On the Large Deviations Properties of the Weighted-Serve-the-Longer-Queue Policy

Paul Dupuis, Kevin Leder and Hui Wang

Abstract. We identify the large deviation rate function for a single server with multi-class arrivals in which the service priority is determined according to the weighted-serve-the-longer-queue policy. The problem setup falls into the general category of systems with discontinuous statistics. Our analysis, which is largely based on a weak convergence approach, does not require any symmetry or dimensional restrictions.

1. Introduction

Consider a single server that must serve multiple queues of customers from different classes. A common service discipline in this situation is the serve-the-longest-queue policy, in which the longest queue is given priority. In this paper we will consider a natural generalization of this discipline, namely, the *weighted-serve-the-longer-queue* (WSLQ) policy. Under WSLQ, each queue length is multiplied by a constant to determine a "score" for that queue, and the queue with the largest score is granted priority. Such service policies are more appropriate than serve-the-longest-queue policy when the different arrival queues or customer classes have different requirements or statistical properties. For example, if there is a finite queueing capacity to be split among the different classes, one may want to choose the partition and the weighting constants in order to optimize a certain performance measure.

Because WSLQ is a frequently proposed discipline for queueing models in communication problems, a large deviations analysis of this protocol is useful [12]. However, service policies such as WSLQ are not smooth functions of the system state and lead to multidimensional stochastic processes with *discontinuous*

Paul Dupuis was supported in part by the National Science Foundation (NSF-DMS-0306070 and NSF-DMS-0404806) and the Army Research Office (DAAD19-02-1-0425).

Kevin Leder was supported in part by the National Science Foundation (NSF-DMS-0404806).

Hui Wang was supported in part by the National Science Foundation (NSF-DMS-0404806).

statistics. In general, large deviation properties of processes with discontinuous statistics are hard to analyze [1, 7, 8, 9]. This is especially true when the discontinuities appear on the interior of the state space, rather than the boundary. In fact, very general results with an explicit identification of the rate function only exist for the case where two regions of smooth statistical behavior are separated by an interface of codimension one [3]. For the WSLQ policy, the large deviation analysis has been limited to special, two-dimensional cases [11]. Large deviations for a weighted-serve-the-longer-workload policy are treated in [10], but considers a particular event rather than the sample path large deviations principle.

The purpose of the present work is to show that a complete large deviation analysis of WSLQ is possible without any symmetry or dimensional assumptions. Given the intrinsic difficulties in models with discontinuous statistics, it is worthwhile to explain what makes such an analysis possible for WSLQ. To this end, we recall the main difficulty in the large deviation analysis of systems with discontinuous statistics. A large deviation upper bound can often be established using the results in [6], which assumes little regularity on the statistical behavior of the underlying processes. However, this upper bound is generally *not* tight, even for the very simple situation of two regions of constant statistical behavior separated by a hyperplane of codimension one [5].

The reason for this gap is most easily identified by considering the corresponding lower bound. When proving a large deviation lower bound, it is necessary to analyze the probability that the process closely follows or tracks a constant velocity trajectory that lies on the interface of two or more regions of smooth statistical behavior. For this one has to consider all changes of measure in these different regions that lead to the desired tracking behavior. The thorny issue is how to characterize such changes of measure. In the case of two regions [5], this can be done in a satisfactory fashion and it turns out that the large deviation rate function is a modified version of the upper bound in [6]. The modification is made to explicitly include certain "stability about the interface" conditions, and part of the reason that everything works out nicely in the setup of [5] is that these stability conditions can be easily characterized. However, the analogous characterization of stability is not known for more elaborate settings such as WSLQ, where the regions of constant statistical behavior are defined according to the partition of the state space by a finite number of hyperplanes of codimension one. Therefore, at present there is no general theory that subsumes WSLQ as a special case.

However, a key observation that makes possible a large deviations analysis for WSLQ is that for this model, the required stability conditions are *implicitly* and *automatically* built into the upper bound rate function of [6]. More precisely, it can be shown that in the lower bound analysis one can restrict, a priori, to a class of changes of measure for which the stability conditions are automatically implied. Thus while it is true that the upper bound of [6] is not tight in general, it is so in this case due to the structural properties of WSLQ policy.

The study of the large deviation properties of WSLQ is partly motivated by the problem of estimating buffer overflow (rare event) probabilities for stable

WSLQ systems using importance sampling. It turns out that the simple form of the large deviation local rate function as exhibited in (3.1) and (3.2) is essential toward constructing simple and asymptotically optimal importance sampling schemes using a game theoretic approach. These results will be reported elsewhere.

The paper is organized as follows. In Section 2, we introduce the single server system with WSLQ policy. In Section 3, we state the main result, whose proof is presented in Section 4. Collected in the appendices are a lengthy technical part of the proof that involves the approximation of continuous trajectories, as well as miscellaneous results.

2. Problem Setup

Consider a server with d customer classes, where customers of class i arrive according to a Poisson process with rate λ_i and are buffered at queue i for $i = 1, \ldots, d$. The service time for a customer of class i is exponentially distributed with rate μ_i.

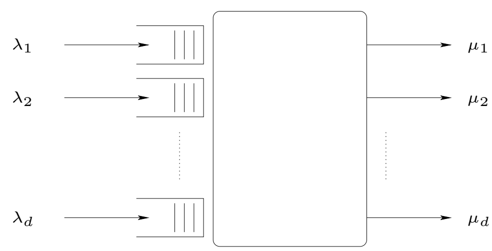

Figure 1: WSLQ system

The service policy is determined according to the WSLQ discipline that can be described as follows. Let c_i be the weight associated with class i. If the size of queue i is q_i, then the “score” of queue i is defined as $c_i q_i$ and service priority will be given to the queue with the maximal score. When there are multiple queues with the maximal score, the assignment of priority among these queues can be arbitrary – the choice is indeed non-essential and will lead to the same rate function. We adopt the convention that when there are ties, the priority will be given to the queue with the largest index.

The system state at time t is the vector of queue lengths and is denoted by $Q(t) \doteq (Q_1(t), \ldots, Q_d(t))$. Then Q is a continuous time pure jump Markov process whose possible jumps belong to the set

$$\Theta = \{\pm e_1, \pm e_2, \ldots, \pm e_d\}.$$

For $v \in \Theta$, let $r(x; v)$ denote the jump intensity of process Q from state x to state $x + v$. Under the WSLQ discipline, these jump intensities are as follows. For

$x = (x_1, \ldots, x_d) \in \mathbb{R}^d_+$ and $x \neq 0$, let $\pi(x)$ denote the indices of queues that have the maximal score, that is,

$$\pi(x) \doteq \left\{ 1 \leq i \leq d : c_i x_i = \max_j c_j x_j \right\}.$$

Then

$$r(x; v) = \begin{cases} \lambda_i, & \text{if } v = e_i \text{ and } i = 1, \ldots, d, \\ \mu_i, & \text{if } v = -e_i \text{ where } i = \max \pi(x), \\ 0\,, & \text{otherwise.} \end{cases}$$

For $x = 0$, there is no service and the jump intensities are

$$r(0; v) = \begin{cases} \lambda_i, & \text{if } v = e_i \text{ and } i = 1, \ldots, d, \\ 0\,, & \text{otherwise.} \end{cases}$$

We also set

$$\pi(0) \doteq \{0, 1, 2, \ldots, d\}.$$

An illustrative figure for the case of two queues in given below.

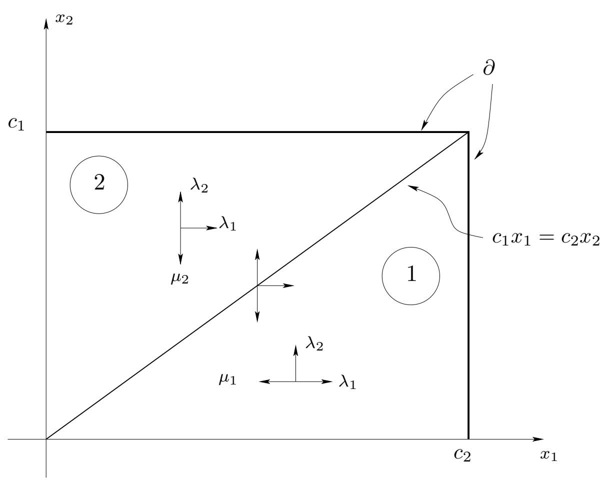

Figure 2: System dynamics for $d = 2$.

Remark 2.1. It is not difficult to see that the system dynamics have constant statistical behavior in the regions where $\pi(\cdot)$ is constant. Discontinuity occurs when $\pi(\cdot)$ changes, and every x with $|\pi(x)| \geq 2$ (i.e., when there is a tie) is indeed a discontinuous point. Therefore, we have various discontinuity interfaces with different dimensions. For example, for every subset $A \subset \{1, 2, \ldots, d\}$ with $|A| \geq 2$ or $A = \{0, 1, 2, \ldots, d\}$, the set $\{x \in \mathbb{R}^d_+ : \pi(x) = A\}$ defines an interface with dimension $d - |A| + 1$.

Remark 2.2. The definition of $\pi(0)$ is introduced to cope with the discontinuous dynamics at the origin. Note that with this definition, $\pi(x)$ can only be $\{0,1,2,\ldots,d\}$ if $x=0$ and a subset of $\{1,2,\ldots,d\}$ if $x\neq 0$.

Remark 2.3. A useful observation is that π is *upper semicontinuous* as a set-valued function. That is, for any $x\in\mathbb{R}^d_+$, $\pi(y)\subset\pi(x)$ for all y in a small neighborhood of x.

3. The main result

In order to state a large deviation principle on path space, we fix arbitrarily $T>0$, and for each $n\in\mathbb{N}$ let $\{X^n(t): t\in[0,T]\}$ be the scaled process defined by

$$X^n(t)\doteq\frac{1}{n}Q(nt).$$

Then X^n is a continuous time Markov process with generator

$$\mathcal{L}^n f(x)=n\sum_{v\in\Theta}r(x;v)\left[f\left(x+v/n\right)-f(x)\right].$$

The processes $\{X^n\}$ live in the space of cadlag functions $\mathcal{D}([0,T]:\mathbb{R}^d)$, which is endowed with the Skorohod metric and thus a Polish space.

3.1. The rate function

For each $i=1,\ldots,d$, let $H^{(i)}$ be the convex function given by

$$H^{(i)}(\alpha)\doteq\mu_i(e^{-\alpha_i}-1)+\sum_{j=1}^{d}\lambda_j(e^{\alpha_j}-1),$$

for all $\alpha=(\alpha_1,\ldots,\alpha_d)\in\mathbb{R}^d$. We also define for $i=0$,

$$H^{(0)}(\alpha)\doteq\sum_{j=1}^{d}\lambda_j(e^{\alpha_j}-1)$$

for $\alpha\in\mathbb{R}^d$.

For each non-empty subset $A\subset\{1,\ldots,d\}$ or $A=\{0,1,\ldots,d\}$, let L^A be the Legendre transform of $\max_{i\in A}H^{(i)}$, that is,

$$L^A(\beta)\doteq\sup_{\alpha\in\mathbb{R}^d}\left[\langle\alpha,\beta\rangle-\max_{i\in A}H^{(i)}(\alpha)\right]$$

for each $\beta\in\mathbb{R}^d$. Clearly, L^A is convex and non-negative. When A is a singleton $\{i\}$, we simply write L^A as $L^{(i)}$. A useful representation of L^A [4, Corollary D.4.3] is

$$L^A(\beta)=\inf\sum_{i\in A}\rho^{(i)}L^{(i)}(\beta^{(i)}),\tag{3.1}$$

where the infimum is taken over all $\{(\rho^{(i)}, \beta^{(i)}) : i \in A\}$ such that

$$\rho^{(i)} \geq 0, \ \sum_{i \in A} \rho^{(i)} = 1, \ \sum_{i \in A} \rho^{(i)} \beta^{(i)} = \beta. \tag{3.2}$$

Furthermore, for $x \in \mathbb{R}^d_+$ let $L(x, \beta) \doteq L^{\pi(x)}(\beta)$.

Now we can define the process level rate function. For $x \in \mathbb{R}^d_+$, define $I_x : \mathcal{D}([0, T] : \mathbb{R}^d) \to [0, \infty]$ by

$$I_x(\psi) \doteq \int_0^T L(\psi(t), \dot{\psi}(t)) dt \tag{3.3}$$

if $\psi(0) = x$, ψ is absolutely continuous, and $\psi(t) \in \mathbb{R}^d_+$ for all $t \in [0, T]$. Otherwise set $I_x(\psi) \doteq \infty$. The family of rate functions $\{I_x : x \in \mathbb{R}^d_+\}$ has compact level sets on compacts in the sense that, for every $M \geq 0$ and every compact set $C \in \mathbb{R}^d_+$, the set

$$\cup_{x \in C} \left\{ \psi \in \mathcal{D}([0, T] : \mathbb{R}^d) : I_x(\psi) \leq M \right\}$$

is compact [6, Theorem 1.1].

3.2. The main theorem

The main result of this paper can be stated as follows. Let E_{x_n} denote the expectation conditioned on $X^n(0) = x_n$.

Theorem 3.1. *The process $\{X^n(t) : t \in [0, T]\}$ satisfies the uniform Laplace principle with rate functions $\{I_x : x \in \mathbb{R}^d_+\}$. That is, for any sequence $\{x_n\} \subset \mathbb{R}^d_+$ such that $x_n \to x$ and any bounded continuous function $h : \mathcal{D}([0, T] : \mathbb{R}^d) \to \mathbb{R}$, we have*

$$\lim_{n \to \infty} -\frac{1}{n} \log E_{x_n} \left\{ \exp \left[-nh(X^n) \right] \right\} = \inf_{\psi \in \mathcal{D}([0,T]:\mathbb{R}^d)} \{ I_x(\psi) + h(\psi) \}.$$

In particular, $\{X^n(t) : t \in [0, T]\}$ satisfies the large deviation principle with rate function $\{I_x : x \in \mathbb{R}^d_+\}$.

4. Proof of the main theorem

Throughout the rest of the paper, we will assume without loss of generality that $T = 1$. We only need to show the uniform Laplace principle which automatically implies the large deviation principle [4, Theorem 1.2.3]. The uniform Laplace principle upper bound is implied by the large deviation upper bound [6, Theorem 1.1] and an argument similar to [4, Theorem 1.2.1]. Therefore, it is only necessary to prove the uniform Laplace principle lower bound. That is,

$$\limsup_{n \to \infty} -\frac{1}{n} \log E_{x_n} \left\{ \exp \left[-nh(X^n) \right] \right\} \leq \inf_{\psi \in \mathcal{D}([0,1]:\mathbb{R}^d)} \{ I_x(\psi) + h(\psi) \}. \tag{4.1}$$

One can a priori restrict to ψ such that $I_x(\psi) < \infty$ since the inequality holds trivially otherwise. Note that such ψ's are necessarily absolutely continuous by the definition of I_x.

4.1. An approximation lemma

A very important step in the proof of (4.1) is the following approximation lemma. A function ψ^* has property $\mathcal{P}$ if there exists a $k \in \mathbb{N}$ and $0 = t_0 < t_1 < \cdots < t_{k-1} < t_k = 1$ such that on each interval $(t_{i-1}, t_i), i = 1, \ldots, k$ $\dot{\psi}^*$ and $\pi(\psi^*)$ are each constant. Then define $\mathcal{N}$ be the collection of functions $\psi^* \in \mathcal{C}([0,1] : \mathbb{R}_+^d)$ that have property $\mathcal{P}$.

Lemma 4.1. *Given any $\psi \in \mathcal{D}([0,1] : \mathbb{R}^d)$ such that $I_x(\psi) < \infty$ and any $\delta > 0$, there exists $\psi^* \in \mathcal{N}$ such that $\|\psi - \psi^*\|_\infty < \delta$ and $I_x(\psi^*) \le I_x(\psi) + \delta$.*

The proof of Lemma 4.1 is lengthy and technical, and is deferred to Appendix A.

We claim that, in order to show inequality (4.1), it suffices to show that for any $x_n \to x$ and $\psi^* \in \mathcal{N}$,

$$\limsup_{n\to\infty} -\frac{1}{n} \log E_{x_n} \{\exp[-nh(X^n)]\} \le I_x(\psi^*) + h(\psi^*). \tag{4.2}$$

This reduction follows easily from Lemma 4.1 and the continuity of h. We omit the details.

4.2. A stochastic control representation

The proof of (4.2) uses the weak convergence approach, and is based on the following formula. Define the function ℓ by

$$\ell(x) = \begin{cases} x\log x - x + 1, & \text{if } x \ge 0, \\ \infty \quad , & \text{if } x < 0, \end{cases}$$

with the convention $0\ell(0/0) \doteq 0$. Let $\bar{r}(x,t;v)$ be non-negative and uniformly bounded, and piecewise constant in t, and also satisfy $\bar{r}(x,t;v) = 0$ whenever $r(x;v) = 0$. Let $\bar{X}^n$ be the non-stationary jump Markov process with generator

$$\bar{\mathcal{L}}^n f(x,t) = n \sum_{v\in\Theta} \bar{r}(x,t;v)\left[f(x+v/n) - f(x)\right].$$

Then

$$\begin{aligned} & -\frac{1}{n}\log E_x\{\exp[-nh(X^n)]\} \qquad (4.3) \\ & \le \quad \inf_{\bar{r}} E_x\left[\int_0^1 \sum_{v\in\Theta} r(\bar{X}^n(t);v)\,\ell\left(\frac{\bar{r}(\bar{X}^n(t),t;v)}{r(\bar{X}^n(t);v)}\right)dt + h(\bar{X}^n)\right]. \end{aligned}$$

In this inequality $\bar{r}$ can be viewed as a control and $\bar{X}^n$ a controlled process.

The proof of (4.3) follows from the relative entropy representation formula for exponential integrals. Let P be a probability measure and let h be a bounded continuous function on $\mathcal{D}([0,1] : \mathbb{R}^d)$. For another probablity measure Q on $\mathcal{D}([0,1] : \mathbb{R}^d)$ let $R(Q\,\|P)$ denote the relative entropy of Q with respect to P. Then [4]

$$-\frac{1}{n}\log\int_{\mathcal{D}([0,1]:\mathbb{R}^d)} e^{-nh}dP = \inf\left[\frac{1}{n}R(Q\,\|P) + \int_{\mathcal{D}([0,1]:\mathbb{R}^d)} hdQ\right],$$

where the infimum is over all Q. If one restricts Q to the probability measures induced by the class of Markov processes described above, one obtains an inequality. Finally, we substitute the explicit form of the Radon-Nikodym derivative dQ/dP (as in [11, Theorem B.6]) into the definition of $R(Q\,\|P)$ to arrive at the inequality in (4.3).

Thanks to the control representation (4.3), the Laplace principle lower bound (4.2) for $\psi^* \in \mathcal{N}$ follows if one can, for an arbitrarily fixed $\varepsilon > 0$, construct a control (abusing the notation) $\bar{r} = \bar{r}^\varepsilon$ such that

$$\limsup_{n\to\infty} E_{x_n}\left[\int_0^1 \sum_{v\in\Theta} r(\bar{X}^n(t);v)\ell\left(\frac{\bar{r}(\bar{X}^n(t),t;v)}{r(\bar{X}^n(t);v)}\right)dt + h(\bar{X}^n)\right] \tag{4.4}$$
$$\leq \quad I_x(\psi^*) + h(\psi^*) + \varepsilon.$$

The details of the construction will be carried out in the next section.

4.3. Properties of the rate function

We give some useful representation formulae for the local rate functions L^A.

Lemma 4.2. *Given $\beta \in \mathbb{R}^d$, the following representation for $L^{(i)}(\beta)$ holds.*

1. *For each $i = 1, 2, \ldots, d$,*

$$L^{(i)}(\beta) = \inf\left\{\mu_i\ell\left(\frac{\bar{\mu}_i}{\mu_i}\right) + \sum_{j=1}^d \lambda_j\ell\left(\frac{\bar{\lambda}_j}{\lambda_j}\right) : -\bar{\mu}_i e_i + \sum_{j=1}^d \bar{\lambda}_j e_j = \beta\right\}.$$

2. *For $i = 0$,*

$$L^{(0)}(\beta) = \inf\left\{\sum_{j=1}^d \lambda_j\ell\left(\frac{\bar{\lambda}_j}{\lambda_j}\right) : \sum_{j=1}^d \bar{\lambda}_j e_j = \beta\right\}.$$

In every case the infimum is attained.

Proof. Note that for every $\lambda > 0$ and $v \in \mathbb{R}^d$, the Legendre transform of the convex function

$$h(\alpha) \doteq \lambda\left(e^{\langle\alpha,v\rangle} - 1\right)$$

is

$$h^*(\beta) \doteq \begin{cases} \lambda\ell(\bar{\lambda}/\lambda), & \text{if } \beta = \bar{\lambda}v \text{ for some } \bar{\lambda} \in \mathbb{R}, \\ \infty \quad , & \text{otherwise} \end{cases}$$

for every $\beta \in \mathbb{R}^d$. The proof of this claim is straightforward computation and we omit the details. Now the representation for $L^{(i)}$ follows directly from [4, Corollary D.4.2]. The attainability of the infimum is elementary. $\square$

Lemma 4.3. *Given $\beta \in \mathbb{R}^d$, we have the following representation for $L^A(\beta)$.*

1. *Assume $A \in \{1, 2, \ldots, d\}$ is non-empty. Then*

$$L^A(\beta) = \inf \left[\sum_{i \in A} \rho^{(i)} \mu_i \ell \left(\frac{\bar{\mu}_i}{\mu_i} \right) + \sum_{j=1}^{d} \lambda_j \ell \left(\frac{\bar{\lambda}_j}{\lambda_j} \right) \right],$$

where the infimum is taken over all collections of $(\rho^{(i)}, \bar{\mu}_i, \bar{\lambda}_j)$ such that

$$\rho^{(i)} \geq 0, \ \sum_{i \in A} \rho^{(i)} = 1, \ -\sum_{i \in A} \rho^{(i)} \bar{\mu}_i e_i + \sum_{j=1}^{d} \bar{\lambda}_j e_j = \beta. \tag{4.5}$$

2. *For $A = \{0, 1, 2, \ldots, d\}$, we have*

$$L^A(\beta) = \inf \left[\sum_{i=1}^{d} \rho^{(i)} \mu_i \ell \left(\frac{\bar{\mu}_i}{\mu_i} \right) + \sum_{j=1}^{d} \lambda_j \ell \left(\frac{\bar{\lambda}_j}{\lambda_j} \right) \right],$$

where the infimum is taken over all collections of $(\rho^{(i)}, \bar{\mu}_i, \bar{\lambda}_j)$ such that

$$\rho^{(i)} \geq 0, \ \sum_{i=0}^{d} \rho^{(i)} = 1, \ -\sum_{i=1}^{d} \rho^{(i)} \bar{\mu}_i e_i + \sum_{j=1}^{d} \bar{\lambda}_j e_j = \beta.$$

Proof. We only present the proof for Part 1. The proof for Part 2 is similar and thus omitted. Thanks to Lemma 4.2 and equations (3.1)-(3.2), we have

$$L^A(\beta) = \inf \sum_{i \in A} \rho^{(i)} \left\{ \mu_i \ell \left(\frac{\bar{\mu}_i^{(i)}}{\mu_i} \right) + \sum_{j=1}^{d} \lambda_j \ell \left(\frac{\bar{\lambda}_j^{(i)}}{\lambda_j} \right) \right\},$$

where the infimum is taken over all $(\rho^{(i)}, \bar{\mu}_i^{(i)}, \bar{\lambda}_j^{(i)})$ such that

$$\rho^{(i)} \geq 0, \ \sum_{i \in A} \rho^{(i)} = 1, \ \sum_{i \in A} \rho^{(i)} \left[-\bar{\mu}_i^{(i)} e_i + \sum_{j=1}^{d} \bar{\lambda}_j^{(i)} e_j \right] = \beta. \tag{4.6}$$

Abusing the notation a bit, write $\bar{\mu}_i = \bar{\mu}_i^{(i)}$ for $i \in A$, and let $\bar{\lambda}_j \doteq \sum_{i \in A} \rho^{(i)} \bar{\lambda}_j^{(i)}$ for $j = 1, 2, \ldots, d$. Thanks to (4.6), the collection $(\rho^{(i)}, \bar{\mu}_i, \bar{\lambda}_j)$ satisfies the constraints (4.5). Observing that, by convexity of ℓ,

$$\sum_{i \in A} \rho^{(i)} \sum_{j=1}^{d} \lambda_j \ell \left(\frac{\bar{\lambda}_j^{(i)}}{\lambda_j} \right) = \sum_{j=1}^{d} \lambda_j \sum_{i \in A} \rho^{(i)} \ell \left(\frac{\bar{\lambda}_j^{(i)}}{\lambda_j} \right) \geq \sum_{j=1}^{d} \lambda_j \ell \left(\frac{\bar{\lambda}_j}{\lambda_j} \right),$$

with equality if $\lambda_j^{(i)} = \lambda_k^{(i)}$ for every j, k. The first part of Lemma 4.3 now follows readily. ☐

Remark 4.4. The representation of L^A in Lemma 4.3 remains valid if we further constrain $\rho^{(i)}$, $\bar{\mu}_i$, and $\bar{\lambda}_i$ to be strictly positive for every $i \in A$. This is an easy consequence of the fact that ℓ is finite and continuous on the interval $[0, \infty)$. We omit the details.

Remark 4.5. Given a non-empty subset $A \subsetneq \{1,2,\ldots,d\}$ and $\beta = (\beta_1,\ldots,\beta_d) \in \mathbb{R}^d$, $L^A(\beta)$ is finite if and only if $\beta_j \geq 0$ for all $j \notin A$. For $A = \{1,2,\ldots,d\}$ or $\{0,1,2,\ldots,d\}$, $L^A(\beta)$ is finite for every $\beta \in \mathbb{R}^d$.

4.4. The construction of controls and the cost

Fix $\varepsilon > 0$. We will use ψ^* to construct a control $\bar{r} = \bar{r}^\varepsilon$ based on the representation of the local rate function as in Lemma 4.3. For notational simplicity, we drop the superscript ε.

Since $\psi^* \in \mathcal{N}$, there exist $0 = t_0 < t_1 < \cdots < t_K = 1$ such that for every k there are β_k and A_k such that $\dot{\psi}^*(t) \equiv \beta_k$ and $\pi(\psi^*(t)) \equiv A_k$ for all $t \in (t_k, t_{k+1})$. We start by defining a suitable collection of $\{(\rho_k^{(i)}, \bar{\mu}_{i,k}, \bar{\lambda}_{j,k}) : 0 \leq i \leq d, 1 \leq j \leq d\}$. We consider the following two cases.

Case 1. Suppose $A_k = \{0,1,2,\ldots,d\}$. Lemma 4.3 and Remark 4.4 imply the existence of a collection $\{(\rho_k^{(i)}, \bar{\mu}_{i,k}, \bar{\lambda}_{j,k}) : 0 \leq i \leq d, 1 \leq j \leq d\}$ such that $\bar{\mu}_{i,k} > 0$, $\bar{\lambda}_{i,k} > 0$ for all i and

$$\rho_k^{(i)} > 0,\ \sum_{i=0}^{d} \rho_k^{(i)} = 1,\ -\sum_{i=1}^{d} \rho_k^{(i)} \bar{\mu}_{i,k} e_i + \sum_{j=1}^{d} \bar{\lambda}_{j,k} e_j = \beta_k, \tag{4.7}$$

$$\sum_{i=1}^{d} \rho_k^{(i)} \mu_i \ell\left(\frac{\bar{\mu}_{i,k}}{\mu_i}\right) + \sum_{j=1}^{d} \lambda_j \ell\left(\frac{\bar{\lambda}_{j,k}}{\lambda_j}\right) \leq L^{A_k}(\beta_k) + \varepsilon. \tag{4.8}$$

Case 2. Suppose $A_k \subset \{1,2,\ldots,d\}$. According to Lemma 4.3 and Remark 4.4, for each k there exist a collection $\{(\rho_k^{(i)}, \bar{\mu}_{i,k}, \bar{\lambda}_{j,k}) : i \in A_k, 1 \leq j \leq d\}$ such that $\bar{\mu}_{i,k} > 0$, $\bar{\lambda}_{i,k} > 0$ for all $i \in A_k$ and

$$\rho_k^{(i)} > 0,\ \sum_{i \in A_k} \rho_k^{(i)} = 1,\ -\sum_{i \in A_k} \rho_k^{(i)} \bar{\mu}_{i,k} e_i + \sum_{j=1}^{d} \bar{\lambda}_{j,k} e_j = \beta_k, \tag{4.9}$$

$$\sum_{i \in A_k} \rho_k^{(i)} \mu_i \ell\left(\frac{\bar{\mu}_{i,k}}{\mu_i}\right) + \sum_{j=1}^{d} \lambda_j \ell\left(\frac{\bar{\lambda}_{j,k}}{\lambda_j}\right) \leq L^{A_k}(\beta_k) + \varepsilon. \tag{4.10}$$

We extend the definition by letting $\rho_k^{(i)} \doteq 0$, $\bar{\mu}_{i,k} \doteq \mu_i$ for $i \notin A_k$ and $i \neq 0$, and letting $\rho_k^{(0)} \doteq 0$.

The control $\bar{r}$ is defined as follows. For $t \in [t_k, t_{k+1})$, let

$$\bar{r}(x,t;v) = \begin{cases} \bar{\lambda}_{j,k}, & \text{if } v = e_j \text{ and } j = 1,\ldots,d, \\ \bar{\mu}_{j,k}, & \text{if } v = -e_j \text{ where } j = \max \pi(x) \text{ and } x \neq 0, \\ 0\ , & \text{otherwise.} \end{cases}$$

Thus, on time interval $[t_k, t_{k+1})$, the system has arrival rates $\{\bar{\lambda}_{1,k}, \ldots, \bar{\lambda}_{d,k}\}$ and service rates $\{\bar{\mu}_{1,k}, \ldots, \bar{\mu}_{d,k}\}$ under this control $\bar{r}$. The corresponding running cost for $t \in [t_k, t_{k+1})$ is

$$\begin{aligned}
&\sum_{v\in\Theta} r(\bar{X}^n(t); v)\ell\left(\frac{\bar{r}(\bar{X}^n(t), t; v)}{r(\bar{X}^n(t); v)}\right) \qquad (4.11)\\
&= \sum_{j=1}^{d} \mu_j \ell\left(\frac{\bar{\mu}_{j,k}}{\mu_j}\right) 1_{\{\max \pi(\bar{X}^n(t))=j, \bar{X}^n(t)\neq 0\}} + \sum_{j=1}^{d} \lambda_j \ell\left(\frac{\bar{\lambda}_{j,k}}{\lambda_j}\right)\\
&= \sum_{i\in A_k, i\neq 0} \mu_i \ell\left(\frac{\bar{\mu}_{i,k}}{\mu_i}\right) 1_{\{\max \pi(\bar{X}^n(t))=i, \bar{X}^n(t)\neq 0\}} + \sum_{j=1}^{d} \lambda_j \ell\left(\frac{\bar{\lambda}_{j,k}}{\lambda_j}\right),
\end{aligned}$$

here the last equality holds since $\bar{\mu}_{j,k} = \mu_j$ for $j \notin A_k$ and $\ell(1) = 0$.

For future use, we also define for each $i = 1, \ldots, d$,

$$\beta_k^{(i)} \doteq -\bar{\mu}_{i,k} e_i + \sum_{j=1}^{d} \bar{\lambda}_{j,k} e_j, \qquad (4.12)$$

which is the law of large number limit of the velocity of the controlled process if queue of class i is served. Analogously, we also define (when none of the queues are being served)

$$\beta_k^{(0)} \doteq \sum_{j=1}^{d} \bar{\lambda}_{j,k} e_j. \qquad (4.13)$$

4.5. Weak convergence analysis

In this section we characterize the limit processes. Below are a few definitions. For each j, define random measures $\{\gamma_j^n\}$ on $[0,1]$ by

$$\begin{aligned}
\gamma_j^n\{B\} &\doteq \int_B 1_{\{\max \pi(\bar{X}^n(t))=j, \bar{X}^n(t)\neq 0\}} dt, \ j = 1, 2, \ldots, d,\\
\gamma_0^n\{B\} &\doteq \int_B 1_{\{\bar{X}^n(t)=0\}} dt,
\end{aligned}$$

for Borel subsets $B \subset [0,1]$, and denote $\gamma^n \doteq (\gamma_0^n, \gamma_1^n, \ldots, \gamma_d^n)$. We also define the stochastic processes

$$S^n(t) \doteq x_n + \sum_{j=0}^{d} \left[\sum_{k=0}^{\kappa(t)-1} \beta_k^{(j)} \gamma_j^n\{[t_k, t_{k+1})\} + \beta_{\kappa(t)}^{(j)} \gamma_j^n\{[t_{\kappa(t)}, t)\} \right],$$

where $\kappa(t) = \max\{0 \leq k \leq K : t_k \leq t\}$.

Proposition 4.6. *Given any subsequence of $(\gamma^n, S^n, \bar{X}^n)$, there exist a subsubsequence, a collection of random measures $\gamma \doteq (\gamma_0, \gamma_1, \ldots, \gamma_d)$ on $[0,1]$, and a continuous process $\bar{X}$ such that*

(a) *The subsubsequence converges in distribution to $(\gamma, \bar{X}, \bar{X})$.*

(b) *With probability one, γ_j is absolutely continuous with respect to the Lebesgue measure on $[0,1]$, and its density, denoted by h_j, satisfies*

$$\sum_{j=0}^{d} h_j(t) = \sum_{j\in\pi(\bar{X}(t))} h_j(t) = 1$$

for almost every t.

(c) *With probability one, the process $\bar{X}$ satisfies*

$$\bar{X}(t) = x + \sum_{j=0}^{d}\left[\sum_{k=0}^{\kappa(t)-1}\beta_k^{(j)}\gamma_j\{[t_k,t_{k+1})\} + \beta_{\kappa(t)}^{(j)}\gamma_j\{[t_{\kappa(t)},t)\}\right]$$

for every t. Therefore, $\bar{X}$ is absolutely continuous with derivative

$$\frac{d\bar{X}(t)}{dt} = \sum_{j=0}^{d}\beta_{\kappa(t)}^{(j)}h_j(t).$$

Proof. The family of random measures $\{\gamma_j^n\}$ is contained in the compact set of sub-probability measures on $[0,1]$ and therefore tight. Furthermore, since $\{S^n\}$ is uniformly Lipschitz continuous, it takes values in a compact subset of $\mathcal{C}([0,1]:\mathbb{R}^d)$, and therefore is also tight. We also observe that for every $\varepsilon>0$,

$$\lim_{n\to\infty} P(\|\bar{X}^n - S^n\|_\infty > \varepsilon) = 0, \tag{4.14}$$

which in turn implies that $\{\bar{X}^n\}$ is tight. Equation (4.14) is trivial since $\bar{X}^n - S^n$ is a martingale (e.g., [11, Appendix B.2]), whence the process $\|\bar{X}^n - S^n\|^2$ is a submartingale. Therefore, by the submartingale inequality and the uniform boundedness of the jump intensity $\bar{r}$

$$P\left(\sup_{0\le t\le 1}\|\bar{X}^n(t) - S^n(t)\| > \varepsilon\right) \le \frac{2}{\varepsilon^2}E\left[\|\bar{X}^n(1) - S^n(1)\|^2\right] \to 0.$$

It follows that there exists a subsubsequence that converges weakly to say (γ,S,S), with $\gamma=(\gamma_0,\gamma_1,\ldots,\gamma_d)$. By the Skorohod representation theorem, we assume without loss of generality that the convergence is almost sure convergence, and everything is defined on some probability space, say $(\bar{\Omega},\bar{\mathcal{F}},\bar{P})$.

Since the $\{\gamma_j^n\}$ are absolutely continuous with respect to Lebesgue measure on $[0,1]$ with uniformly bounded densities (i.e., Radon-Nikodým derivatives) in both n and t, the limit γ_j is also absolutely continuous. Furthermore, if we define the process $\bar{X}$ as in (c), the above consideration yields that, for every $t\in[0,1]$, $S^n(t)$ converges to $\bar{X}(t)$ almost surely. Therefore, with probability one, $S(t)=\bar{X}(t)$ for all these rational $t\in[0,1]$. Since both S and $\bar{X}$ are continuous, $S=\bar{X}$ with probability one.

It remains to show the two equalities of (b). Since $\sum_{j=0}^{d}\gamma_j^n$ equals Lebesgue measure for every n, we have $\sum_{j=0}^{d}h_j(t)=1$ for almost every t. The proof of the second equality is similar to that of [4, Theorem 7.4.4(c)]. Consider an $\omega\in\bar{\Omega}$ such that $\bar{X}(t,\omega)$ is a continuous function of $t \in [0,1]$, $\gamma^n(\omega) \Rightarrow \gamma(\omega)$, and such that $\bar{X}^n(\cdot,\omega)$ converges to $\bar{X}(\cdot,\omega)$ in the Skorohod metric (whence also in sup-norm since $\bar{X}(\cdot,\omega)$ is continuous [4, Theorem A.6.5]). By the upper semicontinuity of $\pi(\cdot)$ [Remark 2.3], it follows that for any $t \in (0,1)$ and $A \subset \{0,1,\ldots,d\}$ such that $\pi(\bar{X}(t,\omega)) \subset A$, there exist an open interval (a,b) containing t and $N \in \mathbb{N}$ such that $\pi(\bar{X}^n(s,\omega)) \subset A$ for all $n \geq \mathbb{N}$ and $s \in (a,b)$. Therefore $\sum_{j \notin A} \gamma_j^n(\omega)\{(a,b)\} = 0$ for all $n \geq \mathbb{N}$. Taking the limit as $n \to \infty$ we have $\sum_{j\notin A}\gamma_j(\omega)\{(a,b)\} = 0$, which in turn implies that

$$\sum_{j\notin A} \gamma_j(\omega)\{t \in (0,1) : \pi(\bar{X}(t,\omega)) \subset A\} = 0,$$

or equivalently,

$$\int_0^1 \sum_{j\notin A} h_j(t,\omega) 1_{\{\pi(\bar{X}(t,\omega))\subset A\}}\, dt = 0.$$

We claim that this implies the desired equality at ω. Otherwise, there exists a subset $D \subset (0,1)$ with positive Lebesgue measure such that for every $t \in D$,

$$\sum_{j\notin \pi(\bar{X}(t,\omega))} h_j(t,\omega) > 0.$$

Since $\pi(\bar{X}(t,\omega))$ can only take finitely many possible values, there exists a subset (abusing the notation) $A \subset \{0,1,\ldots,d\}$ such that the set

$$\bar{D} \doteq \{t \in D : \pi(\bar{X}(t,\omega)) = A\}$$

has positive Lebesgue measure. It follows that

$$\begin{aligned}
\int_0^1 \sum_{j\notin A} h_j(t,\omega) 1_{\{\pi(\bar{X}(t,\omega))\subset A\}}\, dt &\geq \int_{\bar{D}} \sum_{j\notin A} h_j(t,\omega) 1_{\{\pi(\bar{X}(t,\omega))\subset A\}}\, dt \\
&= \int_{\bar{D}} \sum_{j\notin \pi(\bar{X}(t,\omega))} h_j(t,\omega)\, dt \\
&> 0,
\end{aligned}$$

a contradiction. This completes the proof. □

4.6. Stability analysis

In this section we prove a key lemma in the analysis that identifies the weak limit $\bar{X}$ as ψ^*. The proof uses the implied "stability about the interface" in a crucial way.

We discuss the main idea behind this stability property before giving the detailed proof. For the large deviation analysis, it is important to analyze the probability that the process tracks a segment of trajectory that lies on an interface, say $\{x : \pi(x) = A\}$, with a constant velocity, say β. To this end, it is natural to use the change of measure induced by β through the local rate function L as described in Section 4.4. However, for general systems, this very natural construction does *not* guarantee that $\bar{X}$ will follow or track ψ, and certain "stability about the interface" conditions have to be added for this to happen.

For WSLQ system, a stability condition is not explicitly needed since it is implicitly and automatically built into the upper bound local rate function L. To be more precise, denote the arrival and service rates under the change of measure by $\bar{\lambda}_i$ and $\bar{\mu}_i$ respectively. Then for the tracking behavior to take place it is required that the proportion of time that the process $\bar{X}$ spent in the region $\{x : \max \pi(x) = i\}$ equals $\rho^{(i)}$, where $\{\rho^{(i)}\}$ is the (strictly positive) solution to the system of equations

$$c_i(\bar{\lambda}_i - \rho^{(i)}\bar{\mu}_i) = c_j(\bar{\lambda}_j - \rho^{(j)}\bar{\mu}_j),\ i, j \in A, \text{ and } \sum_{i\in A} \rho^{(i)} = 1.$$

Thanks to Proposition 4.6, such proportions are characterized by $\{h_i(t)\}$ where

$$\sum_{i\in\pi(\bar{X}(t))} h_i(t) = 1.$$

Therefore, we would like to show that $\pi(\bar{X}(t)) \equiv A$ and $h_i(t) \equiv \rho^{(i)}$ for $i \in A$. This can be shown, and the argument is based on the simple fact that for any non-empty subset $B \subset \{1, 2, \ldots, d\}$ and any b, the solution $\{x_i : i \in B\}$ to the system of equations

$$c_i(\bar{\lambda}_i - x_i\bar{\mu}_i) = c_j(\bar{\lambda}_j - x_j\bar{\mu}_j),\ i, j \in B, \text{ and } \sum_{i\in B} x_i = b, \tag{4.15}$$

is unique and component-wise strictly increasing with respect to b.

For example, suppose on some time interval (a, b) that $\pi(\bar{X}(t)) \equiv B$ where B is a strict subset of A. It is not difficult to see that $\{h_i(t) : i \in B\}$ is a solution to equation (4.15) because, thanks to Proposition 4.6,

$$\frac{d}{dt}(\bar{X}(t))_i = \bar{\lambda}_i - h_i(t)\bar{\mu}_i, \text{ for } i \in B.$$

Since

$$\sum_{i\in B} \rho^{(i)} = 1 - \sum_{i\in A\setminus B} \rho^{(i)} < 1,$$

the monotonicity implies $h_i(t) > \rho^{(i)}$ for all $i \in B$. This in turn yields [see the proof of (4.25) for details] that for $i \in B$ and $j \in A \setminus B$,

$$\frac{d}{dt}\left[c_i((\bar{X}(t))_i - c_j((\bar{X}(t))_j\right] < 0. \tag{4.16}$$

Thus the differences between weighted queue lengths grows smaller, and the state is "pushed" towards the interface A. This derivation can be used to prove by contradiction that $A \subset \pi(\bar{X}(t))$. The other direction $\pi(\bar{X}(t)) \subset A$ can be shown similarly. Once $\pi(\bar{X}(t)) = A$ is shown, $h_i(t) = \rho^{(i)}$ follows immediately from the uniqueness of the solution to equation (4.15).

We want to point out that in general such stability conditions are equivalent to the existence of certain Lyapunov functions, and for the case of WSLQ, inequality (4.16) indicates that the difference of weighted queue lengths is a Lyapunov function.

Lemma 4.7. *Let $(\gamma, \bar{X}, \bar{X})$ be a limit of any weakly converging subsubsequence $(\gamma^n, S^n, \bar{X}^n)$ as in Proposition 4.6. Then with probability one, $\bar{X}(t) = \psi^*(t)$ for every $t \in [0,1]$, and for each j,*

$$h_j(t) = \sum_{k=0}^{K-1} \rho_k^{(j)} 1_{(t_k, t_{k+1})}(t)$$

for almost every $t \in [0,1]$.

Proof. The proof is by induction. By definition $\bar{X}(0) = x = \psi^*(0)$. Assume that $\bar{X}(t) = \psi^*(t)$ for all $t \in [0, t_k]$. The goal is to show that $\bar{X}(t) = \psi^*(t)$ for all $t \in [0, t_{k+1}]$. Define $A_k = \pi(\psi^*(t))$, $t \in (t_k, t_{k+1})$ and $A = \pi(\psi^*(t_k))$. Note that $A_k \subset A$ thanks to the continuity of ψ^* and the upper semicontinuity of π. Define the random time

$$\tau_k = \inf \left\{ t > t_k : \pi(\bar{X}(t)) \not\subset A \right\}.$$

Since $\pi(\bar{X}(t_k)) = \pi(\psi^*(t_k)) = A$ and $\bar{X}$ is continuous, the upper semicontinuity of π implies that $\tau_k > t_k$ and $\pi(\bar{X}(\tau_k)) \not\subset A$. We claim that it suffices to show $\bar{X}(t) = \psi^*(t)$ and $h_j(t) = \rho_k^{(j)}$ for all $t \in (t_k, t_{k+1} \wedge \tau_k)$. Indeed, if this is the case, we must have $\tau_k \geq t_{k+1}$ with probability one, since otherwise by continuity $\bar{X}(\tau_k) = \psi^*(\tau_k)$ and thus $\pi(\bar{X}(\tau_k)) = A_k \subset A$, a contradiction.

The proof of $\bar{X}(t) = \psi^*(t)$ and $h_j(t) = \rho_k^{(j)}$ for all $t \in (t_k, t_{k+1} \wedge \tau_k)$ proceeds in three steps.

Step 1.. For every $t \in (t_k, t_{k+1} \wedge \tau_k)$, either $A_k \subset \pi(\bar{X}(t))$ or $\pi(\bar{X}(t))$ is a strict subset of A_k.

Step 2.. For every $t \in (t_k, t_{k+1} \wedge \tau_k)$, $\pi(\bar{X}(t)) = A_k$.

Step 3.. For every $t \in (t_k, t_{k+1} \wedge \tau_k)$, $\bar{X}(t) = \psi^*(t)$ and $h_j(t) = \rho_k^{(j)}$.

Note that, for $t \in (t_k, t_{k+1} \wedge \tau_k)$, the definition of τ_k implies $\pi(\bar{X}(t)) \subset A$. It follows from (4.12), (4.13), and Proposition 4.6 that,

$$\frac{d}{dt}\bar{X}(t) = \sum_{i \in \pi(\bar{X}(t))} \beta_k^{(i)} h_i(t) = \sum_{j=1}^{d} \bar{\lambda}_{j,k} e_j - \sum_{i \in \pi(\bar{X}(t)),\ i \neq 0} \bar{\mu}_{i,k} h_i(t) e_i. \tag{4.17}$$

To verify Step 1, we assume that A_k is a strict subset of A and a strict subset of $\{1, 2, \ldots, d\}$ as well, since otherwise the claim is trivial. It follows from the definitions of π, A, and A_k that for every $i \in A_k$ and $j \in A \setminus A_k$ such that $j \neq 0$,

$$c_i \left(\psi^*(t_k)\right)_i - c_j \left(\psi^*(t_k)\right)_j = 0$$

and for $t \in (t_k, t_{k+1})$

$$c_i \left(\psi^*(t)\right)_i - c_j \left(\psi^*(t)\right)_j > 0.$$

Since $\dot{\psi}^*(t) \equiv \beta_k$ for $t \in (t_k, t_{k+1})$, the last display and (4.9) yield

$$0 < c_i(\beta_k)_i - c_j(\beta_k)_j = c_i(\bar{\lambda}_{i,k} - \rho_k^{(i)} \bar{\mu}_{i,k}) - c_j \bar{\lambda}_{j,k}. \tag{4.18}$$

Note that Step 1 amounts to claiming that

$$\pi(\bar{X}(t)) \text{ cannot be written as } B \cup C, \text{ where } B \text{ is a strict subset of } A_k \text{ and } C \subset A \setminus A_k \text{ is non-empty.} \tag{4.19}$$

Indeed, note that $\pi(\bar{X}(t))$ can always be written as $\pi(\bar{X}(t)) = \bar{B} \cup \bar{C}$ where $\bar{B} \subset A_k$ and $\bar{C} \cap A_k = \emptyset$. The sets $\bar{B}, \bar{C}$ are uniquely determined by

$$\bar{B} = \pi(\bar{X}(t)) \cap A_k, \ \bar{C} = \pi(\bar{X}(t)) \cap (A \setminus A_k).$$

Then Step 1 amounts to claiming that either $\bar{B} = A_k$, or $\bar{B}$ is a strict subset of A_k and $\bar{C} = \emptyset$. This is clearly equivalent to (4.19).

We will prove (4.19) by contradiction and assume that there exists an $s \in (t_k, t_{k+1} \wedge \tau_k)$ such that $\pi(\bar{X}(s))$ can be written as such a union $B \cup C$. Note that C must contain at least one non-zero element, since otherwise $C = \{0\}$ and by Remark 2.2 $B \cup C = \{0, 1, 2, \ldots, d\}$ or $B = \{1, 2, \ldots, d\}$, which contradicts the assumption that B is a strict subset of A_k. Let

$$\bar{t} \doteq \sup \left\{ t \leq s : \pi(\bar{X}(t)) \cap (A_k \setminus B) \neq \emptyset \right\}.$$

We claim that $\bar{t} \in [t_k, s)$ and $\pi(\bar{X}(\bar{t})) \cap (A_k \setminus B) \neq \emptyset$. Indeed, $\bar{t} \geq t_k$ is trivial since $\pi(\bar{X}(t_k)) = A \supset A_k$ and $A_k \setminus B$ is non-empty. Thanks to the upper semicontinuity of π and the continuity of $\bar{X}$, there exists a small neighborhood of s such that for any t in this small neighborhood $\pi(\bar{X}(t)) \subset \pi(\bar{X}(s)) = B \cup C$. It follows readily that $\bar{t} < s$. An analogous use of upper semicontinuity shows that $\pi(\bar{X}(\bar{t})) \cap (A_k \setminus B) \neq \emptyset$.

Fix $i \in \pi(\bar{X}(\bar{t})) \cap (A_k \setminus B)$. Note that $i \notin \pi(\bar{X}(t))$ for $t \in (\bar{t}, s)$. Thus, for every $j \in C$ such that $j \neq 0$ and $t \in (\bar{t}, s)$, it follows from equations (4.17) and (4.18) that

$$\begin{aligned}
\frac{d}{dt}\left[c_i((\bar{X}(t))_i - c_j((\bar{X}(t))_j \right] &= c_i \bar{\lambda}_{i,k} - c_j(\bar{\lambda}_{j,k} - \bar{\mu}_{j,k} h_j(t)) \\
&= [c_i(\bar{\lambda}_{i,k} - \rho_k^{(i)} \bar{\mu}_{i,k}) - c_j \bar{\lambda}_{j,k}] \\
&\quad + [c_i \rho_k^{(i)} \bar{\mu}_{i,k} + c_j \bar{\mu}_{j,k} h_j(t)] \\
&> 0.
\end{aligned}$$

Therefore,

$$0 \geq c_i((\bar{X}(s))_i - c_j((\bar{X}(s))_j > c_i((\bar{X}(\bar{t}))_i - c_j((\bar{X}(\bar{t}))_j \geq 0,$$

here the first inequality holds since $j \in C \subset \pi(\bar{X}(s))$ and the last inequality is due to $i \in \pi(\bar{X}(\bar{t}))$. This is a contradiction. Thus (4.19) holds and Step 1 is completed.

We now show Step 2. Thanks to Step 1, it suffices to show that $|\pi(\bar{X}(t))| \geq |A_k|$ for $t \in (t_k, t_{k+1} \wedge \tau_k)$. Let $t^* \in (t_k, t_{k+1} \wedge \tau_k)$ be such that

$$|\pi(\bar{X}(t^*))| = \min\{|\pi(\bar{X}(t))| : t \in (t_k, t_{k+1} \wedge \tau_k)\},$$

and assume that $|\pi(\bar{X}(t^*))| < |A_k|$ for the purpose of getting a contradiction. It follows from Step 1 that $B \doteq \pi(\bar{X}(t^*))$ is a strict subset of A_k. Note that $B \subset \{1, 2, \ldots, d\}$ (otherwise $B = \{0, 1, \ldots, d\}$ by Remark 2.2, a clear contradiction). Thanks to the upper semicontinuity of π and the continuity of $\bar{X}$, there exists an open interval containing t^* such that $\pi(\bar{X}(s)) \subset \pi(\bar{X}(t^*))$ for all s in this interval. We will assume (a, b) to be the largest of such intervals. By the definition of t^*, $\pi(\bar{X}(s)) = \pi(\bar{X}(t^*)) = B$ for every $s \in (a, b) \cap (t_k, t_{k+1} \wedge \tau_k)$. Obviously $a \geq t_k$ and $\pi(\bar{X}(a))$ contains B as a strict subset by the upper semicontinuity of π. Therefore, for $s \in (a, b)$, $c_i(\bar{X}(s))_i = c_j(\bar{X}(s))_j$ if $i, j \in B$. However, (4.17) yields

$$\frac{d}{ds}(\bar{X}(s))_j = \begin{cases} \bar{\lambda}_{j,k} & \text{if } j \in A_k \setminus B, \ j \neq 0, \\ \bar{\lambda}_{j,k} - h_j(s)\bar{\mu}_{j,k} & \text{if } j \in B. \end{cases} \tag{4.20}$$

It follows that

$$c_i(\bar{\lambda}_{i,k} - h_i(s)\bar{\mu}_{i,k}) = c_j(\bar{\lambda}_{j,k} - h_j(s)\bar{\mu}_{j,k}) \text{ for } \ i, \, j \in B. \tag{4.21}$$

In addition, thanks to Proposition 4.6(b),

$$\sum_{i \in B} h_i(s) = 1. \tag{4.22}$$

Solving the system of equations (4.21)-(4.22), we obtain the unique solution

$$h_i(s) = \frac{\bar{\lambda}_{i,k}}{\bar{\mu}_{i,k}} - \frac{1}{c_i \bar{\mu}_{i,k}} \cdot \left(\sum_{j \in B} \frac{1}{c_j \bar{\mu}_{j,k}} \right)^{-1} \cdot \left(-1 + \sum_{j \in B} \frac{\bar{\lambda}_{j,k}}{\bar{\mu}_{j,k}} \right).$$

On the other hand, since $\pi(\psi^*(s)) \equiv A_k \supset B$ and $\dot{\psi}^*(s) \equiv \beta_k$ for $s \in (a, b)$, we have $c_i(\beta_k)_i = c_j(\beta_k)_j$ for every $i, j \in B$. Invoking (4.7) and (4.9), we arrive at

$$c_i(\bar{\lambda}_{i,k} - \rho_k^{(i)}\bar{\mu}_{i,k}) = c_j(\bar{\lambda}_{j,k} - \rho_k^{(j)}\bar{\mu}_{j,k}) \text{ for } \ i, \, j \in B, \tag{4.23}$$

and that

$$b \doteq \sum_{i \in B} \rho_k^{(i)} = 1 - \sum_{i \in A_k \setminus B} \rho_k^{(i)} < 1. \tag{4.24}$$

Similarly, one can solve (4.23)-(4.24) to uniquely determine $\{\rho_k^{(i)}\}$ for $i \in B$:

$$\rho_k^{(i)} = \frac{\bar{\lambda}_{i,k}}{\bar{\mu}_{i,k}} - \frac{1}{c_i \bar{\mu}_{i,k}} \cdot \left(\sum_{j \in B} \frac{1}{c_j \bar{\mu}_{j,k}} \right)^{-1} \cdot \left(-b + \sum_{j \in B} \frac{\bar{\lambda}_{j,k}}{\bar{\mu}_{j,k}} \right).$$

Since $b < 1$ it follows that $h_i(s) > \rho_k^{(i)}$ for all $i \in B$ and $s \in (a, b)$.

Assume for now that there exists $j \in A_k \setminus B$ such that $j \neq 0$. Then for $i \in B$ and $s \in (a,b)$ we have, thanks to (4.20) and (4.23),

$$\begin{aligned} \frac{d}{ds}\left[c_i(\bar{X}(s))_i - c_j(\bar{X}(s))_j\right] &= c_i\left[\bar{\lambda}_{i,k} - h_i(s)\bar{\mu}_{i,k}\right] - c_j\bar{\lambda}_{j,k} \\ &< c_i\left[\bar{\lambda}_{i,k} - \rho_k^{(i)}\bar{\mu}_{i,k}\right] - c_j\bar{\lambda}_{j,k} \\ &= -c_j\rho_k^{(j)}\bar{\mu}_{j,k} \\ &< 0. \end{aligned} \tag{4.25}$$

However, since $\pi(\bar{X}(a)) \neq B$, Step 1 implies that $\pi(\bar{X}(a)) \cap (A_k \setminus B) \neq \emptyset$. Pick any j in this set. It follows from (4.25) that for any $i \in B$,

$$0 \geq c_i(\bar{X}(a))_i - c_j(\bar{X}(a))_j > c_i(\bar{X}(t^*))_i - c_j(\bar{X}(t^*))_j.$$

This contradicts that $i \in B = \pi(\bar{X}(t^*))$. Therefore, we have $|\pi(\bar{X}(t))| \geq |A_k|$. It follows from Step 1 that $A_k \subset \pi(\bar{X}(t))$, for $t \in (t_k, t_{k+1} \wedge \tau_k)$.

It remains to consider the case where $A_k \setminus B = \{0\}$. In this case we necessarily have $B = \{1,2,\ldots,d\}$ and $A_k = \{0,1,\ldots,d\} = A$. It follows that $\psi^*(t) \equiv 0$ for $t \in (t_k, t_{k+1})$ and thus $\beta_k = 0$. Then for any $i \in B$ and $s \in (a,b)$ we have, thanks to (4.20) and (4.7),

$$\frac{d}{ds}\left[c_i(\bar{X}(s))_i\right] = c_i\left[\bar{\lambda}_{i,k} - h_i(s)\bar{\mu}_{i,k}\right] < c_i\left[\bar{\lambda}_{i,k} - \rho_k^{(i)}\bar{\mu}_{i,k}\right] = (\beta_k)_i = 0.$$

However, since $\pi(X(a))$ contains B as a strict subset, we must have $\pi(X(a)) = \{0,1,\ldots,d\}$ or $X(a) = 0$. It follows then

$$0 = c_i(\bar{X}(a))_i > c_i(\bar{X}(t^*))_i.$$

This contradicts the non-negativity of $\bar{X}$, and again we have $|\pi(\bar{X}(t))| \geq |A_k|$. It follows from Step 1 that $A_k \subset \pi(\bar{X}(t))$, for $t \in (t_k, t_{k+1} \wedge \tau_k)$.

If $A = A_k$ then we finish the proof of Step 2 since $\pi(\bar{X}(t)) \subset A$. Consider the case when A_k is a strict subset of A, which in turn implies that $A_k \subset \{1,2,\ldots,d\}$. Following an argument analogous to that leading to equations (4.21)–(4.22) and (4.23)–(4.24), we have, for every $i,j \in A_k$ and almost every $t \in (t_k, t_{k+1} \wedge \tau_k)$,

$$c_i(\bar{\lambda}_{i,k} - h_i(t)\bar{\mu}_{i,k}) = c_j(\bar{\lambda}_{j,k} - h_j(t)\bar{\mu}_{j,k}), \ \sum_{j \in A_k} h_j(t) \leq 1. \tag{4.26}$$

$$c_i(\bar{\lambda}_{i,k} - \rho_k^{(i)}\bar{\mu}_{i,k}) = c_j(\bar{\lambda}_{j,k} - \rho_k^{(j)}\bar{\mu}_{j,k}), \ \sum_{i \in A_k} \rho_k^{(i)} = 1. \tag{4.27}$$

One can solve these two equations like before to obtain $h_i(t) \leq \rho_k^{(i)}$ for every $i \in A_k$ and $t \in (t_k, t_{k+1} \wedge \tau_k)$, whence

$$\frac{d}{dt}\left[c_i(\bar{X}(t))_i - c_i(\psi^*(t))_i\right] = c_i(\bar{\lambda}_{i,k} - h_i(t)\bar{\mu}_{i,k}) - c_i(\bar{\lambda}_{i,k} - \rho_k^{(i)}\bar{\mu}_{i,k}) \geq 0.$$

It follows that for any $i \in A_k$,

$$c_i(\bar{X}(t))_i - c_i(\psi^*(t))_i \geq c_i(\bar{X}(t_k))_i - c_i(\psi^*(t_k))_i = 0, \tag{4.28}$$

or $(\bar{X}(t))_i \geq (\psi^*(t))_i$. However, since $\pi((\psi^*(t)) \equiv A_k \subset \{1,2,\ldots,d\}$, we must have $\bar{X}(t) \neq 0$ and thus $\pi(\bar{X}(t)) \subset \{1,2,\ldots,d\}$.

On the other hand, for any $j \in A \backslash A_k$ and $j \neq 0$, since $(\dot{\psi}^*(t))_j = (\beta_k)_j = \bar{\lambda}_{j,k}$ thanks to (4.9), we have

$$\frac{d}{dt}\left[c_j(\bar{X}(t))_j - c_j(\psi^*(t))_j\right] = c_j(\bar{\lambda}_{j,k} - h_j(t)\bar{\mu}_{j,k}) - c_j\bar{\lambda}_{j,k} \leq 0.$$

It follows, since $\bar{X}(t_k) = \psi^*(t_k)$ and (4.28), that for every $t \in (t_k, t_{k+1} \wedge \tau_k)$ and $i \in A_k$,

$$c_j(\bar{X}(t))_j - c_j(\psi^*(t))_j \leq 0 \leq c_i(\bar{X}(t))_i - c_i(\psi^*(t))_i.$$

But $(\psi^*(t))_i > (\psi^*(t))_j$ by definition of A_k, and therefore $(\bar{X}(t))_i > (\bar{X}(t))_j$. It follows that $j \notin \pi(\bar{X}(t))$. Therefore $\pi(\bar{X}(t)) \equiv A_k$ and we finish Step 2.

The proof of Step 3 is simple. Assume first $A_k \subset \{1,2,\ldots,d\}$. Since $\pi(\bar{X}(t)) = A_k$ we know that $\{h_i(t)\}$ satisfies the equation (4.26) except now $\sum_{j\in A_k} h_j(t) = 1$. Compared with equation (4.27), it follows easily that $h_j(t) = \rho_k^{(j)}$. Note that equations (4.9) and (4.12) imply

$$\sum_{i\in A_k} \beta_k^{(i)}\rho_k^{(i)} = \beta_k,$$

whence

$$\frac{d}{dt}\bar{X}(t) = \sum_{i\in A_k} \beta_k^{(i)} h_i(t) = \sum_{i\in A_k} \beta_k^{(i)}\rho_k^{(i)} = \beta_k = \dot{\psi}^*(t),$$

which implies $\bar{X}(t) = \dot{\psi}^*(t)$ for all $t \in (t_k, t_{k+1} \wedge \tau_k)$.

For the case where $A_k = \{0,1,\ldots,d\}$, we must have $\bar{X}(t) = \psi^*(t) = 0$. It is not difficult to see that equations (4.26) and (4.27) reduce to

$$c_i(\bar{\lambda}_{i,k} - h_i(t)\bar{\mu}_{i,k}) = 0 = c_i(\bar{\lambda}_{i,k} - \rho_k^{(i)}\bar{\mu}_{i,k}),\ i = 1,2,\ldots,d.$$

Thus $h_i(t) = \rho_k^{(i)}$ for all $i = 1,2,\ldots,d$, and also $h_0(t) = \rho_k^{(0)}$ since $\sum_{i=0}^d h_i(t) = \sum_{i=0}^d \rho_k^{(i)} = 1$.

$\square$

4.7. Analysis of the cost

In this section, we prove the Laplace lower bound, or inequality (4.4).

Proof. Thanks to (4.11) and (4.10),

$$\begin{aligned}
&\lim_{n\to\infty} E_{x_n}\left[\int_{t_k}^{t_{k+1}} \sum_{v\in\Theta} r(\bar{X}^n(t);v)\ell\left(\frac{\bar{r}(\bar{X}^n(t),t;v)}{r(\bar{X}^n(t);v)}\right)dt\right] \\
&= \lim_{n\to\infty} E_{x_n}\left[\int_{t_k}^{t_{k+1}} \sum_{i\in A_k, i\neq 0} \mu_i\ell\left(\frac{\bar{\mu}_{i,k}}{\mu_i}\right)\gamma_i^n(dt) + \sum_{j=1}^{d}\lambda_j\ell\left(\frac{\bar{\lambda}_{j,k}}{\lambda_j}\right)dt\right] \\
&= \int_{t_k}^{t_{k+1}}\left[\sum_{i\in A_k, i\neq 0} \mu_i\ell\left(\frac{\bar{\mu}_{i,k}}{\mu_i}\right)h_i(t) + \sum_{j=1}^{d}\lambda_j\ell\left(\frac{\bar{\lambda}_{j,k}}{\lambda_j}\right)\right]dt \\
&= (t_{k+1}-t_k)\left[\sum_{i\in A_k, i\neq 0} \mu_i\ell\left(\frac{\bar{\mu}_{i,k}}{\mu_i}\right)\rho_k^{(i)} + \sum_{j=1}^{d}\lambda_j\ell\left(\frac{\bar{\lambda}_{j,k}}{\lambda_j}\right)\right] \\
&\leq (t_{k+1}-t_k)\cdot[L^{A_k}(\beta_k)+\varepsilon] \\
&= \int_{t_k}^{t_{k+1}} L(\psi^*(t),\dot{\psi}^*(t))dt + (t_{k+1}-t_k)\varepsilon.
\end{aligned}$$

Summing over k and observing

$$\lim_{n\to\infty} E_{x_n} h(\bar{X}^n) = E_x h(\bar{X}) = h(\psi^*),$$

(4.4) follows readily. This completes the proof of inequality (4.4) and also the proof of Theorem 3.1. □

5. Summary

In this paper we analyze the large deviation properties for a class of systems with discontinuous dynamics, namely, the WSLQ policy for a network with multiple queues and a single server. A key observation that allows us to derive explicitly the large deviation rate function on path space is that the stability condition about the interface is automatically implied in the formulation of a general large deviation upper bound.

This is not an unprecedented situation, and indeed analogous results are proved in [2] for processes that model Jackson networks and head-of-the-line processor sharing. These problems also feature discontinuous statistics, though the discontinuities appear on the boundary of the state space rather than the interior, and whence alternative techniques based on Skorokhod mapping can be used. With the inclusion of the WSLQ policy there are now a number of physically meaningful models with discontinuous statistics for which the upper bound of [6] is tight. An intriguing question that will be investigated elsewhere is whether there is a common property of these models that can be easily identified and recognized.

Appendix A. Proof of Lemma 4.1.

Before getting into the details of the construction, it is worth pointing out that the main difficulty in the proof is dealing with the situation where the function ψ may spiral into (or away from) a low dimensional interface while hitting higher dimensional interfaces infinitely many times in the process. More precisely, there can be a time t and a sequence $t_n \uparrow t$ (or $t_n \downarrow t$) such that $|\pi(\psi(t_n))| < |\pi(\psi(t))|$, but with the sets $\{\pi(\psi(t_n))\}$ different for successive n. The problem is how to approximate ψ on a small neighborhood of t with a small cost, which is made difficult by the fact that the domain of finiteness of $L(x, \cdot)$ is not continuous in x.

Proof. Throughout this section we assume that ψ is Lipschitz continuous. This is without loss of generality, since for a given continuous function ψ and any $\delta > 0$, there exists a Lipschitz continuous function ζ such that $\|\zeta - \psi\|_\infty \leq \delta$ and $I_x(\zeta) \leq I_x(\psi) + \delta$. The proof of this claim is based on a time-rescaling argument very much analogous to that of [4, Lemma 6.5.3], and we omit the details.

A.1. Dividing the time interval

The approximation requires a suitable division of the time interval $[0, 1]$. The following results are useful in proving the main result of this section, namely, Lemma A.4. Note that whenever we say two intervals are "non-overlapping", it means that the two intervals cannot have common interiors but may have a same endpoint.

Lemma A.1. *Consider a non-empty closed interval $[a, b]$ and assume*

$$k \doteq \max\{|\pi(\psi(t))| : t \in [a, b]\} \geq 2.$$

Then for arbitrary $\sigma > 0$, there exists a finite collection of non-overlapping intervals $\{[\alpha_j, \beta_j]\}$ such that

1. *$\alpha_j, \beta_j \in [a, b]$ and $0 \leq \beta_j - \alpha_j \leq \sigma$ for every j.*
2. *$|\pi(\psi(t))| \leq k - 1$ for every $t \in [a, b] \setminus \cup_j [\alpha_j, \beta_j]$.*

Lemma A.2. *Consider a non-empty interval (a, b) such that for some $k \geq 2$, $|\pi(\psi(t))| \leq k$ for $t \in (a, b)$ and $|\pi(\psi(a))| \wedge |\pi(\psi(b))| \geq k + 1$. Then for arbitrary $\sigma > 0$ and $\varepsilon > 0$, there exist a finite collection of non-overlapping intervals $\{[\alpha_j, \beta_j]\}$ and $\varepsilon_1, \varepsilon_2 \in [0, \varepsilon)$ such that*

1. *$\alpha_j, \beta_j \in [a + \varepsilon_1, b - \varepsilon_2]$ and $0 \leq \beta_j - \alpha_j \leq \sigma$ for each j,*
2. *for $t \in (\alpha_j, \beta_j)$, $\pi(\psi(t)) \subset \pi(\psi(\alpha_j)) = \pi(\psi(\beta_j))$,*
3. *$|\pi(\psi(\alpha_j))| = |\pi(\psi(\beta_j))| = k$ for each j,*
4. *for any $t \in (a, a + \varepsilon_1]$, $\pi(\psi(t))$ is a strict subset of $\pi(\psi(a))$ and $|\pi(\psi(t))| \leq k \leq |\pi(\psi(a + \varepsilon_1))|$,*
5. *for any $t \in [b - \varepsilon_2, b)$, $\pi(\psi(t))$ is a strict subset of $\pi(\psi(b))$ and $|\pi(\psi(t))| \leq k \leq |\pi(\psi(b - \varepsilon_2))|$,*
6. *$|\pi(\psi(t))| \leq k - 1$ for every $t \notin \cup_j [\alpha_j, \beta_j] \cup (a, a + \varepsilon_1] \cup [b - \varepsilon_2, b)$.*

Lemma A.3. *Consider a non-empty interval $[a, b)$ such that for some $2 \leq k \leq d$, $|\pi(\psi(t))| \leq k$ for $t \in [a, b)$ and $|\pi(\psi(b))| \geq k+1$. Then for arbitrary $\sigma > 0$ and $\varepsilon > 0$, there exist a finite collection of non-overlapping intervals $\{[\alpha_j, \beta_j]\}$ and $\varepsilon_2 \in [0, \varepsilon)$ such that*

1. *$\alpha_j, \beta_j \in [a, b-\varepsilon_2]$ and $0 \leq \beta_j - \alpha_j \leq \sigma$ for each j,*
2. *for $t \in (\alpha_j, \beta_j)$, $\pi(\psi(t)) \subset \pi(\psi(\alpha_j)) = \pi(\psi(\beta_j))$,*
3. *$|\pi(\psi(\alpha_j))| = |\pi(\psi(\beta_j))| = k$ for each j,*
4. *for every $t \in [b-\varepsilon_2, b)$, $\pi(\psi(t))$ is a strict subset of $\pi(\psi(b))$ and $|\pi(\psi(t))| \leq k \leq |\pi(\psi(b-\varepsilon_2))|$,*
5. *$|\pi(\psi(t))| \leq k-1$ for every $t \notin \cup_j[\alpha_j, \beta_j] \cup [b-\varepsilon_2, b)$.*

Symmetric results hold for a non-empty interval $(a, b]$ such that for some $k \geq 2$, $|\pi(\psi(t))| \leq k$ for $t \in (a, b]$ and $|\pi(\psi(a))| \geq k+1$.

We will only provide the details of the proof for Lemma A.2. The proofs for Lemma A.1 and Lemma A.3 are very similar (indeed, simpler versions of the proof for Lemma A.2), and thus omitted.

Proof. If the set $\{t \in (a, b) : |\pi(\psi(t))| = k\}$ is empty, then the claim holds trivially since we can let $\varepsilon_1 = \varepsilon_2 = 0$ and $\{[\alpha_j, \beta_j]\} \doteq \emptyset$. Assume from now on that this set is non-empty. We first define ε_1 and ε_2. Let

$$\bar{a} \doteq \inf\{t \in (a, b) : |\pi(\psi(t))| = k\}, \ \bar{b} \doteq \sup\{t \in (a, b) : |\pi(\psi(t))| = k\}.$$

If $\bar{a} = a$, then the upper semicontinuity of π implies that there exists $0 < \varepsilon_1 < \varepsilon$ such that $|\pi(\psi(a+\varepsilon_1))| = k$ and $\pi(\psi(t))$ is a subset (whence a strict subset) of $\pi(\psi(a))$ for every $t \in (a, a+\varepsilon_1]$. If $\bar{a} > a$ then we let $\varepsilon_1 = 0$. ε_2 is defined in a completely analogous fashion. It is easy to see that parts 4 and 5 of the claim are satisfied. We will define α_j and β_j recursively.

1. Let $\alpha_1 \doteq a + \varepsilon_1$ if $\bar{a} = a$ and $\alpha_1 \doteq \bar{a}$ if $\bar{a} > a$. Clearly, in either case $\alpha_1 \in [a+\varepsilon_1, b-\varepsilon_2]$. Moreover, we have $|\pi(\psi(\alpha_1))| = k$. Indeed, when $\bar{a} = a$ it follows from the definition, and when $\bar{a} > a$, it follows from a simple argument by contradiction, thanks to the upper semicontinuity of π and the assumption that $|\pi(\psi(t))| \leq k$ for every $t \in (a, b)$.
2. Suppose now $\alpha_j \in [a+\varepsilon_1, b-\varepsilon_2]$ is given such that $|\pi(\psi(\alpha_j))| = k$. We wish to define β_j. Let
$$\begin{aligned} t_j &\doteq \inf\{t \in (\alpha_j, b) : \pi(\psi(t)) \not\subset \pi(\psi(\alpha_j))\}, \\ s_j &\doteq \sup\{t \in [\alpha_j, t_j \wedge (\alpha_j+\sigma)) : \pi(\psi(t)) = \pi(\psi(\alpha_j))\}, \\ \beta_j &\doteq s_j \wedge (b-\varepsilon_2). \end{aligned}$$
It is easy to check that $\pi(\psi(s_j)) = \pi(\psi(\alpha_j))$. We claim that $\pi(\psi(\beta_j)) = \pi(\psi(\alpha_j))$. This is trivial when $\beta_j = s_j$. Assume for now that $\beta_j < s_j$. Then we must have $s_j > b-\varepsilon_2$ and $\beta_j = b-\varepsilon_2$. This could happen only if $\varepsilon_2 > 0$, in which case $|\pi(\psi(b-\varepsilon_2))| = k$, or $|\pi(\psi(\beta_j))| = k$. But $\pi(\psi(\beta_j)) \subset \pi(\psi(\alpha_j))$

since $\beta_j < s_j \le t_j$. Thus we must have $\pi(\psi(\alpha_j)) = \pi(\psi(\beta_j))$. It is clear now that parts 1, 2, and 3 of the lemma holds.

3. If $\beta_j = \bar{b}$ (when $\bar{b} < b$) or $\beta_j = b - \varepsilon_2$ (when $\bar{b} = b$), we stop. Otherwise, define $\alpha_{j+1} \doteq \inf\{t \in (\beta_j, b) : |\pi(\psi(t))| = k\}$. Then $\alpha_{j+1} \in [a + \varepsilon_1, b - \varepsilon_2]$ and $|\pi(\psi(\alpha_{j+1}))| = k$. Now repeat step 2.

Note that part 6 holds by the construction. It only remains to show that the construction will terminate in finitely many steps. Observe that $\alpha_{j+1} \ge t_j \wedge (\alpha_j + \sigma)$, since by definition, for every $t \in (\beta_j, t_j \wedge (\alpha_j + \sigma))$, $\pi(\psi(t))$ is a strict subset of $\pi(\psi(\alpha_j))$ and whence $|\pi(\psi(t))| < k$. Therefore, $\alpha_{j+1} - \alpha_j \ge (t_j - \alpha_j) \wedge \sigma$, and it suffices to show that $t_j - \alpha_j$ is uniformly bounded away from 0.

We should first consider the case $k \le d-1$. Observe that in the construction we indeed have $\alpha_j, \beta_j \in \bar{I} \doteq [(a + \varepsilon_1) \vee \bar{a}, (b - \varepsilon_2) \wedge \bar{b}] \subset (a, b)$. Define

$$c \doteq \inf \left\{ \max_{i=1,\ldots,d} c_i(\psi(t))_i - \max_{i \notin \pi(\psi(t))} c_i(\psi(t))_i : t \in \bar{I}, |\pi(\psi(t))| = k \right\}.$$

We claim that $c > 0$. If this is not the case, there exists a sequence of $\{t_n\} \subset \bar{I}$ such that

$$\max_{i=1,\ldots,d} c_i(\psi(t_n))_i - \max_{i \notin \pi(\psi(t_n))} c_i(\psi(t_n))_i \downarrow 0.$$

One can find a subsequence of $\{t_n\}$, still denoted by $\{t_n\}$, such that $t_n \to t^* \in I$ and $\pi(\psi(t_n)) \equiv B$ for some $B \subset \{1, 2, \ldots, d\}$ with $|B| = k$. Thanks to the upper semicontinuity of π, $B \subset \pi(\psi(t^*))$. But $|\pi(\psi(t^*))| \le k$, thus $B = \pi(\psi(t^*))$. However, for every $j \in B$,

$$\begin{aligned} 0 &= \lim_n \left[\max_{i=1,\ldots,d} c_i(\psi(t_n))_i - \max_{i \notin \pi(\psi(t_n))} c_i(\psi(t_n))_i \right] \\ &= \lim_n \left[c_j(\psi(t_n))_j - \max_{i \notin B} c_i(\psi(t_n))_i \right] \\ &= c_j(\psi(t^*))_j - \max_{i \notin B} c_i(\psi(t^*))_i \\ &> 0, \end{aligned}$$

a contradiction. Therefore $c > 0$.

Now, by the definition of t_j, there exists $l \notin \pi(\psi(\alpha_j))$ such that $l \in \pi(\psi(t_j))$. Therefore, for every $i \in \pi(\psi(\alpha_j))$,

$$c_i(\psi(t_j))_i - c_l(\psi(t_j))_l \le 0.$$

But $c_i(\psi(\alpha_j))_i - c_l(\psi(\alpha_j))_l \ge c$ and ψ is Lipschitz continuous. It follows readily that $t_j - \alpha_j$ is uniformly bounded away from 0.

The case $k = d$ can be treated in a completely analogous fashion with

$$c \doteq \inf \left\{ \max_{i=1,\ldots,d} c_i(\psi(t))_i : t \in \bar{I}, |\pi(\psi(t))| = d \right\}.$$

We omit the details. $\square$

The next lemma is the main result of this section, from which one can construct the approximating function in $\mathcal{N}$. The intervals $\{E_i\}$ in the lemma are indeed introduced to take care of the "spiraling" problem.

Lemma A.4. *Consider the interval* $[0,1]$. *Given* $\sigma > 0$, *there exist two finite collections of non-overlapping intervals* $\{I_j\}$ *and* $\{E_i\}$ *such that*

1. $I_j \subset [0,1]$ *is of the form* $[a_j, b_j]$ *with* $0 \le b_j - a_j \le \sigma$ *for each* j,
2. $E_i \subset [0,1]$ *is either of the form* $[z_i, d_i)$ *or* $(z_i, d_i]$, *and* $\sum_i (d_i - z_i) \le \sigma$,
3. $|\pi(\psi(t))| = 1$ *for all* $t \notin (\cup_j I_j) \cup (\cup_i E_i)$,
4. *for every* j *and every* $t \in (a_j, b_j)$, $\pi(\psi(t)) \subset \pi(\psi(a_j)) = \pi(\psi(b_j))$,
5. *for every* j, $|\pi(\psi(a_j))| = |\pi(\psi(b_j))| \ge 2$,
6. *for all* $t \in E_i = (z_i, d_i]$, $\pi(\psi(t))$ *is a strict subset of* $\pi(\psi(z_i))$ *and* $|\pi(\psi(t))| \le |\pi(\psi(d_i))|$,
7. *for all* $t \in E_i = [z_i, d_i)$, $\pi(\psi(t))$ *is a strict subset of* $\pi(\psi(d_i))$ *and* $|\pi(\psi(t))| \le |\pi(\psi(z_i))|$.

Proof. Let $N \doteq \max\{|\pi(\psi(t))| : t \in [0,1]\}$. If $N = 1$ then the claim holds trivially. Assume from now on that $N \ge 2$. The construction can be done in N steps.

Step 1.. Apply Lemma A.1 to the interval $[0,1]$ with $k = N$. This will produce a collection of closed intervals $\{I_j^{(1)}\}$. Let $U_1 \doteq [0,1] \setminus \cup_j I_j^{(1)}$.

Step 2.. Note that U_1 is the union of finitely many non-overlapping intervals. These intervals are of two types. They are either open intervals that satisfy the conditions of Lemma A.2 with $k = N-1$, or intervals of type $[0,a)$ or $(b,1]$ that satisfy the conditions of Lemma A.3. Apply the corresponding lemma to these intervals to generate a collection of closed intervals $\{I_j^{(2)}\}$ and intervals $\{E_i^{(2)}\}$ of type $[a,b)$ or $(b,a]$. Since the number of intervals in $\{E_i^{(2)}\}$ is at most twice the number of intervals in U_1, one can choose ε in Lemma A.2 and Lemma A.3 so small that the total Lebesgue measure of $\cup_i E_i^{(2)}$ is bounded from above by σ/N. Define $U_2 \doteq U_1 \setminus ((\cup_j I_j^{(2)}) \cup (\cup_i E_i^{(2)}))$.

Step m.. For $3 \le m \le N-1$, the procedure of Step m is just like Step 2, except U_1 is replaced by U_{m-1}.

Step N.. Note that U_{N-1} consists of intervals on which $\pi(\psi(\cdot))$ is a singleton. We stop the construction.

Let $\{I_j\} \doteq \cup_l \{I_j^{(l)}\}$ and $\{E_i\} \doteq \cup_l \{E_j^{(l)}\}$, and it is not difficult to see that $\{I_j\}$ and $\{E_i\}$ have the required properties. □

A.2. Construction of the approximating function

In this section we construct an approximating function of ψ parameterized by σ. Note that the set $[0,1] \setminus ((\cup_j I_j) \cup (\cup_i E_i))$ is the union of finitely many non-overlapping intervals, and we will denote these intervals by $\{G_k\}$. Without loss of generality assume that the length of each interval G_k is bounded above by σ (if necessary, we can divide G_k into the union of several smaller intervals each with a length less than σ).

Define a piece-wise constant function u^σ on interval $[0,1]$ as follows. Given any interval $D \in \{I_j\} \cup \{E_i\} \cup \{G_k\}$ with positive length, define for every $t \in D$,

$$u^\sigma(t) \doteq \frac{1}{\text{length of } D} \int_D \dot{\psi}(s)\, ds.$$

Then the candidate approximating function will be

$$\psi^\sigma(t) \doteq \psi(0) + \int_0^t u^\sigma(s) ds.$$

Note that $\psi^\sigma(t)$ coincides with $\psi(t)$ at those t that are end points of intervals $\{I_j\}$, $\{E_i\}$, and $\{G_k\}$. Furthermore, ψ^σ is affine on each of these intervals.

We claim that $\psi^\sigma \in \mathcal{N}$. All we need is to show is that $\pi(\psi^\sigma(t))$ remain unchanged in the interior of each interval. This is immediate from the following result, whose proof is simple and straightforward, and thus omitted.

Lemma A.5. *Let ϕ be an affine function on interval $[a,b]$. If $\pi(\phi(a)) \cap \pi(\phi(b)) \neq \emptyset$, then for every $t \in (a,b)$ we have $\pi(\phi(t)) = \pi(\phi(a)) \cap \pi(\phi(b))$.*

A.3. The analysis of the rate function

Since the length of any interval in $\{I_j\}$, $\{E_i\}$, and $\{G_k\}$ is bounded from above by σ, $\lim_{\sigma \to 0} \|\psi^\sigma - \psi\|_\infty = 0$. Therefore it only remains to show that given $\delta > 0$ we have $I_x(\psi^\sigma) \leq I_x(\psi) + \delta$ for σ small enough. Clearly we can assume $I_x(\psi) < \infty$ hereafter.

The analysis for intervals in $\{I_j\}$ and $\{G_k\}$ is simple. Consider an interval $D \in \{I_j\} \cup \{G_k\}$, and denote the interior of D by (a,b) [D itself could be $[a,b]$, (a,b), $(a,b]$, or $[a,b)$]. Since $\psi(a) = \psi^\sigma(a)$ and $\psi(b) = \psi^\sigma(b)$ by construction, it follows from Lemma A.4 and Lemma A.5 that

$$\pi(\psi(t)) \subset \pi(\psi(a)) \cap \pi(\psi(b)) = \pi(\psi^\sigma(a)) \cap \pi(\psi^\sigma(b)) = \pi(\psi^\sigma(t)) \doteq B.$$

for every $t \in (a,b)$. Note that $L^B(\beta) \leq L^A(\beta)$ for every β whenever $A \subset B$, thanks to Lemma 4.3. Therefore,

$$\int_a^b L(\psi(t), \dot{\psi}(t))\, dt = \int_a^b L^{\pi(\psi(t))}(\dot{\psi}(t))\, dt \geq \int_a^b L^B(\dot{\psi}(t))\, dt.$$

Furthermore, for $t \in (a,b)$, the construction of ψ^σ implies that

$$\psi^\sigma(t) \equiv \frac{\psi(b) - \psi(a)}{b-a} \doteq v.$$

Thanks to the convexity of L^B and Jensen's inequality, we arrive at

$$\int_a^b L^B(\dot{\psi}(t))\, dt \geq (b-a) L^B(v) = \int_a^b L(\psi^\sigma(t), \psi^\sigma(t)) dt,$$

whence

$$\int_a^b L(\psi(t), \dot{\psi}(t))\, dt \geq \int_a^b L(\psi^\sigma(t), \psi^\sigma(t)) dt. \tag{A.1}$$

for any σ.

Now we turn to analyze the intervals in $\{E_i\}$. The next two lemmas claim that over any interval, say $D \in \{E_i\}$, $\dot{\psi}^\sigma$ will always be in the domain of finiteness of the local rate function L. The first lemma considers the case where $D = (a, b]$ and the second considers the case where D is of form $[a, b)$.

Lemma A.6. *Suppose that $I_x(\psi) < \infty$ and there exists an interval $(a, b] \subset [0, 1]$ such that $\pi(\psi(t))$ is a strict subset of $\pi(\psi(a))$ for $t \in (a, b]$. Then $(\psi(a))_j \le (\psi(b))_j$ for all $j \in \{1, \ldots, d\}$.*

Proof. For notational simplicity, we assume $c_1 = c_2 = \cdots = c_d$. The case of general $(c_1, \ldots, c_d)$ is shown in exactly the same fashion with $(\psi(t))_j$ replaced by $c_j(\psi(t))_j$, and whence omitted.

Let $B \doteq \pi(\psi(a))$. For $j \notin B$, we have $j \notin \pi(\psi(t))$ and thus $\dot{\psi}(t) \ge 0$ for almost every $t \in (a, b]$, thanks to Remark 4.5 and the assumption that $I_x(\psi)$ is finite. Therefore for $j \notin B$ the claim holds. Now we consider those $j \in B$. Fix arbitrarily $\varepsilon > 0$ and let

$$t^* \doteq \inf\{t > a : (\psi(t))_j < (\psi(a))_j - \varepsilon \text{ for some } j \in B\} \wedge b.$$

It suffices to show that $t^* = b$ always holds. Indeed, if this is the case, we have $(\psi(b))_j \ge (\psi(a))_j - \varepsilon$ for every $j \in B$. Since ε is arbitrary, we arrive at the desired inequality for $j \in B$.

We will argue by contradiction and assume $t^* < b$. It follows that $(\psi(t^*))_j \ge (\psi(a))_j - \varepsilon$ for every $j \in B$, and there exist $j^* \in B$ and a sequence $t_n \downarrow t^*$ such that

$$(\psi(t_n))_{j^*} < (\psi(a))_{j^*} - \varepsilon. \tag{A.2}$$

In particular, $(\psi(t^*))_{j^*} = (\psi(a))_{j^*} - \varepsilon$.

We claim that $(\psi(t^*))_j = (\psi(a))_j - \varepsilon$ holds for every $j \in B$. If this is not the case, then there exists (abusing the notation) $j \in B$ such that

$$(\psi(t^*))_j > (\psi(a))_j - \varepsilon.$$

Since $j, j^* \in B = \pi(\psi(a))$ we have $(\psi(a))_j = (\psi(a))_{j^*}$, and whence $(\psi(t^*))_j > (\psi(t^*))_{j^*}$. Therefore we can find a small interval, say $[t^*, t^* + \delta)$ such that for any t in this interval we have $(\psi(t))_j > (\psi(t))_{j^*}$, which in turn implies $j^* \notin \pi(\psi(t))$. Thanks to the finiteness of $I_x(\psi)$ and Remark 4.5, $(\dot{\psi}(t))_{j^*} \ge 0$ for almost every $t \in [t^*, t^* + \delta)$. In particular, $(\psi(t))_{j^*} \ge (\psi(t^*))_{j^*} = (\psi(a))_{j^*} - \varepsilon$ for all $t \in [t^*, t^* + \delta)$, which contradicts equation (A.2) for large n. Therefore $(\psi(t^*))_j = (\psi(a))_j - \varepsilon$ holds for every $j \in B$.

Since $B = \pi(\psi(a))$, it follows that $(\psi(t^*))_j$ takes the same value for every $j \in B$. This contradicts the assumption that $\pi(\psi(t^*))$ is a strict subset of B. We complete the proof. □

Lemma A.7. *Suppose that $I_x(\psi) < \infty$ and there exists an interval $[a, b) \subset [0, 1]$ such that $\pi(\psi(t))$ is a strict subset of $\pi(\psi(b))$ for every $t \in [a, b)$. Let $B \doteq \pi(\psi(a))$. If $|\pi(\psi(t))| \le |B|$ for every $t \in [a, b)$, then $(\psi(a))_j \le (\psi(b))_j$ for all $j \notin B$ and $j \ne 0$.*

Proof. As in the proof of Lemma A.6 we assume without loss of generality that $c_1 = c_2 = \cdots = c_d = 1$. Let $A \doteq \pi(\psi(b))$. For $j \notin A$, $(\dot{\psi}(t))_j \geq 0$ for almost every $t \in [a, b)$ since $j \notin \pi(\psi(t))$. Thus the claim holds for $j \notin A$. It remains to show for those $j \in A \setminus B$ such that $j \neq 0$. We will argue by contradiction and assume that there is a $j^* \in A \setminus B$ such that $(\psi(a))_{j^*} > (\psi(b))_{j^*}$.

For each $i \in B$ let $t_i \doteq \inf\{t > a : (\psi(t))_i \leq (\psi(a))_{j^*}\}$. Note that $(\psi(a))_i > (\psi(a))_{j^*}$ since $i \in B = \pi(\psi(a))$ and $j^* \notin B$, whence $t_i > a$. Similarly, since $j^* \in A = \pi(\psi(b))$, $(\psi(b))_i \leq (\psi(b))_{j^*} < (\psi(a))_{j^*}$, whence $t_i < b$. It follows that $(\psi(t_i))_i = (\psi(a))_{j^*}$.

We claim that $i \in \pi(\psi(t_i))$ for all $i \in B$. Otherwise, there exists a small positive number δ such that $i \notin \pi(\psi(t))$ for $t \in (t_i - \delta, t_i + \delta)$. Thanks to the assumption that $I_x(\psi)$ is finite, $(\dot{\psi}(t))_i \geq 0$ for almost every $t \in (t_i - \delta, t_i + \delta)$. In particular, $(\psi(t_i - \delta))_i \leq (\psi(t_i))_i = (\psi(a))_{j^*}$, which contradicts the definition of t_i. Therefore, $i \in \pi(\psi(t_i))$ for all $i \in B$. An immediate consequence is that for any $i, k \in B$, $(\psi(t_i))_k \leq (\psi(t_i))_i = (\psi(a))_{j^*}$, whence $t_k \leq t_i$ by definition. Therefore, for all $i, k \in B$, $t_i = t_k \doteq t^*$ and $B \subset \pi(\psi(t^*))$. However, since $|\pi(\psi(t^*))| \leq |B|$ by assumption, we have necessarily $B = \pi(\psi(t^*))$.

Since $j^* \notin B = \pi(\psi(t^*))$, we have $(\psi(t^*))_{j^*} < (\psi(a))_{j^*}$. Define

$$s \doteq \sup\{t \in [a, t^*) : (\psi(t))_{j^*} \geq (\psi(a))_{j^*}\} \wedge t^*.$$

Clearly $s \in [a, t^*)$ and $(\psi(s))_{j^*} = (\psi(a))_{j^*}$. By the definitions of s and t_i and that $t^* \equiv t_i$ for all $i \in B$, we have, for any $t \in (s, t^*)$ and any $i \in B$,

$$(\psi(t))_{j^*} < (\psi(a))_{j^*} < (\psi(t))_i.$$

Therefore, $j^* \notin \pi(\psi(t))$ and whence $(\dot{\psi}(t))_{j^*} \geq 0$ for almost every $t \in (s, t^*)$. In particular, $(\psi(a))_{j^*} = (\psi(s))_{j^*} \leq (\psi(t^*))_{j^*} < (\psi(a))_{j^*}$, a contradiction. We complete the proof. □

We are now ready to show that $I_x(\psi^\sigma) \leq I_x(\psi) + \delta$ for σ small enough. Let M be the Lipschitz constant for ψ, and define

$$C \doteq \max_{A \subset \{1,\ldots,d\}} \sup\{L^A(\beta) : \beta \in \mathrm{dom}(L^A), \|\beta\| \leq M\}.$$

That C is finite follows easily from Lemma 4.3 and the definition of ℓ.

We now analyze the intervals in $\{E_i\}$. If $E_i = [z_i, d_i)$, then by Lemma A.5 and Lemma A.4, $\pi(\psi^\sigma(t)) = \pi(\psi(z_i)) \doteq B$ for all $t \in (z_i, d_i)$. Since

$$\dot{\psi}^\sigma(t) \equiv \frac{\psi(d_i) - \psi(z_i)}{d_i - z_i} \doteq v,$$

it follows from Lemma A.7 and Remark 4.5 that $v \in \mathrm{dom}(L^B)$. Therefore

$$\int_{z_i}^{d_i} L(\psi^\sigma(t), \dot{\psi}^\sigma(t))\, dt = (d_i - z_i) L^B(v) \leq C(d_i - z_i). \tag{A.3}$$

The same inequality holds for the case $E_i = (z_i, d_i]$, where Lemma A.6 is invoked in place of Lemma A.7.

By (A.1), (A.3), the non-negativity of L, and Lemma A.4, we have

$$\begin{aligned}\int_0^1 L(\psi^\sigma(t),\dot{\psi}^\sigma(t))\,dt &\le \int_0^1 L(\psi(t),\dot{\psi}(t))\,dt + C\sum_i (d_i - z_i)\\ &\le \int_0^1 L(\psi(t),\dot{\psi}(t))\,dt + C\sigma.\end{aligned}$$

Choose $\sigma < \delta/C$ and we complete the proof. □

Acknowledgment. We thank the referee for correcting an error in the first version of this paper and for helpful suggestions.

References

[1] M. Alanyali and B. Hajek. On large deviations of Markov processes with discontinuous statistics. *Ann. Appl. Probab.*, 8:45–66, 1998.

[2] R. Atar and P. Dupuis. Large deviations and queueing networks: methods for rate function identification. *Stoch. Proc. and Their Appl.*, 84:255–296, 1999.

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

[4] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York, 1997.

[5] P. Dupuis and R.S. Ellis. Large deviations for Markov processes with discontinuous statistics, II: Random walks. *Probab. Th. Rel. Fields*, 91:153–194, 1992.

[6] P. Dupuis, R.S. Ellis, and A. Weiss. Large deviations for Markov processes with discontinuous statistics, I: General upper bounds. *Annals of Probability*, 19:1280–1297, 1991.

[7] R. Foley and D. McDonald. Join the shortest queue: stability and exact asymptotics. *Ann. Appl. Probab.*, 11:569–607, 2001.

[8] I. Ignatiouk-Robert. Large deviations for processes with discontinuous statistics. *Ann. Probab.*, 33:1479–1508, 2005.

[9] K. Majewski. Large deviations of the steady-state distribution of reflected processes with applications to queueing systems. *Queueing Systems*, 29:351–381, 1998.

[10] K. Ramanan and S. Stolyar. Largest weighted delay first scheduling: Large deviations and optimality. *The Annals of Applied Probab.*, 11:1–49, 2001.

[11] A. Shwartz and A. Weiss. *Large Deviations for Performance Analysis: Queues, Communication and Computing*. Chapman and Hall, New York, 1995.

[12] L. Ying, R. Srikant, A. Eryilmaz, and G.E. Dullerud. A large deviation analysis of scheduling in wireless networks. *Preprint*, 2005.

Paul Dupuis
Lefshetz Center for Dynamical Systems
Brown University
Providence, RI 02912
USA
e-mail: dupuis@dam.brown.edu

Kevin Leder
Lefshetz Center for Dynamical Systems
Brown University
Providence, RI 02912
USA
e-mail: kleder@dam.brown.edu

Hui Wang
Lefshetz Center for Dynamical Systems
Brown University
Providence, RI 02912
USA
e-mail: huiwang@dam.brown.edu