MATHEMATICS OF OPERATIONS RESEARCH
Vol. 30, No. 2, May 2005, pp. 281–310
ISSN 0364-765X | EISSN 1526-5471 | 05 | 3002 | 0281
DOI 10.1287/moor.1040.0121

Large Deviations with Diminishing Rates

Adam Shwartz
Department of Electrical Engineering, Technion, Israel Institute of Technology,
Haifa 32000, Israel, adam@ee.technion.ac.il

Alan Weiss
Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, apdoo@research.bell-labs.com

The theory of large deviations for jump Markov processes has been generally proved only when jump rates are bounded below, away from zero (Dupuis and Ellis, 1995, The large deviations principle for a general class of queueing systems I. *Trans. Amer. Math. Soc.* **347** 2689–2751; Ignatiouk-Robert, 2002, Sample path large deviations and convergence parameters. *Ann. Appl. Probab.* **11** 1292–1329; Shwartz and Weiss, 1995, *Large Deviations for Performance Analysis*, Chapman-Hall). Yet, various applications of interest do not satisfy this condition. We describe several classes of models where jump rates diminish to zero in a Lipschitz continuous way. Under appropriate conditions, we prove that the sample path large deviations principle continues to hold. Under our conditions, the rate function remains an integral over a local rate function, which retains its standard representation.

Key words: sample path large deviations; large deviations with boundaries; $M/M/\infty$; $M/M/\infty$-like networks
MSC2000 subject classification: Primary: 60F10, 60J27; secondary: 60J50, 60K25
OR/MS subject classification: Primary: Queues: limit, theorems; secondary: mathematics
History: Received March 25, 2002; revised February 28, 2003, December 3, 2003, and March 18, 2004.

1. Introduction. One of the common technical assumptions in existing large deviations theory for jump Markov processes is that jump rates are bounded below, away from zero (Dupuis and Ellis [5], Ignatiouk-Robert [9], Shwartz and Weiss [13]). This is not merely a technical assumption: if the rates may go down to zero, the process may get stuck at a point, or it may or may not be possible to reach certain regions. We illustrate these issues below through examples. From a technical point of view, this condition is required to obtain smoothness of the local rate function.

However, in many applications this condition is violated. In this paper, we prove the large deviations principle for the sample paths of a wide class of such models, under the basic scaling of (2.2). Under our conditions, the usual Markovian integral representation of the rate function continues to hold, with the same variational formula for the integrand, as in the case with rates bounded away from zero. We allow the boundary defined by the region where some rates go to zero to be fairly general; in contrast with existing large deviations theory which deals only with "flat" boundaries, our boundaries are quite general.

Applications for which rates are not bounded away from zero fall roughly into two categories. In the first, the rates are proportional to the occupancy of the system, and thus go to zero when the system empties. In the second, a control is designed to avoid, say, overflows. "Soft controls" are characterized by continuous jump rates, and these controls may have rates become zero. Another example of a soft control is when additional processing power may be gradually added to a system when it grows beyond a certain threshold. Here are some specific examples of these applications.

Example A. In an $M/M/\infty$ queue, the service rate is linear in the queue size, so that as the queue empties, the rate goes to zero. In this case, it is not possible to continue reducing the queue size: the rate diminishes as we approach a boundary of the state space. This type of behavior is one of our main motivations, as there are many multidimensional models exhibiting this behavior (Mandjes and Weiss [11]). To obtain the scaling (2.2), scale the space variable by n and make the arrival rate $n\lambda$.

Example B. Consider an $M/M/1$ queue scaled by n: arrivals occur at rate $n\lambda$, services at rate $n\mu$, and the queue size $x(t)$ is scaled to $z_n(t) = x(t)/n$. Suppose that there is an auxiliary $M/M/\infty$ server, also with parameter μ, that kicks in when the scaled queue size

$z_n(t) > 1$, and that each service of this auxiliary server kicks out a pair of customers. The rate of departure of pairs is thus $\mu \cdot (z_n(t) - 1)$ whenever $z_n(t) > 1$ and 0 otherwise. Because the rate of jumps of size (-2) diminishes to zero at $z_n = 1$, this case falls outside the current theory. One key feature of this model is that the set of possible jump directions (the positive cone spanned by the jump directions) does not change as the jump rate diminishes to zero.

Example C. Consider a two-queue system with "soft priority" to queue 1. We suppose that the two queues behave like independent *M/M/*1 queues, with arrival rate 0.5 and service rate 1, as long as the (scaled) size of queue 1 is below 1. When the first queue is larger than 1, the server of queue 2 enters a processor sharing mode, so that the service rate to queue 1 increases linearly with queue size until its queue size is 2, and the rate remains at 2 when the first queue is larger, while service rate at queue 2 decreases linearly with the size of queue 1 until no service is offered at all. In this case, the service rate at queue 2 goes down and is 0 on $\{\vec{x}\colon x_1 \geq 2\}$. In contrast with Example A, this region can be entered with positive probability, so that it is not delineated by a boundary. In contrast with our other examples, we do *not* develop a theory that covers this case, although the methods we develop suffice for the analysis of simple models of this type.

Example D. We describe a simplified model of connection admission control (CAC), based on an idea by Tse [15]. Suppose that customers arrive at an infinite-server queue with Poisson rate $n\lambda$. Arrivals may be turned away (blocked), according to rules given in the next paragraph. Accepted customers depart the queue at rate μ. We let $q(t)$ represent the number in the system at time t. Customers who are accepted (not blocked) have two states, represented by 0 and 1. Accepted customer i moves from $x_i = 0$ to $x_i = 1$ with rate α, and from $x_i = 1$ to $x_i = 0$ with rate β. The bandwidth customer i uses at time t is $x_i(t)$; the total bandwidth in use at time t is thus $b(t) \triangleq \sum_{i=1}^{q(t)} x_i(t)$.

We now state the problem. Suppose that the capacity of the system is nC, so that there is trouble if $\sum_i x_i(t) > nC$. We try to ensure that this happens with small probability by accepting connections only when the scaled system state $\vec{z}_n(t) \triangleq 1/n(b(t), q(t))$ is in a fixed region G. For some fixed $\delta > 0$, we define the probability that a connection is accepted by

$$\mathbb{P}(\text{accept}) = \begin{cases} 1 & \text{if } \vec{z}_n(t) \in G \text{ and } d[\vec{z}_n(t), \partial G] > \delta, \\ (1/\delta)\,\text{dist}[\vec{z}_n(t), \partial G] & \text{if } \vec{z}_n(t) \in G \text{ and } d[\vec{z}_n(t), \partial G] \leq \delta, \\ 0 & \text{otherwise.} \end{cases}$$

We may try to design the set G so that the cheapest rate function for a path from $\vec{z}_n(0)$ to the set $b(t)/n = C$ is the same for all starting points $\vec{z}_n(0) \in \partial G$. In conclusion, we have described a Markov model of CAC that fits into the theory we develop in this paper, although we do not carry out the analysis and design of the appropriate region G for this model.

Example E. This example is to show some potential pitfalls in the theory of diminishing rates. Consider a pure birth process with $\lambda(x) = |x|$. This process moves to the right and, because $\lambda(0) = 0$, if $x(0) < 0$ is an integer, then $x(t) \leq 0$ for all t. Using the formulas for the local rate functions that hold true when the rates are bounded below, a formal calculation (detailed in §7) of the rate function $I(r)$ for the path $r(t) = t - 0.5$, $t \in [0, 1]$, yields $I(r) < \infty$, implying that $\mathbb{P}(\text{process} \approx r(t)) \approx e^{-nI(r)} > 0$. It is clear that the probability of the process following near this path is exactly 0, so the formal calculation of the rate is incorrect. However, if the process $x(t)$ starts with a noninteger value, then the probability is strictly positive, and it turns out that in this case the formal calculation is correct, although we will not detail that straightforward calculation in this paper. In more generality, if the rates $\lambda(x)$ are bounded below, then the sequence of processes with rates $\lambda^n(x) = n\lambda(x) + 1$ is exponentially equivalent to the case with $\lambda^n(x) = n\lambda(x)$ so that, on the large deviations scale, their behavior is identical. However, for the decreasing rate case described above, it leads to completely different behavior, because now it becomes possible to cross the $x = 0$ barrier. The upshot is that the form of the rate function we derive is in some sense less robust than it is under the usual condition that the rates are bounded away from 0.

We note that our analysis holds for processes confined to a convex set by having jump rates, in directions heading out of the set, diminish to zero at the boundary, as in Example D. Thus we establish, for the first time, a sample-path large deviations principle in the case of curved boundaries. Previous published work (Shwartz and Weiss [13, Chapter 8], Dupuis and Ellis [5], Dupuis et al. [7], Ignatiouk-Robert [9]) dealt only with flat boundaries; there is some unpublished work dealing with curved boundaries with some restrictions. Our general approach is based on Shwartz and Weiss [13]; indeed, we use this as the source of many of our lemmas, and sometimes prove theorems by giving only the changes necessary to use arguments in Shwartz and Weiss [13].

In general, there is no exponential equivalence between processes with log-bounded rates and those whose rates may vanish, as Example E clearly shows. In fact, it is natural to approximate a process with rates that go to zero by imposing a small lower bound, say ε, on the rates, and there is an obvious coupling between these two processes. However, it is not hard to see that this coupling does not provide an exponential approximation. Therefore, we resort to the more technical approach of following the steps of Shwartz and Weiss [13].

Our approach turns the problem of estimating the frequency or manner of occurrence of a rare event into a variational problem. Such problems are not necessarily easy to solve, although many one-dimensional problems have been, and some authors (e.g., Mandjes and Ridder [10]) have solved some specific multidimensional models. The relationship between our finite-time approach and steady-state statistics is given by the Freidlin-Wentzell theory, given for diffusion processes in Freidlin and Wentzell [8], and for queues in Shwartz and Weiss [13, Chapter 6]. There are many other approaches to sample-path large deviations; see, for example, Dupuis and Ellis [6] or Puhalskii [12]. The literature on non-sample-path large deviations is vast. The most relevant early references for the types of models we address are Botvich and Duffield [2], Courcoubetis and Weber [3], and Simonian and Guibert [14], which include tail estimates in steady state.

This paper is organized as follows. In §2, we set up the notation and describe the problem as well as our main results: Theorem 2.1 for processes with boundaries, where some jump rates go to zero at the boundary, and Theorem 2.2 for processes with diminishing rates but where, at each point, the jump directions associated with positive rates span $\mathbb{R}^d$. Corollary 2.1 in §2 shows that these two results can be combined to cover a large class of models exhibiting small rates both near boundaries and in the interior. Finally, in Corollary 2.4 we remove the technical assumption that the set of interest is compact, to obtain our results under the weakest assumptions. This last extension is a consequence of Corollary 2.3, a result of independent interest, on the exponential tightness of models with rates that grow at most linearly.

In §3, we develop some preliminary results. In §4, we prove the large deviations upper bound for the case that rates diminish at a boundary. In §5, we establish the corresponding lower bound. In §6, we state and prove the large deviations principle when some rates become zero in the interior of the region for models such as Example B in the introduction.

In §7, we give a simple sufficient condition for the rate function to be finite for paths that approach a boundary; this shows that many boundaries may be reached with probability that decreases to zero at an exponential rate, as opposed to a superexponential rate. This condition is also virtually necessary in the one-dimensional case. Section 8 summarizes our results and sketches open problems. Appendix A contains a technical lemma that allows us to extend Lipschitz continuous jump rates from regions that are unions of convex sets to all of $\mathbb{R}^d$. Appendix B contains an example illustrating the type of processes we consider and shows how to verify a technical assumption.

2. Assumptions and main results. We study jump Markov processes with boundaries where some jump rates become zero. The structure of the processes is simple and is detailed first. The structure of the boundaries of the regions of interest is more complicated and is deferred for a few paragraphs. Using the notation of Shwartz and Weiss [13], our model of jump processes is specified through k jump directions $\{\vec{e}_j\}_{j=1}^k$ and their respective Poisson jump rates $\{\lambda_j(\vec{x})\}$, which are defined for all $\vec{x}$ in some set G (see Assumption 2.1). We assume that the jump rates are Lipschitz continuous functions and, without loss of generality, that the positive cone (defined in (2.3)) spanned by the $\{\vec{e}_j\}$ is $\mathbb{R}^d$. We call such a process $x(t)$. Previous studies have assumed that the jump rates are uniformly bounded away from 0; the sole novelty of this paper is the relaxation of that condition.

We scale the process $x(t)$ in space and time to $z^n(t)$ as follows. All jump rates are multiplied by a scaling parameter n, and all jump sizes are divided by n; in other words, for $z^n(t)$, jump $\vec{e}_j$ becomes $\vec{e}_j/n$, and occurs at rate $n\lambda_j$. The generators for the original and scaled process are thus given, respectively, by

$$\mathscr{L}f(\vec{x}) \triangleq \sum_j \lambda_j(\vec{x})(f(\vec{x}+\vec{e}_j) - f(\vec{x})), \tag{2.1}$$

$$\mathscr{L}^n f(\vec{x}) \triangleq \sum_j n\lambda_j(\vec{x})\left(f\left(\vec{x}+\frac{\vec{e}_j}{n}\right) - f(\vec{x})\right). \tag{2.2}$$

We now begin our description of the boundaries with some notation and definitions. We denote by x_j the jth coordinate of a vector $\vec{x}$, by ∂S the boundary of a set S, by S^o the interior of S, by $B(x, r)$ the open ball of radius $r > 0$ centered at x, and by $d(\vec{x}, \vec{y})$ the Euclidean distance between $\vec{x}$ and $\vec{y}$. Given a set of vectors $\{\vec{u}_j\}$, the positive cone spanned by the vectors is

$$\mathscr{C}\{\vec{u}_j\} \triangleq \left\{\vec{v}\text{: there exist } \alpha_j \geq 0 \text{ with } \vec{v} = \sum_j \alpha_j \vec{u}_j\right\}. \tag{2.3}$$

The cone generated by the positive jump rates at $\vec{x}$ is denoted by

$$\mathscr{C}_x \triangleq \mathscr{C}\left\{\vec{e}_j\text{: } \lambda_j(\vec{x}) > 0\right\}. \tag{2.4}$$

To motivate our assumptions, suppose G is bounded and convex; then it as well as its boundary are compact. Therefore, they can be covered with a finite number of open balls, and in particular, we can achieve this covering with some balls centered on the boundary, and the rest having no intersection with the boundary. Moreover, at each point of ∂G we can fit a cone which is, at least locally, contained in G: more precisely, we say that G has an *interior cone* property if there are numbers $\varepsilon > 0$, $\beta > 0$ (independent of $\vec{x}$), such that for every $\vec{x} \in \partial G$, there is a vector $\vec{v}$ such that for each $t \in (0, \varepsilon)$, we have $B(\vec{x} + t\vec{v}, \beta t) \subset G$.

Assumption 2.1. *The set G is compact, the closure of its interior. There exist positive $\eta, \varepsilon, \gamma, \delta_0, \beta$, vectors $\vec{v}_i$, and open balls B_i so that*

(i) $\{B_i = B(\vec{x}_i, r_i), i = 1, \ldots I_1\}$ *covers ∂G and $\vec{x}_i \in \partial G$.*

(ii) $\{B_i = B(\vec{x}_i, r_i), i = 1, \ldots I\}$ *covers G with $\vec{x}_i \in \partial G$, $i \leq I_1$ and $\vec{x}_i \in G^o$, $i > I_1$.*

(iii) *G satisfies an interior cone condition with parameters ε and β. The vectors $\vec{v}$ for the interior cone may be taken as constant, $\vec{v}_i$, in each region B_i. Moreover, if $d(\vec{x}, \partial G) < \gamma$, then $d(\vec{x} + t\vec{v}_i, \partial G) > \eta t$ and is monotone increasing for $0 \leq t \leq \varepsilon$.*

(iv) *For any $\vec{x} \in G$, we have $B(\vec{x}, \delta_0) \cap G \subset B_i$ for some i.*

To illustrate that these assumptions are quite weak, we consider the following class. A set S is called star-like with respect to a point $x \in S$ if for any point $y \in S$ the closed line segment between x and y lies in S.

Lemma 2.1. *Suppose that G is compact and the closure of its interior, and that there exists a ball $B \subset G$ so that G is star-like with respect to each $x \in B$. Then Assumption 2.1 holds.*

In particular, any compact convex set with nonempty interior satisfies the assumptions of the lemma.

Proof. The first statement holds because G is compact and contains a ball. Because the boundary is compact, we can cover ∂G with balls B centered on the boundary. Cover the interior with balls contained in G^o. Now extract a finite cover, and (i)–(ii) are established.

Take a ball $B(\vec{x}_0, \delta)$, $\delta \le 2$, contained in B. Without loss of generality we may assume that $\max_i r_i < \delta/2$. Suppose that $\vec{x} \in B_j$. Then, the convex hull of $B_j \cup B(\vec{x}_0, \delta)$ is contained in G, and so the first part of (iii) follows. For the second part, take $\vec{v}_j = (\vec{x}_0 - \vec{x}_j)/(2\|\vec{x}_0 - \vec{x}_j\|)$. The last part of (iii) follows from the first because there is a minimal angle to the (finite number of) cones. Finally, (iv) follows by compactness: assume the contrary and take a sequence of points for which the largest ball is of size 2^{-n}. Then, the limit point cannot belong to any B_i, a contradiction. $\square$

A second property that holds easily when G is convex is that we can extend the jump rates from being defined on G to being defined on $\mathbb{R}^d$ while maintaining their Lipschitz continuity, as follows. Because G is convex, then for each point $\vec{x} \notin G$, there is a unique point $p(\vec{x}) \in \partial G$ that is closest to G: it is the projection of $\vec{x}$ on G. The definition of $\lambda_j(\vec{x})$ can therefore be extended to $\mathbb{R}^d$ by setting $\lambda_j(\vec{x}) \triangleq \lambda_j(p(\vec{x}))$. In Appendix A, we show that if the $\lambda_i(\vec{x})$ are Lipschitz continuous for $\vec{x} \in G$, then they are Lipschitz when extended in this way.

The choice of vectors $\vec{v}_i$ in Assumption 2.1(iii) is obviously not unique. Below we make the assumption that this choice can be made so that the vectors are consistent with the directions of increase of the diminishing rates.

The following assumption concerns the case where rates diminish near a boundary.

Assumption 2.2. *The rates and jump directions satisfy the following*:

A. *There is a constant K_λ such that $|\lambda_j(\vec{x}) - \lambda_j(\vec{y})| \le K_\lambda |\vec{x} - \vec{y}|$. Moreover, the rates can be extended to a δ neighborhood of G, so that the Lipschitz property continues to hold.*

B. *For each $\vec{x} \in \partial G$, there is an $\varepsilon_1 > 0$ so that $\vec{y} \in \mathscr{C}_x$ together with $|\vec{y}| < \varepsilon_1$ implies $\vec{x} + \vec{y} \in G$.*

C. *$\lambda_i(\vec{x}) > 0$ for all i and $\vec{x} \in G^o$. Moreover, γ, ε, and $\vec{v}_i$ of Assumption 2.1 can be chosen so that*

$$\vec{v}_i \in \mathscr{C}\left\{\vec{e}_j\colon \inf_{\vec{x} \in B_i} \lambda_j(\vec{x}) > \gamma\right\}$$

and if $\vec{x} \in B_i$, $d(\vec{x}, \partial G) < \gamma$ and $\lambda_j(\vec{x}) < \gamma$, then $\lambda_j(\vec{x} + \alpha\vec{v})$ is monotone increasing in α for $0 < \alpha < \varepsilon$.

D. $\mathscr{C}\{\vec{e}_j\} = \mathbb{R}^d$.

Note that Assumptions 2C and 2D together show that $\vec{x} \in G^o$ implies $\mathscr{C}_x = \mathbb{R}^d$. Appendix B contains a nontrivial example of a process satisfying the assumptions and sketches how to verify Assumption 2.2C.

When rates diminish in the interior, Assumptions 2.2B and 2C are not relevant, and Assumption 2.2D is replaced with

Assumption 2.3. *$\mathscr{C}_x = \mathbb{R}^d$ for all $\vec{x}$.*

This assumption is used for Theorem 2.2, which does not require all the parts of Assumption 2.2.

Note that our assumptions allow the process to jump out of G; see Appendix B for a worked example. However, by Assumption 2.2, at any given point in G—including the boundary—if the jump size is small enough, then one jump will not cause the process to exit G. In particular, it is not possible to jump from a point in ∂G in a direction parallel to ∂G in directions where G is strictly convex. Finally, note that Assumptions 2.2A and 2B imply that rates for jumps out of G must decrease to 0 as the boundary is approached.

LEMMA 2.2. *Let $G = \bigcup G_i$, $1 \le i \le k$, where each G_i is a convex compact set. Suppose that the G_i satisfy the property that any intersection of collections of G_i is either empty or is the closure of its interior. Then, Assumption 2.1 holds for G. Furthermore, if the first part of Assumption 2.2*A *holds, then the second part does as well* (*Lipschitz extension*).

PROOF. The first claim follows from reasoning like that of Lemma 2.1, detailed below. The second part is proved in Appendix A.

Because any nonempty intersection of G_i is convex and is the closure of its interior, this intersection satisfies the interior cone condition by Lemma 2.1. Consider any point $y \in \partial G$. The intersection of all G_i containing y is nonempty. Thus, y satisfies the interior cone condition. For any such y, there exists a $\delta(y)$ such that the distance between y and any G_i with $y \notin G_i$ is greater than $\delta(y)$ (because all the G_i are closed). Now we can use a single ball contained in $\bigcap_{y\in G_j} G_j$, as in Lemma 2.1, to construct a vector $v(y)$ that serves as a uniform interior cone direction for starting points near y (nearer than $\delta(y)/2$, say). We are now in a position to use the compactness of the set G to extract a finite subcover of neighborhoods that have uniform interior cones. The increasing distance property and single B_i property (part iv) now follow as in the proof of Lemma 2.1. □

Note that in our example of the nonrobustness of the rate function (Introduction, Example E), the set G is the union of the convex sets $(-\infty, 0]$ and $[0, \infty)$ which do not satisfy the intersection property above. The result fails for this case, because for the ball containing 0 there can be no $\vec{v}$ satisfying Assumption 2.1(iii). Consequently, our theory does not apply (which it should not, because Theorem 2.1 does not hold for this case).

We state one more note on where our assumptions apply. Recall that the second part of Assumption 2.2A follows from the first for convex sets by projection. It is not hard to show that if there is a δ_0 so that for each $\vec{x} \in \partial G$ there is a closed ball $B(\vec{y}, \delta_0)$ with $B(\vec{y}, \delta_0) \cap G = \vec{x}$, then the second part of Assumption 2.2A follows from the first, as we can define projection locally.

For $\vec{x}, \vec{y} \in \mathbb{R}^d$ and measurable $\vec{r}\colon [0, T] \to G$, define

$$\ell(\vec{x}, \vec{y}) \triangleq \sup_{\vec{\theta}\in\mathbb{R}^d} \left(\langle \vec{\theta}, \vec{y} \rangle - \sum_j \lambda_j(\vec{x})(e^{\langle \vec{\theta}, \vec{e}_j \rangle} - 1) \right), \tag{2.5}$$

$$I_{[0,T]}(\vec{r}) \triangleq \begin{cases} \displaystyle\int_0^T \ell(\vec{r}(t), \vec{r}\,'(t))\, dt & \text{if } \vec{r}(t) \text{ is absolutely continuous,} \\ \infty & \text{otherwise.} \end{cases} \tag{2.6}$$

We can now state our main results. Let D denote the space of bounded, right-continuous functions from $[0, T]$ to $\mathbb{R}^d$, possessing left-hand limits, and let D_s denote the space D endowed with the sup norm topology. (Note that Corollary 2.2 shows that, under the present scaling, we may use either the sup norm topology or the Skorohod J_1 topology.) For any measurable set S in D_s, denote

$$I_{[0,T]}(S) \triangleq \inf\{I_{[0,T]}(\vec{r})\colon \vec{r} \in S,\ \vec{r}(0) = \vec{x}\},$$

where the dependence on $\vec{x}$ is suppressed. Our main result concerning rates that diminish toward a boundary is

THEOREM 2.1. *Let Assumptions 2.1 and 2.2 hold and consider the sequence $\{\vec{z}^{\,n}\}$ of processes taking values in D_s. This sequence satisfies the large deviations principle with rate function $I_{[0,T]}$. That is, for each $T > 0$, closed set $C \in D_s$, and open set $O \in D_s$, and each point $\vec{x} \in G$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x\big(\vec{z}^{\,n} \in C\big) \le -I_{[0,T]}(C), \tag{2.7}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x\big(\vec{z}^{\,n} \in O\big) \ge -I_{[0,T]}(O). \tag{2.8}$$

Moreover, the nonnegative function $I_{[0,T]}$ is a good rate function, meaning its level sets $\{\vec{r}: I_{[0,T]}(\vec{r}) \le \alpha, \vec{x} \in G\}$ are compact for each α.

Suppose now that $G = \mathbb{R}^d$; the process has no boundaries. When the rates diminish in the interior without changing the positive cone of jump direction, we have

THEOREM 2.2. *Let Assumptions 2.2A and 2.3 hold and consider the sequence $\{\vec{z}^n\}$ of processes taking values in D_s. This sequence satisfies a large deviations principle with good rate function $I_{[0,T]}$.*

The two results can be combined as follows. The assumptions of Theorem 2.1 are imposed near the boundary, while the assumptions of Theorem 2.2 are imposed away from the boundary. Thus, rates may go to zero in the interior, but the cone can only change on the boundary.

COROLLARY 2.1. *Let B be an open set so that $\bar{B} \subset G^o$. Let Assumptions 2.1, 2.2A, and 2.2B hold in G, let Assumption 2.2C hold in $G\backslash B$, and Assumption 2.3 hold for $\vec{x} \in \bar{B}$. Consider the sequence $\{\vec{z}^n\}$ of processes taking values in D_s. This sequence satisfies a large deviations principle with good rate function $I_{[0,T]}$.*

PROOF. Use Theorems 2.1 and 2.2 in each region separately. Then, connect the results exactly as in the proof of Lemma 4.3: see also the comments at the end of the proof of the lower bound. □

Comment. As in Shwartz and Weiss [13], we move freely between a jump process and its piecewise linear interpolation—which is obtained by interpolating linearly every sample path between jump points—without changing notation. These two processes are exponentially equivalent under the sup norm as their distance at any time is at most one jump. We therefore work with interpolated processes, which are piecewise linear (and, in particular, Lipschitz continuous) and for which the notion of compactness is easier to handle.

COROLLARY 2.2. *Theorems 2.1, 2.2, and Corollary 2.1 hold as stated if we endow the space D with the Skorohod J_1 topology.*

PROOF. Consider the linearly interpolated version of the process. The identity map from D_s to D with the Skorohod topology is continuous. Therefore, by the contraction principle, the theorem holds with the same rate function. Because the distance between the jump process and the interpolated process is at most $1/n$, the two are exponentially equivalent under the Skorohod J_1 metric, and the result is established. □

Corollary 2.3, a result of independent interest, shows that linear growth rates guarantee exponential tightness. This implies that is suffices to prove the upper bound for compact sets, namely, to establish the weak large deviations principle. Thus, the compactness assumption in Assumption 2.1 is not necessary: it suffices that the assumptions hold for bounded subsets of G.

LEMMA 2.3. *Assume that the rates $\lambda_i(\vec{x})$ have linear growth: $\lambda_i(\vec{x}) \le K(1+|\vec{x}|)$. Then, uniformly for $\vec{x}$ in compact sets,*

$$\lim_{r\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x \left(\sup_{0\le t\le T} |\vec{z}^n(t)| > r \right) = -\infty. \tag{2.9}$$

PROOF. Fix b_0 and a point $\vec{x}$ with $|\vec{x}| \le b_0$. Define a process y^n by setting $y^n(0) = b_0$ and having y^n increase by $K_e = \max_j |\vec{e}_j|$ with every jump of $\vec{z}^n$. Then, because $y^n(t)$ is increasing,

$$\sup_{0\le t\le T} |\vec{z}^n(t)| \le \frac{y^n(T)}{n}. \tag{2.10}$$

Now take a collection of independent Yule processes x_i, which are pure birth processes, with $x_i(0) = 1$ and jump rates $K_1 x_i$, where $K_1 \triangleq K_\lambda k K_e$. We couple y^n to this collection in the sense that y jumps whenever one of the x_i jumps, so that

$$y^n(t) \le b_0 + K_e \sum_{i=1}^n x_i(t), \qquad n \ge b_0. \tag{2.11}$$

Using Shwartz and Weiss [13, Corollary 14.14], it follows that there exists a good rate function ℓ_T so that

$$\ell_T\left(e^{K_1 T}\right) = 0, \qquad \lim_{m\to\infty} \ell_T(am) = \infty \tag{2.12}$$

for all $a > 0$, and such that $\ell_T(a)$ is the rate function for $x_i(t)$. That is, for each $a > e^{K_1 T}$, we have

$$\mathbb{P}\left(\sum_{i=1}^n x_i(t) \ge ap\right) \le e^{-n\ell_T(a)}. \tag{2.13}$$

Therefore,

$$\limsup_n \frac{1}{n} \log \mathbb{P}_x\left(\sup_{0\le t\le T} |\vec{z}^n(t)| > r\right) \le \limsup_n \frac{1}{n} \log \mathbb{P}_x\left(\sum_{i=1}^n x_i(t) > \frac{(r-b_0)n}{K_2}\right) \tag{2.14}$$

$$= -\ell_T\left(\frac{r-b_0}{K_2}\right), \tag{2.15}$$

which tends to $-\infty$ as $r \to \infty$. □

Note that Assumption 2.2A implies the condition of Lemma 2.3. From this we obtain exponential tightness as follows.

Corollary 2.3. *If $\lambda_i(\vec{x}) \le K(1+|\vec{x}|)$ for all i, then the $\{\vec{z}^n\}$ are exponentially tight. Consequently, it suffices to prove the large deviations estimates for bounded sets, uniformly in the initial conditions over compact sets.*

Proof. Given $\alpha > 0$, by Lemma 2.3 there is a r_α so that

$$\limsup_n \frac{1}{n} \log \mathbb{P}_x\left(\sup_{0\le t\le T} |\vec{z}^n(t)| > r_\alpha\right) \le -\alpha \tag{2.16}$$

for all initial conditions $\vec{x}$ in a bounded set. But then, because the rates are Lipschitz, they are bounded over the region $|\vec{z}^n(t)| \le r_\alpha$. Therefore, by Shwartz and Weiss [13, Lemma 5.58], the $\{\vec{z}^n\}$ are exponentially tight. Thus, by Dembo and Zeitouni [4, Lemma 1.2.18(a)], it suffices to prove the upper bound for compact (and, in particular, bounded) sets. By Shwartz and Weiss [13, (5.31–5.32)], it suffices to prove the lower bound on balls: however, all paths in a ball are by definition bounded. □

Corollary 2.4. *Let G be a closed set and $\lambda_i(\vec{x}) \le K(1+|\vec{x}|)$ for all i. Assume that there are $R_k \to \infty$ and open $B_k \subset G^o$ so that the conditions of Corollary 2.1 hold with B_k replacing B, $G \cap B(0, R_k)$ replacing G, and with $\partial G \cap B(0, R_k)$ replacing ∂G. Then, Theorems 2.1, 2.2, and Corollary 2.1 hold.*

Note that the constants in Assumptions 2.1 and 2.2 need not be uniform: they may depend on R_k. Because it thus suffices to derive our results in the case that G is compact, we shall do so because this simplifies the technicalities. We shall therefore assume henceforth that G is bounded and so $\bar{\lambda} \triangleq \max_j \sup_{\vec{x}} \lambda_j(\vec{x}) < \infty$.

Lemma 2.4. *There exists a continuous monotone "scale function" $s(\delta)$ such that $s(0) = 0$, $s(\delta) > 0$ for any $\delta > 0$, and for any $\vec{x} \in G$, we have*

$$\lambda_i(\vec{x}) \ge s(d(\vec{x}, \partial G)) \quad \textit{for all } i.$$

Proof. Take $s(\delta) \triangleq \inf\{\lambda_i(\vec{x})\colon i;\ \vec{x}\colon d(\vec{x}, \partial G) \ge \delta\}$. The properties follow by continuity of $\lambda_i(\vec{x})$ and compactness of G. □

3. Preliminaries. This section collects some preliminary definitions and estimates. We use an alternative form of ℓ in many of the proofs. Let

$$f(\mu,\lambda) \triangleq \sum_j \lambda_j - \mu_j + \mu_j \log \frac{\mu_j}{\lambda_j}, \tag{3.1}$$

$$K_{\vec{y}} \triangleq \left\{\mu \colon \mu_j \geq 0,\ \sum_j \mu_j \vec{e}_j = \vec{y}\right\}. \tag{3.2}$$

Note that, by definition, $K_{\vec{y}}$ is nonempty if and only if $\vec{y} \in \mathscr{C}_x$.

LEMMA 3.1 (SHWARTZ AND WEISS [13, THEOREM 5.26]). $\ell(x,y) = \inf\{f(\mu,\lambda(x)) \colon \mu \in K_{\vec{y}}\}$.

Note that if $\lambda_j(x) = 0$, then it plays no role in the definition of ℓ. Further, because f is to be minimized, if $\lambda_j(x) = 0$, the corresponding μ_j is necessarily 0, too, so that the j component again plays no role. The lemma also means that either both expressions are finite, or both are infinite.

LEMMA 3.2 (SHWARTZ AND WEISS [13, LEMMA 5.20]). *There exists a constant κ such that for all $\vec{x}$ and all $\vec{y} \in \mathscr{C}_{\vec{x}}$, there exists $\vec{a}$ such that*

$$\vec{y} = \sum_{i=1}^{k} a_i \vec{e}_i, \qquad a_i \geq 0, \qquad |\vec{a}| \leq \kappa\, |\vec{y}|\,.$$

LEMMA 3.3. *Fix λ. The function $f(u,\lambda(x))$ is nonnegative, strictly convex in μ for $\mu_j \geq 0$ with compact level sets. For y in $\mathscr{C}_x$ the following hold: The function f has a minimum $f^*(\vec{y})$ over $K_{\vec{y}}$, attained at a unique point $\mu^*(\vec{y})$, so that $f^*(\vec{y}) = f(\mu^*(\vec{y}),\lambda)$. Both the minimum f^* and μ^* are continuous in $\vec{y}$. In addition, given B, there are constants C_0, C_1, C_2, and D so that for all $|\lambda_i| < B$,*

$$|\mu^*| \leq C_0 \quad \text{for } |\vec{y}| \leq D, \tag{3.3}$$

$$|\mu^*| \leq C_1|\vec{y}| \quad \text{for } |\vec{y}| > D, \tag{3.4}$$

$$|\mu^*| \geq C_2|\vec{y}|. \tag{3.5}$$

PROOF. The function $1 - x + x\log x$ is nonnegative and strictly convex for $x \geq 0$, as seen by differentiation. Level sets are compact because the function grows super-linearly. Existence of a minimum now follows because $K_{\vec{y}}$ is closed. Uniqueness follows from strict convexity because $K_{\vec{y}}$ is a convex set.

The bound (3.3) is a consequence of Lemma 3.2 and the compactness of the level sets. The second bound (3.4) follows from the same argument because f grows faster than linearly in μ. Finally, (3.5) is immediate from the definition of $K_{\vec{y}}$.

Finally, we prove continuity of μ^*, which in turn implies continuity of f^*. Fix a point $\vec{y}$. First, we claim that, given $\varepsilon > 0$ there exists a $\delta(\varepsilon) = O(\varepsilon)$ so that $\vec{y}_i \in \mathscr{C}_x$ together with $|\vec{y}_i - \vec{y}| < \delta(\varepsilon)$ implies

$$\vec{y}_i = (1-\varepsilon_1)\vec{y} + \sum_j a_{ij}\vec{e}_j \quad \text{for some } \varepsilon_1 < \varepsilon \text{ and } 0 \leq a_{ij} \leq \varepsilon_2 < \varepsilon. \tag{3.6}$$

For a proof assume the contrary. Then, there are $\vec{y}$ and $\varepsilon > 0$ and a sequence $\vec{y}_i \to \vec{y}$ for which (3.6) does not hold. Let $\mu^*(\vec{y}_i)$ denote the minimizing points for f over K_{y_i}, so that

$$\vec{y}_i = \sum_j \mu_j^*(y_i)\vec{e}_j.$$

By (3.3), the collection $\{\mu^*(\vec{y}_i)\}$ is bounded. So, take a converging subsequence, and denote the limit μ^*. By taking further subsequences we may assume that for each j, the sequence

$\mu_j^*(\vec{y}_i)$ is monotone in i. So, fix $0 < \varepsilon_1 < \min_j \mu_j^*$ and note that

$$\vec{y}_i = \sum_j \mu_j^*(\vec{y}_i)\vec{e}_j \tag{3.7}$$

$$= \sum_{j:\,\mu_j^*>0} \mu_j^* \vec{e}_j + \sum_j \big(\mu_j^*(\vec{y}_i) - \mu_j^*\big)\vec{e}_j \tag{3.8}$$

$$= (1-\varepsilon_1)\sum_{j:\,\mu_j^*>0} \mu_j^* \vec{e}_j + \sum_j \big(\mu_j^*(\vec{y}_i) - \mu_j^* + \varepsilon_1\mu_j^*\big)\vec{e}_j. \tag{3.9}$$

By definition of ε_1, the coefficients in the first summand are positive. Consider the second summand. If $\mu_j^* > 0$, then, for all large i, we have $0 < (\mu_j^*(\vec{y}_i) - \mu_j^* + \varepsilon_1\mu_j^*) < \varepsilon_1 \max_j \mu_j^*$. If $\mu_j^* = 0$, then necessarily $\mu^*(\vec{y}_i)$ is monotone decreasing to zero and we have

$$0 < \mu_j^*(\vec{y}_i) - \mu_j^* + \varepsilon_1\mu_j^* = \mu_j^*(\vec{y}_i) \to 0.$$

Thus, an approximation as in (3.6) holds, and the claim is established.

To complete the proof of continuity, we retain the preceding notation. Let $\mu^*(\vec{y})$ be the (unique) minimizing point at $\vec{y}$: we need to establish that $\mu^* = \mu^*(\vec{y})$. But if not, then using (3.6), set $\vec{a}_i = \sum_j a_{ij}\vec{e}_j$, and we have

$$f(\mu^*(\vec{y}_i), \lambda) \le f(\mu^*(\vec{y})(1-\varepsilon_i) + \vec{a}_i, \lambda) \tag{3.10}$$

$$\le f(\mu^*(\vec{y}), \lambda) + \eta_i \tag{3.11}$$

by continuity of f, and where $\eta_i \to 0$ as $i \to \infty$. But because μ^* is not the minimum at $\vec{y}$,

$$f(\mu^*(\vec{y}_i), \lambda) \le f(\mu^*(\vec{y}), \lambda) + \eta_i \le f(\mu^*, \lambda) + \eta_i - \gamma$$

so that for all large i, $f(\mu^*(\vec{y}_i), \lambda) \le f(\mu^*, \lambda) - \gamma/2$. But this is a contradiction because by the continuity of f, $f(\mu^*(\vec{y}_i), \lambda) \to f(\mu^*, \lambda)$. □

We now establish that $I_{[0,T]}$, as defined in (2.6), is a good rate function.

THEOREM 3.1 (SHWARTZ AND WEISS [13, PROPOSITION 5.49 AND COROLLARY 5.50]). *Assume that the $\lambda_i(\vec{x})$ are bounded and continuous. Then, for each $\vec{x}$, $I_{[0,T]}(\cdot)$ is a good rate function under either the sup norm or the Skorohod J_1 metric.*

PROOF. Identical to the proof in Shwartz and Weiss [13]; the additional assumption (that the $\log \lambda_i$ are bounded) is not used in the proof (the original theorem was stated with unnecessary conditions). □

LEMMA 3.4 (UNIFORM ABSOLUTE CONTINUITY: SHWARTZ AND WEISS [13, LEMMA 5.18]). *Assume that the $\lambda_i(\vec{x})$ are bounded, let $I_{[0,T]}(\vec{r}) \le K$, and fix some $\varepsilon > 0$. Then, there is a δ, independent of $\vec{r}$, such that for any collection of nonoverlapping intervals*

$$\{[t_j, s_j],\ j = 1, \ldots, J\} \quad \textit{with} \ \sum_{j=1}^{J} s_j - t_j = \delta$$

in $[0, T]$ with total length δ, we have

$$\sum_{j=1}^{J} |\vec{r}(t_j) - \vec{r}(s_j)| < \varepsilon.$$

LEMMA 3.5. *Assume that $\lambda_i(\vec{x}) \le \bar{\lambda}$ for all i and define B_i as in Assumption 2.1. Then, for any $T > 0$, $K > 0$ there is a J such that if $I_{[0,T]}(\vec{r}) \le K$, then there are $0 = t_0 < t_1 < \cdots < t_J = T$ and j_i so that $\vec{r}(t) \in B_{j_i}$, $t_{i-1} \le t \le t_i$.*

PROOF. We first claim that there is an $\alpha > 0$ such that for any $\vec{x} \in G$ there is an i such that $B(\vec{x}, \alpha) \subset B_i$. This is easily proved by contradiction; if false, then there is a sequence of points $\vec{x}_j \to \vec{x}$ with diminishing open balls; but the point $\vec{x}$ is contained in the interior of some B_j.

Now, for any path $\vec{r}(t)$ with $I_{[0,T]}(\vec{r}) \le K$, we break up the path according to the following rules. We initially assign the ball B_i to the path at time 0 if $B(\vec{r}(0), \alpha) \subset B_i$ (we may take any B_i that satisfies this constraint). We maintain the choice i until such time that $B(\vec{r}(t), \alpha/2) \not\subset B_i$. At this time, change to any B_j with $B(\vec{r}(0), \alpha) \subset B_j$. By the uniform absolute continuity of $\vec{r}$, Lemma 3.4, there is a minimum time τ between any change in balls B_i. So $J = T/\tau$ suffices as a bound on the number of pieces. Note that at any switchover time t between balls B_i and B_j we have that both $\vec{r}(t) + (\alpha/4)\vec{v}_i$ and $\vec{r}(t) + (\alpha/4)\vec{v}_j$ are contained in $B_i \cap B_j$. □

4. Upper bound—boundary case. We prove the upper bound (2.7) along the same general lines as Shwartz and Weiss [13, Theorem 5.54]. Here is a rough sketch of the main idea of the proof, ignoring technicalities (some of which are commented on below). As in Shwartz and Weiss [13], we show in Lemma 4.7 that it suffices to estimate the probability that an approximation $\vec{y}_n$ of $\vec{z}_n$ (defined below) lies in a compact set of paths $\mathcal{K}$. Paths in the set $\mathcal{K}$ have a cost I larger than $I_{[0,T]}(C)$; here C is the closed set of paths appearing in the statement of the large deviations principle, Theorem 2.1. We stitch together local estimates, over short segments of time, of the probability that $\vec{y}_n$ follows a segment of a path in $\mathcal{K}$. These estimates are obtained by approximating the process with a constant coefficient one, essentially reducing the problem to estimates for random variables; see Shwartz and Weiss [13, Lemma 5.61].

There are two additional difficulties to overcome in our setting, both technical. The first is that the previous proof used the fact that shifting a path $\vec{r}(t)$ to $\vec{r}(t) + \delta\vec{v}_i$ resulted in a continuous change in the cost $I_{[0,T]}(\vec{r})$. This is not true in the present case. A shift may result in a path being outside the set G, thus having infinite cost. Another difficulty is that the approximating functional ℓ^δ, used in the proof of the upper bound, is difficult to estimate in the present case. Our solution is to make approximating functionals ℓ^Δ and ℓ^m that are finite for all absolutely continuous paths in G. We then need to calculate how well these functionals approximate ℓ, and to show that the requisite bounds obtain. The main technical estimate we need is that every path $\vec{r}$, having approximating cost $I^\Delta(\vec{r})$, has a path $\vec{r}_1$ at a distance no more than ε, with true cost $I(\vec{r}_1) \approx I^\Delta(\vec{r})$.

Although we shall use Assumptions 2.1 and 2.2, for the upper bound we do not need Lipschitz continuity of the rates. For our proof it suffices that the rates are absolutely continuous. Recall that a function $f(x)$ (possibly from one Euclidean space to another) is absolutely continuous if and only if there exists $\varepsilon(\delta) \to 0$ as $\delta \to 0$ so that

$$\sum_{i=1}^{I} |x_i - y_i| < \delta \quad \text{implies} \quad \sum_{i=1}^{I} |f(x_i) - f(y_i)| < \varepsilon(\delta).$$

It follows immediately that Lipschitz functions are absolutely continuous, and that the composition of absolutely continuous functions yields an absolutely continuous function. To emphasize this fact that a Lipschitz property is not required, we denote by K_λ the modulus of continuity of the rates. Recall that for $K_\lambda(\delta)$ to be the modulus of continuity of all $\lambda_i(\vec{x})$ means $|\lambda_i(\vec{x}) - \lambda_i(\vec{y})| \le K_\lambda(|\vec{x} - \vec{y}|)$, where $K_\lambda(\delta)$ is continuous and $K_\lambda(0) = 0$.

Lemma 4.1. *Under Assumptions 2.1 and 2.2C, there is a constant C such that for every $x \in \partial G$ and $x \in B_i$, $i \le I_1$, and every η small enough, the cost of the linear path $\vec{r}(t) \triangleq x + t\vec{v}_i$, $0 \le t \le \eta$, is less than $C\eta$. In other words, with $T = \eta$,*

$$I_{[0,T]}(\vec{r}) \le C\eta. \tag{4.1}$$

Proof. By Assumption 2.2C, we can find a set of $\vec{e}_i$ and a_i with $\sum_i a_i \vec{e}_i = \vec{v}$ and with $\lambda_i(\vec{r}(t)) > \gamma/2 > 0$ for all $t \le \eta$. The a_i are bounded because $\vec{v}$ is, by Lemma 3.2. So, by the representation of ℓ in Lemma 3.1, taking the $\mu_i = a_i$ and λ_i as given yields a bound on ℓ. □

The proof of the upper bound in Shwartz and Weiss [13] uses an "approximate" rate function I^δ: here we require several different approximations. The functional I^Δ is defined via the function g^Δ as follows:

$$g^\Delta(\vec{x},\vec{\theta}) \triangleq \sum_{i=1}^{k} \sup_{\vec{z}_i\in B_\delta(\vec{x})} [\lambda_i(\vec{z}_i)(e^{\langle\vec{\theta},\vec{e}_i\rangle}-1)], \tag{4.2}$$

$$\ell^\Delta(\vec{x},\vec{y}) \triangleq \sup_{\vec{\theta}\in\mathbb{R}^d}(\langle\vec{\theta},\vec{y}\rangle - g^\Delta(\vec{x},\vec{\theta})), \tag{4.3}$$

and $I^\Delta_{[0,T]}(\vec{r})$ is now defined like $I_{[0,T]}(\vec{r})$, with ℓ^Δ replacing ℓ. The functional I^δ was defined in a similar manner (Shwartz and Weiss [13, Definitions 5.36–38]), but starting with

$$g^\delta(\vec{x},\vec{\theta}) \triangleq \sup_{\vec{z}\in B_\delta(\vec{x})} \sum_{i=1}^{k} \lambda_i(\vec{z})\big(e^{\langle\vec{\theta},\vec{e}_i\rangle}-1\big). \tag{4.4}$$

The difference between g^δ and g^Δ is that the supremum and sum are interchanged. In g^Δ there are many different $\vec{z}_i$ where each maximum is attained; in g^δ there is only one $\vec{z}$. We are abusing notation a bit here; the terms labeled with a $^\Delta$ are defined in terms of a parameter δ. Of course, we are supposing that the value of Δ is δ, but the difference in notation should make clear what we mean. Clearly, we have $g^\Delta(\vec{x},\vec{\theta}) \ge g^\delta(\vec{x},\vec{\theta})$ for every δ, $\vec{x}$, and $\vec{\theta}$, so $\ell^\Delta(\vec{x},\vec{y}) \le \ell^\delta(\vec{x},\vec{y})$ and $I^\Delta_{[0,T]}(\vec{r}) \le I^\delta_{[0,T]}(\vec{r})$. We could have avoided the use of I^δ entirely in this paper, except that there are many bounds available for it already, so instead of having to prove all these bounds for I^Δ separately, we will use them in conjunction with the simple relationship between I^δ and I^Δ.

The advantage of using I^Δ instead of I^δ is that I^Δ has a representation as a change of measure as follows. Note that because λ_i is continuous and because the supremum is taken for each i separately, $g^\Delta(\vec{x},\vec{\theta})$ is the supremum over a convex set of $\{\lambda_i\}$ of a function which is linear in the λ_i. Therefore,

$$\ell^\Delta(\vec{x},\vec{y}) \triangleq \sup_{\vec{\theta}\in\mathbb{R}^d}\left(\langle\vec{\theta},\vec{y}\rangle - \sum_{i=1}^{k}\sup_{\vec{z}_i\in B_\delta(\vec{x})}\lambda_i(\vec{z}_i)\big(e^{\langle\vec{\theta},\vec{e}_i\rangle}-1\big)\right) \tag{4.5}$$

$$= \sup_{\vec{\theta}\in\mathbb{R}^d}\inf_{\vec{z}_i\in B_\delta(\vec{x})}\left(\langle\vec{\theta},\vec{y}\rangle - \sum_{i=1}^{k}\lambda_i(\vec{z}_i)\big(e^{\langle\vec{\theta},\vec{e}_i\rangle}-1\big)\right) \tag{4.6}$$

$$= \inf_{\vec{z}_i\in B_\delta(\vec{x})}\sup_{\vec{\theta}\in\mathbb{R}^d}\left(\langle\vec{\theta},\vec{y}\rangle - \sum_{i=1}^{k}\lambda_i(\vec{z}_i)\big(e^{\langle\vec{\theta},\vec{e}_i\rangle}-1\big)\right) \tag{4.7}$$

$$= \inf_{\vec{z}_i\in B_\delta(\vec{x})}\inf_{\mu\in K_y}\sum_{i=1}^{k}\left(\lambda_i(\vec{z}_i) - \mu_i + \mu_i\log\frac{\mu_i}{\lambda_i(\vec{z}_i)}\right) \tag{4.8}$$

$$= \sum_{i=1}^{k}\left(\lambda_i(\vec{z}_i^*) - \mu_i^* + \mu_i^*\log\frac{\mu_i^*}{\lambda_i(\vec{z}_i^*)}\right), \tag{4.9}$$

where the $\vec{z}_i^* \in B_\delta(\vec{x})$. Equation (4.5) is by definition. Equation (4.7) follows from the min-max saddle point theorem, because the function in the definition of $\ell^\Delta(\vec{x},\vec{y})$ is concave in $\vec{\theta}$ and linear, hence convex in the λ_i; furthermore, the infimum is taken over a bounded convex set of λ_i. Equation (4.8) follows by taking a minimizing $\vec{z}_i$, and Equation (4.9) follows from the representation result Lemma 3.1 for $\ell(\vec{x},\vec{y})$.

For technical complexity, but to make easy proofs, we also define a function ℓ^m as follows. Define $\vec{\lambda}^m$ by

$$\lambda_j^m(\vec{x}) \triangleq \max\{\lambda_j(\vec{x}), 1/m\}. \tag{4.10}$$

Define ℓ^m and I^m through (2.5)–(2.6) but using the modified rates $\lambda_j^m(\vec{x})$. To complete this sequence of definitions, we define the set of cheap paths in any of the metrics:

$$\Phi_{\vec{x}}(K) \triangleq \{\vec{r}(t)\colon \vec{r}(0) = \vec{x},\, I_{[0,T]}(\vec{r}) \le K\}, \tag{4.11}$$

$$\Phi_{\vec{x}}^{\Delta}(K) \triangleq \{\vec{r}(t)\colon \vec{r}(0) = \vec{x},\, I_{[0,T]}^{\Delta}(\vec{r}) \le K\}, \tag{4.12}$$

$$\Phi_{\vec{x}}^{m}(K) \triangleq \{\vec{r}(t)\colon \vec{r}(0) = \vec{x},\, I_{[0,T]}^{m}(\vec{r}) \le K\}. \tag{4.13}$$

Lemma 4.2. *Let Assumptions 2.1 and 2.2 hold. Fix i. For each $T > 0$, $K > 0$, and $0 < \varepsilon < K$ there is an $0 < \eta_0 < \varepsilon$ such that for any $0 < \eta < \eta_0$, there is an m_0 with the following property. If $m > m_0$ and if the path $\vec{r}(t)$ takes values in B_i with $I_{[0,T]}^m(\vec{r}) < K - \varepsilon$, then the path $\vec{r}_2(t) \triangleq \vec{r}(t) + \eta \vec{v}_i$ satisfies $I_{[0,T]}(\vec{r}_2) < K$.*

Proof. Let $\vec{v} = \vec{v}_i$,

$$\ell_1(t) = \ell^m(\vec{r}(t), \vec{r}\,'(t)), \tag{4.14}$$

$$\ell_2(t) = \ell(\vec{r}(t) + \eta\vec{v}, \vec{r}\,'(t)), \tag{4.15}$$

$$\lambda_i^1(t) = \lambda_i^m(\vec{r}(t)), \tag{4.16}$$

$$\lambda_i^2(t) = \lambda_i(\vec{r}(t) + \eta\vec{v}), \tag{4.17}$$

$$\mu_i^*(t) = \mu_i^{m,*}(\vec{r}(t), \vec{r}\,'(t)), \tag{4.18}$$

where $\mu^{m,*}$ is optimal (see Lemma 3.3) for jump rates λ_i^m. We denote by $\lambda^1(t)$ the vector with coordinates $\lambda_i^1(t)$ and similarly for λ^2 and μ^*. By Lemma 3.1,

$$\ell_2(t) \le f(\mu^*(t), \lambda_i^2(t)).$$

Choose $m_0 > 1/\eta$. Then, by continuity of λ_i,

$$|\lambda_i^2(t) - \lambda_i^1(t)| \le \frac{1}{m} + |\lambda_i^2(t) - \lambda_i(\vec{r}(t))| \le K_\lambda(\eta) + \eta \triangleq K'_\lambda(\eta).$$

Therefore,

$$\ell_2(t) - \ell_1(t) \le \sum_{i=1}^{k} \lambda_i^2(t) - \lambda_i^1(t) + \mu_i^*(t) \log \frac{\lambda_i^1(t)}{\lambda_i^2(t)} \tag{4.19}$$

$$\le k \cdot K'_\lambda(\eta) + \sum_{i=1}^{k} \mu_i^*(t) \log \frac{\lambda_i^1(t)}{\lambda_i^2(t)}. \tag{4.20}$$

By Assumption 2.2C there exists a $\gamma > 0$ such that if $\lambda_i(\vec{x}) < \gamma$, then $\lambda_i(\vec{x} + \eta\vec{v}) \ge \lambda_i(\vec{x})$. Increase m_0 so that (recall that $s(\eta)$ was defined in Lemma 2.4)

$$m_0 > \max\{1/\gamma, 1/s(\eta)\}.$$

For the rest of the proof let $m > m_0$. Then, if $\lambda_i(\vec{r}(t)) < \gamma$, we have $\lambda_i^1(t) < \gamma$ and $\lambda_i^2(t) \ge \lambda_i^1(t)$. Therefore, if $\lambda_i(\vec{r}(t)) < \gamma$, then $\log \lambda_i^1(t)/\lambda_i^2(t) < 0$. If, however, $\lambda_i(\vec{r}(t)) > \gamma$, then $\lambda_i(\vec{r}(t)) = \lambda_i^1(t)$, and so, with $\vec{x} = \vec{r}(t)$,

$$\log \frac{\lambda_i^1(t)}{\lambda_i^2(t)} = \log \frac{\lambda_i(\vec{x})}{\lambda_i(\vec{x} + \eta\vec{v})} \tag{4.21}$$

$$\le \log \frac{\lambda_i(\vec{x})}{\lambda_i(\vec{x}) - K_\lambda(\eta)} \tag{4.22}$$

$$\leq \log \frac{\gamma}{\gamma - K_\lambda(\eta)} \tag{4.23}$$

$$= \log \frac{1}{1 - K_\lambda(\eta)/\gamma} \tag{4.24}$$

$$\leq \frac{2K_\lambda(\eta)}{\gamma} \tag{4.25}$$

for $K_\lambda(\eta) < \gamma/2$, because $\log(1/(1-x)) < 2x$ for $0 < x \leq 1/2$. By Shwartz and Weiss [13, Lemma 5.17] and Lemma 3.3 there exist constants C_1, C_3, and B_1 such that if $|\vec{y}| > B_1$, then $\ell(\vec{x}, \vec{y}) \geq C_3 |\vec{y}| \log |\vec{y}|$ and $|\mu^*| \leq C_1 |\vec{y}|$. So, if $|\vec{r}\,'(t)| \geq B_1$, then using (4.19)–(4.25),

$$\ell_2(t) - \ell_1(t) \leq kK'_\lambda(\eta) + C_1 |\vec{r}\,'(t)| \frac{2K_\lambda(\eta)}{\gamma}.$$

But by the estimate above, if $|\vec{r}\,'(t)| > B_1$, then $|\vec{r}\,'(t)| < \ell_1(t)/(C_3 \log |\vec{r}\,'(t)|)$. Thus, if $|\vec{r}\,'(t)| > B_1$, we get

$$\ell_2(t) - \ell_1(t) \leq kK'_\lambda(\eta) + C \frac{\ell_1(t)}{C_3 \log |\vec{r}\,'(t)|} \frac{2K_\lambda(\eta)}{\gamma} \tag{4.26}$$

$$\leq kK'_\lambda(\eta) + C \frac{\ell_1(t)}{C_3 \log B_1} \frac{2K_\lambda(\eta)}{\gamma}. \tag{4.27}$$

Now consider the case $|\vec{r}\,'(t)| \leq B_1$. By Lemma 3.3, $|\mu^*(t)| \leq K_\mu B_1$ for some K_μ. Therefore,

$$\ell_2(t) - \ell_1(t) \leq kK'_\lambda(\eta) + K_\mu B_1 \frac{2K_\lambda(\eta)}{\gamma}.$$

Putting together the estimates for large and small values of $\vec{r}\,'(t)$, we obtain

$$\ell_2(t) - \ell_1(t) \leq \left(kK'_\lambda(\eta) + K_\mu B_1 \frac{2K_\lambda(\eta)}{\gamma} + C \frac{\ell_1(t)}{C_1 \log B_1} \frac{2K_\lambda(\eta)}{\gamma} \right). \tag{4.28}$$

Therefore, if $I^m_{[0,T]}(\vec{r}) \leq K - \varepsilon$, then

$$I_{[0,T]}(\vec{r} + \eta \vec{v}) \leq K - \varepsilon + T \left(kK'_\lambda(\eta) + K_\mu B_1 \frac{2K_\lambda(\eta)}{\gamma} + \frac{C(K-\varepsilon) 2K_\lambda(\eta)}{TC_1 \log B_1 \gamma} \right). \tag{4.29}$$

Thus, for small enough η, $I_{[0,T]}(\vec{r} + \eta \vec{v}) \leq K$. □

LEMMA 4.3. *Suppose that Assumptions 2.1 and 2.2 hold. Let A be a compact set in G. For each $T > 0$, $K > 0$, and $0 < \varepsilon < K$ there is an $0 < \eta_0 < \varepsilon$ such that for any $0 < \eta < \eta_0$ there is an m_0 with the following property. For any $x \in A$, $m > m_0$, and any path $\vec{r}(t)$ with $\vec{r}(0) = \vec{x}$ and $I^m_{[0,T]}(\vec{r}) < K - \varepsilon$, there is a path $\vec{r}_2(t)$ with $\vec{r}_2(0) = \vec{x}$, with $I_{[0,T]}(\vec{r}_2) < K$ and $d(\vec{r}, \vec{r}_2) \leq \varepsilon$.*

This lemma corresponds to Shwartz and Weiss [13, Lemma 5.48]. It is based on a direct construction, having the path $\vec{r}_2$ composed of a number of segments. The main segments are made of shifted pieces of $\vec{r}$, using Lemma 4.2 to estimate the costs of the segments, with the initial part of the path estimated by Lemma 4.1, and the segments stitched together with the aid of Lemma 4.1 as well.

PROOF. Given $\vec{r}(t)$ with $\vec{r}(0) = \vec{x}$ and $I^m_{[0,T]}(\vec{r}) < K - \varepsilon$, take J as defined in Lemma 3.5. Recall from Assumption 2.1(iii) that there is a number $\beta > 0$ such that for any $\vec{x} \in B_i$, where B_i is one of the boundary neighborhoods, then for small ε, we have $d(\vec{x} + \varepsilon \vec{v}_i, \partial G) \geq \beta \varepsilon$. Clearly, $\beta \leq 1$.

We now construct the path $\vec{r}_2$ from $\vec{r}$ using a parameter η that will be chosen later. Let $0 = t_0, t_1, \ldots, t_J$ represent the switchover times of $\vec{r}(t)$, as in Lemma 3.5. Recall the definition of α from the proof of Lemma 3.5. If $B(\vec{r}(0), \alpha) \subset B_i$, where B_i is one of the boundary neighborhoods ($i \le I_1$, cf., Assumption 2.1(i)), then take $\vec{r}_2(t) = \vec{r}(0) + \vec{v}_i t$ for $0 \le t \le \eta$. For $\eta \le t \le t_1 + \eta$, let $\vec{r}_2(t) = \vec{r}(t-\eta) + \eta\vec{v}_i$. Then, for $t_1 + \eta \le t \le t_1 + \eta + 3\eta/\beta$, let

$$\vec{r}_2(t) = \vec{r}(t_1) + \eta\vec{v}_i + (t - t_1 - \eta)\vec{v}_j; \tag{4.30}$$

here j is the index of the neighborhood B_j that $\vec{r}(t)$ switches to at time t_1, as defined in Lemma 3.5. Then, from time $t_1 + \eta + 3\eta/\beta$ until time $t_2 + \eta + 3\eta/\beta$ we let $\vec{r}_2(t) = \vec{r}(t - \eta - 3\eta\beta) + \eta\vec{v}_i + (3\eta/\beta)\vec{v}_j$. We continue in this fashion, with switchover k having $\vec{r}_2(t)$ following a linear path of length $\eta(3/\beta)^{k-1}$ in direction $\vec{v}_k$, followed by a segment parallel to $\vec{r}(t)$. Note that we need not include the paths of length $\eta(3/\beta)^{k-1}$ in direction $\vec{v}_k$ if B_k is not a boundary neighborhood.

There are three things to check about this path. First, if η is small enough, does $\vec{r}_2(t)$ stay within ε of $\vec{r}(t)$? Second, does $\vec{r}_2$ remain in G? Third, does it satisfy $I_{[0,T]}(\vec{r}_2) \le K$? If all three items hold, then the lemma will be proved. These estimates are hardest if all the paths of type $\eta(3/\beta)^{k-1}$ in direction $\vec{v}_k$ are included, so we assume without loss of generality that they are.

The first thing is easy to verify. Because $I^m_{[0,T]}(\vec{r}) < K - \varepsilon$ by assumption, $\vec{r}(t)$ is uniformly absolutely continuous, so by choosing η small enough, we can ensure that the time shifts introduced in the definition of $\vec{r}_2$ do not cause the paths to differ by more than $\varepsilon/2$. Furthermore, the number of segments J is bounded, so the difference introduced in the segments $\eta(3/\beta)^k$ adds up to less than $\varepsilon/2$ if η is chosen small enough.

For the second point, this is the reason we chose $3/\beta$ as a multiplier. By definition of β, the point $\vec{r}_2 + \eta(3/\beta)^k\vec{v}_k$ is at least $\beta\eta(3/\beta)^k$ from ∂G. Because $3/\beta > 3$, the sum of all the previous shifts has total length less than $(\eta\beta/2)(3/\beta)^k$. Therefore, the point is at least $(\eta\beta/2)(3/\beta)^k$ from ∂G when the kth shift is finished. The beginning of the shift also occurs in the interior of G by construction, whenever the total shifts are of length less than $\alpha/4$. This demonstrates that $\vec{r}_2 \in G$ when η is small enough.

For the third point, we assume that η has been chosen small enough to satisfy the first two points, and note that, after the initial time η, the path $\vec{r}_2(t)$ remains at least $\beta\eta/2$ away from ∂G. By Lemma 4.1, there is a uniform constant C' such that for $0 \le t \le \eta$,

$$I_{[0,t]}(\vec{r}_2) \le C't. \tag{4.31}$$

We use this estimate on every $\eta(3/\beta)^k$ path; the total cost is linear in η, so can be made less than $\varepsilon/2$. Lemma 4.2 enables us to bound the cost of each other segment of $\vec{r}_2$ uniformly as no more than $\varepsilon_1(\eta)$ plus the I^m-cost of the corresponding segment $\vec{r}$, where $\varepsilon_1(\eta)$ goes to 0 with η. Choose η small enough so that $\varepsilon_1(\eta) \le \varepsilon/2J$. Then, choose m large enough so that Lemma 4.2 applies. Then, the total additional cost is bounded by $\varepsilon/2$ for the segments of $\vec{r}_2$. This concludes the estimate, and hence the proof. $\square$

LEMMA 4.4. *Under the assumption of bounded continuous λ_i, given $\varepsilon > 0$, there exists an $m_0 > 0$ such that for all positive $m > m_0$ there exists a $\delta_0 > 0$ such that for all $\delta < \delta_0$,*

$$\ell^m(\vec{x}, \vec{y}) \le \varepsilon + (1+\varepsilon)\ell^\Delta(\vec{x}, \vec{y}). \tag{4.32}$$

PROOF. We consider separately the cases $|\vec{y}| > B$ and $|\vec{y}| \le B$. Choose positive B, C so that $\ell(\vec{x}, \vec{y}) > CB\log B$ for all $\vec{x}$ and all $\vec{y} > B$; this is possible by Shwartz and Weiss [13, Lemma 5.17].

If $|\vec{y}| > B$, then, using (4.9) and the reasoning leading to Equation (4.25),

$$\ell^m(\vec{x}, \vec{y}) - \ell^\Delta(\vec{x}, \vec{y}) \le K_\lambda(\delta)\left(k + \frac{2C\ell^\Delta(\vec{x}, \vec{y})}{\gamma C_1 \log B}\right). \tag{4.33}$$

Therefore, by choosing δ small enough, we can make the right-hand side of (4.33) smaller than $\varepsilon/3$.

Now, if $|\vec{y}| \le B$, take μ as the optimal jump rate for $\lambda^{\Delta}(\vec{x})$. We have $\mu_i \le CB$ for all i. Then,

$$\ell^m(\vec{x}, \vec{y}) - \ell^{\Delta}(\vec{x}, \vec{y}) \le \sum_i \lambda_i^m(\vec{x}) - \lambda_i^{\Delta}(\vec{x}) + \mu_i \log \frac{\lambda_i^{\Delta}(\vec{x})}{\lambda_i^m(\vec{x})} \tag{4.34}$$

$$\le k(K_\lambda(\delta) + 1/m) + \sum_i \mu_i \log \frac{\lambda_i^{\Delta}(\vec{x})}{\lambda_i^m(\vec{x})}. \tag{4.35}$$

If $\lambda_i(\vec{x}) \le 1/m - K_\lambda(\delta)$, then $\lambda_i^{\Delta}(\vec{x}) \le 1/m$, so, because $\lambda_i^m(\vec{x}) \ge 1/m$ for each $\vec{x}$, $\log(\lambda_i^{\Delta}(\vec{x})/\lambda_i^m(\vec{x})) \le 0$. Furthermore, if $\lambda_i(\vec{x}) > 1/m - K_\lambda(\delta)$, then

$$\log \frac{\lambda_i^{\Delta}(\vec{x})}{\lambda_i^m(\vec{x})} \le \log \frac{1/m + K_\lambda(\delta)}{1/m} \le mK_\lambda(\delta). \tag{4.36}$$

Hence, by choosing δ small enough so that $kmK_\lambda(\delta) < \varepsilon/3$, we obtain, for $|\vec{y}| \le B$, $\ell^m(\vec{x}, \vec{y}) - \ell^{\Delta}(\vec{x}, \vec{y}) < \varepsilon/3$. Combined with the result of the previous paragraph, this finishes the proof. ☐

Corollary 4.1. *Assume that the λ_i are bounded and continuous. Given ε, K, and $T > 0$ there exists an $m_0 > 0$ such that for all positive $m > m_0$, there exists a $\delta_0 > 0$ so that $\delta < \delta_0$ and $I^{\Delta}_{[0,T]}(\vec{r}) \le K - \varepsilon$ imply $I^m_{[0,T]}(\vec{r}) \le I^{\Delta}_{[0,T]}(r) + \varepsilon$.*

Corollary 4.1 and Lemma 4.3 combine to give the following:

Corollary 4.2. *Suppose that Assumptions 2.1 and 2.2 hold. Given $K > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\Phi^{\Delta}_{\vec{x}}(K - \varepsilon) \subset \{\vec{r}\colon d(\vec{r}, \Phi_{\vec{x}}(K)) \le \varepsilon\}. \tag{4.37}$$

The proof of this corollary is immediate, by using $\varepsilon/2$ to replace ε, and choosing a large enough m. (Specifically, if $\vec{r} \in \Phi^{\Delta}_{\vec{x}}(K - \varepsilon)$, then, by Corollary 4.1, $\vec{r} \in \Phi^m_{\vec{x}}(K - \varepsilon/2)$. By Lemma 4.3, there is an $\vec{r}_2$ within $\varepsilon/2$ of $\vec{r}$ satisfying $I_{[0,T]}(\vec{r}_2) \le K$.) ☐

Corollary 4.2 shows that, by choosing δ and η small enough,

$$d(\vec{r}, \Phi_{\vec{y}}(K - 4\varepsilon)) > \eta/2 \quad \text{implies} \quad d(\vec{r}, \Phi^{\Delta}_{\vec{y}}(K - 4\varepsilon - \eta/4)) > \eta/4. \tag{4.38}$$

We now state two technical lemmas in measure theory that are used in the proof of Lemma 4.7. Lemma 4.5 is used only for the proof of Lemma 4.6. We let $\mathrm{Leb}(A)$ denote the Lebesgue measure of a set A.

Lemma 4.5. *Let $u(t)$ be a nonnegative, absolutely continuous function on $[0, T]$. Then, given $\delta > 0$, there exists an $\eta > 0$, a set $A \subset [0, T]$, and a finite collection $\{C_i\}$ of intervals so that $\mathrm{Leb}(A) < \delta$ and, for each i, either $\inf\{u(t)\colon t \in C_i\} > \eta$ or $u(t) = 0$ for all $t \in C_i \backslash A$.*

Proof. If $u(0) > 0$, set $t_1 \triangleq \inf\{t > 0\colon u(t) = 0\}$. Then, by continuity $\inf\{u(t)\colon 0 \le t \le t_1 - \delta/2\} > 0$, and so it suffices to establish the result when $u(0) = 0$. By a similar argument we may assume $u(T) = 0$.

Given t, if $u(t) > 0$, then by continuity there exists an open interval O_t containing t so that $u(s) \in O_t$ for all $s \in O_t$ and $u(s) \to 0$ as $s \to \partial O_t$. Let $m_t \triangleq \sup\{u(s)\colon s \in O_t\}$. Because u is absolutely continuous, there is a finite number of disjoint open intervals with $m_t > 1/m$. Therefore, there exists a countable collection $\{O_i\}$ of disjoint open intervals so that $u(t) > 0$ if and only if $t \in O_i$ for some i. Fix N large so that

$$\mathrm{Leb}\left\{\bigcup_{N+1}^{\infty} O_i\right\} < \frac{\delta}{2}. \tag{4.39}$$

For $i \le N$, let $\widetilde{C}_i \subset O_i$ be a closed interval such that

$$\mathrm{Leb}\{O_i \backslash \widetilde{C}_i\} \le \frac{\delta}{2N+1}.$$

Let C_i be the finite collection of closed intervals that cover $[0, T] \backslash \bigcup_i \widetilde{C}_i$ and let

$$A \triangleq \left\{ \bigcup_{N+1}^{\infty} O_i \right\} \cup \left\{ \bigcup_{i=1}^{N} O_i \backslash \widetilde{C}_i \right\}.$$

Then, by construction, $\mathrm{Leb}(A) < \delta$ and $u(t) = 0$ on $t \in C_i \backslash A$. Because there are only a finite number of $\widetilde{C}_i$, we obtain by continuity $\inf\{u(t)\colon t \in \widetilde{C}_i \text{ for some } i\} > \eta > 0$. $\square$

For vectors $\vec{\theta}$, $\vec{\lambda}$, and $\vec{y}$ in $\mathbb{R}^d$, define

$$\ell(\vec{\theta}, \vec{\lambda}, \vec{y}) \triangleq \langle \vec{\theta}, \vec{y} \rangle - \sum_j \lambda_j \big(e^{\langle \vec{\theta}, \vec{e}_j \rangle} - 1\big). \tag{4.40}$$

Our next result extends Shwartz and Weiss [13, Lemma 5.43] to the case of rates that are not bounded below, but under the assumption that they are absolutely continuous.

Lemma 4.6. *Assume that the $\lambda_j(\vec{x})$ are bounded and absolutely continuous. Then, for any $\vec{r}$ with $I_{[0,T]}(\vec{r}) < \infty$ and any $\varepsilon > 0$, there exists a step function $\vec{\theta}$ so that*

$$\int_0^T \ell(\vec{\theta}(t), \vec{\lambda}(\vec{r}(t)), \vec{r}'(t))\, dt \ge I_{[0,T]}(\vec{r}) - \varepsilon.$$

Proof. Because by definition

$$\ell(\vec{\theta}(t), \vec{\lambda}(\vec{r}(t)), \vec{r}'(t)) \le \ell(\vec{r}(t), \vec{r}'(t)),$$

it suffices to establish the result outside a set of small measure: this approximation and the extension of the step function over this set are derived in the proof of Shwartz and Weiss [13, Lemma 5.43].

We claim that given δ, there is a partition of $[0, T]$ into a finite collection of intervals C_i and a set A with $\mathrm{Leb}(A) < \delta$ so that the lemma holds on each $C_i \backslash A$. Then, the result follows by patching together the step function and dealing with A as above. In the rest of the proof we establish this claim.

Fix j and apply Lemma 4.5 to the absolutely continuous function $\lambda_j(\vec{r}(t))$ with δ/k. We then obtain a collection $\{C_i^j,\ A^j,\ 1 \le j \le k,\ 1 \le i \le N\}$ so that either $\lambda_j(\vec{r}(t)) = 0$ for all $t \in C_i^j \backslash A^j$ or $\lambda_j(\vec{r}(t)) > \eta_j > 0$ for all $t \in C_i^j$. Set $A \triangleq \bigcup_j A^j$ and $\eta = \min_j \eta_j$. Then, $\mathrm{Leb}(A) < \delta$ and $\eta > 0$. Fix a subset $\alpha \subset \{1, \ldots, k\}$. Using intersections of the sets C_i^j we obtain a finite collection of intervals C_i^α with the following properties:

$$\lambda_j(\vec{r}(t)) = 0, \quad t \in C_i^\alpha \backslash A \text{ for all } j \in \alpha, \tag{4.41}$$

$$\lambda_j(\vec{r}(t)) > \eta, \quad t \in C_i^\alpha \text{ for all } j \notin \alpha. \tag{4.42}$$

In particular, $\lambda_j(\vec{r}(t)) > \eta$ for $t \in C_i^\varnothing$. Now fix i and α and consider the process on C_i^α with rates and jump directions $\{\lambda_j(\vec{x}), \vec{e}_j,\ j \notin \alpha\}$. For this process, the assumptions of Shwartz and Weiss [13, Lemma 5.43] hold. Moreover, if we denote the local rate function for this process by ℓ^α, then, by definition, $\ell^\alpha(\vec{r}(t), \vec{r}'(t)) = \ell(\vec{r}(t), \vec{r}'(t))$ for $t \in C_i^\alpha \backslash A$. Thus, our claim is established and the proof is concluded. $\square$

Now we state and prove that the analogue of Shwartz and Weiss [13, Proposition 5.62] holds for I^Δ. By Shwartz and Weiss [13, Lemma 5.61], under our continuity assumptions, if

C is a compact set in $\mathbb{R}^d$ and $\vec{\theta}(t)$ is a fixed step function, then for any $\delta > 0$ and compact set $\mathcal{K} \subset \mathcal{K}(M)$ of functions $\vec{r}(t)$ with $\vec{r}(0) \in C$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{x}}(\vec{y}_n \in \mathcal{K}_{\vec{x}}) \leq -\inf_{\vec{r}\in\mathcal{K}_{\vec{x}}} I^{\delta}_{[0,T]}(\vec{r}, \vec{\theta}),$$

where $\mathcal{K}_{\vec{x}} \triangleq \{\vec{r} \in \mathcal{K}\colon \vec{r}(0) = \vec{x}\}$. Note that $I^{\Delta} > I^{\delta}$; therefore, we have, under the same assumptions, that

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{x}}(\vec{y}_n \in \mathcal{K}_{\vec{x}}) \leq -\inf_{\vec{r}\in\mathcal{K}_{\vec{x}}} I^{\Delta}_{[0,T]}(\vec{r}, \vec{\theta}). \tag{4.43}$$

LEMMA 4.7. *Assume that the $\lambda_i(\vec{x})$ are bounded and absolutely continuous, and C is a compact set in $\mathbb{R}^d$. Then, for each $K > 0$, $\delta > 0$, and $\varepsilon > 0$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{x}} \left(d\left(\vec{y}_n, \Phi^{\Delta}_{\vec{x}}(K)\right) > \varepsilon \right) \leq -(K - \varepsilon)$$

uniformly in $\vec{x} \in C$.

The proof of this lemma follows exactly the proof of Shwartz and Weiss [13, Lemma 5.62], with the following two exceptions. First, the step function $\vec{\theta}(t)$ is the one defined in Lemma 4.6. Second, every superscript δ is replaced by the corresponding Δ. This is obvious throughout the proof, with the help of (4.43). ☐

We now state and prove the large deviations upper bound for our process. Recall that for a set F of paths, $I_{\vec{x}}(F) = \inf(I_{[0,T]}(\vec{r})\colon \vec{r}(0) = \vec{x}, \vec{r} \in F)$.

THEOREM 4.1. *Suppose that Assumptions 2.1 and 2.2 hold. Let F be a closed set in $(D^d[0,T], d_d)$, and let $\vec{x}$ be a point in G. Then,*

$$\lim_{\vec{y}\to\vec{x}} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{y}}(\vec{z}_n \in F) \leq -I_{\vec{x}}(F), \tag{4.44}$$

where the points $\vec{y}$ are in $F \cap G$.

PROOF. Fix a closed set F, and let $K = I_{\vec{x}}(F)$. Suppose for now that $K < \infty$; the same proof will work for $K = \infty$, but some arguments need minor modifications. By Shwartz and Weiss [13, Lemma 5.63], for given ε there is a δ so that if $|x - y| < \delta$, $I_{\vec{y}}(F) \geq K - \varepsilon$.

Let

$$C = \bigcup_{y\in B_\delta(\vec{x})\cap G} \left\{\Phi_{\vec{y}}(K - 4\varepsilon)\right\}, \tag{4.45}$$

$$F_\delta = \{\vec{r} \in F\colon |\vec{r}(0) - \vec{x}| \leq \delta\}. \tag{4.46}$$

Then, C is a compact set. Note that $C \cap F_\delta = \varnothing$. By Shwartz and Weiss [13, Theorem A.19], there is a number $\eta > 0$ such that $d(C, F_\delta) = \eta$.

Define the random path $\vec{y}_n(t)$ as the linear interpolation of $\vec{z}_n(t)$ with time spacing T/n; that is, at times jT/n for $j = 0, \ldots, n$, $\vec{y}_n(t) = \vec{z}_n(t)$, and $\vec{y}_n(t)$ is linear in between these points. In Shwartz and Weiss [13, Lemma 5.57] states that, for each $\varepsilon > 0$, uniformly in $\vec{x}$ in bounded sets,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{x}}(d(\vec{z}_n, \vec{y}_n) > \varepsilon) = -\infty. \tag{4.47}$$

Now,

$$\mathbb{P}_{\vec{y}}(\vec{z}_n \in F) = \mathbb{P}_{\vec{y}}(\vec{z}_n \in F_\delta) \tag{4.48}$$

$$\leq \mathbb{P}_{\vec{y}}(d(\vec{y}_n, F_\delta) < \eta/2) + \mathbb{P}_{\vec{y}}(d(\vec{y}_n, \vec{z}_n) \geq \eta/2). \tag{4.49}$$

Equation (4.47) shows that the second term on the right-hand side of this inequality is negligible.

If $\vec{r}(0)=\vec{y}$, then, by definition of η,

$$d(\vec{r}, F_\delta) < \eta/2 \quad \text{implies} \quad d(\vec{r}, \Phi_{\vec{y}}(K-4\varepsilon)) > \eta/2. \tag{4.50}$$

Therefore, by Equation (4.38) and Lemma 4.7,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{y}}(d(\vec{y}_n, F_\delta) < \eta/2) \tag{4.51}$$

$$\le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{y}}\left(d\left(\vec{y}_n, \Phi^{\Delta}(K-4\varepsilon-\eta/4)\right) \ge \eta/4\right) \tag{4.52}$$

$$\le -(K-4\varepsilon-\eta/4) \tag{4.53}$$

uniformly in $\vec{y}$ in a compact set. Therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}_{\vec{y}}(\vec{z}_n \in F) \le -(K-4\varepsilon-\eta/4)$$

whenever $|\vec{y}-\vec{x}| \le \delta$. Because ε and η can be made arbitrarily small, the theorem is proved. □

5. Lower bound—boundary case. Our approach to the proof of the lower bound, Equation (2.8), is mainly standard. Every path $\vec{r}$ in an open set O can be surrounded by a "sausage," a neighborhood of $\vec{r}$, that is entirely contained in O. If we can show that the probability that $\vec{z}_n$ lies in this sausage is about $\exp(-nI_{[0,T]}(\vec{r}))$, then the lower bound will be proved, for if we take a sequence of $\vec{r}$ whose I functions are approximately minimal in O, then we find that the probability that $\vec{z}_n \in O$ is at least the probability that $\vec{z}_n \in$ the sausage around $\vec{r}$, which is about $\exp(-nI_{[0,T]}(\vec{r}))$.

The novelty in the proof is a twofold estimate. The first step is to show that, for any path $\vec{r}$ that lies entirely in a single neighborhood B_i (see Assumption 2.1(i)), the probability that $\vec{z}_n$ is near $\vec{r}$ is approximately $\exp(-nI_{[0,T]}(\vec{r}))$. This is done by showing that the path $\vec{r}+\delta\vec{v}_i$, which lies strictly away from the boundary, has rate function $I(\vec{r}+\delta\vec{v}_i) \approx I(\vec{r})$. Then, because the $\lambda_j(\vec{x})$ are bounded away from zero on this path, existing lower bound theory shows that the probability of $\vec{z}_n$ being near this new path is at least $\exp(-nI(\vec{r}+\delta\vec{v}_i))$. Using $I(\vec{r}+\delta\vec{v}_i) \approx I(\vec{r})$ proves the result for such paths $\vec{r}$. The second step is to show that every path $\vec{r}$ with finite cost $I_{[0,T]}(\vec{r})$ can be decomposed into a finite number of pieces $\vec{r}_j$, each of which lies entirely within a ball B_j, and that the endpoints of the shifted pieces $\vec{r}_j+\delta\vec{v}_j$ can be connected with asymptotically negligible cost. This is mostly the same as Lemma 4.3.

Lemma 5.1. *Let Assumptions 2.1 and 2.2 hold. Suppose that $\vec{r}(t)$ is a path contained in a single B_i with $I_{[0,T]}(\vec{r}) < \infty$. Let $\vec{r}_\delta(t) = \vec{r}(t)+\delta\vec{v}_i$, where $\vec{v}_i$ is the direction in Assumption 2.1*(iii) *for the region B_i. Then,*

$$\limsup_{\delta\to 0} I_{[0,T]}(\vec{r}_\delta) \le I_{[0,T]}(\vec{r}). \tag{5.1}$$

Proof. This is proved in exactly the same way as Lemma 4.2. Let $\ell_1(t)=\ell(\vec{r}(t),\vec{r}'(t))$ and $\ell_2(t)=\ell(\vec{r}_\delta(t),\vec{r}_\delta'(t))$. Then, letting $\vec{\mu}(t)$ be the optimizing set of jump rates for the path $\vec{r}$, we have the equivalent of (4.19):

$$\ell_2(t)-\ell_1(t) \le k\cdot K'_\lambda(\eta) + \sum_{i=1}^{k} \mu_i^*(t) \log \frac{\lambda_i^1(t)}{\lambda_i^2(t)}.$$

The same reasoning as in Lemma 4.2 then leads to a bound like (4.28), which immediately leads to the result. □

LEMMA 5.2. *Suppose that Assumptions 2.1 and 2.2 hold. Using the notation therein, fix i, $T > 0$, and a path $\vec{r}(t)$ which takes values in B_i such that $I_{[0,T]}(\vec{r}) = K < \infty$. Then (with $\vec{r}(0) = \vec{x}$),*

$$\lim_{\delta \downarrow 0} \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x \left(\vec{z}^n(t) \in B(\vec{r}, \delta) \right) \geq -I_{[0,T]}(\vec{r}). \tag{5.2}$$

PROOF. Let $\vec{v} = \vec{v}_i$ be the direction of the interior cone (see Assumptions 2.1 and 2.2) of B_i and let K_i be the constant such that for $\vec{x} \in B_i$, we have $d(\vec{x} + t\vec{v}, \partial G) \geq K_i t$ for t small. Denote by $\tilde{\eta}$ the modulus of continuity of $\vec{r}$ and set $\eta(a) = \max\{\tilde{\eta}(a), a\}$ so that $\eta^{-1}(a) \leq a$.

Now, fix δ and set $t_\delta = \eta^{-1}(\delta/3)$. Then, $t_\delta \leq \delta$ and for $t \leq t_\delta$,

$$\sup_{0\leq t\leq t_\delta} |t \cdot \vec{v} - \vec{r}(t)| \leq t_\delta \cdot |\vec{v}| + \eta(t_\delta) \leq \frac{2\delta}{3}. \tag{5.3}$$

Therefore, for $0 < \alpha < K_i/6$,

$$\mathbb{P}_x \left(\vec{z}^n \in B(\vec{r}, \delta) \right) \geq \mathbb{P}_x \left(|\vec{z}^n(t) - t \cdot \vec{v}| \leq \alpha\delta,\ 0 \leq t \leq t_\delta,\ \vec{z}^n \in B(\vec{r}, \delta) \right), \tag{5.4}$$

where the last ball is around the restriction of $\vec{r}$ to $[t_\delta, T]$. Now, let $r_\delta(t) \triangleq \vec{r}(t) + t_\delta \vec{v}$ be a function on $[t_\delta, T]$. Then, on this time interval

$$\sup_{t_\delta \leq t \leq T} |\vec{r}(t) - \vec{r}_\delta(t)| \leq \frac{\delta}{3}$$

and, moreover, $d(\vec{r}_\delta(t), \partial G) \geq K_i \delta$. Therefore, for any function $\vec{u}$ on that time interval, $\|\vec{u} - \vec{r}_\delta\| \leq K_i\delta/2$ implies that $\|\vec{u} - \vec{r}\| \leq 2\delta/3$ and $d(\vec{u}(t), \partial G) \geq K_i\delta/2$. Now, let $B_\delta \triangleq B(\vec{x} + t_\delta \vec{v}, \alpha\delta)$ and let $\vec{r}_\delta^y$ be the shift of $\vec{r}_\delta$ so that $\vec{r}_\delta^y(t_\delta) = \vec{y}$. Then,

$$\begin{aligned} \mathbb{P}_x(\vec{z}^n \in B(\vec{r}, \delta)) &\geq \mathbb{P}_x(|\vec{z}^n(t) - t \cdot \vec{v}| \leq \alpha\delta,\ 0 \leq t \leq t_\delta) \\ &\quad \times \inf_{y\in B_\delta} \mathbb{P}_y(\vec{z}^n \in B(\vec{r}_\delta^y, K_i\delta/2)), \end{aligned} \tag{5.5}$$

where the last ball contains paths on $[t_\delta, T]$. Now, the first term is bounded below by the probability that, over $0 \leq t \leq t_\delta$, the rates for jumps in directions outside the cone of Assumption 2.2C are zero, while the rates for jumps in directions within the cone are such that the process proceeds with speed one. However, the second condition satisfies a standard large deviations lower bound, and so

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}_x(|\vec{z}^n(t) - t \cdot \vec{v}| \leq \alpha\delta,\ 0 \leq t \leq t_\delta) \geq -Ct_\delta \tag{5.6}$$

for some constant C. Now consider the second probability in (5.5). Because the paths in $B(\vec{r}_\delta^y, K_i\delta/2)$ are bounded away from the boundary uniformly in $\vec{y} \in B_\delta$, we have by Shwartz and Weiss [13, Theorem 5.51], that the large deviations lower bound holds uniformly over $\vec{y} \in B_\delta$ (where uniformity is the usual sense of analysis). Therefore,

$$\liminf_{n\to\infty} \frac{1}{n} \log \inf_{y\in B_\delta} \mathbb{P}_y(\vec{z}^n \in B(\vec{r}_\delta^y, K_i\delta/2)) \tag{5.7}$$

$$\geq -\inf_{y\in B_\delta} \inf\{I_{[t_\delta, T]}(\vec{w}) \colon \vec{w} \in B(\vec{r}_\delta^y, K_i\delta/2)\} \tag{5.8}$$

$$\geq -I_{[t_\delta, T]}(\vec{r}_\delta) \tag{5.9}$$

$$\geq -I_{[0,T]}(\vec{r}_\delta) \tag{5.10}$$

$$\geq -I_{[0,T]}(\vec{r}), \tag{5.11}$$

where the final inequality comes from Lemma 5.1. Thus, the result follows from (5.6)–(5.7). □

Proof of the Lower Bound of Theorem 2.1. Using Lemma 5.2 the proof of the lower bound is almost identical to that of Lemma 4.3, and so we only give a sketch of the proof. Given a path $\vec{r}$ with $I_{[0,T]}(\vec{r}) = K < \infty$, we break the path into J segments and use the shift-and-stitch argument of Lemma 4.3. Lemma 5.2 shows that the shifted path provides a good approximation for the rate and a lower bound for the probabilities, and that the probabilities of the shifts are bounded below (on the exponential scale) by an arbitrarily small constant. This establishes the lower bound. □

6. Interior zeros. In this section, we state and prove a large deviations principle for processes that have rates that may become zero in the interior of the region G, not at a boundary. Our main assumption about these processes is that the cone $\mathscr{C}_{\vec{x}}$ of jump directions of the process does not change, but remains $\mathbb{R}^d$ for all $\vec{x}$. This is a strong assumption. But the result is strong, too: under Lipschitz continuity of the jump rates $\lambda_i(\vec{x})$, the large deviations principle holds, with the usual rate function. This is, therefore, a strict generalization of the large deviations principle proved in Shwartz and Weiss [13, Chapter 5] where the assumption is that the logarithms of the jump rates are bounded, which itself implies $\mathscr{C}_{\vec{x}} = \mathbb{R}^d$ for all $\vec{x}$. Combined with the results for rates that diminish towards a boundary, we obtain a fairly general theory of diminishing rates, Corollary 2.1.

We state our theorem for Lipschitz continuous jump rates and processes that have no boundaries. But, using the exponential tightness argument of Corollary 2.3, we give proofs only for bounded regions and bounded Lipschitz jump rates. Our proofs are based very heavily on the arguments in Shwartz and Weiss [13]. Rather than reproduce those arguments, we give only the lemmas and arguments needed to extend the previous proof to the present case.

We begin with some notation. For $\delta > 0$, we define

$$\lambda_i^{\delta}(\vec{x}) \triangleq \begin{cases} \lambda_i(\vec{x}) & \text{if } \lambda_i(\vec{x}) > \delta, \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

We define $\mathscr{C}_{\vec{x}}^{\delta}$ as the positive cone spanned by the $\vec{e}_i$ with corresponding $\lambda_i^{\delta}(\vec{x}) > 0$. Let $\mathscr{I}^{\nu}(\vec{x}) = \{i\colon \lambda_i(\vec{x}) > \nu\}$, so that $\mathscr{I}^0(\vec{x}) = \{i\colon \lambda_i(\vec{x}) > 0\}$. Also, for a set $\mathscr{I}$ of indices, we let $\mathscr{C}_{\vec{x}}(\mathscr{I})$ be the positive cone spanned by $\{\vec{e}_i\colon i \in \mathscr{I}, \lambda_i(\vec{x}) > 0\}$.

Lemma 6.1. *Suppose that Assumption 2.3 holds, and that the $\lambda_i(\vec{x})$ are continuous. Then, for every $R < \infty$, there exist $\delta > 0$ and $\nu > 0$ such that for $|\vec{x}| \le R$, we have a set $\mathscr{I}(\vec{x})$ with the following property. Every $\vec{z} \in B(\vec{x}, \delta)$ has $\lambda_i(\vec{z}) > \nu$, $i \in \mathscr{I}(\vec{x})$, and $\mathscr{C}_{\vec{z}}(\mathscr{I}(\vec{x})) = \mathbb{R}^d$.*

Note. We use this ν in following lemmas and proofs.

Proof. This follows easily from compactness. The set S consisting of $\vec{x}$ such that at least one $\lambda_i(\vec{x}) = 0$, is closed. Therefore, the set of such $\vec{x}$ that satisfy $|\vec{x}| \le R$ is compact. Cover each point $\vec{x}$ in S with an open ball with radius δ chosen so that the minimal rate of $\lambda_i(\vec{z})$, $i \in \mathscr{I}^0(\vec{x})$, $\vec{z} \in B(\vec{x}, \delta)$ is at least half the minimal rate of $\lambda_i(\vec{x}), i \in \mathscr{I}^0(\vec{x})$. By assumption, $\mathscr{C}_{\vec{x}} = \mathbb{R}^d$, so $\mathscr{C}_{\vec{z}}(\mathscr{I}^0(\vec{x})) = \mathbb{R}^d$ for all $\vec{z} \in B(\vec{x}, \delta)$. Choose a finite subcover of such balls $\{B(\vec{x}_j, \delta)\}$. Call the resulting union of these sets U. Then,

$$\inf\{\lambda_i(\vec{z})\colon \vec{z} \in B(\vec{x}_j, \delta_j),\ i \in \mathscr{I}^0(\vec{x}_j)\} \triangleq \eta > 0$$

by construction. Then, for $\vec{x} \notin U$, $\lambda_i(\vec{x}) > 0$. In fact, there is a positive bound, which without loss of generality we take to be η such that $\lambda_i(\vec{x}) \ge \eta$ for all $\vec{x}$: $|\vec{x}| \le R, \vec{x} \notin U$, because this set is closed and the rates are continuous and not equal to 0. Now set $\nu = \eta/2$. The proof is concluded by showing that there is a δ so that for each $\vec{x} \in U$,

$$B(\vec{x}, \delta) \cap U \subset B(\vec{x}_j, \delta_j) \quad \text{for some } j; \quad \text{set } \mathscr{I}(\vec{x}) = \mathscr{I}(\vec{x}_j) \tag{6.2}$$

$$\text{or} \quad d(\vec{x}, \partial U) \le \delta; \quad \text{set } \mathscr{I}(\vec{x}) = \{1, \ldots, J\}. \tag{6.3}$$

The first case, Equation (6.2), obviously satisfies the statement of the lemma. In the second case, Equation (6.3), because the rates are continuous over a compact set, they are uniformly continuous, and we choose δ so that the rates change by at most $\eta/2$ over the δ ball. We establish the claim by contradiction; assume the contrary. Then, there is a sequence $\vec{x}_i$ and $\delta_i \downarrow 0$ so that $\vec{x}_i \in U$ and the ball $B(\vec{x}_i, \delta_i)$ is not contained in any $B(\vec{x}_j, \delta_j)$. Take a converging subsequence with limit $\vec{x}$. Then, if $\vec{x} \in U$, we obtain a contradiction, because U is a finite union of balls, so that, for large i, the ball around $\vec{x}_i$ must be in some $B(\vec{x}_j, \delta_j)$. If $\vec{x}$ is on the boundary of U, then for large i, $\vec{x}_i$ is within δ of ∂U. $\square$

We define $\vec{\mu}^*(\vec{x}, \vec{y})$ as the unique optimizer for $f(\vec{\mu}, \vec{\lambda})$; that is, $\vec{\mu}$ causes $\ell(\vec{x}, \vec{y}) = f(\vec{\mu}, \vec{\lambda}(\vec{x}))$, with $\vec{y} = \sum_i \mu_i \vec{e}_i$. We define $\vec{\mu}^*_\delta(\vec{x}, \vec{y})$ to be the optimizer for rates λ^δ. We also define $\ell_\delta(\vec{x}, \vec{y})$ (not to be confused with ℓ^δ!) to be the rate function with jump rates λ^δ.

Lemma 6.2. *Suppose that Assumption 2.3 holds and that the λ_i are continuous. The maximizing $\vec{\theta}$ in the definition of ℓ is bounded uniformly in $|\vec{x}| \le R$ and $|\vec{y}| \le C$.*

Proof. The proof is similar to Shwartz and Weiss [13, Lemma 5.21]. By Lemma 6.1,

$$\max\{\langle \vec{\theta}, \vec{e}_i \rangle \colon i \in \mathcal{I}^\nu(\vec{x})\} \ge \alpha |\vec{\theta}|$$

for some $\alpha > 0$ that depends only on R. Let $\vec{\theta}_n$ be a maximizing sequence. Representing $\vec{y}$ as in Lemma 3.2, and using Lemma 6.1,

$$\ell(\vec{x}, \vec{y}) = \lim_{n\to\infty} \langle \vec{\theta}_n, \vec{y} \rangle - \sum_{i=1}^{k} \lambda_i(\vec{x})(e^{\langle \vec{\theta}_n, \vec{e}_i \rangle} - 1) \tag{6.4}$$

$$\le \lim_{n\to\infty} |\vec{\theta}_n| C - \nu e^{\alpha |\vec{\theta}_n|} + \sum_{i=1}^{k} \lambda_i(\vec{x}). \tag{6.5}$$

However, the last sum is bounded and the function $ax + c - e^x$ diverges to $(-\infty)$ as $x \to \infty$. Because ℓ is nonnegative, we conclude that $|\vec{\theta}_n|$ must be bounded for large n, where the bound depends only on R, C, ν, and $\mathcal{I}^\nu$. $\square$

Corollary 6.1. *Suppose that Assumption 2.3 holds and that the λ_i are continuous. For each B, there exists a C such that for all $|\vec{x}| \le B$ and all $|\vec{y}| \le B$,*

$$\ell(\vec{x}, \vec{y}) \le C. \tag{6.6}$$

Proof. In Shwartz and Weiss [13, Theorem 5.26 and Exercise 5.30] show that the optimizing μ^* can be represented as $\mu_i^* = \lambda_i e^{\langle \vec{\theta}^*, \vec{e}_i \rangle}$, where $\vec{\theta}^*$ is the optimizing $\vec{\theta}$ in the definition of ℓ. Therefore, because Lemma 6.2 shows that $\vec{\theta}^*$ is bounded for bounded $\vec{y}$, we have μ_i^*/λ_i is bounded for bounded $\vec{y}$. Therefore, for optimizing μ^*, there is a constant u such that

$$\lambda_i - \mu_i^* + \mu_i^* \log \frac{\mu_i^*}{\lambda_i} \le \lambda_i u. \quad \square \tag{6.7}$$

Lemma 6.3. *Suppose that Assumption 2.3 holds and that the $\lambda_i(\vec{x})$ are continuous. Then, for each B and $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $|\vec{y}| \le B$, $\vec{x}_\delta \in B(\vec{x}, \delta)$, and $\vec{y}_\delta \in B(\vec{y}, \delta)$, we have*

$$|\ell(\vec{x}, \vec{y}) - \ell_\delta(\vec{x}_\delta, \vec{y}_\delta)| < \varepsilon.$$

Proof. By Lemma 3.1,

$$\ell(\vec{x}, \vec{y}) \triangleq \inf_{\mu \in K_y} \left(\sum_{i=1}^{k} \lambda_i(\vec{x}) - \mu_i + \mu_i \log \frac{\mu_i}{\lambda_i(\vec{x})} \right) \tag{6.8}$$

$$\le \inf_{\mu \in K_y} \left(\sum_{i:\lambda_i^\delta > 0} \lambda_i^\delta(\vec{x}) - \mu_i + \mu_i \log \frac{\mu_i}{\lambda_i^\delta(\vec{x})} \right) + \sum_{i:\lambda_i^\delta = 0} \lambda_i \tag{6.9}$$

$$\le \ell_\delta(\vec{x}, \vec{y}) + k\delta, \tag{6.10}$$

where the first inequality holds because the infimum is taken over a smaller set, where $\lambda_i^\delta = 0$ implies $\mu_i = 0$. We claim that to conclude the proof it suffices to establish that

$$\ell_\delta(\vec{x}_\delta, \vec{y}_\delta) \le \ell(\vec{x}, \vec{y}) + \varepsilon, \tag{6.11}$$

because using (6.8) and interchanging the roles of $\vec{x}, \vec{y}$ with $\vec{x}_\delta, \vec{y}_\delta$ in (6.11) and then using (6.8) again, we get

$$\begin{aligned} \ell_\delta(\vec{x}_\delta, \vec{y}_\delta) &\ge \ell(\vec{x}_\delta, \vec{y}_\delta) - k\delta && (6.12)\\ &\ge \ell_\delta(\vec{x}, \vec{y}) - k\delta - \varepsilon && (6.13)\\ &\ge \ell(\vec{x}_\delta, \vec{y}_\delta) - 2k\delta - \varepsilon. && (6.14)\end{aligned}$$

The difficulty in demonstrating (6.11) is that the $\vec{\mu}_\delta^*$ might be very different from the $\vec{\mu}^*$. But we can bound this difference, using the property that $\mathscr{C}_{\vec{x}} = \mathbb{R}^d$ for all $\vec{x}$. Let $\mathcal{I}^\nu(\vec{x})$ be the set of i with $\lambda_i(\vec{x}_\delta) \ge \nu$ for all $\vec{x}_\delta$ near $\vec{x}$; see Lemma 6.1. Note that for each δ, there is a uniformly bounded vector $\vec{\alpha}$ with $\alpha_i = 0$ for $i \notin \mathcal{I}^\nu(\vec{x})$, such that

$$\sum_i (\mu_i^* + \delta\alpha_i)\vec{e}_i = \vec{y}_\delta. \tag{6.15}$$

There may be some nonzero components of the μ_i for $i \notin \mathcal{I}^\nu(\vec{x})$. However, by Lemma 3.3, because $\vec{y}_\delta$ is bounded, for any $\eta > 0$ we may choose δ small enough such that if any of the $\lambda_i^\delta = 0$, then by continuity, $\lambda_i < \eta$. Take

$$\vec{v}_\delta = \sum_{i \notin \mathcal{I}} \mu_i^* \vec{e}_i. \tag{6.16}$$

Then, $\vec{v}_\delta \in \mathscr{C}_{\vec{z}}^\delta$ for all $\vec{z} \in B(\vec{x}, \delta)$, so by Lemma 3.2 we may write

$$\vec{v}_\delta = \sum_{i \in \mathcal{I}} a_i \vec{e}_i \tag{6.17}$$

for some $a_i \ge 0$, $|a_i| \le \kappa\delta$. Therefore, we may write

$$\vec{y}_\delta = \sum_{i \in \mathcal{I}} (\mu_i^* + a_i + \delta\alpha_i)\vec{e}_i. \tag{6.18}$$

We have

$$\ell(\vec{x}_\delta, \vec{y}_\delta) \le \sum_{i \in \mathcal{I}} \lambda_i(\vec{x}_\delta) - \mu_i^* - a_i - \delta\alpha_i + (\mu_i^* + a_i + \delta\alpha_i) \log \frac{\mu_i^* + a_i + \delta\alpha_i}{\lambda_i(\vec{x}_\delta)}. \tag{6.19}$$

Note that

$$\ell(\vec{x}, \vec{y}) = \sum_i \lambda_i(\vec{x}) - \mu_i^* + \mu_i^* \log \frac{\mu_i^*}{\lambda_i}. \tag{6.20}$$

Both $\vec{\alpha}$ and $\vec{a}$ are of size δ, so the difference between corresponding terms in the sums (6.19) and (6.20) can be bounded by terms that go to zero with δ. Furthermore, by definition of $\mathcal{I}$, $\lambda_i(\vec{z}) \ge \nu$ for all $\vec{z}$ near $\vec{x}$. This finishes the estimates. $\square$

Proof of Theorem 2.2. The large deviations upper bound follows from these arguments based on the argument in Shwartz and Weiss [13]. The key theorem there is Theorem 5.64. It is based on Lemmas 5.57, 5.58, 5.63, 5.62, and 5.48 there. Lemmas 5.57 and 5.58 require only bounded rates λ_i, and Lemma 5.63 requires bounded continuous rates, so these three lemmas continue to hold. Lemma 5.62 is based on Lemma 5.43, which required log-bounded rates, but is established without the lower bound on the rates in our Lemma 4.6, which requires only absolutely continuous jump rates λ_i. Finally, in Shwartz and Weiss [13, Lemma 5.48] is based on Lemma 5.35 there, which also requires log-bounded rates.

Our new Lemma 6.3 replaces Lemma 5.35. So, under the assumption that the $\lambda_i(\vec{x})$ are bounded, absolutely continuous, and that $\mathscr{C}_{\vec{x}} = \mathbb{R}^d$ for all $\vec{x}$, the large deviations upper bound is proved.

The large deviations lower bound follows even more directly. Assuming that the rates $\lambda_i(\vec{x})$ are bounded and Lipschitz continuous, so that Kurtz's theorem (Shwartz and Weiss [13, Theorem 5.3]) applies, essentially the same proof of the lower bound in Theorem 5.51, goes through. The only place that log-boundedness is used is in Corollary 5.53 there, and using Lemma 6.3 it is easily seen to hold without that assumption. Using Lemma 6.3, we see that the rate function $\ell(\vec{x}, \vec{y})$ is jointly continuous in $\vec{x}$ and $\vec{y}$ over bounded regions. Approximating jump rates with λ_i^δ makes a small (δ) difference in the rate function; this can be made arbitrarily small. So, under the assumption that the $\lambda_i(\vec{x})$ are bounded, Lipschitz continuous, and that $\mathscr{C}_{\vec{x}} = \mathbb{R}^d$ for all $\vec{x}$, the large deviations lower bound is proved. □

7. Reachability of the boundary. We now provide a simple condition for showing that a point $\vec{x} \in \partial G$ may be reached with a finite cost (hence, exponentially nonzero probability) via a path from the interior of G. We state a sufficient condition that is far from necessary; nevertheless, it is general enough to cover many cases of interest. For the one-dimensional case, we prove that a similar condition is also necessary. Recall that $s(\delta)$ is the scale function defined in Lemma 2.4.

LEMMA 7.1. *Assume that the rates λ_i are bounded and let Assumption 2.2C hold. Fix $\vec{x} \in \partial G$. If*

$$\int_0^\delta \log \frac{1}{s(t)}\, dt < \infty \tag{7.1}$$

for some $\delta > 0$, then $I_{[0,T]}(\vec{r}) < \infty$ for some $\vec{r}$ with $\vec{r}(T) = \vec{x}$.

PROOF. For convenience we shift time. Let $\vec{r}(t)$ be a path with the following properties: $\vec{r}(0) = \vec{x}$; for some $c > 0$ and $t_0 < 0$, we have $d(\vec{r}(t), \partial G) > c|t|$ for $t_0 \le t \le 0$; and, for some $C > 0$, $|\vec{r}\,'(t)| < C$ for almost all $t \in [t_0, 0)$. Because $|\vec{r}\,'(t)| < C$ for almost all $t \in [t_0, 0)$, by Lemma 3.3, $\mu_i \le C_0$, where $\mu_i(t)$ is the optimal change of measure for $\ell(\vec{r}(t), \vec{r}\,'(t))$. Because we assume $d(\vec{r}(t), \partial G) > c|t|$ for $t_0 \le t \le 0$, we have from (3.1) that

$$\ell(\vec{r}(t), \vec{r}\,'(t)) \le C + C \log \frac{1}{s(d(\vec{r}(t), \partial G))}, \tag{7.2}$$

because both λ_i and μ_i are bounded. But $d(\vec{r}(t), \partial G) > c|t|$. Therefore,

$$\ell(\vec{r}(t), \vec{r}\,'(t)) \le C + C \log \frac{1}{s(ct)}. \tag{7.3}$$

This proves the result. □

Note that the condition for reachability holds for any polynomially decreasing scale function, because $\int_0^1 1/(\log x^j)\, dx < \infty$. In particular, in the case of an infinite server queue, where the rates decrease linearly to zero at the boundary, the boundary may be reached in finite time at finite cost. Note also that the condition is tight under the assumption that $|\vec{r}\,'(t)| < C$: the linear rate of approaching the boundary is optimal under this bound, as can be seen by a change of variable argument.

In the case of one-dimensional processes, the condition on reachability is virtually necessary as well as sufficient. This result is interesting enough that we detail it here. We suppose without loss of generality that the boundary is $x = 0$, and that the interior of G is contained in $x > 0$. Let $\mu(x)$ denote the sum of the jump rates in negative directions, and $\Lambda(x)$ denote the sum in positive directions. We consider a path $r(t) = b - at$ for $t \in [0, T = b/a]$, where we suppose that $[0, b] \in G$.

LEMMA 7.2. *Consider the one-dimensional case. Let Assumption 2.2C hold. Assume that the rates λ_i are bounded and that at least one rate of jump away from the boundary is* *bounded away from* 0. *Then the boundary can be reached by the path* $r(t) = b - at$ *with finite cost under the following condition*:

$$I_{[0,T]}(r) < \infty \quad \textit{if and only if} \quad \int_0^b \log \frac{1}{\mu(x)}\, dx < \infty. \tag{7.4}$$

PROOF.

$$\ell(r, r') = \sup_\theta \left(-a\theta - \sum_i \lambda_i(r)(\exp(\theta e_i) - 1) \right). \tag{7.5}$$

Let $\mathcal{J}^+$ be the i with $e_i > 0$, and let $\mathcal{J}^-$ be the i with $e_i < 0$. Differentiating (7.5) with respect to θ and setting the result equal to zero, we see

$$-a - \sum_{i\in\mathcal{J}^+} \lambda_i(r) e_i e^{\theta e_i} - \sum_{i\in\mathcal{J}^-} \lambda_i(r) e_i e^{\theta e_i} = 0, \tag{7.6}$$

or, equivalently,

$$\sum_{i\in\mathcal{J}^-} \lambda_i(r)(-e_i) e^{\theta e_i} = a + \sum_{i\in\mathcal{J}^+} \lambda_i(r) e_i e^{\theta e_i}. \tag{7.7}$$

Both sides of (7.7) are positive, consisting of all positive terms. As $r \to 0$, we have $\lambda_i(r) \to 0$ for all $i \in \mathcal{J}^-$. Therefore, we have $\theta \to -\infty$ as $r \to 0$; recalling that for at least one $j \in \mathcal{J}^+$ we have $\lambda_j(0) > 0$, we see that the rate at which $\theta \to -\infty$ as $r \to 0$ is bounded below, independent of a.

Let

$$h \triangleq \min_{i\in\mathcal{J}^-} |e_i| \quad \text{and} \quad H \triangleq \max_{i\in\mathcal{J}^-} |e_i|, \tag{7.8}$$

$$\varepsilon(r) \triangleq \sum_{i\in\mathcal{J}^+} \lambda_j(r) e_i e^{\theta e_i}. \tag{7.9}$$

Then, $\varepsilon(r) \to 0$ as $r \to 0$, because $\theta \to -\infty$ as $r \to 0$. By (7.7), we have

$$\mu(r) h e^{\theta h} \le a + \varepsilon(r) \le \mu(r) H e^{\theta H} \tag{7.10}$$

(recall that $\mu(r) = \sum_{i\in\mathcal{J}^-} \lambda_i(r)$). Therefore, for each $\delta > 0$, when r is close enough to zero, by using (7.10) to bound θ, we obtain

$$\ell(r, r') \le \frac{a}{h} \log \frac{a + \varepsilon(r)}{\mu h} + \mu + \Lambda + \delta, \tag{7.11}$$

$$\ell(r, r') \ge \frac{a}{H} \log \frac{a + \varepsilon(r)}{\mu H} + \mu + \Lambda - \delta - \frac{a + \varepsilon(r)}{h}; \tag{7.12}$$

the last term on the right of (7.12) is derived from (7.7) by the estimate

$$\sum_i \lambda_i(r) e^{\theta e_i} \le \frac{a + \varepsilon(r)}{h}.$$

Therefore,

$$I_{[0,T]}(r) = \int_0^T \left(\log \frac{a}{\mu(r(t))} + O(1) \right) dt \tag{7.13}$$

$$= \int_0^b \left(\log \frac{1}{\mu(r)} + O(1) \right) dt. \tag{7.14}$$

This shows that $I_{[0,T]}(r)$ is finite if and only if (7.14) is. $\square$

8. Conclusion. We have shown that the usual sample-path jump Markov large deviations theorem and rate function remain unchanged even when jump rates tend to zero in some cases. The cases include the very important one of infinite server queues, as well as the case where the positive cone of jump directions does not change. Moreover, while existing theories that include boundaries assume flat boundaries, the boundaries here are quite general.

However, our understanding of diminishing rate is not complete, as illustrated in Example C in the introduction. Moreover, a theory combining diminishing rates with discontinuous rates, even with flat boundaries, or finite level boundaries, is still lacking.

Appendix A. Extensions to nonconvex sets. In §2, we claimed that it is possible to extend the rates $\lambda_i(x)$ from $x \in G$ to $x \in \mathbb{R}^d$ when G is convex, in such a way that the extended rates are Lipschitz continuous (below Lemma 2.1). We now establish this for a more general class of sets. Let G be a closed set that is the union of a finite number of closed convex sets G_j. We do not assume that the G_j are compact. We use the following notation. For any point x, denote by $y_j(x)$ the projection of x onto G_j.

Lemma A.1 (Contraction). *For any x, z, and j,*

$$|y_j(x) - y_j(z)| \le |x - z|. \tag{A.1}$$

Proof. Obvious, but can be made precise as follows. The only nontrivial case is where both are outside G_j. But then,

$$|x - y_j(z)|^2 > |x - y_j(x)|^2 + |y_j(x) - y_j(z)|^2 \tag{A.2}$$

because the angle $xy_j(x)y_j(z)$ is necessarily larger than 90 degrees. Write the symmetric expression and sum up to obtain the result. □

Lemma A.2 (Continuity). *Let f be a Lipschitz function on G with Lipschitz constant L_f. Define the function g through*

$$g(x) = \begin{cases} f(x) & \text{for } x \in G, \\ \dfrac{\sum_{j=1}^{J} (d[x, y_j(x)])^{-1} f(y_j(x))}{\sum_{j=1}^{J} (d[x, y_j(x)])^{-1}} & \text{for } x \notin G. \end{cases} \tag{A.3}$$

Then, g is continuous.

Proof. Fix an arbitrary point x and note that it suffices to prove continuity locally at x. This is obvious if x is in the interior of G or in the interior of its complement. Because G is closed, the only case to consider is how g changes between the boundary and an outside point. So let $\{x_n\}$ be a sequence of points outside G converging to $x \in G$. Suppose first that $x \in G_j$, but $x \notin G_k$, $k \neq j$. Then, by Lemma A.1, $y_j(x_n) \to y_j(x) = x$. Because G_k is closed, we have $d(y_k(x_n), x) > \varepsilon_k \ge \varepsilon > 0$ for some ε, all $k \neq j$, and all n large. Continuity then follows because $\{f(x_n)\}$ is bounded. In the general case, by reordering the indices we may assume that, for some ℓ, $x \in G_j$ for all $j \le \ell$ and $x \notin G_j$ for all $j > \ell$. Because f is Lipschitz and $y_j(x) = x$, Lemma A.1 gives

$$f(y_j(x_n)) = f(x) + \varepsilon_{jn}, \qquad |\varepsilon_{jn}| \le L_f |x_n - x| \quad \text{for } j \le \ell, \tag{A.4}$$

while obviously,

$$|f(y_j(x_n)) - f(x)| \le K \quad \text{for some } K, \ \text{for } j > \ell. \tag{A.5}$$

By definition, for some $\varepsilon > 0$, we have, for all large n,

$$\begin{cases} d(x_n, y_j(x_n)) \to 0, & j \le \ell, \ \text{and} \\ d(x_n, y_j(x_n)) > \varepsilon, & j > \ell. \end{cases} \tag{A.6}$$

Let $d_n \triangleq \min_j\{d(x_n, y_j(x_n))\}$. Then, $0 < d_n(d[x_n, y_j(x_n)])^{-1} \le 1$, and

$$\lim_{n\to\infty} d_n(d[x_n, y_j(x_n)])^{-1}\varepsilon_{jn} = 0 \quad \text{for } j \le \ell, \tag{A.7}$$

$$\lim_{n\to\infty} d_n(d[x_n, y_j(x_n)])^{-1}K = 0 \quad \text{for } j > \ell. \tag{A.8}$$

Therefore,

$$\lim_n g(x_n) = \lim_n \frac{\sum_{j=1}^J d_n(d[x_n, y_j(x_n)])^{-1} f(y_j(x_n))}{\sum_{j=1}^J d_n(d[x_n, y_j(x_n)])^{-1}} \tag{A.9}$$

$$= \lim_n \frac{\sum_{j=1}^\ell d_n(d[x_n, y_j(x_n)])^{-1} f(x)}{\sum_{j=1}^\ell d_n(d[x_n, y_j(x_n)])^{-1}} \tag{A.10}$$

$$= f(x). \tag{A.11}$$

Thus, $g(x)$ is continuous at each point x. □

We now turn to a proof of Lipschitz continuity. This, again, is a local property.

Lemma A.3 (Lipschitz). *Under the conditions of Lemma A.2, g as defined in (A.3) is Lipschitz continuous with Lipschitz constant L_f.*

Proof. It suffices to prove Lipschitz continuity with constant L_f locally at x for each point x. This holds by definition for x in the interior of G, and we next establish the result for x in the interior of its complement. So, fix such a point x and define

$$\varepsilon \triangleq \frac{1}{2}\min\{d(x, z) \colon z \in G\}, \tag{A.12}$$

so that by our assumption $\varepsilon > 0$. Fix an arbitrary $0 < \delta < 1/2$ and a point z such that $d(z, x) < \delta\varepsilon$. Then, $d(z, G) > \varepsilon$. Because x is fixed it will be convenient to denote $q_j \triangleq (d[x, y_j(x)])^{-1}$. Without loss of generality we assume that $f(x), f(y_j(x)), f(z)$, and $f(y_j(z))$ are all positive (this amounts to a shift by a constant, and does not influence continuity properties). Now by Lemma A.1,

$$d[x, y_j(x)] \le d[x, z] + d[z, y_j(z)] + d[y_j(z), y_j(x)] \tag{A.13}$$

$$\le d[z, y_j(z)] + 2\varepsilon\delta \tag{A.14}$$

so that, because $d(x, G) > \varepsilon$,

$$\frac{1}{d[z, y_j(z)]} \le \frac{1}{d[x, y_j(x)] - 2\varepsilon\delta} \tag{A.15}$$

$$\le q_j(1 + 2\delta). \tag{A.16}$$

Exchanging the roles in the triangle inequality, we conclude that

$$q_j(1 - 2\delta) \le \frac{1}{d[z, y_j(z)]} \le q_j(1 + 2\delta). \tag{A.17}$$

But then,

$$g(x) - g(z) = \frac{\sum_{j=1}^J q_j f(y_j(x))}{\sum_{j=1}^J q_j} - \frac{\sum_{j=1}^J (d[z, y_j(z)])^{-1} f(y_j(z))}{\sum_{j=1}^J (d[z, y_j(z)])^{-1}} \tag{A.18}$$

$$\le \frac{\sum_{j=1}^J q_j f(y_j(x))}{\sum_{j=1}^J q_j} - \frac{\sum_{j=1}^J q_j(1 - 2\delta) f(y_j(z))}{\sum_{j=1}^J q_j(1 + 2\delta)} \tag{A.19}$$

$$\leq \frac{\sum_{j=1}^{J} q_j[f(y_j(x)) - f(y_j(z))]}{\sum_{j=1}^{J} q_j} + \frac{\sum_{j=1}^{J} q_j f(y_j(z))}{\sum_{j=1}^{J} q_j}\left(1 - \frac{1-2\delta}{1+2\delta}\right) \tag{A.20}$$

$$\leq L_f d(x,z) + \max_j\{|f(y_j(z))|\} \cdot \frac{4\delta}{1+2\delta}. \tag{A.21}$$

Exchanging the roles of x and z, we obtain in exactly the same way,

$$g(z) - g(x) \leq L_f d(x,z) + \max_j\{|f(y_j(x))|\} \cdot \frac{4\delta}{1+2\delta}. \tag{A.22}$$

Because δ was arbitrary, the Lipschitz continuity is established.

It remains to consider the case where x is on the boundary of G. So, fix a direction z: if $x + \eta \cdot z$ is in G for all η small, then there is nothing to prove. We need only consider the case where $x + \eta \cdot z$ is in the (open) complement of G for all $0 < \eta \leq \eta_0$. So, fix an arbitrary ε. By Lemma A.2, we can find a δ and a point u in the complement of G (actually, on the line segment $(x, x + \eta_0 z)$) so that

$$|g(x) - g(u)| \leq \varepsilon, \tag{A.23}$$

$$d(x,u) \leq \varepsilon d(x,z). \tag{A.24}$$

By the Lipschitz property for points in the complement of G, we conclude that

$$|g(x) - g(z)| \leq |g(x) - g(u)| + |g(u) - g(z)| \tag{A.25}$$

$$\leq \varepsilon + L_f d(u,z) \tag{A.26}$$

$$\leq \varepsilon + L_f(1+\varepsilon) d(x,z). \tag{A.27}$$

Because ε is arbitrary, the result is established. □

Appendix B. An example. Consider the region

$$G = \{(x,y)\colon 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1.5\}.$$

As pictured in Figure 1, we take jump directions

$$\vec{e}_1 = (1,0), \quad \vec{e}_2 = (-1,0), \quad \vec{e}_3 = (0,1), \quad \vec{e}_4 = (0,-1), \quad \vec{e}_5 = (-1,1).$$

We suppose that the jump rates are as follows. Let

$$t(x,y) = \min(1-x, 1.5 - y - x);$$

$t(x,y)$ represents a distance from the point (x,y) to the right boundaries of G (the vertical and slanted lines). Similarly, $t(y,x)$ represents a distance from (x,y) to the upper boundaries of G. Now we define jump rates as follows:

$$\lambda_1(\vec{x}) = t(x,y), \quad \lambda_2(\vec{x}) = x, \quad \lambda_3(\vec{x}) = t(y,x), \quad \lambda_4(\vec{x}) = y, \quad \lambda_5(\vec{x}) = x(1-y).$$

Note that our assumptions allow the process to jump out of G. For example, suppose that $\vec{z}_n(0) = (1/2, 1/4)$. Then, for every odd n, the process can exit G. Take $n = 3$. Then, two jumps in a row in direction 1 lead $\vec{z}_n(t)$ first to $(5/6, 1/4)$, then to $(7/6, 1/4)$; the associated jump rates are $1/2$ and $1/6$, respectively. Now the process is outside G. As detailed in Appendix A, we can extend the jump rates of $\vec{z}_n$ to all of $\mathbb{R}^n$ by taking the rates to be those at

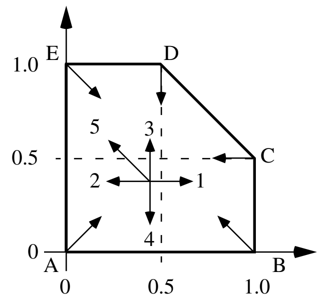

FIGURE 1. Illustration of the example.

the projection of $\vec{z}_n$ on G; that is, as the rates of the point closest to $\vec{z}_n$ in G. Continuing with our example, from $(7/6, 1/4)$ (which projects to $(1, 1/4)$) the process can take a jump in direction 3 to $(7/6, 7/12)$, with rate $3/4$. Now the process can take two jumps in direction 5, from $(7/6, 7/12)$ to $(5/6, 11/12)$ and then to $(1/2, 5/4)$, with rates $1/2$ and $15/98$ (the points project to $(1, 1/2)$ and $(5/7, 11/14)$, respectively). From $(1/2, 5/4)$ the process can jump only to the left or down (directions 2 or 4); the reader may explore how far it may travel before returning to the interior of G.

We claim that the process thus described satisfies all our assumptions. The only part of this claim that is not immediate is Assumption 2C. For this we need to make direction vectors $\vec{v}$ at each corner point (labeled A through E in Figure 1) that make the small jump rates that exist near each corner increase monotonically as we move in directions parallel to the vectors. (We also need to check the flat portions of the boundaries, but this is easy given the corner point calculations.) The appropriate vectors are pictured in each corner. Specifically, at point A we use vector $(1, 1)$, at point B we use vector $(-1, 1)$, at point C we use vector $(-1, 0)$, at point D we use vector $(0, -1)$, and at point E we use vector $(1, -1)$. The reader may check easily that these vectors cause monotone increases in the small jump rates: at point A the small rates are λ_2, λ_4, and λ_5; at point B the small rates are λ_1 and λ_4; at point C the small rates are λ_1 and λ_3; at point D the small rates are λ_1, λ_3, and λ_5; and at point E the small rates are λ_2, λ_3, and λ_5.

To work just one example (they all work the same way), consider a point $\vec{x} = (x, y)$ near point A $= (0, 0)$. Then,

$$\lambda_2(\vec{x} + \alpha\vec{v}) = x + \alpha,$$
$$\lambda_4(\vec{x} + \alpha\vec{v}) = y + \alpha,$$
$$\lambda_5(\vec{x} + \alpha\vec{v}) = (x + \alpha)(1 - y - \alpha) = x(1 - y) + \alpha(1 - y - x) - \alpha^2;$$

it is clear that these functions are all monotone increasing in α for small values of x, y, and α.

Acknowledgments. The authors' research was supported in part by the United States–Israel Binational Science Foundation Grant BSF 1999179. Research of the first author was supported in part by the Fund for Promotion of Research at the Technion, Israel Institute of Technology.

References

[1] Atar, Rami, Paul Dupuis. 1999. Large deviations and queueing networks: Methods for rate function identification. *Stochastic Proc. Appl.* **84** 255–296.

[2] Botvich, D., N. Duffield. 1995. Large deviations, the shape of the loss curve, and economies of scale in large multiplexers. *Queueing Systems* **20** 293–320.

[3] Courcoubetis, C., R. Weber. 1996. Buffer overflow asymptotics for a buffer handling many traffic sources. *J. Appl. Probab.* **33** 886–903.

[4] Dembo, Amir, Ofer Zeitouni. 1998. *Large Deviations Techniques and Applications*, 2nd ed. Springer-Verlag, New York.

[5] Dupuis, Paul, Richard Ellis. 1995. The large deviations principle for a general class of queueing systems I. *Trans. Amer. Math. Soc.* **347** 2689–2751.

[6] Dupuis, Paul, Richard Ellis. 1997. *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley, New York.

[7] Dupuis, Paul, Richard Ellis, Alan Weiss. 1991. Large deviations for Markov processes with discontinuous statistics I: General upper bounds. *Ann. Probab.* **19** 1280–1297.

[8] Freidlin, M. I., A. D. Wentzell. 1998. *Random Perturbations of Dynamical Systems*, 2nd ed. Springer-Verlag, New York.

[9] Ignatiouk-Robert, Irina. 2002. Sample path large deviations and convergence parameters. *Ann. Appl. Probab.* **11** 1292–1329.

[10] Mandjes, M., A. Ridder. 1999. Optimal trajectory to overflow in a queue fed by a large number of sources. *Queueing Systems* **31** 137–170.

[11] Mandjes, Michel, Alan Weiss. 2004. Sample path large deviations of a multiple time-scale queueing model. Submitted for publication.

[12] Puhalskii, Anatolii. 2001. *Large Deviations and Idempotent Probability*. Chapman-Hall/CRC, Boca Raton, FL.

[13] Shwartz, Adam, Alan Weiss. 1995. *Large Deviations for Performance Analysis*. Chapman-Hall, London, UK.

[14] Simonian, A., J. Guibert. 1995. Large deviations approximation for fluid queues fed by a large number of on/off sources. *IEEE J. Selected Areas Comm.* **13** 1017–1027.

[15] Tse, David. 1996. Asymptotic optimality of a measurement-based call admission control scheme. *Proc. 34th Allerton Conf.*, Allerton, IL.