi \in \mathcal{P}(\bar{\mathbb{R}}_+)$ . Then there exists  $\sigma_0 \in \mathcal{P}(\mathbb{R})$  and  $\bar{q} \in \mathcal{S}(\mathbb{R}_+ | \mathbb{R})$  such that

- 1.  $\int_{\mathbb{R}} |z| \sigma_0(dz) < \infty$  and  $\int_{\mathbb{R}} z \sigma_0(dz) \geq 0$ .
- 2.  $\hat{\pi}$  is  $\bar{p}$ -invariant, where  $\bar{p} \in \mathcal{S}(\mathbb{R}_+ | \mathbb{R}_+)$  is defined as

$$\bar{p}(x, A) \doteq \int_{\mathbb{R}} 1_{\{(x+z) \in A\}} \bar{q}(dz|x), \quad A \in \mathcal{B}(\mathbb{R}_+), \quad x \in \mathbb{R}_+ \quad (4.1)$$

3. The infimum in (2.2) (with  $\nu$  replaced by  $\hat{\pi}$ ) and (2.3) are attained at  $\bar{q}$  and  $\sigma_0$ , respectively. I.e.

$$I(\pi) = \pi(\mathbb{R}_+) \int_{\mathbb{R}_+} R(\bar{q}(\cdot|x) \parallel \theta(\cdot)) \hat{\pi}(dx) + (1 - \pi(\mathbb{R}_+)) R(\sigma_0 \parallel \theta).$$

(c) Suppose that Assumption 2.1 is satisfied. Then for all  $M \in [0, \infty)$ , the level set  $\{\pi \in \mathcal{P}(\bar{\mathbb{R}}_+) : I(\pi) \leq M\}$  is a compact set in  $\mathcal{P}(\bar{\mathbb{R}}_+)$ .

**Proof:** Recall that the relative entropy function is convex. Thus to prove the convexity of  $I$ , it suffices to show that  $I_1$  is convex. We begin by noting that for any  $\nu \in \mathcal{P}(\mathbb{R}_+)$ , there is a 1 : 1 correspondence between  $q^* \in \mathcal{S}(\mathbb{R}_+ | \mathbb{R})$  for which  $\nu$  is  $p^*$ -invariant, where  $p^*$  is defined as

$$p^*(x, A) \doteq \int_{\mathbb{R}} 1_{\{(x+z)^+ \in A\}} q^*(dz|x), \quad A \in \mathcal{B}(\mathbb{R}_+), \quad x \in \mathbb{R}_+,$$

and  $\tau \in \mathcal{P}(\mathbb{R}_+ \times \mathbb{R})$  such that

$$\langle f, \nu \rangle = \langle f, (\tau)_1 \rangle = \int_{\mathbb{R}_+ \times \mathbb{R}} f((x+z)^+) \tau(dx dz), \quad \forall f \in C_b(\mathbb{R}_+). \quad (4.2)$$

Define by  $\mathcal{A}(\nu)$  the class of all  $\tau \in \mathcal{P}(\mathbb{R}_+ \times \mathbb{R})$  such that (4.2) holds and let  $\mathcal{A}_1(\nu)$  be as defined below (2.2). The one to one correspondence between  $\mathcal{A}(\nu)$  and  $\mathcal{A}_1(\nu)$  can be described by the relations that if  $\tau \in \mathcal{A}(\nu)$  and  $\tau$  is decomposed as  $\tau(dx dz) = (\tau)_1(dx)q(dz|x)$ , then  $(\tau)_1 = \nu$  and  $q \in \mathcal{A}_1(\nu)$ . Conversely, if  $q \in \mathcal{A}_1(\nu)$  and  $\tau \in \mathcal{P}(\mathbb{R}_+ \times \mathbb{R})$  is defined as  $\tau(dx dz) \doteq \nu(dx)q(dz|x)$ , then  $\tau \in \mathcal{A}(\nu)$ . Thus for  $\nu \in \mathcal{P}(\mathbb{R}_+)$

$$\begin{aligned} I_1(\nu) &= \inf_{q \in \mathcal{A}_1(\nu)} \int_{\mathbb{R}_+} R(q(\cdot|x) \parallel \theta(\cdot)) \nu(dx) \\ &= \inf_{q \in \mathcal{A}_1(\nu)} R(\nu \otimes q \parallel \nu \otimes \theta) \\ &= \inf_{\tau \in \mathcal{A}(\nu)} R(\tau \parallel \nu \otimes \theta), \end{aligned}$$

where  $\nu \otimes q$  denotes the probability measure  $q(dz|x) \nu(dx)$  on  $\mathbb{R}_+ \times \mathbb{R}$ . Now let  $\nu^1, \nu^2 \in \mathcal{P}(\mathbb{R}_+)$ . We would like to show that for all  $\alpha \in [0, 1]$ ,

$$I_1(\alpha\nu^1 + (1-\alpha)\nu^2) \leq \alpha I_1(\nu^1) + (1-\alpha) I_1(\nu^2). \quad (4.3)$$

Observe that if either  $\mathcal{A}(\nu^1)$  or  $\mathcal{A}(\nu^2)$  is empty, then (4.3) holds trivially. Now let  $\tau^1 \in \mathcal{A}(\nu^1)$  and  $\tau^2 \in \mathcal{A}(\nu^2)$ . Clearly,  $\alpha\tau^1 + (1-\alpha)\tau^2 \in \mathcal{A}(\alpha\nu^1 + (1-\alpha)\nu^2)$ . Thus

$$\begin{aligned} I_1(\alpha\nu^1 + (1-\alpha)\nu^2) &\leq R(\alpha\tau^1 + (1-\alpha)\tau^2 \parallel \alpha\nu^1 \otimes \theta + (1-\alpha)\nu^2 \otimes \theta) \\ &\leq \alpha R(\tau^1 \parallel \nu^1 \otimes \theta) + (1-\alpha) R(\tau^2 \parallel \nu^2 \otimes \theta), \end{aligned}$$

where the last step follows from the convexity of the relative entropy function. Since  $\tau^1 \in \mathcal{A}(\nu^1)$  and  $\tau^2 \in \mathcal{A}(\nu^2)$  are arbitrary, we have from the last display that

$$\begin{aligned} I_1(\alpha\nu^1 + (1-\alpha)\nu^2) &\leq \alpha \inf_{\tau^1 \in \mathcal{A}(\nu^1)} R(\tau^1 \parallel \nu^1 \otimes \theta) \\ &\quad + (1-\alpha) \inf_{\tau^2 \in \mathcal{A}(\nu^2)} R(\tau^2 \parallel \nu^2 \otimes \theta) \\ &= \alpha I_1(\nu^1) + (1-\alpha) I_1(\nu^2). \end{aligned}$$

This proves (a). We now consider (b).

We can assume without loss of generality that  $I(\pi) < \infty$ . It suffices to show that the infimum in (2.3) is attained for some  $\sigma_0 \in \mathcal{P}(\mathbb{R})$ , and that for  $\nu \in \mathcal{P}(\mathbb{R}_+)$  with  $I_1(\nu) < \infty$ , the infimum in (2.2) is attained for some  $q \in \mathcal{A}_1(\nu)$ . We consider the latter part of the statement first. Fix  $\nu \in \mathcal{P}(\mathbb{R}_+)$  for which  $I_1(\nu) < \infty$ . Observe that

$$\begin{aligned} I_1(\nu) &= \inf_{\tau \in \mathcal{A}(\nu)} \{R(\tau \parallel (\tau)_1 \otimes \theta)\} \\ &= \inf_{\tau \in \mathcal{A}^*(\nu)} \{R(\tau \parallel (\tau)_1 \otimes \theta)\}, \end{aligned} \tag{4.4}$$

where

$$\mathcal{A}^*(\nu) \doteq \{\tau \in \mathcal{A}(\nu) : R(\tau \parallel (\tau)_1 \otimes \theta) \leq I_1(\nu) + 1\}.$$

We now show that  $\mathcal{A}^*(\nu)$  is compact. This will prove that the infimum in (4.4) is attained. From the one to one correspondence between  $\mathcal{A}(\nu)$  and  $\mathcal{A}_1(\nu)$  we will then have that the infimum in (2.2) is attained for some  $q \in \mathcal{A}_1(\nu)$ . From the lower semi-continuity of the map

$$\mathcal{P}(\mathbb{R}_+ \times \mathbb{R}) \times \mathcal{P}(\mathbb{R}_+ \times \mathbb{R}) \ni (\tau^1, \tau^2) \mapsto R(\tau^1 \parallel \tau^2) \in [0, \infty]$$

and the fact that  $\mathcal{A}(\nu)$  is closed, it follows that  $\mathcal{A}^*(\nu)$  is closed. Hence we need only show that  $\mathcal{A}^*(\nu)$  is relatively compact in  $\mathcal{P}(\mathbb{R}_+ \times \mathbb{R})$ . Since  $(\tau)_1 = \nu$  for all  $\tau \in \mathcal{A}^*(\nu)$ ,  $\{(\tau)_1 : \tau \in \mathcal{A}^*(\nu)\}$  is relatively compact in  $\mathcal{P}(\mathbb{R}_+)$ . Thus it suffices to show the relative compactness of  $\{(\tau)_2 : \tau \in \mathcal{A}^*(\nu)\}$  in  $\mathcal{P}(\mathbb{R})$ . In order to prove this relative compactness, we show that

$$\sup_{\tau \in \mathcal{A}^*(\nu)} \int_{\mathbb{R}} |z| \tau(dx dz) < \infty. \tag{4.5}$$

For  $k \in (0, \infty)$  and  $z \in \mathbb{R}$ , let  $c_k(z) \doteq \min\{\alpha|z|, k\}$ . Then for  $\tau \in \mathcal{A}^*(\nu)$ ,

$$\begin{aligned} \int_{\mathbb{R}} c_k(z) (\tau)_2(dz) &= \int_{\mathbb{R}_+ \times \mathbb{R}} c_k(z) \tau(dx dz) \\ &= \left( \int_{\mathbb{R}_+ \times \mathbb{R}} c_k(z) \tau(dx dz) - \log \int_{\mathbb{R}_+ \times \mathbb{R}} e^{c_k(z)} ((\tau)_1 \otimes \theta)(dx dz) \right) \end{aligned}$$

$$\begin{aligned}
& + \log \int_{\mathbb{R}} e^{c_k(z)} \theta(dz) \\
\leq & R(\tau \parallel (\tau)_1 \otimes \theta) + \log \int_{\mathbb{R}} e^{\alpha|z|} \theta(dz) \\
\leq & I_1(\nu) + 1 + \log \int_{\mathbb{R}} e^{\alpha|z|} \theta(dz),
\end{aligned}$$

where the first inequality uses the Donsker-Varadhan variational formula for relative entropy [7, Lemma 1.4.3(a)]. Letting  $k \rightarrow \infty$ , we have

$$\int_{\mathbb{R}} |z|(\tau)_2(dz) \leq \frac{1}{\alpha} \left( I_1(\nu) + 1 + \log \int_{\mathbb{R}} e^{\alpha|z|} \theta(dz) \right) < \infty.$$

As  $\tau \in \mathcal{A}^*(\nu)$  is arbitrary, this proves (4.5). Thus  $\mathcal{A}^*(\nu)$  is relatively compact, and since it is closed, we have the desired compactness. Thus the infimum in (2.2) is attained for some  $q \in \mathcal{A}_1(\nu)$ .

Now we consider the infimum in (2.3). Observe that

$$\begin{aligned}
J & \doteq \inf_{\sigma \in \mathcal{P}(\mathbb{R})} \left\{ R(\sigma \parallel \theta) : \int_{\mathbb{R}} |z| \sigma(dz) < \infty, \int_{\mathbb{R}} z \sigma(dz) \geq 0 \right\} \\
& = \inf_{\sigma \in \mathcal{P}_{\text{tr}}^*(\mathbb{R})} R(\sigma \parallel \theta),
\end{aligned} \tag{4.6}$$

where

$$\mathcal{P}_{\text{tr}}^*(\mathbb{R}) \doteq \left\{ \sigma \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |z| \sigma(dz) < \infty, \int_{\mathbb{R}} z \sigma(dz) \geq 0, R(\sigma \parallel \theta) \leq J + 1 \right\}.$$

Again using the variational formula for relative entropy, one can show in a manner similar to the proof of (4.5) that

$$\sup_{\sigma \in \mathcal{P}_{\text{tr}}^*(\mathbb{R})} \int_{\mathbb{R}} |z| \sigma(dz) < \infty. \tag{4.7}$$

This proves the relative compactness of  $\mathcal{P}_{\text{tr}}^*(\mathbb{R})$  in  $\mathcal{P}(\mathbb{R})$ . Finally observing that  $\mathcal{P}_{\text{tr}}^*(\mathbb{R})$  is closed, we have that  $\mathcal{P}_{\text{tr}}^*(\mathbb{R})$  is compact. Thus the infimum in (4.6) is attained. This proves (b).

As is often the case when applying weak convergence arguments to prove large deviation results, the compactness of the level sets of  $I$  (item (c) in the theorem) is proved using a deterministic analogue of the argument used to prove the upper bound in Section 3. Details of the argument are given in the Appendix. ■

## 5 Laplace Principle Lower Bound

The main result of this section is the following.



**Theorem 5.1** *Suppose that Assumptions 2.1 and 2.2 hold. Then for all  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$  and  $x \in \mathbb{R}_+$ ,*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_x [\exp(-nF(L^n))] \geq - \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\}.$$

Note that from Theorem 3.1 we have that

$$\inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\} \leq \|F\|_\infty < \infty.$$

Let  $\epsilon \in (0, \infty)$  be arbitrary and let  $\pi \in \mathcal{P}(\bar{\mathbb{R}}_+)$  be such that

$$I(\pi) + F(\pi) < \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\} + \epsilon. \quad (5.1)$$

In view of Lemma 3.2, in order to prove the theorem, it suffices to show that there exists a sequence of stochastic kernels  $\{\nu_j^n, j = 0, \dots\}$  in  $\mathcal{S}(\mathbb{R} | \mathbb{R}_+ \times \mathcal{M}(\mathbb{R}_+))$  and a corresponding controlled sequence of  $\mathbb{R}_+ \times \mathcal{M}(\mathbb{R}_+) \times \mathbb{R}$ -valued random variables  $\{\bar{X}_j^n, \bar{L}_j^n, \bar{W}_j^n\}_{j=0}^n$  on some probability space  $(\Omega, \mathcal{F}, \bar{\mathbb{P}}_x)$  defined as in (3.2) such that

$$\limsup_{n \rightarrow \infty} \bar{\mathbb{E}}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) + F(\bar{L}^n) \right] \leq I(\pi) + F(\pi). \quad (5.2)$$

For the rest of section we will assume, without loss of generality, that  $0 < \pi(\mathbb{R}_+) < 1$ . The cases where  $\pi(\mathbb{R}_+)$  is 0 or 1 are proved using simple modifications of the argument used in this (harder) case.

To prove an inequality like (5.2) we must find controls  $\nu_j^n$  which steer the controlled occupation measures  $\bar{L}^n$  to  $\pi$  and for which the expected mean relative entropy converges to the rate function. An appropriate definition of the control is suggested by  $\bar{q}$  and  $\sigma_0$  in part (b) of Theorem 4.1. In fact, if  $\hat{\pi}$  were the unique invariant measure for  $\bar{p}$  and the mean of  $\sigma_0$  was strictly positive then we could use the following scheme. Consider the time interval  $1, \dots, n$ . For a fraction  $\pi(\mathbb{R}_+)$  of this time (i.e., from 0 till approximately  $n\pi(\mathbb{R}_+)$ ) we use  $\nu_j^n(dz | \bar{X}_j^n, \bar{L}_j^n) = \bar{q}(dz | \bar{X}_j^n)$ . The ergodic theorem would then guarantee the desired convergence of the controlled occupation measures and the expected mean relative entropy. For the remaining time (from  $n\pi(\mathbb{R}_+)$  to  $n$ ) we make the process transient by using  $\nu_j^n = \sigma_0$ , with the associated relative entropy cost. With this partitioning of time, the occupation measures and normalized relative entropies will converge to their proper limits. The main difficulty is that without additional conditions  $\bar{p}$  need not be ergodic. To deal with this, a perturbation argument is used to approximate  $\bar{p}$  by an ergodic probability transition function. In fact, we will perturb  $\bar{p}$  in the direction of an ergodic

transition function  $p_0$ , introduced below, and this will suffice. In addition, the measure  $\sigma_0$  is not transient, and it must also be perturbed slightly. Some of the arguments parallel those under more standard assumptions, and therefore in places where the arguments are more or less identical we refer the reader to [7].

We begin by observing that Assumption 2.2 implies that  $\theta(-\infty, 0) > 0$ . From this and Assumption 2.1, it follows that there exists a  $\theta_0 \in \mathcal{P}(\mathbb{R})$  such that (a)  $\int_{\mathbb{R}} |z| \theta_0(dz) < \infty$ , (b)  $\int_{\mathbb{R}} z \theta_0(dz) < 0$ , (c)  $\theta_0$  and  $\theta$  are mutually absolutely continuous and (d)  $R(\theta_0 \parallel \theta) < \infty$ . In fact, one can define  $\theta_0$  by

$$\frac{d\theta_0}{d\theta}(z) = e^{\alpha z} \Big/ \int_{\mathbb{R}} e^{\alpha z} \theta(dz),$$

where  $\alpha$  is chosen so that  $\int_{\mathbb{R}} z \theta_0(dz)$  is negative and in the support of  $\theta$ . For full details on construction of such a  $\theta_0$  we refer the reader to Lemma 6.2.3 and Appendix C.5 of [7]. In rest of the paper we will fix such a  $\theta_0$ . The Markov chain defined via (2.1) with  $\{W_n\}$  having the common law  $\theta_0$  instead of  $\theta$  will be denoted by  $\{X_n^0, n \in \mathbb{N}_0\}$ , and the transition probability function of this Markov chain will be denoted by  $p_0(x, dy)$ . Note that from (c) above, it follows that Assumption 2.2 is satisfied with  $p$  replaced by  $p_0$ . I.e., with  $l_0, n_0$  as in Assumption 2.2 we have that for all  $x_1, x_2 \in \mathbb{R}_+$

$$\sum_{i=l_0}^{\infty} \frac{1}{2^i} p_0^{(i)}(x_1, dy) \ll \sum_{j=n_0}^{\infty} \frac{1}{2^j} p_0^{(j)}(x_2, dy).$$

We begin with an elementary stability result regarding the random walk.

**Lemma 5.2** *Suppose that Assumptions 2.1 and 2.2 hold. For  $x \in \mathbb{R}_+$  and  $n \in \mathbb{N}_0$ , let  $\mu_n^x \in \mathcal{P}(\mathbb{R}_+)$  be defined by*

$$\mu_n^x(A) \doteq \mathbb{P}_x(X_n^0 \in A), \quad A \in \mathcal{B}(\mathbb{R}_+).$$

*Then for all  $C \in (0, \infty)$ , the family  $\{\mu_n^x : n \in \mathbb{N}_0, |x| \leq C\}$  is relatively compact in  $\mathcal{P}(\mathbb{R}_+)$ .*

**Proof:** Using the Lyapunov function  $V(x) \doteq x$ , the lemma follows from a standard result in stochastic stability [9]. ■

**Lemma 5.3** *Suppose that Assumptions 2.1 and 2.2 hold. Then the following conclusions hold.*

(a) *The  $\mathbb{R}_+$ -valued Markov chain  $X_n^0$  has a unique invariant measure  $\mu^*$ . The Markov chain having  $\mu^*$  as its initial distribution and  $p_0(x, dy)$  as the transition*

probability function is ergodic.

(b) Let  $A \in \mathcal{B}(\mathbb{R}_+)$  be such that  $p^{(l_0)}(x_0, A) > 0$  for some  $x_0 \in \mathbb{R}_+$ , where  $l_0$  is as in Assumption 2.2. Then  $\mu^*(A) > 0$ .

(c) If  $\nu \in \mathcal{P}(\mathbb{R}_+)$  is such that  $I_1(\nu) < \infty$ , then  $\nu \ll \mu^*$ .

**Proof:** Tightness proved in Lemma 5.2 and the Feller property of the Markov chain  $\{X_n^0\}$  implies that there is at least one  $p_0(x, dy)$ -invariant probability measure. The proof of the uniqueness of the invariant measure and part (b) is same as that of Lemma 8.6.2 of [7] on noting that  $p^{(l_0)}(x_0, dy)$  and  $p_0^{(l_0)}(x_0, dy)$  are mutually absolutely continuous. Finally we consider (c). From Theorem 4.1 (b)(3), we have that there exists  $\bar{q} \in \mathcal{S}(\mathbb{R}_+ | \mathbb{R})$  such that

$$I_1(\nu) = \int_{\mathbb{R}_+} R(\bar{q}(\cdot|x) \| \theta(\cdot)) \nu(dx).$$

Since  $I_1(\nu) < \infty$ , we have that  $\Gamma \doteq \{x \in \mathbb{R}_+ | \bar{q}(\cdot|x) \ll \theta(\cdot)\}$  satisfies  $\nu(\Gamma) = 1$ . Since  $\theta$  and  $\theta_0$  are mutually absolutely continuous, we can replace  $\theta$  by  $\theta_0$  in the definition of  $\Gamma$ . Let  $\bar{p} \in \mathcal{S}(\mathbb{R}_+ | \mathbb{R}_+)$  be defined via (4.1). Then the above observations imply that, for all  $x \in \Gamma$ ,

$$\bar{p}(x, dy) \ll p_0(x, dy).$$

Now let  $A \in \mathcal{B}(\mathbb{R}_+)$  be such that  $\nu(A) > 0$ . It follows as in the proof of Lemma 8.6.2 of [7], that  $p_0^{(l_0)}(x_0, A) > 0$  for some  $x_0 \in \mathbb{R}_+$ . Combining this with (b) we have that  $\mu^*(A) > 0$ . This proves (c). ■

**Lemma 5.4** *Let  $\pi \in \mathcal{P}(\bar{\mathbb{R}}_+)$  be as in (5.1) and  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$  be given. Let  $\epsilon_0 \in (0, \infty)$  be arbitrary. Then there exists  $\pi^* \in \mathcal{P}(\bar{\mathbb{R}}_+)$ ,  $q^* \in \mathcal{A}_1(\widehat{\pi}^*)$  and  $\sigma^* \in \mathcal{P}_{tr}(\mathbb{R})$  such that the following hold.*

(a)  $\mu^* \ll \widehat{\pi}^*$  and  $\widehat{\pi}^* \ll \mu^*$ .

(b)

$$I_1(\widehat{\pi}^*) \leq \int_{\mathbb{R}_+} R(q^*(\cdot|x) \| \theta(\cdot)) \widehat{\pi}^*(dx) \leq I_1(\hat{\pi}) + \frac{\epsilon_0}{3}.$$

(c)

$$\begin{aligned} \pi^*(\mathbb{R}_+) & \int_{\mathbb{R}_+} R(q^*(\cdot|x) \| \theta(\cdot)) \widehat{\pi}^*(dx) + (1 - \pi^*(\mathbb{R}_+)) R(\sigma^* \| \theta) + F(\pi^*) \\ & \leq I(\pi) + F(\pi) + \epsilon_0. \end{aligned}$$

(d) *The  $\mathbb{R}_+$ -valued Markov chain with transition probability function  $p^*(x, dy)$  defined as*

$$p^*(x, A) \doteq \int_{\mathbb{R}} 1_{\{(x+z)^+ \in A\}} q^*(dz|x), \quad x \in \mathbb{R}_+, \quad A \in \mathcal{B}(\mathbb{R}_+),$$

*is ergodic with  $\widehat{\pi}^*$  its unique invariant measure.*

$$(e) \int_{\mathbb{R}} z \sigma^*(dz) > 0.$$

**Proof:** Since  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$ , there exists  $\delta_1 \in (0, \infty)$  such that  $|F(\pi) - F(\tilde{\pi})| \leq \frac{\epsilon_0}{3}$  whenever  $\tilde{\pi} \in \mathcal{P}(\bar{\mathbb{R}}_+)$  is such that  $\|\tilde{\pi} - \pi\|_v \leq \delta_1$ . Here,  $\|\cdot\|_v$  denotes the total variation norm. Define

$$\delta_0 \doteq \min \left\{ 1, \delta_1, \frac{\epsilon_0 \pi(\mathbb{R}_+)}{3R(\theta_0 \parallel \theta)}, \frac{2\epsilon_0}{3I_1(\tilde{\pi})} \right\}$$

and let  $\pi^* \in \mathcal{P}(\bar{\mathbb{R}}_+)$  be given by the formula

$$\pi^* \doteq \left(1 - \frac{\delta_0}{2}\right) \pi + \frac{\delta_0}{2} \mu^*.$$

Then clearly

$$F(\pi^*) \leq F(\pi) + \frac{\epsilon_0}{3}. \quad (5.3)$$

A straightforward calculation shows that

$$\widehat{\pi}^* = (1 - \alpha)\widehat{\pi} + \alpha\mu^*, \quad (5.4)$$

where

$$\alpha = \frac{\delta_0}{2((1 - \frac{\delta_0}{2})\pi(\mathbb{R}_+) + \frac{\delta_0}{2})}.$$

From (5.4) it is clear that  $\mu^* \ll \widehat{\pi}^*$ . Also, since  $I_1(\widehat{\pi}) < \infty$ , we have from Lemma 5.3 that  $\widehat{\pi} \ll \mu^*$ . Combining this with (5.4) we have that  $\widehat{\pi}^* \ll \mu^*$ . This proves (a). We now consider (b). From Theorem 4.1 (b)(3), we have that there exists  $\bar{q} \in \mathcal{A}_1(\widehat{\pi})$  such that

$$I_1(\widehat{\pi}) = \int_{\mathbb{R}_+} R(\bar{q}(\cdot|x) \parallel \theta(\cdot)) \widehat{\pi}(dx).$$

Define  $\lambda^* \in \mathcal{P}(\mathbb{R}_+ \times \mathbb{R})$  as

$$\lambda^*(A \times B) \doteq (1 - \alpha) \int_A \bar{q}(B|x) \widehat{\pi}(dx) + \alpha \theta_0(B) \mu^*(A), \quad A \in \mathcal{B}(\mathbb{R}_+), B \in \mathcal{B}(\mathbb{R}).$$

We recall the definition of  $\mathcal{A}$  given right after (4.2). Since  $\bar{q} \in \mathcal{A}_1(\widehat{\pi})$  and  $\mu^*$  is  $p_0$ -invariant, we have that  $\lambda^* \in \mathcal{A}(\widehat{\pi}^*)$ . Let  $q^* \in \mathcal{S}(\mathbb{R}_+ | \mathbb{R})$  be such that

$$\lambda^*(A \times B) = \int_A q^*(B|x) \widehat{\pi}^*(dx).$$

Then  $q^* \in \mathcal{A}_1(\widehat{\pi}^*)$ . This implies that

$$\begin{aligned} I_1(\widehat{\pi}^*) &\leq \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \widehat{\pi}^*(dx) \\ &= R(\lambda^* \parallel \widehat{\pi}^* \otimes \theta) \\ &\leq (1 - \alpha) \int_{\mathbb{R}_+} R(\bar{q}(\cdot|x) \parallel \theta(\cdot)) \widehat{\pi}(dx) + \alpha \int_{\mathbb{R}_+} R(\theta_0 \parallel \theta) \mu^*(dx) \\ &= (1 - \alpha) I_1(\widehat{\pi}) + \alpha R(\theta_0 \parallel \theta), \end{aligned} \quad (5.5)$$

where the second inequality above follows from the convexity of the relative entropy function. Next note that our choice of  $\delta_0$  is such that  $\alpha \leq \frac{\epsilon_0}{3R(\theta_0|\theta)}$ . Using this observation in (5.5) it follows that

$$I_1(\widehat{\pi}^*) \leq \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \widehat{\pi}^*(dx) \leq I_1(\widehat{\pi}) + \frac{\epsilon_0}{3}.$$

This proves (b).

Next, let  $\sigma_0 \in \mathcal{P}(\mathbb{R})$  be as in Theorem 4.1. Then  $J = R(\sigma_0|\theta)$ ,  $\int_{\mathbb{R}} |z| \sigma_0(dz) < \infty$  and  $\int_{\mathbb{R}} z \sigma_0(dz) \geq 0$ . Under Assumption 2.2 it must be true that  $\theta((0, \infty)) > 0$ . Define  $\sigma_1 \in \mathcal{P}(\mathbb{R})$  by  $\sigma_1(A) = \theta(A \cap (0, \infty)) / \theta((0, \infty))$ . Let  $c \in (0, \infty)$  be the constant such that  $\frac{d\sigma_1}{d\theta}(z) = c$  for a.e.  $z \in (0, \infty)$ , and define  $\sigma^* \doteq (1-\kappa)\sigma_0 + \kappa\sigma_1$ . Then

$$\begin{aligned} R(\sigma^*|\theta) &\leq (1-\kappa)R(\sigma_0|\theta) + \kappa R(\sigma_1|\theta) \\ &\leq R(\sigma_0|\theta) + \kappa \log c. \end{aligned}$$

Choosing  $\kappa \in (0, \frac{\epsilon_0}{3 \log c})$  we have

$$R(\sigma^*|\theta) \leq R(\sigma_0|\theta) + \frac{\epsilon_0}{3} = J + \frac{\epsilon_0}{3}.$$

Clearly  $\int_{\mathbb{R}} z \sigma^*(dz) > 0$ , and thus part (e) of the lemma holds. Now observe that

$$\begin{aligned} &\pi^*(\mathbb{R}_+) \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \widehat{\pi}^*(dx) + (1 - \pi^*(\mathbb{R}_+)) R(\sigma^*|\theta) \\ &\leq \pi^*(\mathbb{R}_+) \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \widehat{\pi}^*(dx) + (1 - \pi(\mathbb{R}_+)) R(\sigma^*|\theta) \\ &\leq \pi^*(\mathbb{R}_+) I_1(\widehat{\pi}) + (1 - \pi(\mathbb{R}_+)) J + \frac{\epsilon_0}{3} \\ &\leq \pi(\mathbb{R}_+) I_1(\widehat{\pi}) + (1 - \pi(\mathbb{R}_+)) J + \frac{2\epsilon_0}{3} \\ &\leq I(\pi) + \frac{2\epsilon_0}{3}, \end{aligned} \tag{5.6}$$

where the first inequality follows on noting that  $\pi^*(\mathbb{R}_+) > \pi(\mathbb{R}_+)$  and the third inequality is a consequence of the fact that with our choice of  $\delta_0$ ,  $\pi^*(\mathbb{R}_+) - \pi(\mathbb{R}_+) \leq \frac{\epsilon_0}{3I_1(\widehat{\pi})}$ . Now (c) follows on combining (5.3) and (5.6). Since we assumed that  $\pi(\mathbb{R}_+) > 0$ , it follows that  $\pi^*(\mathbb{R}_+) > 0$  and so we have from (c) that

$$\int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \widehat{\pi}^*(dx) \leq \frac{I(\pi) + \epsilon_0}{\pi^*(\mathbb{R}_+)} < \infty.$$

Using this fact and the definition of  $\lambda^*$  it follows as in the proof of Lemma 8.6.3 of [7] that  $q^*(dy|x)$ ,  $\theta$  and  $\theta_0$  are mutually absolutely continuous,  $\widehat{\pi}^*$ -a.e.  $x$ . By modifying  $q^*$  over a  $\widehat{\pi}^*$  null set we have the above mutual absolute

continuity to hold for all  $x \in \mathbb{R}_+$ . From this it then follows that Assumption 2.2 is satisfied with  $p$  replaced by  $p^*$ . This combined with the fact that  $\widehat{\pi^*}$  is  $p^*$ -invariant, gives (d). ■

Now let  $\pi^*, q^*$  and  $\sigma^*$  be as in the previous lemma. We would like to show that there exists a family of controls  $\{\nu_j^n \in \mathcal{S}(\mathbb{R}_+ \mid \mathbb{R}_+ \times \mathcal{M}(\mathbb{R}_+))\}$ ,  $j = 0, \dots, n$ ,  $n \in \mathbb{N}$  such that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \bar{\mathbb{E}}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot \mid \bar{X}_j^n, \bar{L}_j^n) \parallel \theta(\cdot)) + F(\bar{L}^n) \right] \\ & \leq \pi^*(\mathbb{R}_+) \int_{\mathbb{R}_+} R(q^*(\cdot \mid x) \parallel \theta(\cdot)) \widehat{\pi^*}(dx) + (1 - \pi^*(\mathbb{R}_+)) R(\sigma^* \parallel \theta) + F(\pi^*), \end{aligned} \quad (5.7)$$

where  $\{\bar{X}_j^n, \bar{L}_j^n\}$  are defined as in (3.2). Note that  $\pi^*(\mathbb{R}_+) = (1 - \delta_0/2)\pi(\mathbb{R}_+) + \delta_0/2$ . Since  $\pi(\mathbb{R}_+) \in (0, 1)$  we have that  $\pi^*(\mathbb{R}_+) \in (0, 1)$ . Henceforth, we will denote  $\pi^*(\mathbb{R}_+)$  by  $\rho$ . We introduce the following canonical spaces. Let

$$(\Omega_1, \mathcal{F}_1) \doteq ((\mathbb{R}_+)^{\infty}, \mathcal{B}((\mathbb{R}_+)^{\infty})),$$

and

$$(\Omega_2, \mathcal{F}_2, Q) \doteq (\mathbb{R}^{\infty}, \mathcal{B}(\mathbb{R}^{\infty}), (\sigma^*)^{\otimes \infty}).$$

Denote by  $\tilde{\mathbb{P}}_x$  the probability measure on  $(\Omega_1, \mathcal{F}_1)$  under which the canonical sequence

$$\xi_n(\omega_1) \doteq \omega_1(n), \quad \omega_1 \in \Omega_1, \quad n \in \mathbb{N}_0,$$

is a Markov chain with transition probability function  $p^*(x, dy)$  and  $\xi_0 \equiv x$ . From the ergodicity of this Markov chain it follows that there exists a  $\Phi \in \mathcal{B}(\mathbb{R}_+)$  such that  $\widehat{\pi^*}(\Phi) = 1$  (or equivalently, from Lemma 5.4,  $\mu^*(\Phi) = 1$ ) and such that for all  $x \in \Phi$

$$\tilde{L}_n \doteq \frac{1}{n} \sum_{j=0}^n \delta_{\xi_j} \Rightarrow \widehat{\pi^*}, \quad a.s. [\tilde{\mathbb{P}}_x].$$

On the probability space  $(\Omega_2, \mathcal{F}_2, Q)$  define the canonical sequence  $\eta_n$  as

$$\eta_n(\omega_2) \doteq \omega_2(n), \quad \omega_2 \in \Omega_2, \quad n \in \mathbb{N}_0$$

Now define

$$(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}_x) \doteq (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \tilde{\mathbb{P}}_x \otimes Q).$$

We will continue to denote the canonical sequences  $\{\xi_n\}, \{\eta_n\}$  on this extended space by the same symbols. Now define, for  $n \in \mathbb{N}$  and  $j = 0, \dots, n$ , the random variables

$$\bar{X}_j^n \doteq \begin{cases} \xi_j & j = 0, \dots, [n\rho] \\ (\bar{X}_{j-1}^n + \eta_{j-[n\rho]})^+ & j = [n\rho] + 1, \dots, n, \end{cases}$$

where  $[\cdot]$  denotes the greatest integer function. Clearly,  $\{\bar{X}_j^n, j = 0, \dots, n\}$  is a controlled Markov chain, as in (3.2) with

$$\nu_j^n(dz | x, \gamma) = \begin{cases} q^*(dz|x) & j = 0, \dots, [n\rho] - 1 \\ \sigma^*(dz) & j = [n\rho], \dots, n. \end{cases} \quad (5.8)$$

In what follows, we will show that with this choice of  $\{\nu_j^n\}$  and  $\{\bar{X}_j^n\}$ , (5.7) holds. Note that, once this is proved, Theorem 5.1 follows from Lemma 5.4, (5.1), Lemma 3.2, and the fact that  $\epsilon_0$  in Lemma 5.4 and  $\epsilon$  in (5.1) are arbitrary. We begin with the following lemma.

**Lemma 5.5** *Let  $\Phi$  be as above. Then for all  $x \in \Phi$*

$$\bar{L}^n \Rightarrow \pi^*, \quad a.s. [\bar{\mathbb{P}}_x].$$

**Proof:** It suffices to show that for all  $x \in \Phi$  and continuous and bounded  $g \in C_b(\bar{\mathbb{R}}_+)$ ,

$$\bar{L}^n(g) \rightarrow \langle g, \pi^* \rangle, \quad a.e. [\bar{\mathbb{P}}_x].$$

Now fix  $x \in \Phi$  and let  $g \in C_b(\bar{\mathbb{R}}_+)$ . Let  $\epsilon > 0$  be arbitrary, and let  $\Omega_1^x \in \mathcal{F}_1$  be such that  $\bar{\mathbb{P}}_x(\Omega_1^x) = 1$ , and for all  $\omega_1 \in \Omega_1^x$

$$\frac{1}{n} \sum_{j=0}^n g(\xi_j(\omega_1)) \rightarrow \langle g, \widehat{\pi^*} \rangle$$

as  $n \rightarrow \infty$ . Fix such an  $\omega_1 \in \Omega_1$ , and let  $N_0 \in \mathbb{N}$  be such that

$$\left| \frac{1}{n} \sum_{j=0}^{[n\rho]} g(\xi_j(\omega_1)) - \rho \langle g, \widehat{\pi^*} \rangle \right| \leq \epsilon \text{ for all } n \geq N_0. \quad (5.9)$$

Since  $g \in C_b(\bar{\mathbb{R}}_+)$ , there exists  $L \in (0, \infty)$  such that

$$|g(x) - g(\infty)| < \epsilon \text{ for all } x > L. \quad (5.10)$$

Next let  $\Omega_2^* \in \mathcal{F}_2$  be such that  $Q(\Omega_2^*) = 1$  and for all  $\omega_2 \in \Omega_2^*$ , as  $n \rightarrow \infty$ ,

$$\frac{1}{n} \sum_{j=0}^n \eta_j(\omega_2) \rightarrow \int_{\mathbb{R}} z \sigma^*(dz) > 0..$$

Fix such an  $\omega_2 \in \Omega_2^*$  and let  $J_0 \in \mathbb{N}$  be such that

$$\inf_{n \geq J_0} \frac{1}{n} \sum_{j=0}^n \eta_j(\omega_2) \doteq \kappa > 0.$$

Without loss of generality we can assume that  $J_0 \kappa \geq L$ . Now define

$$N_1 \doteq \max \left\{ N_0, \frac{J_0 + 1}{1 - \rho} \right\}.$$

Let  $\omega \doteq (\omega_1, \omega_2)$ . For  $n \geq N_1$

$$\begin{aligned}
\left| \frac{1}{n} \sum_{j=0}^n g(\bar{X}_j^n(\omega)) - \langle g, \pi^* \rangle \right| &= \left| \frac{1}{n} \sum_{j=0}^n g(\bar{X}_j^n(\omega)) - \rho \langle g, \widehat{\pi}^* \rangle - (1-\rho)g(\infty) \right| \\
&\leq \left| \frac{1}{n} \sum_{j=0}^{[n\rho]} g(\bar{X}_j^n(\omega)) - \rho \langle g, \widehat{\pi}^* \rangle \right| \\
&\quad + \left| \frac{1}{n} \sum_{j=[n\rho]+1}^n g(\bar{X}_j^n(\omega)) - (1-\rho)g(\infty) \right| \\
&\leq \epsilon + \frac{1}{n} \sum_{j=[n\rho]+J_0}^n |g(\bar{X}_j^n(\omega)) - g(\infty)| \\
&\quad + |g|_\infty \left[ \frac{J_0}{n} + \frac{n - [n\rho] - J_0 + 1 - n(1-\rho)}{n} \right] \\
&\leq \epsilon + \sup_{[n\rho]+J_0 \leq j \leq n} |g(\bar{X}_j^n(\omega)) - g(\infty)| \\
&\quad + \frac{J_0 + 2}{n}, \tag{5.11}
\end{aligned}$$

where the second inequality follows from (5.9) and the observation that since  $n \geq \frac{J_0+1}{1-\rho}$ , we have that  $n \geq [n\rho] + J_0$ . Next note that for  $n \geq N_1$  and  $j \geq [n\rho] + J_0$ , since

$$\bar{X}_j^n(\omega) \geq \xi_{[n\rho]}(\omega_1) + \sum_{k=1}^{j-[n\rho]} \eta_k(\omega_2),$$

we have

$$\begin{aligned}
\frac{\bar{X}_j^n(\omega)}{j - [n\rho]} &\geq \frac{\xi_{[n\rho]}(\omega_1)}{j - [n\rho]} + \frac{\sum_{k=1}^{j-[n\rho]} \eta_k(\omega_2)}{j - [n\rho]} \\
&\geq \frac{\sum_{k=1}^{j-[n\rho]} \eta_k(\omega_2)}{j - [n\rho]} \\
&\geq \kappa.
\end{aligned}$$

Thus

$$\bar{X}_j^n \geq \kappa(j - [n\rho]) \geq \kappa J_0 \geq L.$$

In view of (5.10), for such  $n$  and  $j$

$$|g(\bar{X}_j^n(\omega)) - g(\infty)| \leq \epsilon.$$

Substituting this in (5.11) we have

$$\left| \frac{1}{n} \sum_{j=0}^n g(\bar{X}_j^n(\omega)) - \langle g, \pi^* \rangle \right| \leq 2\epsilon + \frac{J_0 + 2}{n}.$$



We send  $n \rightarrow \infty$  in the last display. Since  $\epsilon > 0$  is arbitrary,  $\omega = (\omega_1, \omega_2) \in \Omega_1^x \times \Omega_2^*$  is arbitrary, and  $\bar{P}_x(\Omega_1^x \times \Omega_2^*) = 1$ , the result follows. ■

Finally, we prove Theorem 5.1.

**Proof of Theorem 5.1.** From Corollary 1.2.5 of [7], it follows that we can assume without loss of generality that  $F$  is Lipschitz continuous on  $\mathcal{P}(\bar{\mathbb{R}}_+)$  with respect to the Levy-Prohorov metric. By Lemma 3.2

$$\begin{aligned} & \limsup_{n \rightarrow \infty} -\frac{1}{n} \log \bar{E}_x [\exp(-nF(L^n))] \\ &= \limsup_{n \rightarrow \infty} \inf_{\{\nu_j^n\}} \bar{E}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) + F(\bar{L}^n) \right] \\ &\leq \limsup_{n \rightarrow \infty} \bar{E}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) + F(\bar{L}^n) \right], \end{aligned}$$

where  $\{\nu_j^n\}$  are as defined in (5.8). Now suppose that  $x \in \Phi$ , where  $\Phi$  is as in Lemma 5.5. Then from Lemma 5.5,

$$\bar{E}_x[F(\bar{L}^n)] \rightarrow F(\pi^*). \quad (5.12)$$

Also,

$$\begin{aligned} & \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) \\ &= \frac{1}{n} \sum_{j=0}^{[n\rho]-1} R(q^*(\cdot | \xi_j) \| \theta(\cdot)) + \frac{1}{n} \sum_{j=[n\rho]}^{n-1} R(\sigma^* \| \theta) \end{aligned}$$

Thus

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \bar{E}_x \left[ \frac{1}{n} \sum_{j=0}^{n-1} R(\nu_j^n(\cdot | \bar{X}_j^n, \bar{L}_j^n) \| \theta(\cdot)) \right] \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{[n\rho]} \bar{E}_x [R(q^*(\cdot | \xi_j) \| \theta(\cdot))] + (1 - \rho)R(\sigma^* \| \theta). \quad (5.13) \end{aligned}$$

Since

$$\int_{\mathbb{R}_+} R(q^*(\cdot | y) \| \theta(\cdot)) \widehat{\pi}^*(dy) \leq \frac{I(\pi^*)}{\rho} < \infty,$$

we have by the  $L^1$  ergodic theorem and a standard argument (cf. [7], pages 314-315) that there exists  $\Phi_1 \in \mathcal{B}(\mathbb{R}_+)$  such that  $\widehat{\pi}^*(\Phi_1) = 1$  and for all  $x \in \Phi_1$

$$\bar{E}_x \left[ \frac{1}{n} \sum_{j=0}^n R(q^*(\cdot | \xi_j) \| \theta(\cdot)) \right] \rightarrow \int_{\mathbb{R}_+} R(q^*(\cdot | y) \| \theta(\cdot)) \widehat{\pi}^*(dy) \quad (5.14)$$

as  $n \rightarrow \infty$ . Combining (5.12), (5.13), (5.14), for all  $x \in \Phi \cap \Phi_1$

$$\begin{aligned}
\limsup_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{E}_x [\exp(-nF(L^n))] &\leq \rho \int_{\mathbb{R}_+} R(q^*(\cdot|y) \parallel \theta(\cdot)) \widehat{\pi}^*(dy) \\
&\quad + (1 - \rho)R(\sigma^* \parallel \theta) + F(\pi^*) \\
&\leq I(\pi) + F(\pi) + \epsilon_0 \\
&\leq \inf_{\mu \in \mathcal{P}(\overline{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\} + \epsilon + \epsilon_0.
\end{aligned} \tag{5.15}$$

The last two inequalities follow from Lemma 5.4 and (5.1), respectively. Now an argument, as on pages 316-318 of [7], using the Lipschitz property of  $F$  and the transitivity condition (Assumption 2.2) shows that the above inequality, in fact, holds for all  $x \in \mathbb{R}_+$ . Letting  $\epsilon \rightarrow 0$  and  $\epsilon_0 \rightarrow 0$  in (5.15) completes the proof. ■

## 6 Reflected Brownian motion and the M/M/1 queue:

In this section we use the large deviation results for the discrete time Markov chain considered in Sections 3, 4 and 5 to obtain the empirical measure LDP for two very basic continuous time models: reflected Brownian motion and the M/M/1 queue.

We begin with the study of reflected Brownian motion. Let  $\{W(t), t \in [0, \infty)\}$  be a standard Brownian motion on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $b \in \mathbb{R}$  and  $\sigma \in \mathbb{R}_+$  be fixed. Define

$$Y(t) \doteq Y(0) + bt + \sigma W(t), \quad t \in [0, \infty).$$

We will only consider the case when  $Y(0) = x$  for some point  $x \in \mathbb{R}_+$ , in which case expectation will be denoted by  $\mathbb{E}_x$ . For a Polish space  $\mathcal{S}$ , let  $D([0, \infty) : \mathcal{S})$  be the Skorohod space of  $\mathcal{S}$ -valued right continuous functions with left hand limits. We consider this space with the Skorohod topology. Denote by  $\mathcal{D}_+$  the subspace of  $D([0, \infty) : \mathbb{R})$  consisting of functions  $f$  satisfying  $f(0) \in \mathbb{R}_+$ . Reflected Brownian motion  $X(\cdot)$  is defined by the relation

$$X(t) \doteq \Gamma(Y)(t), \quad t \in [0, \infty),$$

where  $\Gamma : \mathcal{D}_+ \mapsto D([0, \infty) : \mathbb{R}_+)$  is the Skorohod map, and is given by

$$\Gamma(z)(t) \doteq z(t) - \left( \inf_{0 \leq s \leq t} z(s) \right) \wedge 0.$$

Elementary calculations show that for  $z, z' \in \mathcal{D}_+$

$$\sup_{0 \leq t < \infty} |\Gamma(z)(t) - \Gamma(z')(t)| \leq 2 \sup_{0 \leq t < \infty} |z(t) - z'(t)|. \quad (6.1)$$

It is well known that  $X(\cdot)$  is a Feller Markov process with values in  $\mathbb{R}_+$ . For  $T \in (0, \infty)$  define the empirical measure  $L^T$  corresponding to this Markov process as the  $\mathcal{P}(\mathbb{R}_+)$ -valued random variable given by

$$L^T(A) \doteq \frac{1}{T} \int_0^T \delta_{X_s}(A) ds, \quad A \in \mathcal{B}(\mathbb{R}_+). \quad (6.2)$$

We now introduce the rate function that will govern the large deviation probabilities of  $\{L^T, T \in (0, \infty)\}$ .

Let

$$\mathcal{H}^+ \doteq \left\{ u \in C_b^2(\mathbb{R}_+) : u'(0) = 0, \inf_{x \in \mathbb{R}_+} u(x) > 0 \right\}.$$

For  $\nu \in \mathcal{P}(\mathbb{R}_+)$ , let

$$I_1(\nu) \doteq - \inf_{u \in \mathcal{H}^+} \int_{\mathbb{R}_+} \left( \frac{Au}{u} \right) (x) \nu(dx), \quad (6.3)$$

where

$$(Au)(x) \doteq \frac{\sigma^2}{2} u''(x) + bu'(x).$$

The rate function  $I(\cdot)$  for the empirical measure LDP for the reflected Brownian motion  $X(\cdot)$  is given as

$$I(\nu) \doteq \nu(\mathbb{R}_+) I_1(\hat{\nu}) + (1 - \nu(\mathbb{R}_+)) \frac{[(b)^-]^2}{2\sigma^2}, \quad \nu \in \mathcal{P}(\bar{\mathbb{R}}_+), \quad (6.4)$$

where for  $x \in \mathbb{R}$ ,  $(x)^- \doteq -\min\{x, 0\}$ .

**Theorem 6.1** *Let  $I(\cdot)$  be defined as in (6.4). Then for all  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$  and  $x \in \mathbb{R}_+$*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{E}_x [\exp(-TF(L^T))] = - \inf_{\mu \in \mathcal{P}(\mathbb{R}_+)} \{I(\mu) + F(\mu)\}. \quad (6.5)$$

Furthermore,  $I(\cdot)$  is a rate function on  $\mathcal{P}(\bar{\mathbb{R}}_+)$ .

For  $h \in (0, \infty)$ , define

$$X^h((n+1)h) \doteq (X^h(nh) + \xi_n^h)^+, \quad n \in \mathbb{N}_0, \quad (6.6)$$

where

$$\xi_n^h \doteq \sigma(W((n+1)h) - W(nh)) + bh, \quad n \in \mathbb{N}_0,$$

and  $X^h(0) \doteq Y(0)$ . Define the occupation measures for the approximating family by the relation

$$L_h^n(A) \doteq \frac{1}{n} \sum_{j=0}^{n-1} \delta_{X^h(jh)}(A), \quad A \in \mathcal{B}(\mathbb{R}_+), \quad n \in \mathbb{N}. \quad (6.7)$$

The key step in the proof of the theorem is the following lemma. The proof of the lemma is reduced to the corresponding proof for unconstrained processes given in Lemma 3.4 of [5] by using the Lipschitz property (6.1) of the Skorohod map.

**Lemma 6.2** *For every  $\epsilon > 0$ .*

$$\limsup_{h \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{nh} \log \mathbb{P}_x \{d(L^{nh}, L_h^n) > \epsilon\} = -\infty,$$

where  $d$  is the Levy-Prohorov metric on  $\mathcal{P}(\bar{\mathbb{R}}_+)$ .

**Sketch of the Proof of Theorem 6.1:** It suffices to prove (6.5) for functions  $F$  which are Lipschitz continuous. Fix such an  $F$  and denote the Lipschitz constant of  $F$  by  $M$ . From Theorem 2.6 we have that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_x [\exp(-nF(L_h^n))] = - \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I_h(\mu) + F(\mu)\},$$

where

$$I_h(\mu) \doteq \mu(\mathbb{R}_+) I_{h,1}(\hat{\mu}) + (1 - \mu(\mathbb{R}_+)) J_h. \quad (6.8)$$

Here

$$I_{h,1}(\nu) \doteq \inf_{q^* \in \mathcal{A}_1(\nu)} \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta_h(\cdot)) \nu(dx), \quad (6.9)$$

$\theta_h$  is a normal distribution with mean  $bh$  and variance  $\sigma^2 h$ , and

$$J_h \doteq \inf_{\gamma \in \mathcal{P}_{\text{tr}}(\mathbb{R})} R(\gamma \parallel \theta_h).$$

Using the property  $u'(0) = 0$  for  $u \in \mathcal{H}^+$ , a minor modification of the proof of Theorem 2.1 of [6] shows that for  $\nu \in \mathcal{P}(\mathbb{R}_+)$

$$I_{h,1}(\nu) = - \inf_{u \in \mathcal{H}^+} \int_{\mathbb{R}_+} \log \left( \left( \frac{T_h u}{u} \right) (x) \right) \nu(dx),$$

where

$$(T_h u)(x) \doteq \int_{\mathbb{R}} u((x+z)^+) \theta_h(dz).$$

Now as in the proof of Lemma 3.1 of [5], it follows that

$$\lim_{h \rightarrow 0} \frac{1}{h} I_{h,1}(\nu) = I_1(\nu), \quad \forall \nu \in \mathcal{P}(\mathbb{R}_+).$$

Also,  $J_h$  can be rewritten as

$$J_h = \inf_{a \geq 0} \inf_{\gamma \in \mathcal{P}(\mathbb{R}): \int_{\mathbb{R}} x\gamma(dx) = a} R(\gamma \parallel \theta_h).$$

From Lemma 6.2.3(f) of [7] it follows that

$$\inf_{\gamma \in \mathcal{P}(\mathbb{R}): \int_{\mathbb{R}} x\gamma(dx) = a} R(\gamma \parallel \theta_h) = \sup_{\alpha \in \mathbb{R}} \left\{ \alpha a - bh\alpha - \frac{\sigma^2 h \alpha^2}{2} \right\} = \frac{(a - bh)^2}{2\sigma^2 h}.$$

Hence

$$J_h = \inf_{a \geq 0} \frac{(a - bh)^2}{2\sigma^2 h} = \frac{((b)^-)^2 h}{2\sigma^2}.$$

It follows that

$$\lim_{h \rightarrow 0} \frac{1}{h} J_h = \frac{((b)^-)^2}{2\sigma^2},$$

and therefore

$$\lim_{h \rightarrow 0} \frac{1}{h} I_h(\nu) = I(\nu), \text{ for all } \nu \in \mathcal{P}(\bar{\mathbb{R}}_+). \quad (6.10)$$

Next, using the boundedness and Lipschitz continuity of  $F$ , we have

$$\frac{1}{T} \log \mathbb{E}_x[\exp(-TF(L^T))] = \frac{1}{h[T/h]} \log \mathbb{E}_x[\exp(-[T/h]hF(L^{[T/h]h}))] + O(h/T),$$

where for  $x \in \mathbb{R}_+$ ,  $[x]$  denotes the integer part of  $x$ . Thus,

$$\begin{aligned} & \liminf_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{E}_x[\exp(-TF(L^T))] \\ &= \liminf_{n \rightarrow \infty} \frac{1}{nh} \log \mathbb{E}_x[\exp(-nhF(L^{nh}))] \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{nh} \log \mathbb{E}_x[1_{d(L^{nh}, L_h^n) \leq \delta} \exp(-nhF(L^{nh}))] \\ &\geq \liminf_{n \rightarrow \infty} \frac{1}{nh} \log \{ \mathbb{E}_x[\exp(-nh(F(L_h^n) - M\delta)) \\ &\quad - \exp(nh(\|F\|_\infty + M\delta)) \mathbb{P}_x(d(L^{nh}, L_h^n) > \delta)] \}. \end{aligned}$$

Taking the limit as  $h \rightarrow 0$  and using Lemma 6.2 gives

$$\begin{aligned} & \liminf_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{E}_x[\exp(-TF(L^T))] \\ &\geq \liminf_{h \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{nh} \log \mathbb{E}_x[\exp(-n(hF(L_h^n) - Mh\delta))] \\ &\geq \liminf_{h \rightarrow 0} -\frac{1}{h} \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I_h(\mu) + hF(\mu)\} - M\delta \\ &= \liminf_{h \rightarrow 0} -\inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \left\{ \frac{1}{h} I_h(\mu) + F(\mu) \right\} - M\delta \\ &= -\inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\} - M\delta, \end{aligned}$$

where the last step follows from the fact that

$$\lim_{h \rightarrow 0} \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \left\{ \frac{1}{h} I_h(\mu) + F(\mu) \right\} = \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\}. \quad (6.11)$$

The equality in (6.11) is a consequence of (6.10) and follows as in [5] (cf. pages 38-41). The lower bound now follows since  $\delta > 0$  is arbitrary.

In much the same way, one shows that

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{E}_x [\exp(-TF(L^T))] \leq - \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\}.$$

This proves the theorem. ■

We now consider the empirical measure LDP for the M/M/1 queue. Strictly speaking, the discrete time approximations to this model do not satisfy the transitivity condition (Assumption 2.2) for all  $x \in \mathbb{N}_+$ . However, if  $X(0)$  is an integer then  $X(t)$  is an integer for all  $t \in [0, \infty)$ , and it is easy to check that the corresponding condition is satisfied for all  $x_1$  and  $x_2$  in  $\mathbb{N}_0$ .

Let  $N_1, N_2$  be independent Poisson processes with constant rates  $\lambda$  and  $\mu$  respectively. Define

$$X(t) \doteq \Gamma(N_1 - N_2)(t), \quad t \in [0, \infty),$$

where  $\Gamma$  is, as before, the Skorohod map. Define the occupation measure  $L^T$  via (6.2). Let  $A$  be the generator of the Markov process  $X(\cdot)$ . For  $f : \mathbb{N}_0 \mapsto \mathbb{R}$ , a bounded function,  $(Af) : \mathbb{N}_0 \mapsto \mathbb{R}$  is the map given as

$$(Af)(x) = \begin{cases} \lambda(f(x+1) - f(x)) + \mu(f(x-1) - f(x)) & \text{if } x \in \mathbb{N} \\ \lambda(f(x+1) - f(x)) & \text{if } x = 0. \end{cases}$$

Define

$$\mathcal{H}^+ \doteq \left\{ u \in BM(\mathbb{N}_0) : \inf_{u \in \mathbb{N}_0} u(x) > 0 \right\}.$$

For  $\nu \in \mathcal{P}(\bar{\mathbb{R}}_+)$  with  $\nu(\mathbb{N}_0) = 1$ , define  $I_1(\nu)$  via (6.3). Now the rate function  $I(\cdot)$  governing the empirical measure LDP for  $X(\cdot)$  is given by

$$I(\nu) = \begin{cases} \nu(\mathbb{R}_+) I_1(\hat{\nu}) + (1 - \nu(\mathbb{R}_+)) ((\sqrt{\lambda} - \sqrt{\mu})^-)^2 & \text{if } \hat{\nu}(\mathbb{N}_0) = 1 \\ \infty & \text{otherwise.} \end{cases} \quad (6.12)$$

The large deviation result for  $\{L^T, T \in (0, \infty)\}$  is now given as follows.

**Theorem 6.3** *Let  $I(\cdot)$  be defined as in (6.12). Then for all  $F \in C_b(\mathcal{P}(\bar{\mathbb{R}}_+))$  and  $x \in \mathbb{N}_0$*

$$\lim_{T \rightarrow \infty} \frac{1}{T} \log \mathbb{E}_x [\exp(-TF(L^T))] = - \inf_{\mu \in \mathcal{P}(\bar{\mathbb{R}}_+)} \{I(\mu) + F(\mu)\}. \quad (6.13)$$

Furthermore,  $I$  is a rate function.

**Sketch of the Proof:** For  $h \in (0, \infty)$ , define  $X^h(nh)$  and  $L_h^n$  as in (6.6) and (6.7) with  $\xi_n^h$  there replaced by

$$\xi_n^h \doteq (N_1 - N_2)((n+1)h) - (N_1 - N_2)(nh).$$

Define  $I_h(\cdot)$  via (6.8), where  $I_{h,1}(\nu)$  is defined via (6.9) when  $\nu(\mathbb{N}_0) = 1$  and is set to be  $\infty$  otherwise; and  $J_h$  is defined as before, with  $\theta_h$  being the probability law of  $N_1(h) - N_2(h)$ . The only difference in the proof from that of Theorem 6.1 is the treatment of the  $J_h$  term. We now briefly comment on the distinction. Using Lemma 6.2.3(f) of [7] again, one finds

$$J_h = \inf_{a \geq 0} G_h(a),$$

where  $G_h$  is the Legendre-Fenchel transform

$$G_h(a) = \sup_{\alpha \in \mathbb{R}} \left\{ \alpha a - \log \left( \int_{\mathbb{R}} \exp(\alpha y) \theta_h(dy) \right) \right\}.$$

The function  $G_h$  is non-negative and vanishes at the mean of  $\theta_h$ , hence  $G_h((\lambda - \mu)h) = 0$ . Thus if  $\lambda \geq \mu$ ,  $J_h = 0$ . In the case when  $\lambda < \mu$ , the convexity of  $G_h$  implies that

$$\inf_{a \geq 0} G_h(a) = G_h(0).$$

A straightforward calculation shows that

$$G_h(a) = (\lambda + \mu)h + h(\alpha^{\max} a - \lambda e^{\alpha^{\max} h} - \mu e^{-\alpha^{\max} h}),$$

where

$$\alpha^{\max} = \frac{1}{h} \log \left( \frac{a + \sqrt{a^2 + 4\mu\lambda h^2}}{2\lambda h} \right).$$

Substituting  $a = 0$  in these expressions, we have

$$G_h(0) = h(\sqrt{\lambda} - \sqrt{\mu})^2,$$

and thus

$$J_h = h((\sqrt{\lambda} - \sqrt{\mu})^-)^2.$$

We conclude that  $\lim_{h \rightarrow 0} \frac{J_h}{h} = ((\sqrt{\lambda} - \sqrt{\mu})^-)^2$ . The rest of the proof is same as the proof of Theorem 6.1, so we omit the details. ■

**Remark 6.4** We expect that the techniques developed in this paper can also be applied to analogous processes on  $\mathbb{R}$  (rather than  $\mathbb{R}_+$ ), which leads to the following conjecture. Let  $X$  be a Brownian motion with drift  $b$  and diffusion  $\sigma$ , and without loss of generality assume  $b \leq 0$ . Let  $L^T$  denote the empirical measure for  $X$ . We consider the two-point compactification of  $\mathbb{R}$ , and denote a probability measure on this space by  $\nu = (\hat{\nu}, a_1, a_2)$ , where  $\hat{\nu}$  is a sub-probability

measure on  $\mathbb{R}$ ,  $a_1$  is the mass  $\nu$  places on  $-\infty$ , and  $a_2$  is the mass placed on  $\infty$ . Our conjecture is that  $\{L^T, T \in (0, \infty)\}$  satisfies a LDP, with the rate function

$$I(\nu) \doteq \inf_{u \in C_b^2(\mathbb{R}) : \inf_{x \in \mathbb{R}} u(x) > 0} \int_{\mathbb{R}} \left( \frac{Au}{u} \right) (x) \hat{\nu}(dx) + a_2 \frac{b^2}{\sigma^2},$$

where  $A$  is the generator of  $X$ .

## 7 Appendix

**Proof of part (c) of Theorem 4.1.** Since  $\mathcal{P}(\bar{\mathbb{R}}_+)$  is compact, the set  $\{\pi \in \mathcal{P}(\bar{\mathbb{R}}_+) : I(\pi) \leq M\}$  is relatively compact for all  $M \in (0, \infty)$ . Thus compactness of this set follows if the rate function is lower semi-continuous. Let  $\pi_n \in \mathcal{P}(\bar{\mathbb{R}}_+)$  be such that  $\pi_n \rightarrow \pi \in \mathcal{P}(\bar{\mathbb{R}}_+)$ . We need to show that

$$I(\pi) \leq \liminf_{n \rightarrow \infty} I(\pi_n).$$

We can assume without loss of generality that

$$\liminf_{n \rightarrow \infty} I(\pi_n) < \infty.$$

Choose a subsequence  $\{n_k\}_{k \in \mathbb{N}}$  such that

$$\sup_k I(\pi_{n_k}) \doteq L < \infty$$

and

$$\lim_{k \rightarrow \infty} I(\pi_{n_k}) = \liminf_{n \rightarrow \infty} I(\pi_n).$$

Henceforth, we will denote the subsequence  $\{\pi_{n_k}\}$  by  $\{\pi_k\}$ . Let  $\pi_k$  have the decomposition given as

$$\pi_k = \rho_k \hat{\pi}_k + (1 - \rho_k) \delta_\infty.$$

From (b), there exists  $q_k \in \mathcal{A}_1(\hat{\pi}_k)$  and  $\sigma \in \mathcal{P}_{\text{tr}}(\mathbb{R})$  such that

$$\begin{aligned} I(\pi_k) &= \rho_k \int_{\mathbb{R}_+} R(q_k(\cdot|x) \parallel \theta(\cdot)) \hat{\pi}_k(dx) + (1 - \rho_k) R(\sigma \parallel \theta) \\ &= R(\tau_k \parallel \pi_k \otimes \theta), \end{aligned} \tag{7.1}$$

where  $\tau_k \in \mathcal{P}(\bar{\mathbb{R}}_+ \times \mathbb{R})$  is defined as

$$\tau_k(A \times B) = \rho_k \int_A q_k(B|x) \hat{\pi}_k(dx) + (1 - \rho_k) \sigma(B) \delta_\infty(A)$$

and the last equality in (7.1) follows from Lemma 1.4.3 (f) of [7]. Since  $\bar{\mathbb{R}}_+$  is compact and  $\sup_k R(\tau_k \parallel \pi_k \otimes \theta) < \infty$ , we have as in the proof of (b) that



$\{\tau_k\}_{k \geq 1}$  is relatively compact in  $\mathcal{P}(\bar{\mathbb{R}}_+ \times \mathbb{R})$ . Now assume without loss of generality that  $\tau_k \rightarrow \tau \in \mathcal{P}(\bar{\mathbb{R}}_+ \times \mathbb{R})$  as  $k \rightarrow \infty$ . Observe that since  $(\tau_k)_1 = \pi_k$ , we have that  $(\tau)_1 = \pi$ . Using the lower semi-continuity of relative entropy we have from (7.1) that

$$L \geq \lim_{k \rightarrow \infty} I(\pi_k) \geq R(\tau \parallel \pi \otimes \theta). \quad (7.2)$$

Now let  $q^* \in \mathcal{S}(\bar{\mathbb{R}}_+ \parallel \mathbb{R})$  be such that  $\tau(dx \, dz) = q^*(dz|x)\pi(dx)$  and let the decomposition for  $\pi$  be given as

$$\pi = \rho \hat{\pi} + (1 - \rho) \delta_\infty.$$

Using Lemma 1.4.3 (f) of [7], once more, we have that

$$R(\tau \parallel \pi \otimes \theta) = \rho \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \hat{\pi}(dx) + (1 - \rho) R(q^*(\cdot|\infty) \parallel \theta(\cdot)).$$

An argument similar to the one in the proof of Lemma 3.10 shows that if  $\rho > 0$  then  $q^* \in \mathcal{A}_1(\hat{\pi})$ . We claim now that if  $\rho < 1$  then  $q^*(\cdot|\infty) \in \mathcal{P}_{\text{tr}}(\mathbb{R})$ . Observe that once the claim is proved, the result follows on noting that

$$\begin{aligned} & \lim_{k \rightarrow \infty} I(\pi_k) \\ & \geq R(\tau \parallel \pi \otimes \theta) \\ & = \rho \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \hat{\pi}(dx) + (1 - \rho) R(q^*(\cdot|\infty) \parallel \theta(\cdot)) \\ & \geq \rho \inf_{q^* \in \mathcal{A}_1(\hat{\pi})} \int_{\mathbb{R}_+} R(q^*(\cdot|x) \parallel \theta(\cdot)) \hat{\pi}(dx) + (1 - \rho) \inf_{\sigma \in \mathcal{P}_{\text{tr}}(\mathbb{R})} R(\sigma \parallel \theta) \\ & = \rho I_1(\hat{\pi}) + (1 - \rho) J \\ & = I(\pi). \end{aligned}$$

We now prove the claim. Using the observation that  $\sup_k R(\tau_k \parallel \pi_k \otimes \theta) \leq L < \infty$  we have as in the proof of Lemma 3.9 (cf. (3.26)) that

$$(1 - \rho) \int_{\mathbb{R}} z q^*(dz|\infty) = \lim_{k \rightarrow \infty} \int z F'_{c_k}(x) \tau(dx \, dz), \quad (7.3)$$

where  $\{c_k\}$  is a sequence of positive reals increasing to  $\infty$  and  $F_c$  is as in (3.12).

Also, an argument parallel to the one in the proof of Lemma 3.7 shows that for  $c \in (0, \infty)$ ,

$$\int_{\mathbb{R}_+ \times \mathbb{R}} z F'_c(x) \hat{\tau}_k(dx \, dz) \geq -3 \int_{\mathbb{R}} |z| 1_{|z| \geq c} (\hat{\tau}_k)_2(dz) - \frac{1}{c} \int_{\mathbb{R}} |z| (\hat{\tau}_k)_2(dz).$$

Hence

$$\int_{\mathbb{R}_+ \times \mathbb{R}} z F'_c(x) \tau_k(dx \, dz) \quad (7.4)$$

$$\begin{aligned}
&= \rho_k \int_{\mathbb{R}_+ \times \mathbb{R}} z F'_c(x) \hat{\tau}_k(dx dz) + (1 - \rho_k) \int_{\mathbb{R}} z \sigma(dz) \\
&\geq -3\rho_k \int_{\mathbb{R}} |z| 1_{|z| \geq c} (\hat{\tau}_k)_2(dz) - \frac{\rho_k}{c} \int_{\mathbb{R}} |z| (\hat{\tau}_k)_2(dz) + (1 - \rho_k) \int_{\mathbb{R}} z \sigma(dz) \\
&\geq -3 \int_{\mathbb{R}} |z| 1_{|z| \geq c} (\tau_k)_2(dz) - \frac{1}{c} \int_{\mathbb{R}} |z| (\tau_k)_2(dz),
\end{aligned} \tag{7.5}$$

where the last step follows on recalling that  $\int_{\mathbb{R}} z \sigma(dz) \geq 0$  for  $\sigma \in \mathcal{P}_{\text{tr}}(\mathbb{R})$ . Now it follows via arguments similar to those in the proof of Lemma 3.8 that there exists a sequence of positive reals  $\{c_m\}$  such that  $c_m \rightarrow \infty$  as  $m \rightarrow \infty$ , and for all  $m \in \mathbb{N}$ ,

$$\begin{aligned}
\int_{\mathbb{R}_+ \times \mathbb{R}} z F'_{c_m}(x) \tau_k(dx dz) &\rightarrow \int_{\mathbb{R}_+ \times \mathbb{R}} z F'_{c_m}(x) \tau(dx dz), \\
\int_{\mathbb{R}} |z| 1_{|z| \geq c_m} (\tau_k)_2(dz) &\rightarrow \int_{\mathbb{R}} |z| 1_{|z| \geq c_m} (\tau)_2(dz)
\end{aligned}$$

and

$$\int_{\mathbb{R}} |z| (\tau_k)_2(dz) \rightarrow \int_{\mathbb{R}} |z| (\tau)_2(dz)$$

as  $k \rightarrow \infty$ . Now taking the limit as  $k \rightarrow \infty$  in (7.5) with  $c$  replaced by  $c_m$ , we have that

$$\int_{\mathbb{R}_+ \times \mathbb{R}} z F'_{c_m}(x) \tau(dx dz) \geq -3 \int_{\mathbb{R}} |z| 1_{|z| \geq c_m} (\tau)_2(dz) - \frac{1}{c_m} \int_{\mathbb{R}} |z| (\tau)_2(dz).$$

Finally taking limit as  $m \rightarrow \infty$ , we have from (7.3) that

$$(1 - \rho) \int_{\mathbb{R}} z q^*(dz | \infty) \geq 0.$$

This proves the claim. ■

## References

- [1] T.S. Chiang. A lower bound of the asymptotic behavior of some markov processes. *Annals of Probability*, 10:955–967, 1982.
- [2] A. de Acosta. Large deviations of empirical measures of markov chains. *J. Theoret. Prob.*, 3:395–431, 1990.
- [3] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Academic Press, San Diego, Calif., 1989.

- [4] J-D. Deuschel and D. Stroock. *Large Deviations*. Academic Press, San Diego, Calif., 1989.
- [5] M. Donsker and S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, I. *Comm. Pure Appl. Math.*, 28:1–45, 1975.
- [6] M. Donsker and S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time, III. *Comm. Pure Appl. Math.*, 29:389–461, 1976.
- [7] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley and Sons, Inc., 1997.
- [8] N. Jain. Large deviation lower bounds for additive functionals of markov processes. *Annals of Probability*, 18:1071–1098, 1990.
- [9] H.J. Kushner. *Stochastic stability and control*. Academic Press, New York-London, 1967.
- [10] P. Ney and E. Nummelin. Markov additive processes, i: eigenvalue properties and limit theorems. *Annals of Probability*, 15:561–592, 1987.
- [11] P. Ney and E. Nummelin. Markov additive processes, i: large deviations. *Annals of Probability*, 15:593–609, 1987.