Most likely paths to error when estimating the mean of a reflected random walk

Ken R. Duffy[a], Sean P. Meyn[b]

[a] *Hamilton Institute, National University of Ireland Maynooth, Ireland*

[b] *Dept. of Electrical and Computer Engineering, and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, U.S.A.*

Abstract

It is known that simulation of the mean position of a Reflected Random Walk (RRW) $\{W_n\}$ exhibits non-standard behavior, even for light-tailed increment distributions with negative drift. The Large Deviation Principle (LDP) holds for deviations below the mean, but for deviations at the usual speed above the mean the rate function is null. This paper takes a deeper look at this phenomenon. Conditional on a large sample mean, a complete sample path LDP analysis is obtained. Let I denote the rate function for the one dimensional increment process. If I is coercive, then given a large simulated mean position, under general conditions our results imply that the most likely asymptotic behavior, ψ^*, of the paths $n^{-1}W_{\lfloor tn \rfloor}$ is to be zero apart from on an interval $[T_0, T_1] \subset [0, 1]$ and to satisfy the functional equation

$$\nabla I \left(\tfrac{d}{dt}\psi^*(t) \right) = \lambda^*(T_1 - t) \quad \text{whenever } \psi(t) \neq 0.$$

If I is non-coercive, a similar, but slightly more involved, result holds.

These results prove, in broad generality, that Monte Carlo estimates of the steady-state mean position of a RRW have a high likelihood of over-estimation. This has serious implications for the performance evaluation of queueing systems by simulation techniques where steady state expected queue-length and waiting time are key performance metrics. The results show that naïve estimates of these quantities from simulation are highly likely to be conservative.

Keywords: reflected random walks, queue-length, waiting time, simulation mean position, large deviations, most likely paths.

Email addresses: `ken.duffy@nuim.ie` (Ken R. Duffy), `meyn@illinois.edu` (Sean P. Meyn)

1. Introduction

Consider $\boldsymbol{W} = \{W_n, n \geq 0\}$, a random walk that starts at zero, is reflected at the origin, and has increments process $\boldsymbol{X} = \{X_n, n \geq 0\}$. The Reflected Random Walk (RRW) is governed by Lindley's recursion [25],

$$W_0 := 0 \text{ and } W_{n+1} = [W_n + X_n]^+ \text{ for } n \geq 1. \tag{1}$$

This recursion plays a fundamental rôle in queueing systems and has long been an important object of study in evaluating their performance. With $\boldsymbol{X}$ being the difference between service times and inter-arrival times of customers, the RRW $\boldsymbol{W}$ describes the evolution of waiting times at a single server first-come first-served queue with infinite waiting space. Lindley's recursion also governs the evolution of the queue-length at certain single server queues, such as the M/M/1 queue [2].

Since the 1980s, large deviation techniques have been brought to bear on the analysis of equation (1) and the distribution of an element of its stationary solution, which exists whenever $\boldsymbol{X}$ is stationary [26]. For example, using a one-dimensional large deviations approach it has been established in broad generality that the stationary distribution possesses logarithmic asymptotics, see [20][12][13][15][23] and references therein. This fact is exploited in the theory effective bandwidths [22] and in the development of on-the-fly estimation schemes from observations of the queueing behavior for the determination of quality of service performance metrics [11][18][21][36][14][30]. Moreover, through the use of functional large deviation techniques, assuming $\boldsymbol{X}$ is i.i.d., the seminal paper [1] proved the significant, broadly applicable result that the most likely path to a large value of the transient RRW is piece-wise linear. This deduction has since been extended (e.g. [37]), including the establishment of results in the stationary regime (e.g. [10][19][16]). All of these papers report piece-wise linear most likely paths to a large value of the RRW or an element of its stationary solution.

The exclusive focus of all of the research cited above is to garner understanding of likelihood of large values of the RRW either in the transient or stationary regime, and the determination of the most likely paths to these large values. In the present article we employ a functional large deviation approach to analyze the estimation of a fundamental quantity for the performance evaluation of a RRW that has so far been overlooked: its mean value. This study reveals substantially richer structure than the study of large values of the RRW, leading to non-convex rate functions and concave most likely paths. Despite this, perhaps surprisingly, general qualitative and quantitative deductions can still be made.

The starting point of the present paper is the following qualitative result: It was observed recently that simulation of the mean position of a RRW,

$$\overline{W}_n := \frac{1}{n}\sum_{i=1}^{n} W_i,$$

exhibits non-standard behavior, even for light-tailed increments with negative drift. For example, if $\boldsymbol{X}$ is i.i.d., then the probability that $\overline{W}_n$ underestimates the long run expected value decays exponentially in n, but the probability of an over-estimate decays sub-exponentially. This is shown in the following proposition, which is taken from [32]. Part (i) follows from Theorem 11.2.3 and part (ii) from Proposition 11.3.4 (see also [31]).

Proposition 1. *Consider the RRW where $\boldsymbol{X}$ is i.i.d. with $\mathsf{E}[X_0] < 0$ and $\mathsf{E}[X_0^2] < \infty$. Then the Markov chain $\boldsymbol{W}$ has a unique invariant probability measure with finite steady-state mean $\overline{W}$, and the simulated averages have the following properties.*

(i) *The lower error-probability decays exponentially: For each $r < \overline{W}$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathsf{P}\big\{\overline{W}_n \le r\big\} < 0.$$

(ii) *The upper error-probability decays sub-exponentially: For each $r > \overline{W}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathsf{P}\big\{\overline{W}_n \ge r\big\} = 0.$$

This paper takes a deeper look at the latter phenomenon, providing a detailed understanding of why it is hard to simulate the mean position of a RRW and, therefore, why care must be taken drawing deductions regarding average queueing performance from the output of a simulation. We establish that, in broad generality, the process $\{n^{-2}\sum_{i=1}^n W_i\}$ satisfies a Large Deviation Principle (LDP) with a non-trivial rate function. As a consequence, the likelihood that the sample-mean estimate of a RRW is an overestimate decays on a slower than exponential scale. The rate function in question is non-convex and this LDP could not, therefore, be established by asymptotic analysis of scaled cumulant generating functions, an approach commonly employed in queueing theory and used in the Gärtner-Ellis method. Unlike the most likely paths of the RRW that lead to a large position which are piece-wise linear [1][17], we ascertain that that the most likely paths associated with a large simulated mean possess more complex features: they are concave, with a possible discontinuity when the path first becomes deviant. A number of examples of these general results are presented to demonstrate the range of qualitative possibilities.

The results contained in this article clearly indicate that significant statistical care must be taken when using estimates from simulation of the mean position of a RRW. This has serious implications for the performance evaluation of queueing systems by simulation techniques where steady state expected queue-length and waiting time are key performance metrics. Our results show, in broad generality, that the most natural estimation scheme, Monte Carlo estimates, of these expected values suffer a likelihood of over-estimation that is, approximately speaking, Weibull-like with shape parameter 1/2. Consequently, naïve estimation of these quantities from simulation is likely to underestimate system performance.

As a concrete illustration of these general results, one example in which the rate function and most likely paths are explicitly computable can be found in the following proposition.

Proposition 2. *Consider the RRW in which the increments process* $\boldsymbol{X}$ *consists of i.i.d Gaussian random variables with mean* $-\delta < 0$ *and variance* σ^2*. Then* $\{n^{-1}\overline{W}_n\}$ *satisfies the LDP in* $[0, \infty)$ *with rate function*

$$I_{\overline{W}}(z) = \begin{cases} \dfrac{4\delta}{\sigma^2}\sqrt{\dfrac{z\delta}{6}} & \text{if } z \in [0, \delta/6], \\ \dfrac{3}{2\sigma^2}\left(z + \dfrac{\delta}{2}\right)^2 & \text{if } z \in [\delta/6, \infty). \end{cases} \tag{2}$$

As n *tends to infinity, the most likely paths of* $n^{-1}W_{\lfloor nt \rfloor}$ *leading to* $\overline{W}_n \geq nz$*, which we denote* ψ^* *are as follows.*

(i) *If* $z \in (0, \delta/6]$*, then for any* $T_0 \in [0, 1 - \sqrt{6z/\delta}]$

$$\psi^*(t) = \begin{cases} 0 & \text{if } t \in [0, T_0] \cup [T_0 + \sqrt{6z/\delta}, 1], \\ \delta(t - T_0) - \delta\sqrt{\dfrac{\delta}{6z}}(t - T_0)^2 & \text{for } t \in [T_0, T_0 + \sqrt{6z/\delta}]. \end{cases}$$

(ii) *If* $z \in [\delta/6, \infty)$*, then*

$$\psi^*(t) = 3\left(z + \frac{\delta}{2}\right)\left(t - \frac{t^2}{2}\right) - \delta t \quad \text{for } t \in [0, 1].$$

We return to this example in Section 4.1 where the rate function in equation (2) and two most likely paths are illustrated in Figure 4.

Proposition 2 and other results that follow concern asymptotics of the doubly scaled sum $n^{-1}\overline{W}_n = n^{-2}\sum_{i=1}^{n} W_i$. The n^2 scaling is similar to [4, Theorem 4.1], concerning asymptotics for the GI/G/1 queue in the light tailed setting. This result proves that the tail of the busy time distribution decays more slowly than exponentially: $\lim_{n\to\infty} n^{-1}\log P\left(B > n^2 z\right) = -K\sqrt{z}$ for each $z > 0$ and some $K > 0$, where B denotes the busy time in steady-state. The form of the limit can be predicted through scaling arguments: If $\{n^{-2}B\}$ satisfies the LDP, it must do so with a rate function of the form $K\sqrt{z}$, as can be seen by considering the substitution $m = n\sqrt{z/y}$. The rate function for the large deviations of the process $\{n^{-1}\overline{W}_n\}$ is necessarily more complex and, in general, the rate function will diverge more rapidly than $\sqrt{z}$ as $z \to \infty$.

The rest of this paper is organized as follows. In Section 2 we prove in broad generality that the sample paths of the rescaled simulated mean $n^{-1}\overline{W}_n$ satisfy a functional LDP. Using this LDP, in Section 3 we characterize properties of the most likely paths of the RRW given that the rescaled simulated mean is large. In Section 4 we present examples including the RRW with i.i.d. Gaussian increments, the M/D/1 queue, the D/M/1 queue and the M/M/1 queue, as these exhibit the full range of theoretically possible behaviors.

2. Functional LDP for the rescaled simulated mean position

We assume that the reader is familiar with the basics of large deviation theory, such as the definition of the LDP and the statement of the Contraction Principle, as can be found in [9][7][8][17]. The notation in this paper is as follows. Let $C[0,1]$ denote the set of continuous $\mathbb{R}$-valued functions on $[0,1]$ equipped with the topology induced by the supremum norm, $\|\phi\| = \sup_{t\in[0,1]} |\phi(t)|$. Let $D[0,1]$ denote the space of $\mathbb{R}$-valued cádlág functions on $[0,1]$ equipped with the Skorohod (J_1) topology [39][3][40] induced by the following metric: for any two functions $\phi, \psi \in D[0,1]$, define

$$d(\phi,\psi) := \inf_{\lambda\in\Lambda} \left\{ \max\left(\|\phi\circ\lambda - \psi\|, \|\lambda - e\| \right) \right\},$$

where e is the identity ($e(t) = t$), and Λ is the set of strictly increasing functions λ from $[0,1]$ to $[0,1]$ that are continuous, with a continuous inverse. Finally, let $L[0,1] \subset D[0,1]$ denote the set of functions that have finite variation. Each $\phi \in L[0,1]$ has a Lebesgue decomposition with respect to Lebesgue measure whose absolutely continuous part we denote $\phi^{(a)}$ and whose singular component we denote $\phi^{(s)}$, so that $\phi(t) = \int_0^t \dot\phi^{(a)}(s)ds + \phi^{(s)}(t)$. Furthermore, we decompose $\phi^{(s)}$ into its positive $\phi^{\uparrow}$ and negative $\phi^{\downarrow}$ parts by the Hahn Decomposition Theorem.

For each $n \in \mathbb{N}$ and all $t \in [0,1]$, we define the following scaled sample paths:

$$x^n(t) := \frac{1}{n}\sum_{i=0}^{\lfloor nt\rfloor-1} X_i, \quad w^n(t) := \frac{1}{n} W_{\lfloor nt\rfloor} \quad\text{and}\quad \bar w^n(t) := \frac{1}{n^2}\sum_{i=1}^{\lfloor nt\rfloor} W_i + \frac{1}{n^2}(nt - \lfloor nt\rfloor])W_{\lfloor nt\rfloor+1}.$$

The first two of these are elements of $D[0,1]$ and correspond to the paths for the simulated position of the unconstrained random walk and for the simulated position of the RRW, respectively. The sample path $\bar w^n$ is an element of $C[0,1]$ and is the polygonally approximated continuous path for the rescaled simulated mean location of the RRW. In particular, note that $\bar w^n(1) = n^{-1}\overline{W}_n = n^{-2}\sum_{i=1}^n W_i$ is the rescaled simulated mean of the RRW.

For the general qualitative theorem we make the following assumption.

Assumption 1. *The sample paths for the unconstrained random walks $\{x^n\}$ satisfy the LDP in D[0,1] with good rate function I_X.*

This assumption is known to hold for a large collection of processes. For example, if $\boldsymbol{X}$ is an i.i.d. sequence, then define $\theta^{\downarrow} := \sup\{\theta > 0 : \mathsf{E}[\exp(-\theta X_0)] < \infty\}$ and $\theta^{\uparrow} := \sup\{\theta > 0 : \mathsf{E}[\exp(\theta X_0)] < \infty\}$. If $\min\{\theta^{\downarrow}, \theta^{\uparrow}\} > 0$, then by Cramér's Theorem [5][7] the partial sums of $\{x^n(1)\}$ satisfy the LDP in $\mathbb{R}$ with the good, convex (local) rate function

$$I(y) := \sup_{\theta} \left(\theta y - \log \mathsf{E}[\exp(\theta X_0)]\right), \text{ for } y \in \mathbb{R}, \tag{3}$$

and Mogul'skii's Theorem [33] proves that Assumption 1 holds true. The rate function is typically of the form (e.g [28][34][37][38]):

$$I_X(\gamma) = \begin{cases} \int_0^1 I(\dot\gamma^{(a)}(s))ds + \theta^{\downarrow}\gamma^{\downarrow}(1) + \theta^{\uparrow}\gamma^{\uparrow}(1) & \text{if } \gamma \in L[0,1], \\ +\infty & \text{otherwise,} \end{cases} \tag{4}$$

where $\theta^{\downarrow}\gamma^{\downarrow}(1) := 0$ if $\theta^{\downarrow} = \infty$ and $\gamma^{\downarrow}(1) = 0$, and $\theta^{\uparrow}\gamma^{\uparrow}(1) := 0$ if $\theta^{\uparrow} = \infty$ and $\gamma^{\uparrow}(1) = 0$. Dembo and Zajic [6] have generalized Mogul'skii's Theorem to include sequences $\boldsymbol{X}$ that need not be i.i.d, but that satisfy a uniform super-exponential tail condition that ensures that the generalization of $\min(\theta^{\downarrow}, \theta^{\uparrow})$ is $+\infty$, as well as a mixing condition that encompasses, for example, Markov chains that are uniformly ergodic. The resulting rate function for these processes is also of the form in equation (4), but the cumulant generating function $\log \mathsf{E}[\exp(\theta X_0)]$ in equation (3) is replaced with the scaled cumulant generating function $\lim n^{-1} \log \mathsf{E}[\exp(\theta(X_0 + \cdots + X_{n-1}))]$.

Theorem 3. *The following hold under Assumption 1:*

(i) *The sequence of rescaled paths of the simulated mean of the RRW $\{\bar w^n\}$ satisfies the LDP in $C[0,1]$ with rate function*

$$I_{\overline{W}}(\phi) = \inf_{\gamma \in L[0,1]} \left\{ I_X(\gamma) : \int_0^t \sup_{s \le t} (\gamma(t) - \gamma(s))\, dt = \phi(t) \ \textit{ for all } \ t \in [0,1] \right\}.$$

(ii) *If I_X is of the form in equation (4), then $I_{\overline{W}}$ is only finite at those functions ϕ such that $\dot\phi$ exists, $\dot\phi$ is non-negative and $\dot\phi$ is an element of $L[0,1]$, in which case*

$$I_{\overline{W}}(\phi) = \int_0^1 \left(I(\ddot\phi^{(a)}(s)) 1_{\{\dot\phi(s) > 0\}} + \inf_{y \le 0} I(y) 1_{\{\dot\phi(s) = 0\}} \right) ds + \theta^{\downarrow}\dot\phi^{\downarrow}(1) + \theta^{\uparrow}\dot\phi^{\uparrow}(1). \tag{5}$$

Proof. The proof of the first assertion follows from the contraction principle (e.g. [7, Theorem 4.2.16]) after noting the following. The Skorohod map, $f(\gamma)(t) = \gamma(t) - \inf_{s \le t} \gamma(s)$, is continuous from $D[0,1]$ to $D[0,1]$ (e.g. [40, Theorem 13.5.1]) and $f(x^n)(t) = w^n(t)$. The integration map, $g(\psi)(t) = \int_0^t \psi(s)\, ds$ is continuous from $D[0,1]$ to $C[0,1]$ (e.g. [40, Theorem 11.5.1]) and $g(w^n)(t) = \bar w^n(t)$.

For the second assertion, if ϕ is such that $\dot\phi$ does not exist, takes negative values or is not an element of $L[0,1]$, then $I_{\overline{W}}(\phi) = +\infty$, as can be seen from the contraction principle. If $\dot\phi$ exists, is non-negative and is an element of $L[0,1]$, then

$$I_{\overline{W}}(\phi) = \inf_{\gamma \in L[0,1]} \left\{ \int_0^1 I(\dot\gamma^{(a)}(s))\, ds + \theta^{\downarrow}\gamma^{\downarrow}(1) + \theta^{\uparrow}\gamma^{\uparrow}(1) : \sup_{s \le t} (\gamma(t) - \gamma(s)) = \dot\phi(t) \text{ for all } t \in [0,1] \right\}.$$

If $\dot{\phi}(t) > 0$, then γ must satisfy $\ddot{\phi}^{(a)}(t) = \dot{\gamma}^{(a)}(t)$. As $\ddot{\phi}^{(a)}(t) = 0$ for almost all t such that $\dot{\phi}(t) = 0$, if $\dot{\phi}(t) = 0$ we are free to choose $\dot{\gamma}^{(a)}(t) = \inf_{y\leq 0} I(y)$ to minimize the rate function. The singular parts $\dot{\phi}^{\downarrow}$ and $\dot{\phi}^{\uparrow}$ must be mimicked by $\gamma^{(s)}$. In order to minimize the rate function, $\gamma^{(s)}$ is unchanging everywhere else, leading to the result. $\square$

The rate function in equation (5) can be understood as follows. In order to see the rescaled simulated mean sample path ϕ, in the integral one must locally pay for changes in the increments process so long as the location is positive. If the location is zero, then the increments can take their most likely value less than or equal to zero. The singular parts of the location are matched by singular parts in the increments.

3. Most likely RRW paths to a large simulated mean position

Considering Theorem 3 in conjunction with the contraction principle and the projection $\phi \mapsto \phi(1)$, roughly speaking, we can deduce that

$$\mathsf{P}\left\{\frac{1}{n^2}\sum_{i=1}^{n} W_i \approx z\right\} \sim \exp\left(-n \inf_{\phi\in C[0,1]} \left\{I_{\overline{W}}(\phi) : \phi(1) = z\right\}\right).$$

Thus consider the following minimization problem:

$$\inf_{\phi\in C[0,1]} \left\{I_{\overline{W}}(\phi) : \phi(1) = z\right\}.$$

If I_X is of the form in equation (4), then this problem can be rewritten in terms of the fluid limit paths of the RRW:

$$\begin{aligned} &\textbf{minimize} && J(\psi) \\ &\textbf{subject to} && \psi \in L^+[0,1] \text{ and } \int_0^1 \psi(s)\,ds = z, \end{aligned} \tag{6}$$

where $L^+[0,1]$ is the set of non-negative elements of $L[0,1]$ and the objective function is

$$J(\psi) := \int_0^1 \left(I(\dot{\psi}^{(a)}(s)) 1_{\{\psi(s)>0\}} + \inf_{y\leq 0} I(y) 1_{\{\psi(s)=0\}} \right) ds + \theta^{\downarrow}\psi^{\downarrow}(1) + \theta^{\uparrow}\psi^{\uparrow}(1). \tag{7}$$

The evaluation of the optimization (6) and the identification of properties of its infimal argument (or arguments) are the subject of the rest of this paper. That is, we wish to identify properties of the most likely fluid simulated RRW paths that give rise to the simulated mean position being unusually large. These optimizers are most likely paths in the sense that if G is any measurable neighborhood of the set of minimizing arguments to (6) and $\overline{G} = \{\phi : \phi(t) = \int_0^t \psi(s)\,ds$ for some $\psi \in G\}$, then $\lim_{n\to\infty} P\{\bar{w}^n \in \overline{G}\} = 1$. This follows, for example, by [24, Theorem 2.2].

In addition to Assumption 1, the following assumption is in force throughout the rest of this article.

Assumption 2. *The rate function I_X is of the form in equation* (4)*, where I is a good, convex rate function, and there exists $\delta > 0$ such that $I(-\delta) = 0$, so that the RRW is stable.*

Note that as I is a rate function, it is lower semi-continuous. The maximal value for which it is finite is denoted by

$$\bar{r} := \sup\{r : I(r) < \infty\}. \tag{8}$$

Suppose that I is a non-coercive function: $\bar{r} < \infty$ and $\lim_{r\uparrow\bar{r}} I(r) < \infty$. Then the limit must coincide with $I(\bar{r})$, which is thus finite. This is needed to ensure the existence of optimal paths. Note also that I being non-coercive is mutually exclusive with $\theta^{\uparrow} < \infty$, which requires $I(r) < \infty$ for all positive r.

Theorem 4. *An optimal solution to the optimization problem* (6) *exists, and any optimal solution ψ^* satisfies the following properties: There exists $0 \leq T_0 < T_1 \leq 1$ such that,*

(i) *$\psi^*(t) > 0$ on the open interval (T_0, T_1) and $\psi^*(t) = 0$ for $t \in [0,1] \setminus [T_0, T_1]$;*

(ii) *ψ^* is concave on $[T_0, T_1]$;*

(iii) *ψ^* is continuous on $(T_0, T_1]$, with a possible jump at $t = T_0$.*

The proofs of this theorem and the two that follow are postponed to the end of this section.

The time T_0 is taken to be the minimal time that a path is non-zero and T_1 the maximal time that it is non-zero:

$$T_0 := \inf\{t \geq 0 : \psi(t) > 0\} \quad \text{and} \quad T_1 := \sup\{t \leq 1 : \psi(t) > 0\}. \tag{9}$$

If $0 < \bar{r} < \infty$, then we define

$$T_0^0 := \sup\{t \geq 0 : \tfrac{d}{dt}^{-}\psi(t) = \bar{r}\}. \tag{10}$$

If the supremum is over an empty set then we take $T_0^0 = T_0$; hence the inclusion $T_0^0 \in [T_0, T_1]$ follows from the definitions. The following theorem identifies the structure of the most likely path for t between T_0^0 and T_1. The one that follows it identifies how most likely paths must end.

Theorem 5. *Let $0 \leq T_0 \leq T_0^0 < T_1 \leq 1$ denote the values given in* (9) *and* (10)*. Then for ψ^* to be an optimal path, there must exist constants $b \in \mathbb{R}$ and $\lambda^* > 0$ such that*

$$\nabla I(\dot{\psi}^*(t)) = b - \lambda^* t \quad \textit{for all } T_0^0 < t < T_1. \tag{11}$$

In particular, if I is coercive, then $T_0^0 = T_0$ and equation (11) *is satisfied for all t such that $\psi^*(t) > 0$.*

Theorem 6. *Suppose that* $T_0^0 < T_1$. *Then* $\frac{d}{dt}^{-}\psi^*(t)|_{t=T_1} = -\delta$ *and* $b = \lambda^* T_1$ *in equation* (11).

As well as providing insight into the structure of the most likely paths, Theorems 4, 5 and 6 enable the reduction of the problem (6) from an infinite dimensional optimization to a finite dimensional optimization problem that can be readily solved numerically, if not analytically.

Proposition 7. *Given* $z > 0$, *define the subset* $S_z \subset L^+[0,1]$ *of* potential solutions *to be the collection of functions* ψ° *such that, for some* $0 \le T_0 < T_1 \le 1$ *and* $T_0^0 \in [T_0, T_1]$:

(i) $\psi^\circ(t) = 0$ *for* $t \in [0, T_0) \cup (T_1, 1]$;

(ii) *if* $T_1 < 1$, *then* $\psi^\circ(T_1) = 0$;

(iii) $\int_{T_0}^{T_1} \psi^\circ(t)dt = z$.

(iv) *if* $T_0^0 < T_1$, *then* $\frac{d}{dt}^{-}\psi^*(t)|_{t=T_1} = -\delta$.

(v) $\nabla I(\dot{\psi}^*(t)) = \lambda^*(T_1 - t)$ *for all* $t \in (T_0^0, T_1)$.

If I *is coercive and* $\theta^\uparrow = \infty$, *then in addition to (i)-(v):*

- $T_0^0 = T_0$;
- ψ° *has no discontinuities.*

If I *is coercive and* $\theta^\uparrow < \infty$, *then in addition to (i)-(v):*

- $T_0^0 = T_0$;
- *if* ψ° *has a discontinuity, it is at* T_0.

If I *is non-coercive (which ensures that* $\theta^\uparrow = \infty$*), then in addition to (i)-(v):*

- $\dot{\psi}^\circ(t) = \bar{r}$ *for* $t \in [T_0, T_0^0)$;
- ψ° *has no discontinuities.*

The problem (6) *is then equivalent to*

$$\begin{aligned} &\textbf{minimize} \quad I(\bar{r})(T_0^0 - T_0) + \int_{T_0^0}^{T_1} I(\dot{\psi}^\circ(s))dt + \psi^{\circ\uparrow}(T_0)\theta^\uparrow \\ &\textbf{subject to} \quad \psi^\circ \in S_z. \end{aligned} \tag{12}$$

After proving Theorems 4, 5 and 6, in Section 4 we will use the reduced representation of the problem defined in equation (12) in the consideration of illustrative examples.

The following lemma will be used to establish properties of an optimal fluid trajectory.

Lemma 8. *Suppose that ψ^0 is a fluid trajectory satisfying $z(0) := \int_0^1 \psi^0(s)\,ds < \infty$.*

(i) *For any $d \geq 0$ and $t \in [0,1]$ define $\psi^d(t) = \max(0, \psi^0(t) - d)$, and $z(d) := \int_0^1 \psi^d(s)\,ds$. Then $z(\,\cdot\,)$ is convex and non-increasing as a function of d.*

(ii) *If $\psi^{\circ\uparrow}(1) = 0$ then $J(\psi^d)$ is non-increasing as a function of d.*

Proof. For each t, the function of d given by $\psi^d(t) = \max(0, \psi^0(t) - d)$ is convex and non-increasing. It follows that its integral over time is also concave and non-increasing.

Part (ii) then follows from the definition of J given in equation (7). $\square$

Proof of Theorem 4.. We first establish the existence of an optimizer, which follows from topological arguments. The objective function $J : D[0,1] \mapsto [0,1]$ is defined for elements of $L^+[0,1]$ in equation (7); set $J(\psi) = +\infty$ for $\psi \notin L^+[0,1]$. The function J is lower semi-continuous and has compact level sets as it is the good rate function for the LDP of the sample path process $\{w^n\}$. With domain $D[0,1]$, the mapping $\psi \mapsto \int_0^1 \psi(s)\,ds$ is continuous (e.g. [40, Theorem 11.5.1]), so that the set $\{\psi \in D[0,1] : \int_0^1 \psi(s)\,ds = z\}$ is closed. In a Hausdorff space, the infimum of a lower semi-continuous function with compact level sets is attained on closed sets (e.g. [17, Lemma 4.1]). Thus if the infimum in (6) is finite, it is attained at some $\psi^* \in L^+[0,1]$ such that $\int_0^1 \psi^*(s)\,ds = z$.

Regarding the properties of an optimizer ψ^*, note first that it is obvious that $\psi^{\downarrow}(1) = 0$: By removing downward jumps we reduce $J(\psi)$, while increasing the area $\int_0^1 \psi(s)\,ds$. On letting ψ^0 denote the new trajectory, and setting $z(d) := \int_0^1 \psi^d(s)\,ds$, Lemma 8 then implies that $z(d) = z$ for some $d \geq 0$, with $J(\psi^d) \leq J(\psi)$.

We assume without loss of generality that the closure of $\{t : \psi(t) > 0\}$ is equal to the interval $[T_0, T_1]$ (where the endpoints are defined in (9)): If there exist times $t_0 < t_1$ satisfying $t_0 \geq T_0$, $t_1 \leq T_1$, and $\psi(t) = 0$ for $t \in (t_0, t_1)$, then the trajectory can be shifted as follows,

$$\psi^0(t) = \begin{cases} \psi(t) & t \in [T_0, t_0] \\ \psi(t + (t_1 - t_0)) & t \in [t_0, 1 - (t_1 - t_0)] \\ \max(0, \psi(1) + (1 - (t_1 - t_0) - t)\delta) & t \in (1 - (t_1 - t_0), 1). \end{cases}$$

Once again, on setting $z(d) := \int_0^1 \psi^d(s)\,ds$, we have $z(0) \geq z$, and on applying Lemma 8 we have $z(d) = z$ for some $d \geq 0$, with $J(\psi^d) \leq J(\psi)$.

Next, we assume without loss of generality that $T_0 = 0$: We can replace ψ by,

$$\psi^0(t) = \begin{cases} \psi(t - T_0) & t \in [0, 1 - T_0] \\ \max(0, \psi(1 - T_0) + (1 - T_0 - t)\delta) & t \in (1 - T_0, 1). \end{cases}$$

An application of Lemma 8 again shows that $J(\psi^d) \leq J(\psi^0) = J(\psi)$, and $z(d) = z = \int_0^1 \psi(t)\, dt$ for some $d \geq 0$.

We can now prove (iii): Figure 1 illustrates why a jump following time T_0 cannot be optimal. A formal proof can be performed through construction as in the previous steps. We define, for any ψ, the new trajectory with $\psi^0(0) = \psi(0)$, and

$$\psi^0(t) = \psi^{\uparrow}(1) + \psi(t), \qquad 0 < t \leq 1.$$

We have $J(\psi^0) = J(\psi)$, and $\int_0^1 \psi^0(s)\, ds \geq \int_0^1 \psi(s)\, ds$. Applying Lemma 8 we have $z(d) = z$ for some $d \geq 0$, with $J(\psi^d) \leq J(\psi)$. This proves (iii).

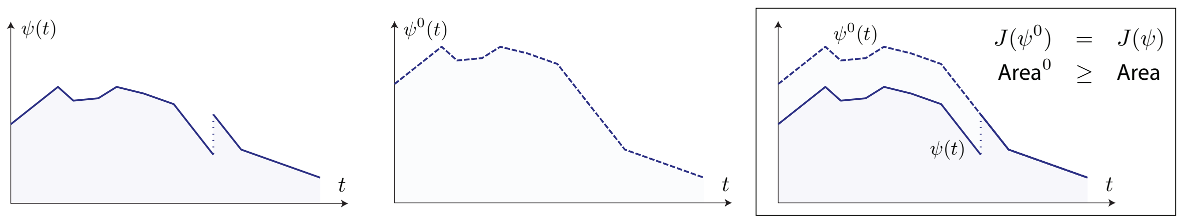


Figure 1: A jump for $t > T_0$ cannot be optimal. The rate functional evaluated at the two paths ψ and ψ^0 is the same, yet the area is greater using ψ^0.

Similar reasoning establishes concavity of an optimal path — Figure 2 shows a transformation of a given trajectory to form a new trajectory with reduced value $J(\psi^0)$, but strictly greater area. Applying Lemma 8 we obtain (ii). Part (i) then follows from (ii). $\square$

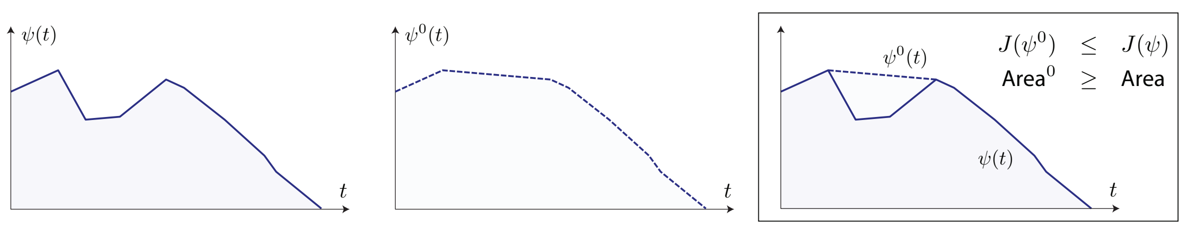


Figure 2: An optimal path is concave on (T_0, T_1). Convexity of $I(r)$ implies that $J(\psi^0) \leq J(\psi)$, yet the area is greater using ψ^0.

Proof of Theorem 5.. From Theorem 4, if ψ^* is an optimal solution for (6) then it is continuous apart from at T_0; if $T_1 < 1$, then $\psi^*(T_1) = 0$. Thus (6) can be considered as identifying

$$\inf_{\psi(T_0)\geq 0} \inf_{T_1\leq 1} \inf_{T_0^0\leq z/\bar{r}} \Bigg\{ (T_0^0 - T_0) I(\bar{r}) + \int_{T_0^0}^{T_1} I(\dot{\psi}(t))dt + \theta^{\uparrow}\psi(T_0) \tag{13}$$

$$: \; (T_0^0 - T_0)^2 \frac{\bar{r}}{2} + (T_1 - T_0)\psi(T_0) + \int_{T_0^0}^{T_1} \psi(t)dt = z, \, \psi(T_1) = 0 \text{ if } T_1 < 1 \Bigg\},$$

For fixed T_0, $\psi(T_0)$, T_1 and T_0^0, we are left to consider finding the solution of a problem of the following kind:

$$\begin{aligned} \textbf{minimize} \quad & \int_0^T I(\dot{\psi}(t))\, dt \\ \textbf{subject to} \quad & \int_0^T \psi(s)\, ds = z'. \end{aligned}$$

If ψ is feasible path, then integration by parts gives

$$\int_0^T t\dot{\psi}(t)\, dt = T\psi(T) - z'. \tag{14}$$

Introduce the Lagrangian

$$\mathcal{L}(\psi, \lambda) = \int_0^T I(\dot{\psi}(t))\, dt + \lambda \Big(\int_0^T t\dot{\psi}(t)\, dt + z' - T\psi(T) \Big).$$

There exists $\lambda = \lambda^*$ so that complementary slackness holds. Hence the optimizer ψ^* of (6) also minimizes $\mathcal{L}(\psi, \lambda^*)$ over all ψ. The constant λ^* exists by [27, (Theorem 1 of Section 8.3)], which only requires feasibility of (14) for z'' in a neighborhood of z' (which is true when $T_0^0 < T_1$. If $T_0^0 = T_1$ then there is nothing to prove).

If ψ^* minimizes the Lagrangian, and if δ represents a perturbation satisfying $\delta(t) = 0$ for $t \in (T_0^0, T_1)^c$, then,

$$0 = \frac{d}{d\theta} \mathcal{L}(\psi^* + \theta\delta, \lambda^*) \Big|_{\theta=0} = \int_0^T [\nabla I(\dot{\psi}(t)) + \lambda^* t] \dot{\delta}(t)\, dt$$

It follows that there exists a constant b such that

$$\nabla I(\dot{\psi}(t)) = b - \lambda^* t \quad \text{ for } a.e. \; t \in (T_0^0, T_1).$$

Returning to problem (13), this implies that irrespective of the optimal values of T_0, $\psi(T_0)$, T_0^0 or T_1, the optimal path satisfies $\nabla I(\dot{\psi}^*(t)) = b - \lambda^* t$ for $t \in (T_0^0, T_1)$. □

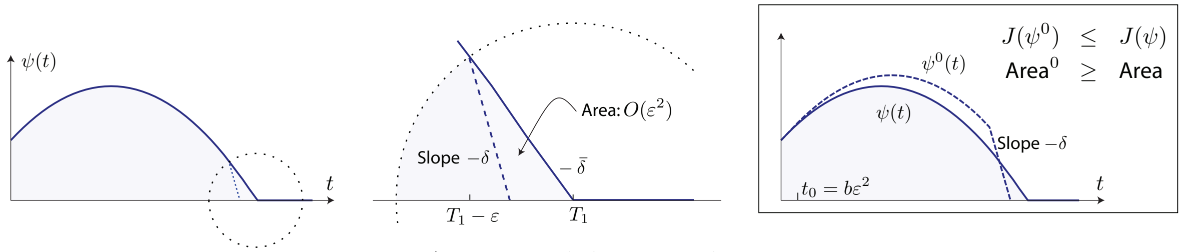


Figure 3: An unnatural slope is costly near $t = T_1$.

Proof of Theorem 6.. Figure 3 illustrates the idea of the proof in the special case $T_1 < 1$. Let $\bar{\delta} = -\frac{d}{dt}^{-}\psi(t)|_{t=T_1}$, and suppose that $\bar{\delta} < \delta$. We will construct a new trajectory ψ^0 with a lower value of $J(\psi^0)$, and increased area. An application of Lemma 8 will then show that ψ cannot be optimal.

We henceforth assume without loss of generality that $T_0 = 0$. Concavity of ψ on $[0, T_1]$ implies that $\frac{d}{dt}\psi(t) \ge -\bar{\delta}$ for all $t \in (T_0, T_1)$. The main idea of the proof (as illustrated in the figure) is as follows: For given $\varepsilon > 0$, the cost contribution to $J(\psi)$ over $[T_1 - \varepsilon, T_1]$ is greater than $\varepsilon I(-\bar{\delta}) = O(\varepsilon)$ if $\bar{\delta} < \delta$. However, the additional area obtained is bounded by $O(\varepsilon^2)$.

We let $-\delta^0$ denote the right derivative, following a possible jump:

$$\delta^0 = -\lim_{t\downarrow 0} \frac{d}{dt}\psi(t)$$

For fixed $\varepsilon > 0$, $b > 0$, let $t_0 = b\varepsilon^2$, and let ψ^0 denote the concave function defined by $\psi^0(0+) = \psi(0+)$, with derivatives for $t > 0$ defined as follows:

$$\frac{d}{dt}\psi^0(t) = \begin{cases} -\delta^0 & t \in (0, t_0] \\ \frac{d}{dt}\psi(t - t_0) & t \in (t_0, T_1 + t_0 - \varepsilon] \\ -\delta & t > T_1 + t_0 - \varepsilon \quad \text{(provided } \psi^0(t) > 0). \end{cases} \tag{15}$$

For b sufficiently large, we have $\int_0^1 \psi^0(s)\, ds \ge \int_0^1 \psi(s)\, ds$ for all $\varepsilon > 0$ sufficiently small. For the same constant b we also have $J(\psi^0) \le J(\psi) - O(\varepsilon) + O(\varepsilon^2)$, so that $J(\psi^0) \le J(\psi)$ for sufficiently small $\varepsilon > 0$. Fixing b and ε so that these bounds hold, we then apply Lemma 8 to conclude that $z(d) = z$ for some $d \ge 0$, with $J(\psi^d) \le J(\psi^0) \le J(\psi)$. The second statement of the theorem follows from $\frac{d}{dt}^{-}\psi^*(t)|_{t=T_1} = -\delta$ and Theorem 5 on noting that $\nabla I(-\delta) = 0$.

$\square$

4. Examples

4.1. Coercive rate function: continuous paths

We present two examples with coercive rate functions. One is the RRW with Gaussian increments. Here the rate function and most likely path can be determined explicitly. The

RRW in the second example corresponds to the queue-length at departures of an M/D/1 queue with batch services. In this case, identification of the rate function and the most likely paths requires the solution of two transcendental equations, which can be readily obtained numerically.

Gaussian increments. Let $\boldsymbol{X}$ be i.i.d. Gaussian, with X_0 having mean $-\delta < 0$ and variance σ^2. Then the conditions of Mogul'skii's Theorem are met with $\theta^{\downarrow} = \theta^{\uparrow} = +\infty$ and the local rate function is

$$I(x) = \frac{1}{2\sigma^2}(x+\delta)^2.$$

As $\theta^{\downarrow} = \theta^{\uparrow} = +\infty$ the sample path rate function $I_{\boldsymbol{X}}$ is only finite at absolutely continuous functions.

Without loss of generality, assume that $T_0 = 0$ and define $T = T_1$ (so that T represents $T_1 - T_0$). By Proposition 7, to solve the problem (12) for a given $z > 0$, for each $T \in (0,1]$ we first identify paths ψ° that satisfy $\nabla I(\dot{\psi}^\circ(t)) = \lambda^*(T-t)$ in $[0,T]$ and $\dot{\psi}^\circ(T) = -\delta$, which leads to candidate solutions satisfying

$$\dot{\psi}^\circ(t) = \sigma^2\lambda^*(T-t) - \delta.$$

If $T < 1$, then in addition we have that $\psi^\circ(T) = 0$ and $\int_0^T \psi^\circ(t)\,dt = z$ giving $\sigma^2\lambda^* = 2\delta/T$ and $T = \sqrt{6z/\delta}$. Note that $T < 1$ only if $6z < \delta$ and therefore the optimal path is

$$\psi^*(t) = \delta t - \delta\sqrt{\frac{\delta}{6z}}\frac{t^2}{2} \text{ for } t \in [0,T] \text{ if } z < \delta/6. \tag{16}$$

If $T = 1$, then we have that $c := \psi^\circ(1) \geq 0$ and $\int_0^1 \psi^\circ(t)\,dt = z$, giving $\sigma^2\lambda^* = 2(c+\delta)$ and $c = 3/2(z - \delta/6)$. Note that $c \geq 0$ only if $6z \geq \delta$ and therefore the optimal path is

$$\psi^*(t) = 3\left(z + \frac{\delta}{2}\right)\left(t - \frac{t^2}{2}\right) - \delta t \text{ for } t \in [0,1] \text{ if } z \geq \delta/6. \tag{17}$$

Evaluating $\int_0^T I(\dot{\psi}^*(t))\,dt$, for $\psi^*(t)$ defined in equations (16) and (17), we obtain the rate function $I_{\overline{W}}(z)$ presented in equation (2), Proposition 2, which can be found in Section 1.

With $\delta = 1$ and $\sigma^2 = 1$, the rate function defined in equation (2) is plotted on the left in Figure 4. It is concave for $z \leq 1/6$ and then convex for $z \geq 1/6$. Thus it is not possible that this LDP could be proved, or its rate function identified, by Gärtner-Ellis style methods that rely on convexity. The transition at $z = 1/6$ occurs when the most likely paths change from returning to 0 within the interval to paths that end with a non-zero position. This explains the dramatic change in shape of the rate function at that point.

Two most likely fluid RRW paths given that $\overline{W}_n \approx nz$ for are shown on the right in Figure 4. The higher path has $z = 1/3$, while the lower path has $z = 1/7$. Note that, for the lower

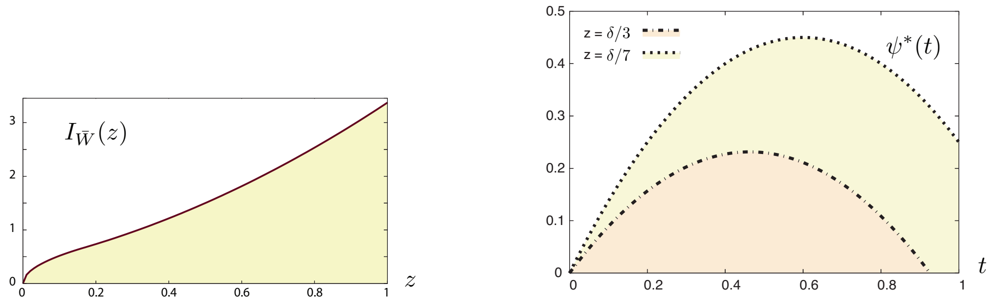


Figure 4: Coercive rate function example: i.i.d. Gaussian increments. On the left hand side is $I_{\overline{W}}(z)$ versus z. Shown on the right hand side are most likely RRW paths, $\psi^*(t) \approx n^{-1} W_{\lfloor nt \rfloor}$, given that $\overline{W}_n \approx nz$.

path, all paths of this shape that start in the interval $[0, 1 - \sqrt{6z}]$ are most likely paths to the deviation. That is, there is no single most likely path, just a most likely shape that can occur anywhere within in $[0, 1]$.

M/D/1 queue-lengths. A second coercive example, albeit one that requires numerics for its ultimate solution, is when $\boldsymbol{X}$ is i.i.d., with X_0 having Poisson distribution with rate α and mean $-\delta = \alpha - \mu$ ($\mu \in \mathbb{N}$), $P\{X_0 = k\} = e^{-\alpha}\alpha^{k+\mu}/(k+\mu)!$ for $k = -\mu, -\mu+1, \ldots$. In this setting, the RRW defined in equation (1) corresponds to the queue-length of an M/D/1 queue with batch μ services, where the queue-length is observed at customer departures. That is, at each service the minimum of the current queue-length and μ customers are processed in a single batch. Between services, a Poisson(α) number of customers arrive to the queue.

The conditions of Mogul'skii's Theorem are met with $\theta^{\downarrow} = \theta^{\uparrow} = +\infty$ and the local rate function is

$$I(x) = \alpha - (x + \mu) + (x + \mu) \log \left(\frac{x + \mu}{\alpha} \right) \text{ if } x > -\mu$$

and $I(x) = \infty$ if $x \le -\mu$. The sample path rate function $I_{\boldsymbol{X}}$ is only finite at absolutely continuous functions.

Again, without loss of generality assume that $T_0 = 0$ and define $T = T_1$. For a given $z > 0$, for each $T \in (0, 1]$ we first identify paths ψ° that satisfy $\nabla I(\dot{\psi}^\circ(t)) = \lambda^*(T - t)$ in $[0, T]$ and $\dot{\psi}^\circ(T) = -\delta$, which leads to candidate optimal paths of the following form:

$$\dot{\psi}^\circ(t) = \alpha e^{-\lambda^*(t-T)} - \mu.$$

Integrating, we have that

$$\psi^\circ(t) = \frac{\alpha}{\lambda^*} e^{\lambda^* T} \left(1 - e^{-\lambda^* t} \right) - \mu t.$$

If $T < 1$, using the constraint $\psi^\circ(T) = 0$ gives the following equation for λ^* in terms of T

$$\frac{\alpha}{\lambda^*} \left(e^{\lambda^* T} - 1 \right) - \mu T = 0.$$

If $T = 1$ and $c := \psi^\circ(1)$ we have the following equation for λ^* in terms of c:

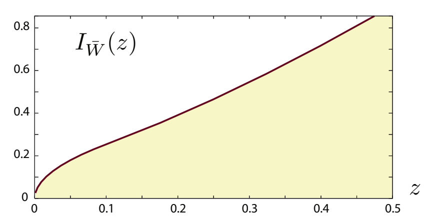
$$\frac{\alpha}{\lambda^*}\left(e^{\lambda^*} - 1\right) - \mu - c = 0.$$

Both of these are transcendental equations, but can be readily solved numerically for λ^*. Once λ^* is known, the constraint

$$z = \int_0^T \psi^\circ(t)\, dt = \frac{\alpha}{\lambda^*} e^{\lambda^* T}\left(T - \frac{1}{\lambda^*}\right) + \frac{\alpha}{(\lambda^*)^2} - \frac{\mu T^2}{2},$$

gives a transcendental equation for T or c ($= \psi^\circ(1)$) in terms of z and identifies the solution ψ^* which determines the rate function

$$I_{\overline{W}}(z) = \int_0^T I(\dot{\psi}^*(t))\, dt = (\alpha - \mu)T - \psi(T) + \frac{\alpha}{\lambda^*}\left(e^{\lambda^* T}(\lambda^* T - 1) + 1\right).$$



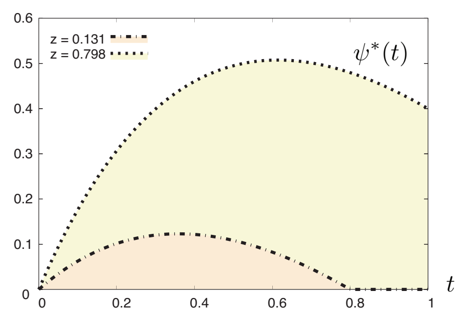


Figure 5: Coercive rate function example: M/D/1 queue-lengths, Poisson increments. On the left hand side is $I_{\overline{W}}(z)$ versus z. Shown on the right hand side are most likely RRW paths, $\phi^*(t) \approx n^{-1}W_{\lfloor nt \rfloor}$, given that $\overline{W}_n \approx nz$.

With $\alpha = 0.5$ and $\mu = 1.0$, the numerically calculated rate function is plotted on the left in Figure 5. The transition from concave to convex again occurs when the most likely path has both $T = 1$ and $c := \psi^\circ(1) = 0$. Two example most likely paths are plotted on the right hand side of Figure 5, which display similar features to the most likely paths as for the Gaussian increments RRW.

4.2. Coercive rate function: paths with jumps

D/M/1 waiting times. Let $\boldsymbol{X}$ be i.i.d. with $P(X_0 \geq x) = \exp(-\alpha(x + \mu^{-1}))$ for $x \geq -\mu^{-1}$ giving $\mathbb{E}(X_0) = \alpha^{-1} - \mu^{-1} := -\delta$. We assume that $\mu < \alpha$ so that $\delta > 0$. Then the RRW defined in equation (1) corresponds to waiting times at a stable D/M/1 queue, where customers arrive at regular intervals of length μ^{-1} and experience i.i.d. exponentially distributed service times with mean α^{-1}. Cramér's Theorem holds for $\{x^n(1)\}$ with rate function

$$I(x) = \begin{cases} \alpha\left(x+\frac{1}{\mu}\right) - \log\left(\alpha\left(x+\frac{1}{\mu}\right)\right) - 1 & \text{if } x \in (-\mu^{-1}, \infty), \\ \infty & \text{otherwise,} \end{cases}$$

which is coercive, so that $T_0^0 = T_0$ for the optimal path. However, $\theta^{\uparrow} = \alpha$ so that the possibility of an initial jump in the most likely path cannot be discounted.

Again, without loss of generality, assume that $T_0 = 0$ and define $T = T_1$. Using $\nabla I(\dot{\psi}^{\circ}(t)) = \lambda^*(T-t)$ and the constraint $\dot{\psi}^{\circ}(T) = -\delta = 1/\alpha - 1/\mu$, candidate solutions must satisfy

$$\dot{\psi}^{\circ}(t) = \frac{1}{\alpha + \lambda^*(t-T)} - \frac{1}{\mu} \text{ for all } t \in (0, T)$$

and hence, for some initial jump $\psi^{\circ}(0) = a \geq 0$,

$$\psi^{\circ}(t) = a + \frac{1}{\lambda^*}\log\left(\frac{\alpha+\lambda^*(t-T)}{\alpha - \lambda^* T}\right) - \frac{t}{\mu} \text{ for } t \in [0, T].$$

If $T < 1$, then in addition we have that $\psi^{\circ}(T) = 0$, which gives the following equation:

$$\alpha - \lambda^* T - \alpha \exp\left(\lambda^*\left(a - \frac{T}{\mu}\right)\right) = 0.$$

If $T = 1$, then we have that $c := \psi^{\circ}(T) \geq 0$, which implies that

$$\alpha - \lambda^* - \alpha \exp\left(\lambda^*\left(a - c - \frac{1}{\mu}\right)\right) = 0.$$

Treating a $(= \psi^{\circ}(0))$ as given, these two transcendental equations can be readily solved for λ^*. Finally we have the constraint that $\int_0^T \psi^{\circ}(t) = z$, so that

$$z = T\left(a - \frac{1}{\lambda^*}\right) + \frac{\alpha}{(\lambda^*)^2}\log\left(\frac{\alpha}{\alpha - \lambda^* T}\right) - \frac{T^2}{2\mu}$$

and

$$a\alpha + \int_0^T I(\dot{\psi}^{\circ}(t))\,dt = \alpha\left(a + \psi^{\circ}(T) + \frac{T}{\mu}\right) - 2T + \left(\frac{\alpha - \lambda^* T}{\lambda^*}\right)\log\left(\frac{\alpha}{\alpha - \lambda^* T}\right).$$

For given $a \geq 0$, having solved the transcendental equations, the most likely path and its associated rate can be calculated. Optimization over a can then be performed numerically.

For example, with $\alpha = 2$ and $\mu = 1$, the rate function is plotted on the left hand side of Figure 6. It looks similar to the earlier examples, but is asymptotically linear with slope

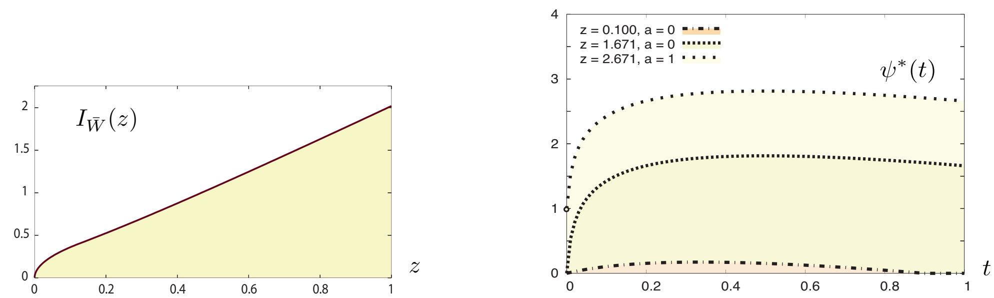


Figure 6: Coercive rate function with jumps example: D/M/1 waiting times, i.i.d. Exponential increments. On the left hand side is $I_{\overline{W}}(z)$ versus z. Shown on the right hand side are most likely RRW paths, $\phi^*(t) \approx n^{-1}W_{\lfloor nt \rfloor}$, given that $\overline{W}_n \approx nz$.

α. The reason for this is best explained by considering the most likely path shown on the right hand side of Figure 6. For small z, $\psi^\circ(0) = a = 0$ and no jump occurs at the start of the most likely path. However, once z is sufficiently large (approximately 1.67 for these parameters), the most likely path has a jump at 0 followed by a vertically shifted version of the largest-area most likely path that doesn't have a jump, as illustrated in Figure 6. The increase in the rate function for the shift of height $\psi^\circ(0) = a$ (gaining area a over the interval $[0, 1]$), is $a\alpha$, which is why the rate function is ultimately linear with slope α.

4.3. Non-coercive rate function: rate-constrained paths

M/M/1 queue-length. Let $\boldsymbol{X}$ be a Bernoulli sequence taking values -1 and $+1$ with $\alpha = P\{X_0 = +1\} < P\{X_0 = -1\} = 1 - \alpha$. The RRW in equation (1) corresponds to the queue-length of an M/M/1 queue observed at arrivals and departures. We have $\theta^\uparrow = \theta^\downarrow = +\infty$. The increments rate function I is infinite outside $[-1, 1]$ and is non-coercive with $\bar{r} = 1$:

$$I(x) = \frac{1+x}{2}\log\left(\frac{1+x}{2\alpha}\right) + \frac{1-x}{2}\log\left(\frac{1-x}{2(1-\alpha)}\right) \text{ and } \bar{r} = 1.$$

Note that $I_{\overline{W}}(z) = +\infty$ if $z > 1/2$; if $z = 1/2$, then $T_0^0 = T = 1$ and $\psi^*(t) = t$ so that $I_{\overline{W}}(z) = -\log(\alpha)$. Without loss of generality, let $T_0 = 0$ and define $T = T_1$. Assume that $T_0^0 < T$. The equation $\nabla I(\dot{\psi}^\circ(t)) = \lambda^*(T - t)$ with the boundary condition $\dot{\psi}^\circ(T) = -\delta = 2\alpha - 1$ gives

$$\dot{\psi}^\circ(t) = \begin{cases} 1 & \text{if } t \in [0, T_0^0] \\ \dfrac{\alpha\exp(2\lambda^*(T-t)) - (1-\alpha)}{\alpha\exp(2\lambda^*(T-t)) + 1 - \alpha} & \text{if } t \in (T_0^0, T]. \end{cases}$$

By integrating, we are looking at proposed solutions

$$\psi^\circ(t) = \begin{cases} t & \text{if } t \in [0, T_0^0] \\ 2T_0^0 - t + \dfrac{1}{\lambda^*} \log\left(\dfrac{\alpha e^{2\lambda^*(T-T_0^0)} + 1 - \alpha}{\alpha e^{2\lambda^*(T-t)} + 1 - \alpha}\right) & \text{if } t \in (T_0^0, T]. \end{cases}$$

If $T < 1$, then $\psi^\circ(T) = 0$ gives

$$\alpha e^{2\lambda^*(T-T_0^0)} - e^{\lambda^*(T-2T_0^0)} + 1 - \alpha = 0$$

and, in particular, if $T_0^0 = 0$, then $\lambda^* = \log((1-\alpha)/\alpha)/T$. While if $T = 1$, then $c{:=}\psi^\circ(T) \geq 0$ ($c < 1$) gives the following equation for λ^*

$$\alpha e^{2\lambda^*(1-T_0^0)} - e^{\lambda^*(c+1-2T_0^0)} + 1 - \alpha = 0. \tag{18}$$

Note that this equation only has a positive solution for $c \in (\max\{0, 2\alpha - 1 + 2(1-\alpha)T_0^0\}, 1)$. The lower bound embodies the fact that the optimal path has $\dot{\psi}^*(t) \geq 2\alpha - 1$ and therefore $c = \psi^*(1) = T_0^0 + \int_{T_0^0}^1 \dot{\psi}^*(t)\,dt \geq 2\alpha - 1 + 2(1-\alpha)T_0^0$. Once λ^* is identified, one can numerically evaluate the integral

$$z = \int_0^T \psi^\circ(t)\,dt.$$

For example, with $\alpha = 1/3$, Figure 7 plots the numerically evaluated rate function $I_{\overline{W}}(z)$ versus z. The initial shape of the rate function is similar to $\sqrt{x}$, but, as can be seen clearly in the graph, once z is sufficiently large that the optimal path has $c > 0$, the rate function increases dramatically.

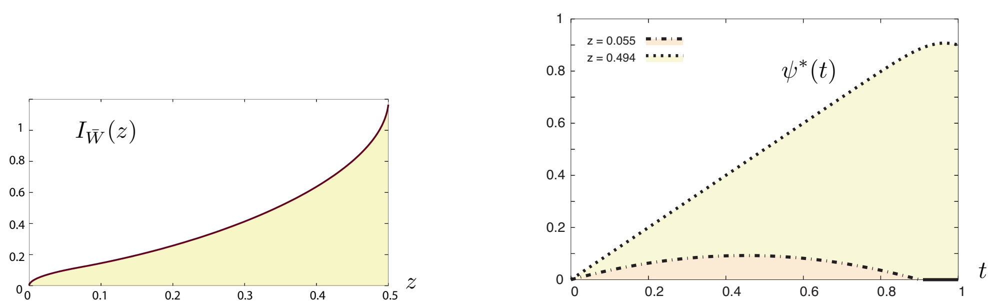


Figure 7: Non-coercive rate function example: M/M/1 queue-lengths, Bernoulli increments. On the left hand side is $I_{\overline{W}}(z)$ versus z. Shown on the right hand side are most likely RRW paths, $\phi^*(t) \approx n^{-1}W_{\lfloor nt \rfloor}$, given that $\overline{W}_n \approx nz$.

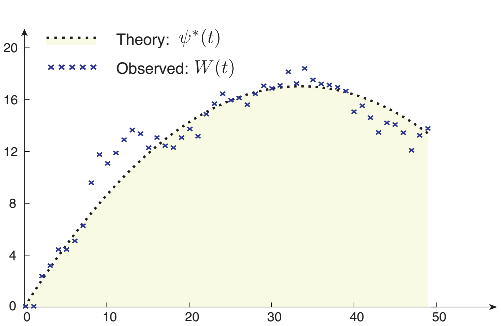
For $z < 0.0682$, the optimal value of T_1 is less than 1 and the numerically-identified most likely path occurs with $T_0^0 = 0$. For $z > 0.0682$, $T = 1$ and the optimal path also has $T_0^0 = 0$

apart from, possibly, as $z \to 1/2$. The reason for this caveat is that the cost of a path with $T_0^0 = 0$ becomes numerically indistinguishable from those with $T_0^0 > 0$ if $z \approx 1/2$. To see this, note that as $z \uparrow 1/2$, $c \uparrow 1$, so that if $T_0^0 = 0$, then λ^*, the solution of equation (18), tends to 1 and the most likely path to large simulated mean has slope close to 1 for a substantial range of t. This is illustrated in the higher paths on the right in Figure 7 corresponding to the most likely path for the deviation $z \approx 0.45$. For this path $\lambda^* \approx 0.941$. Note that the slope is nearly 1 until near $t = 0.85$, even though technically $T_0^0 = 0$.

4.4. Simulations

One of the strong deductions of these sample-path arguments is the prediction of the most likely path that gives rise to a large simulated mean. To illustrate the merits of these predictions we conducted simulations of the RRW in two settings: Gaussian i.i.d. increments, and the M/M/1 queue example introduced in Section 4.3.

In each case, the RRW was simulated for a fixed number of steps $n \geq 1$, and the simulation was repeated 2×10^8 times. Of these simulated RRWs, the one with the largest simulated mean was recorded and compared with the theory laid out in Section 4. This theory predicts the approximation $W_{[t]} \approx n\,\phi^*(t/n)$ for $t \in [0, n]$. The results from two experiments are illustrated in Figure 8.



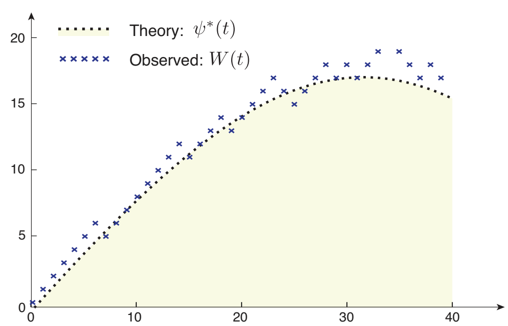


Figure 8: RRW with i.i.d. increments. The figure shown on the left hand side shows experiments obtained with Gaussian increments. On the right hand side are results obtained for the M/M/1 queue model in which the increments take on values ± 1. In each case, the observed path has the largest simulated mean out of 2×10^8 sampled paths. Also, shown in each figure is the corresponding theoretical prediction of the most likely path, given the observed simulated mean.

In the first experiment illustrated on the left hand side, the increments of the random walk were taken to be i.i.d. Gaussian with $\delta = 0.5$, $\sigma^2 = 1$ and the time-horizon $n = 50$. The second experiment used the M/M/1 queue example found in Section 4.3 with $\alpha = 0.3$ and $n = 40$. In each experiment, the observed sample path is plotted along with the theoretical prediction corresponding to the observed simulated mean. The theory's quantitative power in predicting the shape of the most likely path is apparent in these figures.

5. Discussion

As a final remark, we have mentioned that our fundamental hypothesis, Assumption 1, encompasses the light tailed setting in the absence of long range dependence. However, by changing the speed of the LDP, this assumption also holds for certain long range dependent processes. For example, in continuous time it is known, e.g. [9][35][29], that fractional Brownian Motion (fBM) with Hurst parameter H satisfies the LDP at speed $n^{2(1-H)}$ in $D[0,1]$. As the nature of the speed does not enter the proof of Theorem 3, the first conclusion of that theorem holds for these processes. However, even for the canonical example of fBM the rate function is not of the integral form in equation (4) and thus, in the long range dependent setting, it is hard to deduce if any general properties exist for the most likely paths to a large simulated mean.

Acknowledgment. Financial support from the National Science Foundation (NSF CCF 07-29031), AFOSR (FA9550-09-1-0190) and Science Foundation Ireland (07/IN.1/1901) is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, AFOSR or SFI.

References

[1] V. Anantharam, *How large delays build up in a $GI/G/1$ queue*, Queueing Systems Theory Appl. **5** (1989), no. 4, 345–367.

[2] S. Asmussen, *Applied probability and queues*, second ed., Applications of Mathematics (New York), vol. 51, Springer-Verlag, New York, 2003, Stochastic Modelling and Applied Probability.

[3] P. Billingsley, *Convergence of probability measures*, John Wiley & Sons Inc., New York, 1968.

[4] A. A. Borovkov, O. J. Boxma, and Z. Palmowski, *On the integral of the workload process of the single server queue*, J. Appl. Probab. **40** (2003), no. 1, 200–225.

[5] H. Cramér, *Sur un nouveau théorème-limite de la théorie des probabilités*, Actualits Sci. Indust. **736** (1938), 5–23.

[6] A. Dembo and T. Zajic, *Large deviations: from empirical mean and measure to partial sums process*, Stochastic Process. Appl. **57** (1995), no. 2, 191–224.

[7] A. Dembo and O. Zeitouni, *Large deviation techniques and applications*, Springer, 1998.

[8] F. den Hollander, *Large deviations*, Fields Institute Monographs, vol. 14, American Mathematical Society, Providence, RI, 2000.

[9] J.-D. Deuschel and D. W. Stroock, *Large deviations*, Pure and Applied Mathematics, vol. 137, Academic Press Inc., Boston, MA, 1989.

[10] R. L. Dobrushin and E. A. Pecherskiĭ, *Large deviations for random processes with independent increments on an infinite interval*, Problemy Peredachi Informatsii **34** (1998), no. 4, 76–108.

[11] N. G. Duffield, J. T. Lewis, N. O'Connell, R. Russell, and F. Toomey, *Entropy of ATM traffic streams: a tool for estimating QoS parameters*, IEEE J. Sel. Area Comm. **13** (1995), no. 6, 981–990.

[12] N. G. Duffield and N. O'Connell, *Large deviations and overflow probabilities for the general single server queue, with applications*, Math. Proc. Cam. Phil. Soc. **118** (1995), no. 2, 363–374.

[13] K. Duffy, J. T. Lewis, and W. G. Sullivan, *Logarithmic asymptotics for the supremum of a stochastic process*, Ann. Appl. Probab. **13** (2003), no. 2, 430–445.

[14] K. Duffy and A. P. Metcalfe, *The large deviations of estimating rate functions*, J. Appl. Probab. **42** (2005), no. 1, 267–274.

[15] K. Duffy and W. G. Sullivan, *Logarithmic asymptotics for unserved messages at a FIFO*, Markov Process. Related Fields **10** (2004), no. 1, 175–189.

[16] K. R. Duffy and D. Malone, *Logarithmic asymptotics for a single-server processing distinguishable sources*, Math. Methods Oper. Res. **67** (2008), no. 3, 509–537.

[17] A. Ganesh, N. O'Connell, and D. Wischik, *Big queues*, Lecture Notes in Mathematics, vol. 1838, Springer-Verlag, Berlin, 2004.

[18] A. J. Ganesh, *Bias correction in effective bandwidth estimation*, Performance Evaluation **27** (1996), no. 8, 319–330.

[19] A. J. Ganesh and N. O'Connell, *A large deviation principle with queueing applications*, Stochastics and Stochastic Reports **73** (2002), no. 1-2, 25–35.

[20] P. Glynn and W. Whitt, *Logarithmic asymptotics for steady-state tail probabilities in a single-server queue*, J. Appl. Probab. **31A** (1994), 413–430.

[21] L. Györfi, A. Rácz, K. Duffy, J. T. Lewis, and F. Toomey, *Distribution-free confidence intervals for measurement of effective bandwidths*, J. Appl. Probab. **37** (2000), 1–12.

[22] F. P. Kelly, *Notes on effective bandwidths*, Stochastic Networks: Theory and Applications (F. P. Kelly, S. Zachary, and I. B. Ziedins, eds.), Oxford University Press, 1996, pp. 141–168.

[23] M. Lelarge, *Tail asymptotics for discrete event systems*, Discrete Event Dyn. Syst. **18** (2008), no. 4, 563–584.

[24] J. T. Lewis, C.-E. Pfister, and W. G. Sullivan, *Entropy, concentration of probability and conditional limit theorems*, Markov Process. Related Fields **1** (1995), no. 3, 319–386.

[25] D. V. Lindley, *The theory of queues with a single server*, Proc. Cambridge Philos Soc. **48** (1952), 277–289.

[26] R. M. Loynes, *The stability of a queue with non-independent inter-arrival and service times*, Proc. Cam. Phil. Soc. **58** (1962), 497–520.

[27] D. G. Luenberger, *Optimization by vector space methods*, John Wiley & Sons Inc., New York, 1969.

[28] J. Lynch and J. Sethuraman, *Large deviations for processes with independent increments*, Ann. Probab. **15** (1987), no. 2, 610–627.

[29] K. Majewski, *Large deviations for multi-dimensional reflected fractional Brownian motion*, Stoch. Stoch. Rep. **75** (2003), no. 4, 233–257.

[30] M. Mandjes and R. Van De Meent, *Resource dimensioning through buffer sampling*, IEEE/ACM Trans. Network. **17** (2009), no. 5, 1631–1644.

[31] S. P. Meyn, *Large deviation asymptotics and control variates for simulating large functions*, Ann. Appl. Probab. **16** (2006), no. 1, 310–339.

[32] ______, *Control techniques for complex networks*, Cambridge University Press, Cambridge, 2008.

[33] A. A. Mogul′skiĭ, *Large deviations for trajectories of multi-dimensional random walks*, Th. Prob. Appl **21** (1976), 300–315.

[34] ______, *Large deviations for processes with independent increments*, Ann. Probab. **21** (1993), no. 1, 202–215.

[35] I. Norros, *Busy periods of fractional brownian storage: a large deviations approach*, Adv. Performance Anal. **2** (1999), no. 1, 1–19.

[36] I. Ch. Paschalidis and S. Vassilaras, *On the estimation of buffer overflow probabilities from measurements*, IEEE Trans. Inform. Theory **47** (2001), no. 1, 178–191.

[37] A. Puhalskii, *Large deviation analysis of the single server queue*, Queueing Systems **21** (1995), no. 1–2, 5–66.

[38] A. A. Puhalskii and W. Whitt, *Functional large deviation principles for waiting and departure processes*, Probab. Engrg. Inform. Sci. **12** (1998), no. 4, 479–507.

[39] A. V. Skorohod, *Limit theorems for stochastic processes*, Teor. Veroyatnost. i Primenen. **1** (1956), 289–319.

[40] W. Whitt, *Stochastic-process limits*, Springer-Verlag, New York, 2002.