

Large deviations of mean-field interacting particle systems in a fast varying environment

Sarath Yasodharan*[†] and Rajesh Sundaresan*
Indian Institute of Science

August 18, 2020

Abstract

This paper studies large deviations of a “fully coupled” finite state mean-field interacting particle system in a fast varying environment. The empirical measure of the particles evolves in the slow time scale and the random environment evolves in the fast time scale. Our main result is the path-space large deviation principle for the joint law of the empirical measure process of the particles and the occupation measure process of the fast environment. This extends previous results known for two time scale diffusions to two time scale mean-field models with jumps. Our proof is based on the method of stochastic exponentials. We characterise the rate function by studying a certain variational problem associated with an exponential martingale.

MSC 2010 subject classifications: Primary 60F10; Secondary 60K37, 60K35, 60J75

Keywords: Mean-field interaction, large deviations, time scale separation, averaging principle, metastability

1 Introduction

Let \mathcal{X}, \mathcal{Y} be finite sets and $(\mathcal{X}, \mathcal{E}_{\mathcal{X}})$ and $(\mathcal{Y}, \mathcal{E}_{\mathcal{Y}})$ be directed graphs on \mathcal{X} and \mathcal{Y} respectively. Let $M_1(\mathcal{X})$ denote the space of probability measures on \mathcal{X} . For each $N \geq 1$, we consider Markov processes with infinitesimal generators acting on functions f on $M_1^N(\mathcal{X}) \times \mathcal{Y}$ of the form

$$\sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} N \xi(x) \lambda_{x, x'}(\xi, y) \left[f \left(\xi + \frac{\delta_{x'}}{N} - \frac{\delta_x}{N}, y \right) - f(\xi, y) \right] + N \sum_{y' : (y, y') \in \mathcal{E}_{\mathcal{Y}}} (f(\xi, y') - f(\xi, y)) \gamma_{y, y'}(\xi),$$

$\xi \in M_1^N(\mathcal{X})$ and $y \in \mathcal{Y}$; here $M_1^N(\mathcal{X}) \subset M_1(\mathcal{X})$ denotes the set of probability measures on \mathcal{X} that can arise as empirical measures of N -particle configurations on \mathcal{X}^N , $\lambda_{x, x'}(\cdot, y) : M_1(\mathcal{X}) \rightarrow \mathbb{R}_+$, $(x, x') \in \mathcal{E}_{\mathcal{X}}$ and $y \in \mathcal{Y}$, and $\gamma_{y, y'} : M_1(\mathcal{X}) \rightarrow \mathbb{R}_+$, $(y, y') \in \mathcal{E}_{\mathcal{Y}}$, are given functions. Such processes arise in the context of weakly interacting Markovian mean-field particle systems in a fast varying environment where the empirical measure of the particle system evolves in the slow time scale and the environment process evolves in the fast time scale. An important feature of such processes is that they are “fully coupled”, i.e., the evolution of the empirical measure depends on the state of the environment, and the environment itself changes its state depending on the empirical measure of the particle system. This paper establishes a process-level large deviation principle (LDP) for the joint law of the empirical measure process and the occupation measure of the fast environment for

*Supported by the Indo-French Centre for Applied Mathematics.

[†]Supported by a fellowship grant from the Centre for Networked Intelligence (a Cisco CSR initiative) of the Indian Institute of Science, Bangalore.

such fully coupled two time scale mean-field models (see Section 2.2 for the precise mathematical model and Theorem 2.2 for the statement of the main result).

Our study of the LDP for such a two time scale mean-field model is motivated by the metastability phenomenon in networked systems. Many networked systems that arise in practice can be modelled using a two time scale mean-field model; see Appendix A for details of a retrial queueing system with N orbit queues, and a wireless local area network with local interactions. In such networks, there could be multiple seemingly “stable points of operation”, or metastable points. Some of these may be desirable but some others undesirable in terms of some performance metrics. One is often interested in understanding the following metastable phenomena: (i) the mean time spent by the network near an operating point, (ii) the mean time required for transiting from one stable operating point to another, (iii) the mean time for the system to be sufficiently close to stationarity, etc. The process level large deviations result established in this paper helps to answer such questions on the large time behaviour of these systems.

The above two time scale mean-field model is an example of a stochastic process with time scale separation where a certain component of the process evolves in the slow time scale (i.e. $O(1)$ -change in a given $O(1)$ time duration) and another component evolves in the fast time scale (i.e. $O(N)$ -change in a given $O(1)$ time duration). Such processes that evolve on multiple time scales have been well studied in the past, and it is known that, under mild conditions, they exhibit the “averaging principle”: when the time scale separation N becomes large, the slow component tracks the solution to a certain dynamical system whose driving function is “averaged” over the stationary behaviour of the fast component. In his seminal work, Khasminskii [21] first proved the averaging principle for two time scale diffusions. Freidlin and Wentzell [15, Chapter 7, Section 9] studied the averaging phenomenon in a fully coupled system of diffusions where both the drift and the diffusion coefficients of the slow component depend on the fast component and vice-versa. Their proof is based on discretisation arguments. The averaging phenomenon has also been studied in the context of jump processes with applications to performance analysis of various computer communication systems and queueing networks – Castiel et al. [6] studied a carrier sense multiple access algorithm in the context of wireless networks, Bordenave et al. [3] studied performance analysis of wireless local area networks, Hunt and Kurtz [16] studied scaling limits of loss networks, Hunt and Laws [17] studied analysis of trunk reservation policy in the context of loss networks; also see Kelly [20] and the references therein for other works on loss networks in the two time scale framework. While the above works on jump processes study the averaging principle in the large- N limit, this paper focuses on process-level large deviations from the large- N limit.

Various authors have studied process level large deviations of diffusion processes evolving on multiple time scales under various assumptions – see Freidlin [15], Veretennikov [31, 32], Liptser [25], Puhalskii [28] and the references therein. Liptser [25] established the large deviation principle for the joint law of the slow process and the occupation measure of the fast process for one-dimensional diffusions when the fast process does not depend on the slow variable. More recently, Puhalskii [28] extended this for multidimensional diffusions when the slow and fast processes are fully coupled. His approach is based on the method of stochastic exponentials for large deviations [26], where one identifies a suitable exponential martingale associated with the process and characterises the rate function in terms of this exponential martingale. In identifying the rate function, the main ingredient in the proof is to study a certain variational problem and show certain continuity property of its solution.

In this paper, our proof of the process-level large deviation result is based on the method of stochastic exponentials, see Puhalskii [26, 28], but the main difficulty lies in extending the approach of Puhalskii [28] to our two time scale mean-field model with jumps. In particular, our setting requires us to study certain variational problems in an Orlicz space, instead of the usual L^2 space in the context of diffusions, to characterise the rate function; see Theorem 5.3 and Theorem 6.2.

While Puhalskii [28] uses tools from the theory of elliptic partial differential equations for the characterisation of the rate function, we use tools from convex analysis and parametric continuity of optimisation problems. Also, our mean-field setting makes the solutions to these variational problems blow up near the boundary of the state space, and one of the main novelties of our work is the methodology to obtain a characterisation of the rate function in such cases via suitable approximations – see Section 7.

Other works in the two time scale regime include Budhiraja et al. [5] who studied the case where the slow process is a diffusion and the fast process is a Markov chain on a finite set; their proof is based on the weak convergence approach to large deviations where one establishes the LDP by studying certain controlled versions of the processes. Kumar and Popovic [22] established the LDP for two time scale jump-diffusions under some general conditions via convergence of nonlinear semigroups, but their approach requires verification of the comparison principle for a certain nonlinear operator. While this is a possible alternative approach for the mean-field problem under consideration, we have used the more probabilistic stochastic exponentials approach.

Let us also mention some works on large deviations of mean-field models that do not involve the fast environment. Dawson and Gärtner [8] established process-level large deviations of interacting diffusions of mean-field type where each particle evolves as a diffusion process with coefficients that depend on the other particles via the empirical measure of the states of all the particles. Léonard [24, 23] extended this to the case of jump processes. Our work can be viewed as an extension of Léonard [23] to the case of finite state mean-field interacting particle systems with a fully coupled fast varying environment. In the stationary regime, Borkar and Sundaresan [4] studied large deviations of the stationary measure of finite state mean-field interacting particle systems using tools from Freidlin and Wentzell [15, Chapter 6], and the authors [33] studied large time behaviour, metastability and convergence to stationarity in such systems using tools from Hwang and Sheu [18]. Our results in this paper, along with the results in [33], can be used to study the large time behaviour and metastability of two time scale mean-field models; see Section 2.3.2.

The rest of the paper is organised as follows. We start with a formal description of our fully coupled two time scale mean-field model and state our main result and its implications in Section 2. The proof of the main result is carried out in Sections 3–8. Section 3 establishes exponential tightness of the joint law of the empirical measure process and the occupation measure process of the fast environment. In Section 4, we define a certain exponential martingale and show a necessary condition that holds for every subsequential rate function. In Section 5, we define our candidate rate function using the above exponential martingale and study its relevant properties. In Section 6, we obtain a characterisation of subsequential rate functions for sufficiently regular elements in the space and Section 7 extends this to the whole space using certain approximation arguments. Finally we complete the proof of the main result in Section 8.

2 System model and main result

2.1 Notation

We summarise the frequently used notation in the paper. Let $\langle \cdot, \cdot \rangle$ denote inner product and $\| \cdot \|$ denote the norm on Euclidean spaces. Given a complete separable metric space S , let $B(S)$ denote the space of bounded Borel-measurable functions on S equipped with the uniform topology. Let $M(S)$ denote the space of finite measures on S equipped with the topology of weak convergence. Let $M_1(S)$ denote the space of probability measures on S equipped with the Lévy-Prohorov metric (which generates the topology of weak convergence). (If S is a finite set, then $M_1(S)$ can be viewed as an $(|S| - 1)$ -dimensional subset of the Euclidean space $\mathbb{R}^{|S|}$; in this case, for $\nu \in M_1(S)$, we shall denote the density of ν with respect to the counting measure on S by ν). Given $N \in \mathbb{N}$,

$M_1^N(S) \subset M_1(S)$ denotes the set of probability measures that can arise as empirical measures of N independent S -valued random variables. Given $T > 0$, let $D([0, T], S)$ (resp. $D(\mathbb{R}_+, S)$) denote the space of càdlàg functions on $[0, T]$ (resp. \mathbb{R}_+) equipped with the Skorohod- J_1 topology (see, for example, Ethier and Kurtz [13, Chapter 3]). Similarly, given a finite set \mathcal{Y} , $D_\uparrow([0, T], M(\mathcal{Y})) \subset D([0, T], M(\mathcal{Y}))$ denotes the space of càdlàg functions θ on $[0, T]$ such that for each $0 \leq s \leq t \leq T$, $\theta_t - \theta_s$ is an element of $M(\mathcal{Y})$ and $\theta_t(\mathcal{Y}) = t$. This equipped with its subspace topology is a complete and separable metric space, and is closed in $D([0, T], M(\mathcal{Y}))$. If X is an element of $D([0, T], S)$, $D([0, \infty), S)$ or $D_\uparrow([0, T], M(\mathcal{Y}))$, let X_t and $X(t)$ denote the coordinate projection of X at time t .

Denote the moment generating function of the centred unit rate Poisson law by $\tau(u) := e^u - u - 1$, $u \in \mathbb{R}$, and its convex dual by

$$\tau^*(u) := \begin{cases} +\infty & \text{if } u < -1 \\ 1 & \text{if } u = -1 \\ (u + 1) \log(u + 1) - u & \text{if } u \geq -1. \end{cases}$$

Given a complete separable metric space S and a finite measure ϑ on S , let $L^\tau(S, \vartheta)$ and $L^{\tau^*}(S, \vartheta)$ denote the Orlicz spaces corresponding to the functions τ and τ^* , respectively (see, for example, Rao and Ren [29, Chapter 3] for an introduction to Orlicz spaces). The Orlicz norms on these spaces are denoted by $\|\cdot\|_{L^\tau(S, \vartheta)}$ and $\|\cdot\|_{L^{\tau^*}(S, \vartheta)}$, respectively. Given a directed and connected graph (V, E) and $\Delta = (u, v) \in E$, let $u + \Delta$ denote v . Given a function f on $[0, T] \times S \times V$, let Df denote the function on $[0, T] \times S \times V \times E$ defined by $f(t, s, u, \Delta) = f(t, s, v) - f(t, s, u)$ where $\Delta = (u, v) \in E$. Given a subset W of a Euclidean space and $T > 0$, let $C^{1,1}([0, T] \times W \times S)$ (resp. $C^\infty([0, T] \times W \times S)$) denote the space of functions on $f(t, u, s)$, $(t, u, s) \in [0, T] \times W \times S$, that is continuously differentiable (resp. infinitely differentiable) in both t and u . For any function X on $[0, T] \times S$, let $X_t(s)$ and $X(t, s)$ denote the evaluation of X at $(t, s) \in [0, T] \times S$.

We finally recall the definition of a large deviation principle. Let S be a metric space. We say that a sequence $\{X_N\}_{N \geq 1}$ of S -valued random variables defined on a probability space (Ω, \mathcal{F}, P) satisfies the large deviation principle (LDP) with rate function $I : S \rightarrow [0, +\infty]$ if

- the lower level sets of I are compact, i.e., for each $M > 0$, $\{x \in S : I(x) \leq M\}$ is a compact subset of S ;
- for each open set $G \subset S$,

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \log P(X_N \in G) \geq - \inf_{x \in G} I(x);$$

- for each closed set $F \subset S$,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(X_N \in F) \leq - \inf_{x \in F} I(x).$$

We say that $I : S \rightarrow [0, +\infty]$ is a subsequential rate function for the family $\{X_N\}_{N \geq 1}$ if there exists a subsequence $\{N_k\}_{k \geq 1}$ of \mathbb{N} such that the sequence $\{X_{N_k}\}_{k \geq 1}$ satisfies the large deviation principle with rate function I .

2.2 System model

We describe our model of the mean-field interacting particle system in a fast environment. Let there be N particles and an environment. There is a state associated with each particle as well as the environment at all times; the particle states come from a finite set \mathcal{X} and the environment state comes from a finite set \mathcal{Y} . The state of the n th particle at time t is denoted by $X_n^N(t) \in \mathcal{X}$, and the state of the environment at time t is denoted by $Y_N(t) \in \mathcal{Y}$. To describe the evolution of the states of the particles, we consider a directed graph $(\mathcal{X}, \mathcal{E}_{\mathcal{X}})$ on the vertex set \mathcal{X} with the interpretation

that whenever $(x, x') \in \mathcal{E}_X$, a particle at state x can transit to state x' . Similarly, to describe the evolution of the environment, we consider a directed graph $(\mathcal{Y}, \mathcal{E}_Y)$; $(y, y') \in \mathcal{E}_Y$ implies that the environment can transit from state y to state y' .

To describe the particle transitions, we define, for each $y \in \mathcal{Y}$ and $(x, x') \in \mathcal{E}_X$, a function $\lambda_{x,x'}(\cdot, y) : M_1(\mathcal{X}) \rightarrow \mathbb{R}_+$, and for each $y \in \mathcal{Y}$, we consider the generator $Q_{N,y}$ acting on functions on \mathcal{X}^N by

$$Q_{N,y}f(\mathbf{x}^N) = \sum_{n=1}^N \sum_{x'_n : (x_n, x'_n) \in \mathcal{E}_X} \lambda_{x_n, x'_n}(\overline{\mathbf{x}}^N, y) (f(\mathbf{x}_{n, x_n, x'_n}^N) - f(\mathbf{x}^N)),$$

where $\overline{\mathbf{x}}^N := \frac{1}{N} \sum_{n=1}^N \delta_{x_n}$ denotes the empirical measure associated with the configuration \mathbf{x}^N , and $\mathbf{x}_{n, x_n, x'_n}^N$ denotes the resultant configuration of particles when the n th particles changes its state from x_n to x'_n in \mathbf{x}^N . To describe the transitions of the environment, for each $(y, y') \in \mathcal{Y}$, we define a function $\gamma_{y,y'}(\cdot) : M_1(\mathcal{X}) \rightarrow \mathbb{R}_+$, and for each $\xi \in M_1(\mathcal{X})$, we consider the generator L_ξ acting on functions on \mathcal{Y} by

$$L_\xi g(y) = \sum_{y' : (y, y') \in \mathcal{E}_Y} (g(y') - g(y)) \gamma_{y,y'}(\xi).$$

Finally, we consider the generator Ψ_N acting on functions f on $\mathcal{X}^N \times \mathcal{Y}$ by

$$\Psi_N f(\mathbf{x}^N, y) = Q_{N,y} f(\cdot, y)(\mathbf{x}^N) + N L_{\overline{\mathbf{x}}^N} f(\mathbf{x}^N, \cdot)(y),$$

where $Q_{N,y} f(\cdot, y)(\mathbf{x}^N)$ (resp. $L_{\overline{\mathbf{x}}^N} f(\mathbf{x}^N, \cdot)(y)$) indicates that the operator $Q_{N,y}$ (resp. $L_{\overline{\mathbf{x}}^N}$) acts on the first variable (resp. second variable) of f and the resultant function is evaluated at \mathbf{x}^N (resp. y).

We make the following assumptions on the particle system:

- (A1) The graph $(\mathcal{X}, \mathcal{E}_X)$ is irreducible;
- (A2) For each $y \in \mathcal{Y}$ and $(x, x') \in \mathcal{E}_X$, the function $\lambda_{x,x'}(\cdot, y)$ is Lipschitz continuous on $M_1(\mathcal{X})$ and $\inf_{\xi \in M_1(\mathcal{X})} \lambda_{x,x'}(\xi, y) > 0$;

and the following assumptions on the environment:

- (B1) The graph $(\mathcal{Y}, \mathcal{E}_Y)$ is irreducible;
- (B2) For each $(y, y') \in \mathcal{E}_Y$, the function $\gamma_{y,y'}(\cdot)$ is continuous on $M_1(\mathcal{X})$ and $\inf_{\xi \in M_1(\mathcal{X})} \gamma_{y,y'}(\xi) > 0$.

As a consequence of the assumptions (A2) and (B2), we see that the transition rates of the particles as well as that of the environment are bounded, i.e.,

$$\sup_{\xi \in M_1(\mathcal{X})} \lambda_{x,x'}(\xi, y) < +\infty \quad \forall (x, x') \in \mathcal{E}_X \text{ and } \forall y \in \mathcal{Y}$$

and

$$\sup_{\xi \in M_1(\mathcal{X})} \gamma_{y,y'}(\xi) < +\infty \quad \forall (y, y') \in \mathcal{E}_Y,$$

and hence the $D([0, \infty), \mathcal{X}^N \times \mathcal{Y})$ -valued martingale problem for Ψ_N is well-posed (see, for example, Ethier and Kurtz [13, Section 4.1, Exercise 15]). Therefore, given an initial configuration of particles $(X_n^N(0), 1 \leq n \leq N) \in \mathcal{X}^N$ and an initial state of the environment $Y_N(0) \in \mathcal{Y}$, we have a Markov process $\{((X_n^N(t), 1 \leq n \leq N), Y_N(t)), t \geq 0\}$ whose sample paths are elements of $D([0, \infty), \mathcal{X}^N \times \mathcal{Y})$.

To describe the process $\{(X_n^N(t), 1 \leq n \leq N), Y_N(t), t \geq 0\}$ in words, consider the mapping

$$\begin{aligned} \{((X_n^N(t), 1 \leq n \leq N), Y_N(t)), t \geq 0\} &\mapsto \left\{ \frac{1}{N} \sum_{n=1}^N \delta_{X_n^N(t)}, t \geq 0 \right\} \\ &=: \{\mu_N(t), t \geq 0\} \in D([0, \infty), M_1^N(\mathcal{X})) \end{aligned}$$

that takes the process $\{(X_n^N(t), 1 \leq n \leq N), Y_N(t), t \geq 0\}$ and maps it to the empirical measure process $\{\mu_N(t), t \geq 0\}$. Note that, if the environment were frozen to be y , then μ_N is Markov with infinitesimal generator

$$\Phi_{N,y}f(\xi) = \sum_{(x,x') \in \mathcal{E}_{\mathcal{X}}} N\xi(x)\lambda_{x,x'}(\xi, y) \left[f\left(\xi + \frac{\delta_{x'}}{N} - \frac{\delta_x}{N}\right) - f(\xi) \right].$$

We see that a particle in state x at time t makes a transition to state x' at rate $\lambda_{x,x'}(\mu_N(t), Y_N(t))$ independent of everything else. Similarly, the environment makes a transition from state y to y' at time t at rate $N\gamma_{y,y'}(\mu_N(t))$ independent of everything else. Thus, the evolution of each particle depends on the empirical measure of the states of all the particles and the environment, and the evolution of the environment depends on the empirical measure of the states of all the particles. Note that the factor N in the second term of the generator Ψ_N indicates that the process Y_N makes $O(N)$ many transitions while each particle makes $O(1)$ transitions in a given $O(1)$ duration of time. Therefore, we have a “fully coupled” system where the particles evolve in a fast varying environment. Also, the empirical measure process μ_N makes $O(N)$ transitions over a given duration of time, but each of those transitions are of size $O(1/N)$ on the probability simplex $M_1(\mathcal{X})$. We shall refer to μ_N as the slow process and Y_N as the fast process.

Remark 2.1. Throughout the paper, we assume that all stochastic processes are defined on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$. We denote integration with respect to P by E .

Fix $T > 0$. We now describe the typical behaviour of our two time scale mean-field system for large N over the time duration $[0, T]$. Towards this, we define the occupation measure of the fast process Y_N by

$$\theta_N(t) := \int_0^t 1_{\{Y_N(s) \in \cdot\}} ds, \quad 0 \leq t \leq T.$$

Note that $\theta_N \in D_{\uparrow}([0, T], M(\mathcal{Y}))$, $\theta_{N,t}(\mathcal{Y}) = t$ and we can view θ_N as a measure on $[0, T] \times \mathcal{Y}$. For a fixed empirical measure of the particles $\xi \in M_1(\mathcal{X})$, assumptions (B1) and (B2) imply that there exists a unique invariant probability measure for the Markov process on \mathcal{Y} with infinitesimal generator L_{ξ} (we denote this by π_{ξ}). Therefore, when the empirical measure at time t is at a fixed state μ_t , since the fast process Y_N makes $O(N)$ transitions, we expect that the occupation measure of Y_N for large N becomes “close” to π_{μ_t} , the unique invariant probability measure associated with L_{μ_t} . Due to this ergodic behaviour of the fast process, we anticipate that a particle in state x at time t moves to state x' , where $(x, x') \in \mathcal{E}_{\mathcal{X}}$, at rate $\int_{\mathcal{Y}} \lambda_{x,x'}(\mu_t, y) \pi_{\mu_t}(dy)$, i.e., the average of $\lambda_{x,x'}(\mu_t, \cdot)$ over π_{μ_t} (for any $\xi \in M_1(\mathcal{X})$, $(x, x') \in \mathcal{E}_{\mathcal{X}}$ and $m \in M_1(\mathcal{Y})$, we define $\bar{\lambda}_{x,x'}(\xi, m) := \int_{\mathcal{Y}} \lambda_{x,x'}(\xi, y) m(dy)$).

More precisely, for large enough N , we anticipate the following averaging principle for the empirical measure process μ_N . If we assume that the initial conditions $\mu_N(0) \rightarrow \nu$ weakly for some deterministic element $\nu \in M_1(\mathcal{X})$, then we anticipate that μ_N converges in probability, in $D([0, T], M_1(\mathcal{X}))$, to the solution to the McKean-Vlasov ODE

$$\dot{\mu}_t = \bar{\Lambda}_{\mu_t, \pi_{\mu_t}}^* \mu_t, \quad t \geq 0, \quad \mu_0 = \nu, \quad (2.1)$$

where $\bar{\Lambda}_{\mu_t, \pi_{\mu_t}}$ denotes the $|\mathcal{X}| \times |\mathcal{X}|$ rate matrix of the slow process when the empirical measure is μ_t and the occupation measure of the fast process is π_{μ_t} , i.e., $\bar{\Lambda}_{\mu_t, \pi_{\mu_t}}(x, x') = \bar{\lambda}_{x, x'}(\mu_t, \pi_{\mu_t})$ when $(x, x') \in \mathcal{E}_{\mathcal{X}}$, $\bar{\Lambda}_{\mu_t, \pi_{\mu_t}}(x, x') = 0$ when $(x, x') \notin \mathcal{E}_{\mathcal{X}}$, $\bar{\Lambda}_{\mu_t, \pi_{\mu_t}}(x, x) = -\sum_{x' \neq x} \bar{\lambda}_{x, x'}(\mu_t, \pi_{\mu_t})$, and $\bar{\Lambda}_{\mu_t, \pi_{\mu_t}}^*$ denotes its transpose. Note that the above ODE is well-posed, thanks to the Lipschitz assumption on the transition rates (A2). See Bordenave et al. [3] for the study of averaging phenomena of a slightly general two time scale model in which each particle has a fast varying environment associated with it.

2.3 Main result

Our main result is on the large deviations of $\{(\mu_N, \theta_N)\}_{N \geq 1}$, the joint empirical measure process associated with the particle system and the occupation measure process associated with the environment Y_N , on $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$. Our main result is the following theorem.

Theorem 2.2. *Assume (A1), (A2), (B1), (B2), and fix $T > 0$. Suppose that $\{\mu_N(0)\}_{N \geq 1}$ satisfies the LDP on $M_1(\mathcal{X})$ with rate function I_0 . Then the sequence $\{(\mu_N(t), \theta_N(t)), 0 \leq t \leq T\}_{N \geq 1}$ satisfies the LDP on $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ with rate function*

$$I(\mu, \theta) := I_0(\mu(0)) + J(\mu, \theta),$$

where J is defined by

$$\begin{aligned} J(\mu, \theta) := & \int_{[0, T]} \left\{ \sup_{\alpha \in \mathbb{R}^{|\mathcal{X}|}} \left(\langle \alpha, (\dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t) \rangle \right. \right. \\ & \left. \left. - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) \right) \right. \\ & \left. + \sup_{g \in B(\mathcal{Y})} \int_{\mathcal{Y}} \left(-L_{\mu_t} g(y) \right. \right. \\ & \left. \left. - \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg(y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) \right) m_t(dy) \right\} dt \end{aligned} \quad (2.2)$$

whenever the mapping $[0, T] \ni t \mapsto \mu_t \in M_1(\mathcal{X})$ is absolutely continuous and θ , when viewed as a measure on $[0, T] \times \mathcal{Y}$, admits the representation $\theta(dt dy) = m_t(dy) dt$ for some $m_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$, and $J(\mu, \theta) = +\infty$ otherwise.

Note that our rate function consists of two parts – one corresponding to the empirical measure process μ_N and the other corresponding to the occupation measure of the fast process Y_N . The form of the first part of the rate function in (2.2) corresponding to the empirical measure process μ_N appears in the literature on large deviations of mean-field models (see Léonard [23, Theorem 3.3], [10, Theorem 1]). The form of the second part is related to the rate function that appears in the study of occupation measure of Markov processes (see Donsker and Varadhan [11, Theorem 1]). Here, the canonical form of the rate function is $\int_{[0, T]} \sup_{h > 0} \int_{\mathcal{Y}} -\frac{L_{\mu_t} h(y)}{h(y)} m_t(dy) dt$ and this form of the second part of our rate function in (2.2) can be obtained by taking supremum over functions of the form e^g , $g \in B(\mathcal{Y})$. We see that the first part of the rate function corresponding to the empirical measure process μ_N has parameters of the mean-field model “averaged” by the fast variable. Further the second part corresponding to the occupation measure of the fast process has parameters “frozen” at the current value of the slow variable. The form of our rate function is similar in spirit to that obtained by Puhalskii [28] in the case of coupled diffusions.

Note that, when μ is the solution to the McKean-Vlasov equation (2.1) starting at $\mu(0)$ and θ , when viewed as a measure on $[0, T] \times \mathcal{Y}$, is given by $\theta(dy dt) = \pi_{\mu_t}(dy) dt$ where π_{μ_t} is the

unique invariant probability measure associated with the infinitesimal generator L_{μ_t} , it is easy to see that the suprema in (2.2) are attained at the identically 0 functions $\alpha \equiv 0$ and $g \equiv 0$ and hence $J(\mu, \theta) = 0$. Therefore, we recover the typical behaviour of our fully coupled system – at each time $t > 0$, the empirical measure process μ_N tracks the solution to the McKean-Vlasov equation μ_t starting at $\mu(0)$ and the occupation measure of the fast process θ_N tracks the invariant probability measure of the fast process Y_N when the empirical measure is frozen at μ_t . Our result on the large deviations of the joint empirical measure process and the occupation measure of the fast process $\{(\mu_N, \theta_N)\}$ enables us to estimate the probabilities of two kinds of deviations from the typical behaviour – one where, for a given μ , the occupation measure of the fast process deviates from its typical behaviour (which at time t is $\pi_{\mu_t}(dy)dt$) and the other where μ deviates from its typical behaviour (which is the solution to (2.1) starting at $\mu(0)$).

We now provide an outline of the proof of Theorem 2.2. Our proof is broadly built upon the methodology of stochastic exponentials for large deviations by Puhalskii [26, 27, 28], where one shows the large deviation principle by first obtaining an equation for a subsequential rate function in terms of a suitable exponential martingale and then obtaining a characterisation of this subsequential rate function. Towards this, we first show that the sequence $\{(\mu_N, \theta_N)\}_{N \geq 1}$ is exponentially tight in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ (see Theorem 3.3); this is shown using standard martingale arguments and Doob’s inequality. Exponential tightness of the sequence $\{(\mu_N, \theta_N)\}_{N \geq 1}$ implies that there exists a subsequence $\{N_k\}_{k \geq 1}$ of \mathbb{N} such that the family $\{(\mu_{N_k}, \theta_{N_k})\}_{k \geq 1}$ satisfies the LDP (see, for example, Dembo and Zeitouni [9, Lemma 4.1.23]); let \tilde{I} denote the rate function that governs the LDP for the family $\{(\mu_{N_k}, \theta_{N_k})\}_{k \geq 1}$. In Sections 4-7, we obtain a characterisation of \tilde{I} when \tilde{I} is such that, for some $\nu \in M_1(\mathcal{X})$, $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$; specifically we show that $\tilde{I}(\mu, \theta)$ is given by the right hand side of (2.2). In some more detail, in Section 4, we define an exponential martingale associated with the Markov process (μ_N, Y_N) for a class of functions $\alpha : [0, T] \times M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$ and $g : [0, T] \times M_1(\mathcal{X}) \times \mathcal{Y} \rightarrow \mathbb{R}$ with certain properties, and we obtain an equation that the rate function \tilde{I} must satisfy in terms of this exponential martingale (see Theorem 4.1). In Section 5, we define our candidate rate function I^* in terms of this exponential martingale as a variational problem over functions α and g , and we then show that I^* coincides with the RHS of (2.2), and provide a nonvariational expression for I^* using elements from suitable Orlicz spaces (see Theorem 5.3). In Section 6, using the properties of the solution to the variational problem established in Section 5 and an extension of the equation of \tilde{I} to a larger class of functions α and g , we are able to obtain a characterisation of the rate function \tilde{I} for sufficiently regular elements in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ (see Theorem 6.2). In Section 7, we extend the above characterisation of \tilde{I} to the whole space $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ via certain approximation arguments. We finally complete the proof of Theorem 2.2 in Section 8, by removing the restriction that, for some $\nu \in M_1(\mathcal{X})$, $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$.

Our setting of mean-field interaction with jumps introduces some difficulties in characterising a subsequential rate function. One of them is in obtaining regularity properties of the solution to the variational problem appearing in the definition of $J(\mu, \theta)$ in (2.2) when (μ, θ) possesses some good properties. In the recent work of Puhalskii [28] on large deviations of fully coupled diffusions, the author uses tools from the theory of elliptic partial differential equations for this purpose whereas we resort to tools from convex analysis (Léonard [24, Sections 4-6]) and parametric continuity of optimisation problems (Sundaram [30, Chapter 9]) – see Theorem 5.3 and Theorem 6.2. Also, unlike in the case of Gaussian noise in Puhalskii [28], our Poissonian noise prevents us from obtaining an explicit form of the solution to the variational problem appearing in the rate function (2.2). Yet another difficulty is in obtaining a characterisation of $\tilde{I}(\mu, \theta)$ when the path μ hits the boundary of $M_1(\mathcal{X})$. In such cases, the solution to the variational problem that appears in (2.2) blows up near the boundary and hence the condition on \tilde{I} established in Theorem 6.1 cannot be directly used. We demonstrate how to approximate (μ, θ) via a sequence of regular elements $\{(\mu^i, \theta^i)\}_{i \geq 1}$ so that the solution to the variational problem in $J(\mu^i, \theta^i)$ is well-behaved. We can then use the conclusion of

Theorem 6.1 on the above sequence and show that $\tilde{I}(\mu^i, \theta^i) \rightarrow \tilde{I}(\mu, \theta)$ as $i \rightarrow \infty$; see Theorem 7.5.

2.3.1 Marginal μ_N

The above result on large deviations of the joint law of the empirical measure process of the particles and the occupation measure of the fast process enables us to easily obtain large deviations of the empirical measure process μ_N by using the contraction principle (see, for example, Dembo and Zeitouni [9, Theorem 4.2.1]).

Corollary 2.3. *Assume (A1), (A2), (B1), (B2), and fix $T > 0$. Suppose that $\{\mu_N(0)\}_{N \geq 1}$ satisfies the LDP in $M_1(\mathcal{X})$ with rate function I_0 . Then $\{\mu_N\}_{N \geq 1}$ satisfies the LDP in $D([0, T], M_1(\mathcal{X}))$ with rate function J_T defined as follows. If $[0, T] \ni t \mapsto \mu_t$ is absolutely continuous, then*

$$J_T(\mu) = I_0(\mu_0) + \int_{[0, T]} \left\{ \sup_{\alpha \in \mathbb{R}^{|\mathcal{X}|}} \left(\langle \alpha, \dot{\mu}_t \rangle - \sup_{m \in M_1(\mathcal{Y})} \left[\langle \alpha, \bar{\Lambda}_{\mu_t, m}^* \mu_t \rangle + \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m) \mu_t(dx) \right] \right) - \sup_{g \in B(\mathcal{Y})} \int_{\mathcal{Y}} \left(-L_{\mu_t} g(y) - \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg(y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) m(dy) \right) m(dy) \right\} dt,$$

where θ , when viewed as a measure on $[0, T] \times \mathcal{Y}$, admits the representation $\theta(dydt) = m_t(dy)dt$ for some $m_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$, and $J_T(\mu) = +\infty$ otherwise.

2.3.2 Large time behaviour

Using the result on the finite duration LDP for the process $\{\mu_N\}_{N \geq 1}$ in Corollary 2.3, we can employ the tools of Freidlin and Wentzell [15, Chapter 6] and Hwang and Sheu [18] to study the large time behaviour of the process μ_N . The programme to understand the large time behaviour is carried out in [33, Section 3]. The two crucial properties needed to establish large time behaviour of μ_N are: (i) the continuity of the Freidlin-Wentzell quasipotential (see [33, Section 3] for its definition) and (ii) uniform large deviations of μ_N , uniformly with respect to the initial condition $\mu_N(0)$ lying in a given closed set. One can show that the Freidlin-Wentzell quasipotential is continuous on $M_1(\mathcal{X}) \times M_1(\mathcal{X})$ by constructing constant velocity trajectories between any two given points in $M_1(\mathcal{X})$ and estimating the corresponding J_T for that path; see Borkar and Sundaresan [4, Lemma 3.4]. Since the space $M_1(\mathcal{X})$ is compact, one can also establish uniform large deviation estimates, see [33, Corollary 2.1]. Using the above two properties and the fact that (μ_N, Y_N) is strong Markov, one can establish results on the large time behaviour of μ_N such as (i) the mean exit time from a neighbourhood of an ω -limit set of (2.1), (ii) the probability of reaching a given ω -limit set starting from another, etc. – we refer the reader to [33, Section 3] for such results.

3 Exponential tightness

In this section, we prove the exponential tightness of the sequence $\{(\mu_N(t), \theta_N(t)), 0 \leq t \leq T\}_{N \geq 1}$ in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$. Towards this, we shall use the following results (Theorems 3.1-3.2). The proof of these results are standard and will be omitted here (see Feng and Kurtz [14, Theorem 4.4] and Puhalskii [26, Theorem B]).

Theorem 3.1. *A sequence $\{X_N\} = \{X_{N,t}, 0 \leq t \leq T\}$ taking values in $D([0, T], S)$ is exponentially tight if and only if*

(i) for each $M > 0$, there exists a compact set $K_M \subset S$ such that

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(\exists t \in [0, T] \text{ such that } X_{N,t} \notin K_M) \leq -M,$$

(ii) there exists a family of functions $F \subset C(S)$ that is closed under addition and separates points on S such that for each $f \in F$, $\{f(X_N)\}$ is exponentially tight in $D([0, T], \mathbb{R})$.

See Feng and Kurtz [14, Theorem 4.4] for a proof. We also need the following sufficient condition for exponential tightness in $D([0, T], \mathbb{R})$.

Theorem 3.2. *Let $\{X_N\}$ be a sequence taking values in $D([0, T], \mathbb{R})$. Suppose that*

(i) *we have*

$$\lim_{M \rightarrow \infty} \limsup_{N \rightarrow \infty} \frac{1}{N} \log P(\exists t \in [0, T] \text{ such that } |X_{N,t}| > M) = -\infty,$$

(ii) *for each $\varepsilon > 0$,*

$$\lim_{\delta \downarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{t_1 \in [0, T]} P(\sup_{t_2 \in [t_1, t_1 + \delta]} |X_{N,t_2} - X_{N,t_1}| > \varepsilon) = -\infty.$$

Then $\{X_N\}$ is exponentially tight in $D([0, T], \mathbb{R})$.

See Puhalskii [26, Theorem B] for a proof.

We now show the main result of this section, namely exponential tightness of the sequence $\{(\mu_N, \theta_N)\}_{N \geq 1}$.

Theorem 3.3. *The sequence of random variables $\{(\mu_N(t), \theta_N(t)), t \in [0, T]\}_{N \geq 1}$ is exponentially tight in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, i.e., given any $M > 0$, there exists a compact set $K_M \subset D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ such that*

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(\{(\mu_N(t), \theta_N(t)), 0 \leq t \leq T\} \notin K_M) \leq -M$$

Proof. It suffices to show that μ_N and θ_N are individually exponentially tight in $D([0, T], M_1(\mathcal{X}))$ and $D_{\uparrow}([0, T], M(\mathcal{Y}))$ respectively (see, for example, Feng and Kurtz [14, Lemma 3.6]).

Consider θ_N . Note that, for $0 \leq t \leq T$, we have $|\theta_{N,t}(Y)| \leq t$ for any subset $Y \subset \mathcal{Y}$. Therefore, using the compact set $K_M = \{y \in \mathbb{R}^{|\mathcal{Y}|} : 0 \leq y_i \leq t \forall i\} \subset M(\mathcal{Y})$, condition (i) of Theorem 3.1 holds. To verify condition (ii), define the collection of functions $F := \{f : M(\mathcal{Y}) \rightarrow \mathbb{R} : f(\theta) = \langle \alpha, \theta \rangle, \alpha \in \mathbb{R}^{|\mathcal{Y}|}\}$. Clearly, F is closed under addition and separates points on $M(\mathcal{Y})$. For any f of the form $f(\theta) = \langle \alpha, \theta \rangle$ for some $\alpha \in \mathbb{R}^{|\mathcal{Y}|}$, note that, with $X_{N,t} = f(X_{N,t})$, condition (i) of Theorem 3.2 holds since $|X_{N,t}| \leq t \max_{i \in \mathcal{Y}} |\alpha_i|$. To verify condition (ii) of Theorem 3.2, note that, for any $0 \leq s \leq t \leq T$, we have $|\theta_{N,t}(Y) - \theta_{N,s}(Y)| \leq t - s$ for any $Y \subset \mathcal{Y}$ and hence $|X_{N,t} - X_{N,s}| \leq (t - s) \max_i |\alpha_i|$. Thus, by choosing a sufficiently small $\delta > 0$, it is easy to see that condition (ii) of Theorem 3.2 holds. This establishes the exponential tightness of θ_N in $D_{\uparrow}([0, T], M(\mathcal{Y}))$.

We now show that μ_N is exponentially tight in $D([0, T], M_1(\mathcal{X}))$. Since for each $t > 0$, $\mu_{N,t}$ takes values in a compact space, condition (i) of Theorem 3.1 holds trivially. Again, to show condition (ii) in Theorem 3.1, we shall make use of Theorem 3.2. For this, we fix the class of functions $F := \{f : M_1(\mathcal{X}) \rightarrow \mathbb{R}_+, f(\xi) = \langle \alpha, \xi \rangle, \alpha \in \mathbb{R}^{|\mathcal{X}|}\}$, which is clearly closed under addition and separates points on $M_1(\mathcal{X})$. Fix $f \in F$ such that $f(\xi) = \langle \alpha, \xi \rangle$ for some $\alpha \in \mathbb{R}^{|\mathcal{X}|}$ and let

$X_{N,t} = f(\mu_{N,t}) = \langle \alpha, \mu_{N,t} \rangle$. Note that, we have $|X_{N,t}| \leq \max_x |\alpha_x|$ for all $t \geq 0$ and $N \geq 1$, hence condition (i) of Theorem 3.2 holds. To check condition (ii), note that, for each $t_1 \geq 0$ and $\beta > 1$,

$$M_t := \exp \left\{ N \left(\beta X_{N,t} - \beta X_{N,t_1} - \beta \int_{t_1}^t \Phi_{Y_{N,s}} f(\mu_{N,s}) ds - \int_{t_1}^t \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(\beta D\alpha(x, \Delta)) \lambda_{x, x+d\Delta}(\mu_{N,s}, Y_{N,s}) \mu_{N,s}(dx) ds \right) \right\}, t \geq t_1,$$

is an \mathcal{F}_t -martingale (see Léonard [24, Lemma 3.3]; alternatively, this can be easily checked using the Doléans-Dade exponential formula, see, for example, Jacod and Shiryaev [19, Chapter I, Theorem 4.61]). Therefore, given $\varepsilon > 0$, $\delta > 0$ and $t_1 > 0$, we have

$$\begin{aligned} & P \left(\sup_{t_2 \in [t_1, t_1 + \delta]} (X_{N,t_2} - X_{N,t_1}) > \varepsilon \right) \\ &= P \left(\sup_{t_2 \in [t_1, t_1 + \delta]} \exp\{N\beta(X_{N,t_2} - X_{N,t_1})\} > \exp\{N\beta\varepsilon\} \right) \\ &= P \left(\sup_{t_2 \in [t_1, t_1 + \delta]} M_t \times \exp \left\{ N\beta \int_{t_1}^t \Phi_{Y_{N,s}} f(\mu_{N,s}) \right. \right. \\ &\quad \left. \left. + N \int_{t_1}^t \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(\beta D\alpha(x, d\Delta)) \lambda_{x, x'}(\mu_{N,s}, Y_{N,s}) \mu_{N,s}(dx) ds \right\} > \exp\{N\beta\varepsilon\} \right) \\ &\leq P \left(\sup_{t_2 \in [t_1, t_1 + \delta]} M_t \exp\{N\delta c_{\alpha, \beta}\} > \exp\{N\beta\varepsilon\} \right) \\ &\leq \exp\{-N(\beta\varepsilon - \delta c_{\alpha, \beta})\} \end{aligned}$$

where $c_{\alpha, \beta}$ is a constant depending on α and β ; here the first inequality follows from the boundedness of the transition rates which is a consequence of the Lipschitz assumption (A2), and the second inequality follows from Doob's martingale inequality and the fact that $EM_t = EM_{t_1} = 1$. Thus, we obtain

$$\lim_{\delta \downarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{t_1 \in [0, T]} P \left(\sup_{t_2 \in [t_1, t_1 + \delta]} (X_{N,t_2} - X_{N,t_1}) > \varepsilon \right) \leq -\beta\varepsilon,$$

and hence, letting $\beta \rightarrow \infty$, we have

$$\lim_{\delta \downarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{t_1 \in [0, T]} P \left(\sup_{t_2 \in [t_1, t_1 + \delta]} (X_{N,t_2} - X_{N,t_1}) > \varepsilon \right) = -\infty.$$

We can now replace α with $-\alpha$ and repeat the above arguments to conclude that

$$\lim_{\delta \downarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \log \sup_{t_1 \in [0, T]} P \left(\sup_{t_2 \in [t_1, t_1 + \delta]} |X_{N,t_2} - X_{N,t_1}| > \varepsilon \right) = -\infty.$$

We have thus verified condition (ii) of Theorem 3.2 and hence it follows that $\{\mu_N\}_{N \geq 1}$ is exponentially tight in $D([0, T], M_1(\mathcal{X}))$. This completes the proof of the theorem. \square

4 An equation for the subsequential rate function

Let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ denote a subsequential rate function for the family $\{(\mu_N, \theta_N)\}_{N \geq 1}$, i.e., for some sequence $\{N_k\}_{k \geq 1}$ of \mathbb{N} , the family $\{(\mu_{N_k}, \theta_{N_k})\}_{k \geq 1}$ satisfies the large deviation principle with rate function \tilde{I} . In this section, we obtain a condition that every such subsequential rate function must satisfy.

We start with some definitions. Given $g \in C^{1,1}([0, T] \times M_1(\mathcal{X}) \times \mathcal{Y})$, define

$$\begin{aligned} V_t^g(\mu_N, Y_N) &:= g_t(\mu_N(t), Y_N(t)) - g_0(\mu_N(0), Y_N(0)) - \int_0^t \frac{\partial g_s}{\partial s}(\mu_N(s), Y_N(s)) ds \\ &\quad - \int_0^t \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \left[g_s \left(\mu_N(s) + \frac{\delta_{x'} - \delta_x}{N}, Y_N(s) \right) \right. \\ &\quad \quad \left. - g_s(\mu_N(s), Y_N(s)) \right] \times N \mu_{N,s}(x) \lambda_{x, x'}(\mu_N(s), Y_N(s)) ds \\ &\quad - \int_0^t \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \tau \left(\left[g_s \left(\mu_N(s) + \frac{\delta_{x'} - \delta_x}{N}, Y_N(s) \right) \right. \right. \\ &\quad \quad \left. \left. - g_s(\mu_N(s), Y_N(s)) \right] \right) \times N \mu_{N,s}(x) \lambda_{x, x'}(\mu_N(s), Y_N(s)) ds \end{aligned} \quad (4.1)$$

Let $n \in \mathbb{N}$. Given the time points $0 = t_0 < t_1 < \dots < t_n < T$, $\alpha = (\alpha_{t_i})_{i=0}^n$ where $\alpha_{t_i} : M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$ is continuous for each $0 \leq i \leq n$, and $\mu \in D([0, T], M_1(\mathcal{X}))$, define

$$\int_0^t \alpha_s(\mu_s) d\mu_s := \sum_{i=1}^n \langle \alpha_{t \wedge t_{i-1}}(\mu_{t_{i-1}}), (\mu_{t \wedge t_i} - \mu_{t \wedge t_{i-1}}) \rangle, t \in [0, T]; \quad (4.2)$$

note that this object is an element of $D([0, T], \mathbb{R})$. Given $x \in \mathcal{X}$ and $\Delta = (x, x') \in \mathcal{E}_{\mathcal{X}}$, define

$$D\alpha_s(\mu_s)(x, \Delta) := \alpha_s(\mu_s)(x') - \alpha_s(\mu_s)(x).$$

Similarly, given $y \in \mathcal{Y}$ and $\Delta = (y, y') \in \mathcal{E}_{\mathcal{Y}}$, define

$$Dg_s(\mu_s, y, \Delta) := g_s(\mu_s, y') - g_s(\mu_s, y).$$

Finally, given $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y}))$, time points $0 = t_0 < t_1 < \dots < t_n < T$, $\alpha = (\alpha_{t_i})_{i=0}^n$ and g that satisfy the above requirements, define

$$\begin{aligned} U_t^{\alpha, g}(\mu, \theta) &:= \int_0^t \alpha_s(\mu_s) d\mu_s - \int_0^t \left\langle \alpha_s(\mu_s), \int_{\mathcal{Y}} \Lambda_{\mu_s, y}^* \mu_s m_s(dy) \right\rangle ds \\ &\quad - \int_0^t \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}} \times \mathcal{Y}} \tau(D\alpha_s(\mu_s)(x, \Delta)) \lambda_{x, x+d\Delta}(\mu_s, y) \mu_s(dx) m_s(dy) ds \\ &\quad - \int_0^t \int_{\mathcal{Y}} \left(L_{\mu_s} g_s(\mu_s, \cdot)(y) \right. \\ &\quad \quad \left. + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_s(\mu_s, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_s) \right) m_s(dy) ds; \end{aligned} \quad (4.3)$$

here θ , when viewed as a measure on $[0, T] \times \mathcal{Y}$, admits the representation $\theta(dy dt) = m_t(dy) dt$ for some $m_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$, which follows from the existence of the regular conditional distribution (see, for example, Ethier and Kurtz [13, Theorem 8.1, page 502]).

We prove the following result, a condition that \tilde{I} must satisfy in terms of the functions $U^{\alpha, g}$.

Theorem 4.1. *Let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ denote a rate function and suppose that there is a subsequence $\{(\mu_{N_k}, \theta_{N_k})\}_{k \geq 1}$ of $\{(\mu_N, \theta_N)\}_{N \geq 1}$ that satisfies the LDP with rate function \tilde{I} . Then, for each α and g that satisfy the requirements of the definition of U and V in (4.3) and (4.1) respectively, we have*

$$\sup_{(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))} (U_T^{\alpha, g}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0. \quad (4.4)$$

Proof. Note that, since the transition rates are bounded (which is a consequence of the assumptions (A2) and (B2)),

$$N \left(\int_0^t \alpha_s d\mu_{N,s} - \int_0^t \left\langle \alpha_s(\mu_s), \int_{\mathcal{Y}} \Lambda_{\mu_{N,s}, y}^* \mu_{N,s} \theta_N(dy ds) \right\rangle \right), t \geq 0,$$

is an \mathcal{F}_t -martingale. Also, by Itô's formula,

$$\begin{aligned} & g_t(\mu_N(t), Y_N(t)) - g_0(\mu_N(0), Y_N(0)) - \int_0^t \frac{\partial g_s}{\partial s}(\mu_N(s), Y_N(s)) ds \\ & - \int_0^t \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \left[g_s \left(\mu_N(s) + \frac{\delta_{x'} - \delta_x}{N}, Y_N(s) \right) \right. \\ & \quad \left. - g_s(\mu_N(s), Y_N(s)) \right] \times N \mu_{N,s}(x) \lambda_{x, x'}(\mu_N(s), Y_N(s)) ds \\ & - N \int_0^t L_{\mu_N(s)} g_s(\mu_N(s), \cdot)(Y_N(s)) ds, t \geq 0, \end{aligned}$$

is an \mathcal{F}_t -martingale. Therefore, using the Doléans-Dade exponential formula, it follows that

$$\exp\{NU_t^{\alpha, g}(\mu_N, \theta_N) + V_t^g(\mu_N, Y_N)\}, t \geq 0,$$

is an \mathcal{F}_t -martingale, and hence

$$E \exp\{NU_T^{\alpha, g}(\mu_N, \theta_N) + V_T^g(\mu_N, Y_N)\} = 1.$$

Clearly, $U_T^{\alpha, g}(\cdot, \cdot)$ is continuous on $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, and since g is continuously differentiable in the second argument, $V_T^g(\mu_N, Y_N)$ is bounded, and hence $V_T^g(\mu_N, Y_N)/N$ goes to 0 P -a.s. Therefore, the result follows from an application of Varadhan's lemma along the subsequence $\{N_k\}_{k \geq 1}$ (see, for example, [9, Theorem 4.3.1]). \square

5 The variational problem in J

Motivated by the duality relation (4.4), we define our candidate rate function

$$I^*(\mu, \theta) := \sup_{\alpha, g} U_T^{\alpha, g}(\mu, \theta), \quad (5.1)$$

where the supremum is taken over all functions α and g that satisfy the conditions in Theorem 4.1.

In this section, we study the above variational problem and show that, whenever $I^*(\mu, \theta) < +\infty$, $I^*(\mu, \theta)$ coincides with the RHS of (2.2) and that $I^*(\mu, \theta)$ can be expressed in a non-variational form using elements from suitable Orlicz spaces. We begin with a necessary condition on the elements in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ whose I^* is finite.

Lemma 5.1. *If $I^*(\mu, \theta) < +\infty$, then the mapping $[0, T] \ni t \mapsto \mu_t \in M_1(\mathcal{X})$ is absolutely continuous.*

Proof. Take $g \equiv 0$ and α to be a function of only time (and denote this by α_t) in the definition of $U_t^{\alpha, g}$ in (4.3). Then (5.1) becomes

$$\begin{aligned} I^*(\mu, \theta) &= \sup_{\alpha, g} U_T^{\alpha, g}(\mu, \theta) \\ &\geq \int_0^T \alpha_t d\mu_t - \int_0^T \langle \alpha_t, \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle dt \\ &\quad - \int_0^T \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt. \end{aligned}$$

Therefore,

$$\begin{aligned} \int_0^T \alpha_t d\mu_t &\leq I^*(\mu, \theta) + \int_0^T \langle \alpha_t, \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle dt \\ &\quad + \int_0^T \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt. \end{aligned}$$

Replacing $c\alpha_t$ in place of α_t in the above equation, dividing throughout by c and choosing $c = 1/\|D\alpha\|_{L^\tau([0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)}$ (i.e. the inverse of the norm of the function $\alpha_t(x, x + \Delta)$ in the Orlicz space $L^\tau([0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)$), we have

$$\begin{aligned} \int_0^T \alpha_t d\mu_t &\leq \|D\alpha\|_{L^\tau([0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)} (I^*(\mu, \theta) + 1) \\ &\quad + \int_0^T \langle \alpha_t, \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle dt. \end{aligned}$$

Since α_t is arbitrary, from the definition of $\int_0^t \alpha_t d\mu_t$ in (4.2), it is clear that the mapping $[0, T] \ni t \mapsto \mu_t \in M_1(\mathcal{X})$ is absolutely continuous. \square

We also need the following lemma, whose proof can be found in Puhalskii [27, Lemma A.2, page 460].

Lemma 5.2. *Let \mathcal{V} be a complete separable metric space, and let \mathcal{U} be a dense subspace of \mathcal{V} . Let $f(t, v)$ be a function defined on $[0, T] \times \mathcal{V}$ that is measurable in t and continuous in v . Further, if $f(t, \beta(t))$ is locally integrable with respect to the Lebesgue measure on $[0, T]$ for all measurable functions $\beta : [0, T] \rightarrow \mathcal{U}$, then*

$$\sup_{\beta(\cdot)} \int_0^T f(t, \beta(t)) dt = \int_0^T \sup_{y \in \mathcal{U}} f(t, y) dt,$$

where the supremum in the LHS is taken over all \mathcal{U} -valued measurable functions $\beta(\cdot)$.

Let us introduce some notations. Let $DC_{\mathcal{X}}$ (resp. $DC_{\mathcal{Y}}$) denote the space of functions $D\alpha$ (resp. Dg) on $[0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}$ (resp. $[0, T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}$) such that $\alpha \in C^1([0, T] \times \mathcal{X})$ (resp. $g \in C^1([0, T] \times \mathcal{Y})$). (For economy of notation in the sequel, we shall also view \mathbb{R} -valued functions on $[0, T] \times \mathcal{X}$ as $\mathbb{R}^{|\mathcal{X}|}$ -valued functions on $[0, T]$.) Given $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, let $\mathcal{H}_{\mathcal{X}}(\mu, \theta)$ denote the $L^{\tau^*}([0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)$ -closure of functions of the form $\{\exp\{D\alpha\} - 1, D\alpha \in DC_{\mathcal{X}}\}$ and let $\mathcal{H}_{\mathcal{Y}}(\mu, \theta)$ denote the $L^{\tau^*}([0, T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}, \gamma_{y, y+d\Delta}(\mu_t) m_t(dy) dt)$ -closure of functions of the form $\{\exp\{Dg\} - 1, Dg \in DC_{\mathcal{Y}}\}$, where θ admits the representation $\theta(dy dt) = m_t(dy) dt$ for some $m_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$. We now prove the main result of this section.

Theorem 5.3. *Suppose that $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ is such that $I^*(\mu, \theta) < \infty$. Then, we have*

$$\begin{aligned}
I^*(\mu, \theta) = \int_{[0, T]} \left\{ \sup_{\alpha \in \mathbb{R}^{|\mathcal{X}|}} \left(\langle \alpha, (\dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t) \rangle \right. \right. \\
\left. \left. - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) \right) \right. \\
\left. + \sup_{g \in B(\mathcal{Y})} \int_{\mathcal{Y}} \left(-L_{\mu_t} g(y) \right. \right. \\
\left. \left. - \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg(y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) m_t(dy) \right) m_t(dy) \right\} dt, \tag{5.2}
\end{aligned}$$

where $m_t \in M_1(\mathcal{Y})$ is such that θ , when viewed as a measure on $[0, T] \times \mathcal{Y}$, admits the representation $\theta(dy ds) = m_t(dy) ds$ for almost all $t \in [0, T]$. Moreover, there exist functions $h_{\mathcal{X}} \in \mathcal{H}_{\mathcal{X}}(\mu, \theta)$ and $h_{\mathcal{Y}} \in \mathcal{H}_{\mathcal{Y}}(\mu, \theta)$ that satisfy

$$\begin{aligned}
\int_{[0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}} h_{\mathcal{X}} D\alpha \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt \\
= \int_{[0, T]} \langle \alpha_t, (\dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t) \rangle dt, \quad \forall \alpha \in B([0, T] \times \mathcal{X}), \tag{5.3}
\end{aligned}$$

and

$$\begin{aligned}
\int_{[0, T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}} h_{\mathcal{Y}} Dg \bar{\gamma}_{y, y+d\Delta}(\mu_t) m_t(dy) dt \\
= - \int_{[0, T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}} Dg \bar{\gamma}_{y, y+d\Delta}(\mu_t) m_t(dy) dt, \quad \forall g \in B([0, T] \times \mathcal{Y}), \tag{5.4}
\end{aligned}$$

respectively, $h_{\mathcal{X}} \in L^*([0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}, \bar{\lambda}_{x, x+d\Delta} \mu_t(dx) dt)$ and $h_{\mathcal{Y}} \in L^*([0, T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}, \gamma_{y, y+d\Delta}(\mu_t) m_t(dy) dt)$, and $I^*(\mu, \theta)$ admits the representation

$$\begin{aligned}
I^*(\mu, \theta) = \int_{[0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau^*(h_{\mathcal{X}}) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt \\
+ \int_{[0, T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}} \tau^*(h_{\mathcal{Y}}) \gamma_{y, y+d\Delta}(\mu_t) m_t(dy) dt. \tag{5.5}
\end{aligned}$$

Furthermore, if $\inf_{t \in [0, T]} \min_{x \in \mathcal{X}} \mu_t(x) > 0$ and $\inf_{t \in [0, T]} \min_{y \in \mathcal{Y}} m_t(y) > 0$, the suprema in (5.2) over α and g are attained by $\hat{\alpha}_t \in \mathbb{R}^{|\mathcal{X}|}$ and $\hat{g}_t \in B(\mathcal{Y})$ that satisfy

$$\begin{aligned}
\dot{\mu}_t(x) - (\bar{\Lambda}_{\mu_t, m_t}^* \mu_t)(x) \\
+ \mu_t(x) \sum_{\substack{x' \in \mathcal{X}: \\ (x, x') \in \mathcal{E}_{\mathcal{X}}}} (\exp\{\hat{\alpha}_t(x') - \hat{\alpha}_t(x)\} - 1) \bar{\lambda}_{x, x'}(\mu_t, m_t) \\
- \sum_{\substack{x_0 \in \mathcal{X}: \\ (x_0, x) \in \mathcal{E}_{\mathcal{X}}}} \mu_t(x_0) (\exp\{\hat{\alpha}_t(x) - \hat{\alpha}_t(x_0)\} - 1) \bar{\lambda}_{x_0, x}(\mu_t, m_t) = 0, \quad \forall x \in \mathcal{X}, \tag{5.6}
\end{aligned}$$

and

$$\begin{aligned}
m_t(y) \sum_{\substack{y' \in \mathcal{Y}: \\ (y, y') \in \mathcal{E}_{\mathcal{Y}}}} \exp\{\hat{g}_t(y') - \hat{g}_t(y)\} \gamma_{y, y'}(\mu_t) \\
- \sum_{\substack{y_0 \in \mathcal{Y}: \\ (y_0, y) \in \mathcal{E}_{\mathcal{Y}}}} m_t(y_0) \exp\{\hat{g}_t(y) - \hat{g}_t(y_0)\} \gamma_{y_0, y}(\mu_t) = 0, \quad \forall y \in \mathcal{Y}, \tag{5.7}
\end{aligned}$$

for almost all $t \in [0, T]$, respectively.

Proof. For the first part of the theorem, we shall make use of Lemma 5.2. Note that, by Lemma 5.1, we have that the mapping $[0, T] \ni t \mapsto \mu_t \in M_1(\mathcal{X})$ is absolutely continuous and θ admits the representation $\theta(dydt) = m_t(dy)dt$ where $m_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$. Therefore, for each $t \geq 0$, $U_t^{\alpha, g}$ in (4.3) can be written as

$$\begin{aligned} U_t^{\alpha, g}(\mu, \theta) &= \int_0^t \langle \alpha_s(\mu_s), \dot{\mu}_s \rangle ds - \int_0^t \langle \alpha_s(\mu_s), \bar{\Lambda}_{\mu_s, m_s}^* \mu_s \rangle ds \\ &\quad - \int_0^t \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_s(\mu_s)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_s, m_s) \mu_s(dx) ds \\ &\quad - \int_0^t \int_{\mathcal{Y}} \left(L_{\mu_s} g_s(\mu_s, \cdot)(y) \right. \\ &\quad \left. + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_s(\mu_s, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_s) \right) m_s(dy) ds, \end{aligned}$$

where α and g be satisfy the requirements in the definition of $U_t^{\alpha, g}$ in (4.3). Thus,

$$\begin{aligned} I^*(\mu, \theta) &= \sup_{\alpha} \int_{[0, T]} \left(\langle \alpha_t(\mu_t), \dot{\mu}_t \rangle - \langle \alpha_t(\mu_t), \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle \right. \\ &\quad \left. - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(\mu_t)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) \right) dt \\ &\quad + \sup_g \int_{[0, T]} \int_{\mathcal{Y}} \left(-L_{\mu_t} g_t(\mu_t, \cdot)(y) \right. \\ &\quad \left. - \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(\mu_t, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) \right) m_t(dy) dt \end{aligned}$$

where the supremum is taken over all functions α and g that satisfy the conditions in the definition of $U_t^{\alpha, g}$ in (4.3). Note that, since μ is kept fixed, an approximation argument using mollifiers implies that the above supremum over α can be replaced by supremum over α_s , where α_s is any $\mathbb{R}^{|\mathcal{X}|}$ -valued bounded measurable function on $[0, T]$. Once again, since μ is fixed, we can replace the supremum over $g \in C^{1,1}([0, T], M_1(\mathcal{X}) \times \mathcal{Y})$ with the supremum over g where g is any bounded measurable function on $[0, T] \times \mathcal{Y}$. Therefore,

$$\begin{aligned} I^*(\mu, \theta) &= \sup_{\alpha} \int_{[0, T]} \left(\langle \alpha_t(\mu_t), \dot{\mu}_t \rangle - \langle \alpha_t(\mu_t), \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle \right. \\ &\quad \left. - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(\mu_t)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) \right) dt \\ &\quad + \sup_g \int_{[0, T]} \int_{\mathcal{Y}} \left(-L_{\mu_t} g_t(\mu_t, \cdot)(y) \right. \\ &\quad \left. - \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(\mu_t, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) \right) m_t(dy) dt \end{aligned}$$

where the supremum is taken over bounded measurable functions $\alpha : [0, T] \rightarrow \mathbb{R}^{|\mathcal{X}|}$ and $g : [0, T] \times \mathcal{Y} \rightarrow \mathbb{R}$. We can now apply Lemma 5.2 to conclude that $I^*(\mu, \theta)$ is given by (5.2).

We obtain the existence of functions $h_{\mathcal{X}} \in \mathcal{H}_{\mathcal{X}}(\mu, \theta)$ and $h_{\mathcal{Y}} \in \mathcal{H}_{\mathcal{Y}}(\mu, \theta)$ that satisfy the conditions (5.3) and (5.4) and the non-variational representation of I^* in (5.5) by carrying out the convex analytic programme of Léonard [24, Sections 5-6] to the bounded linear functionals

$$\alpha \mapsto \int_{[0, T]} \langle \alpha, (\dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t) \rangle dt$$

and

$$g \mapsto \int_{[0,T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}} (g(y + \Delta) - g(y)) \gamma_{y,y+d\Delta}(\mu_t) m_t(dy) dt$$

on the closure of $\{D\alpha, \alpha \in B([0, T] \times \mathcal{X})\}$ and $\{Dg, g \in B([0, T] \times \mathcal{Y})\}$ in the Orlicz spaces $L^\tau([0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}, \bar{\lambda}_{x,x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)$ and $L^\tau([0, T] \times \mathcal{Y} \times \mathcal{E}_{\mathcal{Y}}, \gamma_{y,y+d\Delta}(\mu_t) m_t(dy) dt)$ respectively; the proof follows verbatim from Léonard [24] to our case, and we omit the details here.

Finally, to show the existence of supremisers $\hat{\alpha}_t$ and \hat{g} in (5.2) and the conditions (5.6) and (5.7) in the case when $\inf_{t \in [0, T]} \min_{x \in \mathcal{X}} \mu_t(x) > 0$ and $\inf_{t \in [0, T]} \min_{y \in \mathcal{Y}} m_t(y) > 0$, note that, for each $t \in [0, T]$ for which $\dot{\mu}_t$ exists, the mappings

$$\alpha_t \mapsto \langle \alpha_t, (\dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t) \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x,x+d\Delta}(\mu_t, m_t) \mu_t(dx) \quad (5.8)$$

and, viewing g_t as an element of $\mathbb{R}^{|\mathcal{Y}|}$,

$$g_t \mapsto - \int_{\mathcal{Y}} \left(L_{\mu_t} g_t(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y,y+d\Delta}(\mu_t) \right) m_t(y) \quad (5.9)$$

are concave on $\mathbb{R}^{|\mathcal{X}|}$ and $\mathbb{R}^{|\mathcal{Y}|}$ respectively. Therefore, there is an $\hat{\alpha}_t$ and a \hat{g}_t that attain the suprema in (5.2); the conditions in (5.6) and (5.7) on $\hat{\alpha}_t$ and \hat{g}_t easily follow by writing down the first order conditions for optimality of the mappings in (5.8) and (5.9) respectively. \square

6 Characterisation of the subsequential rate function for sufficiently regular elements

Let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function for the family $\{(\mu_N, \theta_N)\}_{N \geq 1}$, i.e., for some sequence $\{N_k\}_{k \geq 1}$ of \mathbb{N} , $\{(\mu_{N_k}, \theta_{N_k})\}_{k \geq 1}$ satisfies the large deviation principle with rate function \tilde{I} . In addition suppose that, for some $\nu \in M_1(\mathcal{X})$, $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. In this section, we characterise \tilde{I} for sufficiently regular elements in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, i.e., we show that $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$ for all elements $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ that satisfy certain regularity properties, where I^* is given by (5.2) (see Theorem 6.2).

6.1 An extension of Theorem 4.1

We first extend the conclusion of Theorem 4.1 to a larger class of functions α and g . Let $\Gamma \subset D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ denote the set of points (μ, θ) such that the mapping $[0, T] \ni t \mapsto \mu_t \in M_1(\mathcal{X})$ is absolutely continuous, and θ , when viewed as a measure on $[0, T] \times \mathcal{Y}$ admits the representation $\theta(dydt) = m_t(dy)dt$ where $m_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$. In particular, $(\mu, \theta) \in \Gamma$ implies that the mapping $t \mapsto \mu_t$ is differentiable for almost all $t \in [0, T]$. Given bounded measurable functions $\alpha : [0, T] \times M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$ and $g : [0, T] \times M_1(\mathcal{X}) \times \mathcal{Y} \rightarrow \mathbb{R}$ such that for all $t \in [0, T]$ and $y \in \mathcal{Y}$ both $\alpha(t, \cdot)$ and $g(t, \cdot, y)$ are continuous on $M_1(\mathcal{X})$, we define, with a slight

abuse of notation, for $(\mu, \theta) \in \Gamma$ and $t \in [0, T]$,

$$\begin{aligned}
U_t^{\alpha, g}(\mu, \theta) := & \int_{[0, t]} \left\{ \langle \alpha_s(\mu_s), \dot{\mu}_s - \bar{\Lambda}_{\mu_s, m_s}^* \mu_s \rangle \right. \\
& - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_s(\mu_s)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_s, m_s) \mu_s(dx) \\
& - \int_{\mathcal{Y}} \left(L_{\mu_s} g_s(\mu_s, \cdot)(y) \right. \\
& \left. \left. + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_s(\mu_s, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_s) \right) m_s(dy) \right\} ds. \tag{6.1}
\end{aligned}$$

Note that the boundedness of α and g in the above definition implies that $D\alpha \in L^\tau([0, T] \times \mathcal{X} \times \mathcal{E}_{\mathcal{X}}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)$, and $Dg \in L^\tau([0, T] \times \mathcal{E}_{\mathcal{Y}} \times \mathcal{Y}, \gamma_{y, y+d\Delta}(\mu_t) m_t(dy) dt)$.

Let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function for the family $\{(\mu_N, \theta_N)\}_{N \geq 1}$. Note that, by Theorem 4.1 and the definition of I^* in (5.1), we have that $\tilde{I}(\mu, \theta) \geq I^*(\mu, \theta)$ for all $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y}))$. Given $\delta > 0$, define

$$K_\delta = \{(\mu, \theta) : \tilde{I}(\mu, \theta) \leq \delta\};$$

since \tilde{I} has compact level sets, K_δ is compact in $D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y}))$. By Lemma 5.1 and the fact that $\tilde{I} \geq I^*$, we have that $K_\delta \subset \Gamma$. We now prove the following extension to Theorem 4.1.

Theorem 6.1. *Let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function. Let $\alpha : [0, T] \times M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$, $g : [0, T] \times M_1(\mathcal{X}) \times \mathcal{Y} \rightarrow \mathbb{R}$ be bounded and measurable functions such that both α and g are continuous on $M_1(\mathcal{X})$. Then,*

$$\sup_{(\mu, \theta) \in \Gamma} (U_T^{\alpha, g}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0.$$

Moreover, there exists some $\delta > 0$ (depending on α and g) such that

$$\sup_{(\mu, \theta) \in K_\delta} (U_T^{\alpha, g}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0, \tag{6.2}$$

and the above supremum is attained.

Proof. We first define certain approximations of functions α and g that meet the requirements of Theorem 4.1 and prove certain convergence properties of these approximations. We then use the conclusion of Theorem 4.1 for these approximations and pass to the limit to obtain (6.2). Our proof is inspired by ideas from Puhalskii [27, Lemma 7.2 and Theorem 7.1], with necessary modifications to our mean-field with jumps setting.

Since α is a Carathéodory function, using the Scorza-Dragoni theorem, for each $i \geq 1$, there exists a compact set $F_i \subset [0, T]$ and a measurable function $\bar{\alpha}_i : [0, T] \times M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$ such that $\bar{\alpha}_i = \alpha$ on $F_i \times M_1(\mathcal{X})$, $\bar{\alpha}_i$ is continuous on $F_i \times M_1(\mathcal{X})$, and $\text{Leb}([0, T] \setminus F_i) \leq 1/i$ (see, for example, Ekeland and Temam [12, page 235]). Since $[0, T] \setminus F_i$ is open in $[0, T]$, we can write it as a countable union of disjoint open intervals, and hence we can extend $\bar{\alpha}_i$ to a continuous function on $[0, T] \times M_1(\mathcal{X})$ by a linear interpolation between the two endpoints of the above open intervals; we again denote this function by $\bar{\alpha}_i$. Put $\alpha_i(t, \mu_t) = \bar{\alpha}_i(\frac{\lfloor tn(i) \rfloor}{n(i)}, \mu_{\lfloor \frac{tn(i)}{n(i)} \rfloor})$, where $n(i) \rightarrow \infty$ as $i \rightarrow \infty$. By continuity of τ , boundedness of α and α_i , boundedness of transition rates of the particles (which

is a consequence of assumption (A2)), we have that, for each $\delta > 0$,

$$\begin{aligned}
& \sup_{(\mu, \theta) \in K_\delta} \left| \int_{[0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha_i(t, \mu_t)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt - \right. \\
& \quad \left. \int_{[0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha(t, \mu_t)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt \right| \\
& = \sup_{(\mu, \theta) \in K_\delta} \left| \int_{K_i^c \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha_i(t, \mu_t)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt - \right. \\
& \quad \left. \int_{K_i^c \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha(t, \mu_t)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt \right| \\
& \leq \text{Leb}(K_i^c) \times c_\alpha \rightarrow 0
\end{aligned} \tag{6.3}$$

as $i \rightarrow \infty$, where $c_\alpha > 0$ is a constant depending on α . Furthermore, given $\delta > 0$ and $(\mu, \theta) \in K_\delta$, by Lemma 5.1, the mapping $[0, T] \ni t \mapsto \mu_t \in M_1(\mathcal{X})$ is absolutely continuous. Hence, noting that μ is kept fixed, by (5.3) in Theorem 5.3, there exists $h_\mathcal{X} \in \mathcal{H}(\mu, \theta)$ such that

$$\begin{aligned}
& \int_{[0, T]} \langle \alpha(t, \mu_t), (\dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t) \rangle dt \\
& = \int_{[0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} h_\mathcal{X} D\alpha \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt,
\end{aligned}$$

and

$$\begin{aligned}
& \int_{[0, T]} \langle \alpha_i(t, \mu_t), (\dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t) \rangle dt \\
& = \int_{[0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} h_\mathcal{X} D\alpha_i \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \left| \int_{[0, T]} \langle \alpha_i(t, \mu_t) - \alpha(t, \mu_t), \dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle dt \right| \\
& = \left| \int_{[0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} h_\mathcal{X} (D\alpha_i - D\alpha) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt \right| \\
& \leq \int_{[0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}} |h_\mathcal{X} (D\alpha_i - D\alpha)| \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt \\
& \leq 2 \|h_\mathcal{X}\|_{L^{\tau^*}([0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)} \\
& \quad \times \|D\alpha_i - D\alpha\|_{L^\tau([0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)} \\
& \leq 2 \max\{1, \delta + T\} \\
& \quad \times \|D\alpha_i - D\alpha\|_{L^\tau([0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)},
\end{aligned}$$

where the second inequality follows from Hölder's inequality in Orlicz spaces and the third inequality follows from the non-variational representation of the candidate rate function in I^* in (5.5), which gives that $\|h_\mathcal{X}\|_{L^{\tau^*}([0, T] \times \mathcal{X} \times \mathcal{E}_\mathcal{X}, \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) dt)} \leq \max\{1, I^*(\mu, \theta) + T\}$, along with the fact that $(\mu, \theta) \in K_\delta$ and $I^*(\mu, \theta) \leq \tilde{I}(\mu, \theta)$. Hence,

$$\sup_{(\mu, \theta) \in K_\delta} \left| \int_{[0, T]} \langle \alpha_i(t, \mu_t) - \alpha(t, \mu_t), \dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle dt \right| \rightarrow 0 \tag{6.4}$$

as $i \rightarrow \infty$. Similarly, by standard arguments using mollifiers and the Scorza-Dragoni theorem, we can show that there exist functions g_i on $[0, T] \times M_1(\mathcal{X}) \times \mathcal{Y}$ such that $g_i(\cdot, \cdot, y) \in C^\infty([0, T] \times M_1(\mathcal{X}))$ for all $y \in \mathcal{Y}$ and $\text{Leb}\{t \in [0, T] : g_i(t, \cdot, \cdot) \neq g(t, \cdot, \cdot)\} \leq 1/i$ for each $i \geq 1$. Therefore, using boundedness of the functions g , g_i , $i \geq 1$, and boundedness of the transition rates of the fast process (which is a consequence of assumption (B2)), we see that

$$\begin{aligned} & \sup_{(\mu, \theta) \in K_\delta} \left| \int_{[0, T] \times \mathcal{Y}} \left(L_{\mu_t} g_i(t, \mu_t, \cdot)(y) \right. \right. \\ & \quad + \int_{\mathcal{E}_y} \tau(Dg_i(t, \mu_t, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) \Big) m_t(dy) dt \\ & \quad - \int_{[0, T] \times \mathcal{Y}} \left(L_{\mu_t} g(t, \mu_t, \cdot)(y) \right. \\ & \quad \left. \left. + \int_{\mathcal{E}_y} \tau(Dg(t, \mu_t, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) \right) m_t(dy) dt \right| \rightarrow 0 \end{aligned} \quad (6.5)$$

as $i \rightarrow \infty$. Since α_i and g_i , $i \geq 1$, satisfy the conditions on α and g respectively in the definitions of U in (4.3) and V in (4.1), Theorem 4.1 implies that

$$\sup_{(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y}))} (U_T^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0.$$

By Lemma 5.1 and the fact that $\tilde{I}(\mu, \theta) \geq I^*(\mu, \theta)$, we see that $\tilde{I}(\mu, \theta) = +\infty$ whenever $(\mu, \theta) \notin \Gamma$, and hence we immediately get

$$\sup_{(\mu, \theta) \in \Gamma} (U_T^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0. \quad (6.6)$$

Let us now show that

$$\sup_{(\mu, \theta) \in K_\delta} (U_T^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0 \quad (6.7)$$

holds for a suitable $\delta > 0$ and all $i \geq 1$. Note that, using the boundedness of the functions α , g , α_i and g_i , $i \geq 1$, and boundedness of the transition rates (as a consequence of assumptions (A2) and (B2)), we have

$$\begin{aligned} U_T^{2\alpha_i, 2g_i}(\mu, \theta) &= \int_{[0, T]} \left\{ 2\langle \alpha_i(t, \mu_t), \dot{\mu}_t - \bar{\Lambda}_{\mu_t, m_t}^* \mu_t \rangle \right. \\ & \quad - \int_{\mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(2D\alpha_i(t, \mu_t)(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\mu_t, m_t) \mu_t(dx) \\ & \quad - \int_{\mathcal{Y}} \left(2L_{\mu_t} g_t(\mu_t, \cdot)(y) \right. \\ & \quad \left. \left. + \int_{\mathcal{E}_y} \tau(2Dg_i(t, \mu_t, y, \Delta)) \gamma_{y, y+d\Delta}(\mu_t) \right) m_t(dy) \right\} dt \\ &\geq 2U^{\alpha_i, g_i}(\mu, \theta) - 2Tc_{\alpha, g} \end{aligned}$$

for all $i \geq 1$, where $c_{\alpha, g} > 0$ is a constant depending on α and g . Therefore, for a fixed $M > 0$, we

have

$$\begin{aligned}
& \sup_{(\mu, \theta): U_T^{\alpha_i, g_i}(\mu, \theta) \geq M} (U^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) \\
& \leq \sup_{(\mu, \theta): U_T^{\alpha_i, g_i}(\mu, \theta) \geq M} (2U^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) - M \\
& \leq \sup_{(\mu, \theta): U_T^{\alpha_i, g_i}(\mu, \theta) \geq M} (U^{2\alpha_i, 2g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) + 2Tc_{\alpha, g} - M \\
& \leq 2Tc_{\alpha, g} - M.
\end{aligned}$$

Therefore the above implies that,

$$\begin{aligned}
& \sup_{(\mu, \theta) \in \Gamma} (U^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) \\
& \leq \sup_{(\mu, \theta) \in K_\delta} (U^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) \\
& \quad \vee \sup_{(\mu, \theta): U^{\alpha_i, g_i}(\mu, \theta) \geq M} (U^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) \\
& \quad \vee (M - \delta) \\
& \leq \sup_{(\mu, \theta) \in K_\delta} (U^{\alpha_i, g_i}(\mu, \theta) - \tilde{I}(\mu, \theta)) \vee (2Tc_{\alpha, g} - M) \vee (M - \delta).
\end{aligned}$$

Hence, choosing $M = 1 + 2Tc_{\alpha, g}$ and $\delta = M + 1$, the above and (6.6) imply (6.7). Letting $i \rightarrow \infty$, using convergences (6.3)-(6.4) for the slow process, and (6.5) for the fast process, (6.7) becomes

$$\sup_{(\mu, \theta) \in K_\delta} (U_T^{\alpha, g}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0. \tag{6.8}$$

Since the functions $U_T^{\alpha_i, g_i}$ (defined in (4.3)), $i \geq 1$, are continuous on Γ and since for all $\delta' > 0$

$$\lim_{i \rightarrow \infty} \sup_{(\mu, \theta) \in K_{\delta'}} |U_T^{\alpha_i, g_i}(\mu, \theta) - U_T^{\alpha, g}(\mu, \theta)| \rightarrow 0$$

as $i \rightarrow \infty$, it follows that, for all $\delta' > 0$, $U_T^{\alpha, g}$ (defined in (6.1)) is continuous on $K_{\delta'}$. Hence, using the compactness of the level sets of \tilde{I} , we see that the supremum in (6.8) is attained. This completes the proof of the theorem. \square

6.2 Characterisation of \tilde{I} for regular elements

We now prove the main result of this section, namely $\tilde{I}(\mu, \theta) = I^*(\mu, \theta)$ for all $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y}))$ that satisfy certain regularity properties.

Theorem 6.2. *Let $\nu \in M_1(\mathcal{X})$ and let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function such that $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. Suppose that $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y}))$ is such that*

- $\inf_{t \in [0, T]} \min_{x \in \mathcal{X}} \hat{\mu}_t(x) > 0$,
- the mapping $[0, T] \ni t \mapsto \hat{\mu}_t \in M_1(\mathcal{X})$ is Lipschitz continuous,
- $\hat{\theta}$, when viewed as a measure on $[0, T] \times \mathcal{Y}$, admits the representation $\hat{\theta}(dydt) = \hat{m}_t(dy)dt$ for some $\hat{m}_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$, and $\inf_{t \in [0, T]} \min_{y \in \mathcal{Y}} \hat{m}_t(y) > 0$.

Then $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.

Proof. Let $\delta = \inf_{t>0} \min_{x \in \mathcal{X}} \hat{\mu}_t(x)$. For each $t \in [0, T]$, consider the parametrised optimisation problems

$$\sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{u, \hat{m}_t}^* u \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(u, \hat{m}_t) u(dx) \right\}, \quad (6.9)$$

$u \in M_1(\mathcal{X})$ is such that $u(x) \geq \delta/2$ for all $x \in \mathcal{X}$, and

$$\sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_u g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(u) \right) \hat{m}_t(dy) \right\}, \quad (6.10)$$

$u \in M_1(\mathcal{X})$. Note that the mappings

$$\alpha_t \mapsto \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{u, \hat{m}_t}^* u \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(u, \hat{m}_t) u(dx), \quad (6.11)$$

where u is such that $u(x) \geq \delta/2$ for all $x \in \mathcal{X}$, and since $\inf_{t \in [0, T]} \min_{y \in \mathcal{Y}} \hat{m}_t(y) > 0$, viewing g_t as an element of $\mathbb{R}^{|\mathcal{Y}|}$,

$$g_t \mapsto - \int_{\mathcal{Y}} \left(L_u g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(u) \right) \hat{m}_t(dy) \quad (6.12)$$

are concave on $\mathbb{R}^{|\mathcal{X}|}$ and $\mathbb{R}^{|\mathcal{Y}|}$ respectively. Therefore, we see that there exist an $\hat{\alpha}_t(u) \in \mathbb{R}^{|\mathcal{X}|}$ and a $\hat{g}_t(u) \in \mathbb{R}^{|\mathcal{Y}|}$ that solve (6.9) and (6.10) respectively. Guided by (5.6) and (5.7), $\hat{\alpha}_t(u)$ and $\hat{g}_t(u)$ satisfy the first order optimality conditions

$$\begin{aligned} & \dot{\hat{\mu}}_t(x) - (\bar{\Lambda}_{u, \hat{m}_t}^* u)(x) \\ & + u(x) \sum_{\substack{x' \in \mathcal{X}: \\ (x, x') \in \mathcal{E}_{\mathcal{X}}}} (\exp\{\hat{\alpha}_t(u)(x') - \hat{\alpha}_t(u)(x)\} - 1) \bar{\lambda}_{x, x'}(u, \hat{m}_t) \\ & - \sum_{\substack{x_0 \in \mathcal{X}: \\ (x_0, x) \in \mathcal{E}_{\mathcal{X}}}} u(x_0) (\exp\{\hat{\alpha}_t(u)(x) - \hat{\alpha}_t(u)(x_0)\} - 1) \bar{\lambda}_{x_0, x}(u, \hat{m}_t) = 0, \quad \forall x \in \mathcal{X}, \end{aligned} \quad (6.13)$$

where $t \in [0, T]$ and $u \in M_1(\mathcal{X})$ is such that $u(x) \geq \delta/2$ for all $x \in \mathcal{X}$, and

$$\begin{aligned} & \hat{m}_t(y) \sum_{\substack{y' \in \mathcal{Y}: \\ (y, y') \in \mathcal{E}_{\mathcal{Y}}}} \exp\{\hat{g}_t(u, y') - \hat{g}_t(u, y)\} \gamma_{y, y'}(u) \\ & - \sum_{\substack{y_0 \in \mathcal{Y}: \\ (y_0, y) \in \mathcal{E}_{\mathcal{Y}}}} \hat{m}_t(y_0) \exp\{\hat{g}_t(u, y) - \hat{g}_t(u, y_0)\} \gamma_{y_0, y}(u) = 0, \quad \forall y \in \mathcal{Y}, \end{aligned} \quad (6.14)$$

where $t \in [0, T]$ and $u \in M_1(\mathcal{X})$, respectively.

We now define bounded measurable functions $\hat{\alpha} : [0, T] \times M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$ and $\hat{g} : [0, T] \times M_1(\mathcal{X}) \times \mathcal{Y} \rightarrow \mathbb{R}$ that are continuous on $M_1(\mathcal{X})$ such that $\hat{\alpha}(u)$ (resp. $\hat{g}(u)$) solves the optimisation problem in (6.9) (resp. (6.10)). Note that the objective function in (6.10) is uniquely determined by $\{g(t, y') - g(t, y), (y, y') \in \mathcal{E}_{\mathcal{Y}}\}$, and by assumption (A1), the objective function in (6.9) is uniquely determined by $\{\alpha_t(x') - \alpha_t(x), (x, x') \in \mathcal{E}_{\mathcal{X}}\}$. Since $\inf_{t \in [0, T]} \min_{x \in \mathcal{X}} \hat{\mu}_t(x) > 0$, the mapping $t \mapsto \hat{\mu}_t$ is Lipschitz continuous and the transition rates of the slow process are bounded (which is a consequence of assumption (A2)), we see that we can restrict the supremum over α_t in (6.9) to a single compact and convex subset of $\mathbb{R}^{|\mathcal{X}|}$, regardless of $t \in [0, T]$ and $u \in M_1(\mathcal{X})$ with $u(x) \geq \delta/2$ for all $x \in \mathcal{X}$. Similarly, since $\inf_{t \in [0, T]} \min_{y \in \mathcal{Y}} \hat{m}_t(y) > 0$ and the transition rates of the fast process are

bounded (which follows from assumption (B2)), we see that we can restrict the supremum in (6.10) to a single compact and convex subset of $\mathbb{R}^{\mathcal{Y}}$, regardless of $t \in [0, T]$ and $u \in M_1(\mathcal{X})$. Also, note that the mappings (6.9) and (6.10), when viewed as

$$\begin{aligned} & \{\alpha_t(x') - \alpha_t(x), (x, x') \in \mathcal{E}_{\mathcal{X}}\} \\ & \mapsto \langle \alpha_t, \dot{\mu}_t - \bar{\Lambda}_{u, \hat{m}_t}^* u \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(u, \hat{m}_t) u(dx) \end{aligned}$$

and,

$$\begin{aligned} & \{g_t(y') - g_t(y), (y, y') \in \mathcal{E}_{\mathcal{Y}}\} \\ & \mapsto - \int_{\mathcal{Y}} \left(L_u g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(u) \right) \hat{m}_t(dy) \end{aligned}$$

are strictly concave on $\mathbb{R}^{|\mathcal{E}_{\mathcal{X}}|}$ and $\mathbb{R}^{|\mathcal{E}_{\mathcal{Y}}|}$ respectively; hence there exists a unique $\{\hat{\alpha}_t(u)(x') - \hat{\alpha}_t(u)(x), (x, x') \in \mathcal{E}_{\mathcal{X}}\}$ and a unique $\{\hat{g}_t(u, y') - \hat{g}_t(u, y), (y, y') \in \mathcal{E}_{\mathcal{Y}}\}$ that solve (6.9) and (6.10) respectively. Fixing $\hat{\alpha}_t(u)(x_0) = 0$ for some $x_0 \in \mathcal{X}$, where $t \in [0, T]$ and $u \in M_1(\mathcal{X})$ with $u(x) \geq \delta/2$ for all $x \in \mathcal{X}$, fixing $g_t(u, y) = 0$ for some $y_0 \in \mathcal{Y}$, where $t \in [0, T]$ and $u \in M_1(\mathcal{X})$, defining $\hat{\alpha}_t(u)(x) = 0 \forall x \in \mathcal{X}$ whenever $u \in M_1(\mathcal{X})$ is such that $u(x) < \delta/4$ for some $x \in \mathcal{X}$, and defining $\hat{\alpha}(u)$ whenever u is such that $u(x) \in [\delta/4, \delta/2]$ for some $x \in \mathcal{X}$ using a linear interpolation, we obtain bounded functions $\hat{\alpha} : [0, T] \times M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$ and $\hat{g} : [0, T] \times M_1(\mathcal{X}) \times \mathcal{Y} \rightarrow \mathbb{R}$. By a measurable selection theorem (see, for example, Ekeland and Temam [12, Theorem 1.2, page 236]), it follows that the mappings $[0, T] \times M_1(\mathcal{X}) \ni (t, u) \mapsto \hat{\alpha}_t(u) \in \mathbb{R}^{|\mathcal{X}|}$ and $[0, T] \times M_1(\mathcal{X}) \times \mathcal{Y} \ni (t, u, y) \mapsto \hat{g}_t(u, y) \in \mathbb{R}$ are measurable. By the Berge's maximum theorem (see, for example, Sundaram [30, Theorem 9.17, page 237]) it follows that the functions $\hat{\alpha}$ and \hat{g} are continuous on $M_1(\mathcal{X})$.

Since $\hat{\alpha}$ and \hat{g} satisfy the assumptions of Theorem 6.1, there exists $(\tilde{\mu}, \tilde{\theta}) \in \Gamma$ that attains the supremum in (6.2) with $\hat{\alpha}$ and \hat{g} in place of α and g , respectively. That is,

$$U_T^{\hat{\alpha}, \hat{g}}(\tilde{\mu}, \tilde{\theta}) = \tilde{I}(\tilde{\mu}, \tilde{\theta}).$$

On the other hand, by (5.2) and the above,

$$I^*(\tilde{\mu}, \tilde{\theta}) \geq U_T^{\hat{\alpha}, \hat{g}}(\tilde{\mu}, \tilde{\theta}) = \tilde{I}(\tilde{\mu}, \tilde{\theta}),$$

and since $\tilde{I}(\tilde{\mu}, \tilde{\theta}) \geq I^*(\tilde{\mu}, \tilde{\theta})$, we have that

$$U_T^{\hat{\alpha}, \hat{g}}(\tilde{\mu}, \tilde{\theta}) = I^*(\tilde{\mu}, \tilde{\theta}) = \tilde{I}(\tilde{\mu}, \tilde{\theta}). \quad (6.15)$$

Note that $\tilde{\mu}_0 = \nu$ since $\tilde{I}(\tilde{\mu}, \tilde{\theta}) < +\infty$. We now proceed to show that $\tilde{m}_t = \hat{m}_t$ for almost all $t \in [0, T]$ and $\tilde{\mu} = \hat{\mu}$. This would establish $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.

By (6.15), we have

$$\begin{aligned} \tilde{m}_t(y) & \sum_{\substack{y' \in \mathcal{Y}: \\ (y, y') \in \mathcal{E}_{\mathcal{Y}}}} \exp\{\hat{g}_t(\tilde{\mu}_t, y') - \hat{g}_t(\tilde{\mu}_t, y)\} \gamma_{y, y'}(\tilde{\mu}_t) \\ & - \sum_{\substack{y_0 \in \mathcal{Y}: \\ (y_0, y) \in \mathcal{E}_{\mathcal{Y}}}} \tilde{m}_t(y_0) \exp\{\hat{g}_t(\tilde{\mu}_t, y) - \hat{g}_t(\tilde{\mu}_t, y_0)\} \gamma_{y_0, y}(\tilde{\mu}_t) = 0, \quad \forall y \in \mathcal{Y}, \end{aligned} \quad (6.16)$$

for almost all $t \in [0, T]$. By assumption (B2), the Markov process on \mathcal{Y} with transition rates $\exp\{\hat{g}_t(\tilde{\mu}_t, y') - \hat{g}_t(\tilde{\mu}_t, y)\} \gamma_{y, y'}(\tilde{\mu}_t)$, $(y, y') \in \mathcal{E}_{\mathcal{Y}}$, possesses a unique invariant probability measure; comparing (6.14) with $u = \tilde{\mu}_t$ and (6.16), we get

$$\tilde{m}_t = \hat{m}_t \quad (6.17)$$

for almost all $t \in [0, T]$.

On one hand, by using the first order optimality condition in (6.13) with $u = \hat{\mu}_t$, and the just established fact that $\tilde{m}_t = \hat{m}_t$ for almost all $t \in [0, T]$, we get

$$\begin{aligned} & \dot{\hat{\mu}}_t(x) - (\bar{\Lambda}_{\hat{\mu}_t, \tilde{m}_t}^* \hat{\mu}_t)(x) \\ & + \hat{\mu}_t(x) \sum_{\substack{x' \in \mathcal{X}: \\ (x, x') \in \mathcal{E}_{\mathcal{X}}}} (\exp\{\hat{\alpha}_t(\hat{\mu}_t)(x') - \hat{\alpha}_t(\hat{\mu}_t)(x)\} - 1) \bar{\lambda}_{x, x'}(\hat{\mu}_t, \tilde{m}_t) \\ & - \sum_{\substack{x_0 \in \mathcal{X}: \\ (x_0, x) \in \mathcal{E}_{\mathcal{X}}}} \hat{\mu}_t(x_0) (\exp\{\hat{\alpha}_t(\hat{\mu}_t)(x) - \hat{\alpha}_t(\hat{\mu}_t)(x_0)\} - 1) \bar{\lambda}_{x_0, x}(\hat{\mu}_t, \tilde{m}_t) = 0, \quad \forall x \in \mathcal{X}, \end{aligned} \quad (6.18)$$

for almost all $t \in [0, T]$. On the other hand, by (6.15), we get

$$\begin{aligned} & \dot{\tilde{\mu}}_t(x) - (\bar{\Lambda}_{\tilde{\mu}_t, \tilde{m}_t}^* \tilde{\mu}_t)(x) \\ & + \tilde{\mu}_t(x) \sum_{\substack{x' \in \mathcal{X}: \\ (x, x') \in \mathcal{E}_{\mathcal{X}}}} (\exp\{\hat{\alpha}_t(\tilde{\mu}_t)(x') - \hat{\alpha}_t(\tilde{\mu}_t)(x)\} - 1) \bar{\lambda}_{x, x'}(\tilde{\mu}_t, \tilde{m}_t) \\ & - \sum_{\substack{x_0 \in \mathcal{X}: \\ (x_0, x) \in \mathcal{E}_{\mathcal{X}}}} \tilde{\mu}_t(x_0) (\exp\{\hat{\alpha}_t(\tilde{\mu}_t)(x) - \hat{\alpha}_t(\tilde{\mu}_t)(x_0)\} - 1) \bar{\lambda}_{x_0, x}(\tilde{\mu}_t, \tilde{m}_t) = 0, \quad \forall x \in \mathcal{X}, \end{aligned} \quad (6.19)$$

for almost all $t \in [0, T]$. Note that, by the optimality condition (6.13) and by (6.17), the mapping

$$\begin{aligned} u \mapsto & \left((\bar{\Lambda}_{u, \tilde{m}_t}^* u)(x) + u(x) \sum_{\substack{x' \in \mathcal{X}: \\ (x, x') \in \mathcal{E}_{\mathcal{X}}}} (\exp\{\hat{\alpha}_t(u)(x') - \hat{\alpha}_t(u)(x)\} - 1) \bar{\lambda}_{x, x'}(u, \tilde{m}_t) \right. \\ & \left. - \sum_{\substack{x_0 \in \mathcal{X}: \\ (x_0, x) \in \mathcal{E}_{\mathcal{X}}}} u(x_0) (\exp\{\hat{\alpha}_t(u)(x) - \hat{\alpha}_t(u)(x_0)\} - 1) \bar{\lambda}_{x_0, x}(u, \tilde{m}_t), \quad x \in \mathcal{X} \right) \in \mathbb{R}^{|\mathcal{X}|} \end{aligned}$$

on $\{u \in M_1(\mathcal{X}) : u(x) \geq \delta/2 \forall x \in \mathcal{X}\}$ is identically equal to $\dot{\hat{\mu}}_t$ for almost all $t \in [0, T]$. Hence, by (6.18) and (6.19), and noting that $\tilde{\mu}_0 = \hat{\mu}_0 = \nu$, Gronwall inequality implies that $\tilde{\mu}_t = \hat{\mu}_t$ for all $t \in [0, T]$.

We have thus shown that $(\tilde{\mu}, \tilde{\theta}) = (\hat{\mu}, \hat{\theta})$, and the second equality in (6.15) implies that $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$. This completes the proof of the theorem. \square

7 Approximating the subsequential rate function

Let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function for the family $\{(\mu_N, \theta_N)\}_{N \geq 1}$, and suppose that, for some $\nu \in M_1(\mathcal{X})$, $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. In this section, we show that $\tilde{I}(\mu, \theta) = I^*(\mu, \theta)$ for all $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$. We shall proceed through a sequence of lemmas. In each lemma, we shall extend the conclusion $\tilde{I}(\mu, \theta) = I^*(\mu, \theta)$ to a larger class of elements (μ, θ) by producing a sequence (μ^i, θ^i) such that $\tilde{I}(\mu^i, \theta^i) = I^*(\mu^i, \theta^i)$ for all $i \geq 1$, $(\mu^i, \theta^i) \rightarrow (\mu, \theta)$ in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ as $i \rightarrow \infty$, and $I^*(\mu^i, \theta^i) \rightarrow I^*(\mu, \theta)$ as $i \rightarrow \infty$. Using these approximations, we finally show that $\tilde{I}(\mu, \theta) = I^*(\mu, \theta)$ for all $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ (see Theorem 7.5).

We start with an extension of the conclusion of Theorem 6.2 to all initial conditions ν .

Lemma 7.1. *Let $\nu \in M_1(\mathcal{X})$ and let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function such that $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. Suppose that $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ is such that*

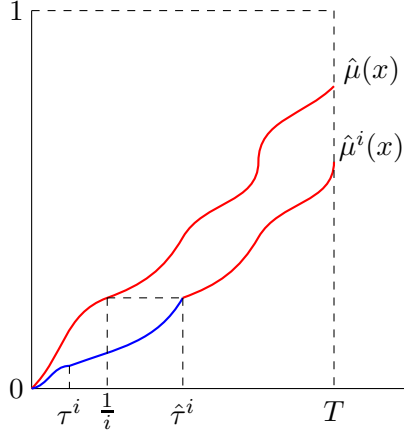


Figure 1: Figure depicting the idea of construction of $\hat{\mu}^i$ in the proof of Lemma 7.1

- $I^*(\hat{\mu}, \hat{\theta}) < +\infty$,
- $\inf_{t \in [\delta, T]} \min_{x \in \mathcal{X}} \hat{\mu}_t(x) > 0$ for all $\delta > 0$,
- the mapping $[0, T] \ni t \mapsto \hat{\mu}_t \in M_1(\mathcal{X})$ is Lipschitz continuous,
- $\hat{\theta}$, when viewed as a measure on $[0, T] \times \mathcal{Y}$, admits the representation $\hat{\theta}(dydt) = \hat{m}_t(dy)dt$ for some $\hat{m}_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$, and $\inf_{t \in [0, T]} \min_{y \in \mathcal{Y}} \hat{m}_t(y) > 0$.

Then $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.

Proof. We begin with some notations. Let $\mathcal{X}_0 = \{x \in \mathcal{X} : \hat{\mu}_0(x) = 0\}$. For each $x \in \mathcal{X}_0$, let $\{x_k^x, 1 \leq k \leq l(x)\}$ be such that $\hat{\mu}_0(x_1^x) \geq 1/|\mathcal{X}_0|$ (in particular, $x_1^x \notin \mathcal{X}_0$), $(x_k^x, x_{k+1}^x) \in \mathcal{E}\mathcal{X}$ for all $1 \leq k \leq l(x) - 1$ and $(x_{l(x)}^x, x) \in \mathcal{E}\mathcal{X}$, i.e., the collection of edges $\{(x_k^x, x_{k+1}^x), 1 \leq k \leq l(x) - 1\} \cup (x_{l(x)}^x, x)$ form a directed path of length $l(x)$ from x_1^x to x . Also, for the given $\nu \in M_1(\mathcal{X})$, let $\mu(\nu, \hat{\theta}) \in D([0, \infty), M_1(\mathcal{X}))$ denote the unique solution to the ODE $\dot{\mu}_t = \bar{\Lambda}_{\mu_t, \hat{m}_t}^* \mu_t$ with initial condition $\mu_0 = \nu$.

For each $i \geq 1$, we define a path $\hat{\mu}^i \in D([0, T], M_1(\mathcal{X}))$ as follows. Define $\hat{\mu}_t^i = \mu_t(\hat{\mu}_0, \hat{\theta})$ for $t \in [0, \tau^i]$ where $\tau^i = \inf\{t > 0 : \mu_t(\hat{\mu}_0, \hat{\theta})(x) = \hat{\mu}_{1/i}(x)/2 \text{ for some } x \in \mathcal{X}_0\}$. Note that $\tau^i < +\infty$ for i sufficiently large. Also note that $\hat{\mu}_{\tau^i}^i(x) > 0$ for all $x \in \mathcal{X}$, and that the supremum over α_t in the definition of $I^*(\hat{\mu}^i, \hat{\theta})$ (see (5.2)) is attained at $\alpha_t = 0$ for all $t \in [0, \tau^i]$. Let $\varepsilon_i(x) = \hat{\mu}_{1/i}(x) - \hat{\mu}_{\tau^i}^i(x)$ for $x \in \mathcal{X}$ and $i \geq 1$. Since the mapping $t \mapsto \hat{\mu}_t$ is Lipschitz continuous, we see that $\tau^i \rightarrow 0$ as $i \rightarrow \infty$, and $\varepsilon_i(x) \rightarrow 0$ as $i \rightarrow \infty$ for all $x \in \mathcal{X}$. For each $x \in \tilde{\mathcal{X}}_0 := \mathcal{X}_0 \cap \{x \in \mathcal{X}_0 : \varepsilon_i(x) > 0\}$, we shall now move the mass $\varepsilon_i(x)$ from the vertex x_1^x to x via the edges defined in the previous paragraph using a piecewise constant velocity path. Denote the elements of $\tilde{\mathcal{X}}_0$ by $x_1, x_2, \dots, x_{|\tilde{\mathcal{X}}_0|}$, let $l(x_0) = 0$ and $\varepsilon_i(x_0) = 0$. Given $r \in \{0, 1, \dots, |\tilde{\mathcal{X}}_0| - 1\}$, $s \in \{0, 1, \dots, l(x_{r+1}) - 1\}$, and $t \in [\tau^i + \sum_{m=0}^r l(x_m)\varepsilon_i(x_m) + s\varepsilon_i(x_{r+1}), \tau^i + \sum_{m=0}^r l(x_m)\varepsilon_i(x_m) + (s+1)\varepsilon_i(x_{r+1})]$, define

$$\dot{\hat{\mu}}_t^i(x) := \begin{cases} -1 & \text{if } x = x_{s+1}^{x_{r+1}} \\ 1 & \text{if } x = x_{s+2}^{x_{r+1}} \\ 0 & \text{otherwise,} \end{cases}$$

i.e., we transport a mass of $\varepsilon_i(x_{r+1})$ at unit rate from the node $x_{s+1}^{x_{r+1}}$ to $x_{s+2}^{x_{r+1}}$ during the above time interval. Note that we have $\hat{\mu}_t^i(x) = \hat{\mu}_t(x)$ for all $x \in \tilde{\mathcal{X}}_0$ at time $t = \tau^i + \sum_{m=1}^{|\tilde{\mathcal{X}}_0|} l(x_m)\varepsilon_i(x_m)$. Similarly, for $x \in \mathcal{X} \setminus \tilde{\mathcal{X}}_0$ with $\varepsilon_i(x) > 0$, one defines a sequence of edges from a suitable $x' \in \mathcal{X} \setminus \tilde{\mathcal{X}}_0$ (possibly from multiple $x' \in \mathcal{X} \setminus \tilde{\mathcal{X}}_0$) with $\varepsilon_i(x') < 0$ and moves the mass $\varepsilon_i(x)$ to x through similar piecewise constant velocity trajectories defined above. For each $x \in \mathcal{X} \setminus \tilde{\mathcal{X}}_0$ with $\varepsilon_i(x) < 0$, we similarly move the mass $\varepsilon_i(x)$ from x to suitable vertices in $\mathcal{X} \setminus \tilde{\mathcal{X}}_0$ via piecewise constant velocity

trajectories. At the end of this procedure, we have $\hat{\mu}_{\hat{\tau}^i}^i = \hat{\mu}_{1/i}$ for some $\hat{\tau}^i \geq \tau^i$. We now define $\hat{\mu}_t^i = \hat{\mu}_{t+1/i-\hat{\tau}^i}$ for all $t \in [\hat{\tau}^i, T]$ (see Figure 1 for a pictorial representation of $\hat{\mu}^i$). Since $\varepsilon_i(x) \rightarrow 0$ as $i \rightarrow \infty$ for all $x \in \mathcal{X}$, we have that $\hat{\tau}^i \rightarrow 0$ as $i \rightarrow \infty$.

Also, for each $i \geq 1$ and $t \in [0, T]$, define the probability measure \hat{m}_t^i on \mathcal{Y} by

$$\hat{m}_t^i(y) := \begin{cases} \hat{m}_t(y) & \text{if } t \in [0, \tau^i], \\ \hat{m}_{\tau^i}(y) & \text{if } t \in [\tau^i, \hat{\tau}^i], \\ \hat{m}_{t+1/i-\hat{\tau}^i}(y) & \text{if } t \in (\hat{\tau}^i, T], \end{cases}$$

for all $y \in \mathcal{Y}$, and define the measure $\hat{\theta}^i$ on $[0, T] \times \mathcal{Y}$ by $\hat{\theta}^i(dydt) = \hat{m}_t^i(dy)dt$. Clearly, $\hat{\theta}^i \in D_{\uparrow}([0, T], M(\mathcal{Y}))$.

Thanks to the fact that $\hat{\mu}_{\tau^i}^i(x) > 0$ for all $x \in \mathcal{X}$ and the fact that $\alpha_t = 0$ attains the supremum in the definition of $I^*(\hat{\mu}^i, \hat{\theta}^i)$ for all $t \in [0, \tau^i]$, using arguments similar to those used in the proof of Theorem 6.2, one can now construct a bounded measurable function $\hat{\alpha}^i : [0, T] \times M_1(\mathcal{X}) \rightarrow \mathbb{R}^{|\mathcal{X}|}$ such that $\hat{\alpha}_t^i(\hat{\mu}_t^i)$ attains the supremum over α_t in the definition of $I^*(\hat{\mu}^i, \hat{\theta}^i)$ (in (5.2)) and $\hat{\alpha}_t^i(\cdot)$ is continuous on $M_1(\mathcal{X})$ for all $t \in [0, T]$. Similarly, since $\hat{\theta}^i$ satisfies the conditions of Theorem 6.2, one can construct a bounded measurable function $\hat{g}^i : [0, T] \times M_1(\mathcal{X}) \times \mathcal{Y} \rightarrow \mathbb{R}$ such that $\hat{g}_t^i(\hat{\mu}_t^i, \cdot)$ attains the supremum over g_t in the definition of $I^*(\hat{\mu}^i, \hat{\theta}^i)$ and $\hat{g}_t^i(\cdot)$ is continuous on $M_1(\mathcal{X})$ for each $t \in [0, T]$. Hence, using arguments similar to those used in the proof of Theorem 6.2, one concludes that $\tilde{I}(\hat{\mu}^i, \hat{\theta}^i) = I^*(\hat{\mu}^i, \hat{\theta}^i)$ for all $i \geq 1$.

Let us now show that $I^*(\hat{\mu}^i, \hat{\theta}^i) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$. For the fast component, since $\hat{\tau}^i \rightarrow 0$, we see that $\hat{\theta}^i \rightarrow \hat{\theta}$ in $D_{\uparrow}([0, T], M(\mathcal{Y}))$ as $i \rightarrow \infty$. By assumption (B2), we see that

$$0 \leq \sup_{i \geq 1, t \in [0, T]} \left\{ \sup_{g_t \in \mathbb{R}^{|\mathcal{Y}|}} - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t^i} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t^i) \right) \hat{m}_t^i(dy) \right\} < +\infty,$$

and hence the bounded convergence theorem immediately yields

$$\int_{[0, \hat{\tau}^i]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t^i} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t^i) \right) \hat{m}_t^i(dy) \right\} dt \rightarrow 0$$

and

$$\int_{[T+1/i-\hat{\tau}^i, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) \hat{m}_t(dy) \right\} dt \rightarrow 0$$

as $i \rightarrow \infty$. Noting that $\hat{m}_t^i = \hat{m}_{t+1/i-\hat{\tau}^i}$ and $\hat{\mu}_t^i = \hat{\mu}_{t+1/i-\hat{\tau}^i}$ for all $t \in [\hat{\tau}^i, T]$, the above convergences imply that

$$\begin{aligned} & \int_{[0, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t^i} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t^i) \right) \hat{m}_t^i(dy) \right\} dt \\ & \rightarrow \int_{[0, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) \hat{m}_t(dy) \right\} dt \end{aligned}$$

as $i \rightarrow \infty$.

For the slow component, since $\hat{\tau}^i \rightarrow 0$ as $i \rightarrow \infty$, using the absolute continuity of the mapping $t \mapsto \hat{\mu}_t$ and the definition of the paths $\hat{\mu}^i$, it follows from the dominated convergence theorem that

$\hat{\mu}_t^i \rightarrow \hat{\mu}_t$ as $i \rightarrow \infty$ uniformly in $t \in [0, T]$ and hence we have that $\hat{\mu}^i \rightarrow \hat{\mu}$ in $D([0, T], M_1(\mathcal{X}))$ as $i \rightarrow \infty$. Let us first show that

$$\int_{[0, \hat{\tau}^i]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} dt$$

converges to 0 as $i \rightarrow \infty$. Towards this, let $t \in [\tau^i + \sum_{m=0}^r l(x_m)\varepsilon_i(x_m) + s\varepsilon_i(x_{r+1}), \tau^i + \sum_{m=0}^r l(x_m)\varepsilon_i(x_m) + (s+1)\varepsilon_i(x_{r+1})]$ where $r \in \{0, 1, \dots, |\tilde{\mathcal{X}}_0| - 1\}$, and $s \in \{0, 1, \dots, l(x_{r+1}) - 1\}$. Note that, we have

$$\begin{aligned} & \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} \\ & \leq \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left((\alpha_t(x_{s+2}^{x_{r+1}}) - \alpha_t(x_{s+1}^{x_{r+1}})) - (\exp\{\alpha_t(x_{s+2}^{x_{r+1}}) - \alpha_t(x_{s+1}^{x_{r+1}})\} - 1) \right. \\ & \quad \left. \times \bar{\lambda}_{x_{s+1}^{x_{r+1}}, x_{s+2}^{x_{r+1}}}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(x_{s+1}^{x_{r+1}}) \right) \\ & \quad - \inf_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \sum_{\substack{(x, x') \in \mathcal{E}_{\mathcal{X}}: \\ (x, x') \neq (x_{s+1}^{x_{r+1}}, x_{s+2}^{x_{r+1}})}} (\exp\{\alpha_t(x') - \alpha_t(x)\} - 1) \bar{\lambda}_{x, x'}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(x) \\ & \leq \log \frac{1}{c \hat{\mu}_t^i(x_{s+1}^{x_{r+1}})} + c_1 \end{aligned}$$

where $c = \min_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \min_{y \in \mathcal{Y}} \min_{\xi \in M_1(\mathcal{X})} \lambda_{x, x'}(\xi, y)$ and $c_1 > 0$ is a suitable constant to bound the extra additive terms. Hence, using a variable change $u = c \hat{\mu}_t^i(x_{s+1}^{x_{r+1}})$, we see that

$$\begin{aligned} & \int \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} dt \\ & \leq -\frac{1}{c} (u \log u - u) \Big|_{c \hat{\mu}_{t_1}^i(x_{s+1}^{x_{r+1}})}^{c \hat{\mu}_{t_2}^i(x_{s+1}^{x_{r+1}})} + c_1 \varepsilon_i(x_{r+1}) \\ & = o(1) \end{aligned}$$

as $i \rightarrow \infty$, where $t_1 = \tau^i + \sum_{m=0}^r l(x_m)\varepsilon_i(x_m) + s\varepsilon_i(x_{r+1})$, $t_2 = t_1 + \varepsilon_i(x_{r+1})$ and the above integral is evaluated over the time interval $[\tau^i + \sum_{m=0}^r l(x_m)\varepsilon_i(x_m) + s\varepsilon_i(x_{r+1}), \tau^i + \sum_{m=0}^r l(x_m)\varepsilon_i(x_m) + (s+1)\varepsilon_i(x_{r+1})]$. Hence, repeating the above calculation for each constant velocity section of the path $\hat{\mu}^i$ during the time interval $[\tau^i, \hat{\tau}^i]$, we see that

$$\int_{[0, \hat{\tau}^i]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} dt$$

converges to 0 as $i \rightarrow \infty$. Therefore, noting that $\hat{\mu}_t^i = \hat{\mu}_{t+1/i-\hat{\tau}^i}$ and $\hat{m}_t^i = \hat{m}_{t+1/i-\hat{\tau}^i}$ for $t \in [\hat{\tau}^i, T]$,

and $\hat{\mu}_t^i = \mu_t(\hat{\mu}_0, \hat{\theta})$ on $t \in [0, \tau^i]$, we have

$$\begin{aligned}
& \left| \int_{[0, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \hat{\mu}_t \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\} dt \right. \\
& \quad \left. - \int_{[0, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} dt \right| \\
& \leq \int_{[0, 1/i]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \hat{\mu}_t \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\} dt \\
& \quad + \int_{[T+1/i-\hat{\tau}^i, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \hat{\mu}_t \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\} dt \\
& \quad + \int_{[0, \hat{\tau}^i]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} dt \\
& \rightarrow 0
\end{aligned}$$

as $i \rightarrow \infty$. We have thus shown that $I^*(\hat{\mu}^i, \hat{\theta}^i) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$.

Since $(\hat{\mu}^i, \hat{\theta}^i) \rightarrow (\hat{\mu}, \hat{\theta})$ in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ as $i \rightarrow \infty$, the lower semicontinuity of \tilde{I} implies that $\liminf_{i \rightarrow \infty} \tilde{I}(\hat{\mu}^i, \hat{\theta}^i) \geq \tilde{I}(\hat{\mu}, \hat{\theta})$. Therefore, using the above convergence and the fact that $\tilde{I}(\hat{\mu}^i, \hat{\theta}^i) = I^*(\hat{\mu}^i, \hat{\theta}^i)$ for all $i \geq 1$, we see that $\tilde{I}(\hat{\mu}, \hat{\theta}) \leq I^*(\hat{\mu}, \hat{\theta})$. On the other hand, since $\tilde{I}(\hat{\mu}, \hat{\theta}) \geq I^*(\hat{\mu}, \hat{\theta})$, it follows that $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$. This completes the proof of the lemma. \square

Remark 7.2. We shall repeatedly use the immediately preceding argument; starting with an element $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, we shall produce a sequence $(\hat{\mu}^i, \hat{\theta}^i) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, $i \geq 1$, such that $\tilde{I}(\hat{\mu}^i, \hat{\theta}^i) = I^*(\hat{\mu}^i, \hat{\theta}^i)$ for all $i \geq 1$, $(\hat{\mu}^i, \hat{\theta}^i) \rightarrow (\hat{\mu}, \hat{\theta})$ in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ as $i \rightarrow \infty$ and $I^*(\hat{\mu}^i, \hat{\theta}^i) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$, and use the above argument to conclude that $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.

We now extend the conclusion of the previous lemma to all elements $\hat{\theta} \in D_{\uparrow}([0, T], M(\mathcal{Y}))$.

Lemma 7.3. *Let $\nu \in M_1(\mathcal{X})$ and let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function such that $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. Suppose that $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ is such that*

- $I^*(\hat{\mu}, \hat{\theta}) < +\infty$,
- $\inf_{t \in [\delta, T]} \min_{x \in \mathcal{X}} \hat{\mu}_t(x) > 0$ for all $\delta > 0$,
- the mapping $[0, T] \ni t \mapsto \hat{\mu}_t \in M_1(\mathcal{X})$ is Lipschitz continuous.

Then $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.

Proof. Let $\hat{\theta}$, when viewed as a measure on $[0, T] \times \mathcal{Y}$, admit the representation $\hat{\theta}(dydt) = \hat{m}_t(dy)dt$, where $\hat{m}_t \in M_1(\mathcal{Y})$ for almost all $t \in [0, T]$. For each $i \geq 1$ and for each $t \in [0, T]$, define the probability measure \hat{m}_t^i on \mathcal{Y} by

$$\hat{m}_t^i(y) = \frac{\hat{m}_t(y) + 1/i}{1 + |\mathcal{Y}|/i}, y \in \mathcal{Y}, \quad (7.1)$$

and, for each $i \geq 1$, define the measure $\hat{\theta}^i(dydt)$ on $[0, T] \times M(\mathcal{Y})$ by $\hat{\theta}^i(dydt) := \hat{m}_t^i(dy)dt$. Clearly, $\hat{\theta}^i \in D_{\uparrow}([0, T], M(\mathcal{Y}))$ for all $i \geq 1$, and $\hat{\theta}^i \rightarrow \hat{\theta}$ in $D_{\uparrow}([0, T], M(\mathcal{Y}))$ as $i \rightarrow \infty$. Since $(\hat{\mu}, \hat{\theta}^i)$ satisfies the assumptions of Lemma 7.1, we have $\tilde{I}(\hat{\mu}, \hat{\theta}^i) = I^*(\hat{\mu}, \hat{\theta}^i)$.

Since, for each $t \in [0, T]$, the mapping

$$(g_t, m_t) \mapsto \max \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) m_t(dy), 0 \right\}$$

on $(\mathbb{R} \cup \{+\infty, -\infty\})^{|\mathcal{Y}|} \times M_1(\mathcal{Y})$ is bounded and continuous (thanks to assumption (B2)), by an application of the Berge's maximum theorem, it follows that the mapping

$$m_t \mapsto \sup_{g_t \in \mathbb{R}^{|\mathcal{Y}|}} - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_y} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) m_t(dy) \quad (7.2)$$

is continuous on $M_1(\mathcal{Y})$. Similarly, for each $t \geq 0$, by assumption (A2), it follows that the mapping

$$(\alpha_t, m_t) \mapsto \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, m_t}^* \rangle - \int_{\mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, m_t) \hat{\mu}_t(dx)$$

is bounded and continuous on $\mathbb{R}^{|\mathcal{X}|} \times M_1(\mathcal{Y})$. Again, by the Berge's maximum theorem,

$$m_t \mapsto \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, m_t}^* \rangle - \int_{\mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, m_t) \hat{\mu}_t(dx) \right\}$$

is continuous on $M_1(\mathcal{Y})$. Therefore, for each $t \in [0, T]$, we see that

$$\begin{aligned} & \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t^i}^* \rangle - \int_{\mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t^i) \hat{\mu}_t(dx) \right\} \\ & \rightarrow \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \rangle - \int_{\mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\}, \end{aligned}$$

and

$$\begin{aligned} & \sup_{g_t \in B(\mathcal{Y})} - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_y} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) \hat{m}_t^i(dy) \\ & \rightarrow \sup_{g_t \in B(\mathcal{Y})} - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_y} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) \hat{m}_t(dy) \end{aligned}$$

as $i \rightarrow \infty$. Noting that

$$\begin{aligned} 0 \leq & \sup_{i \geq 1, t \in [0, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t^i}^* \rangle \right. \\ & \left. - \int_{\mathcal{X} \times \mathcal{E}_\mathcal{X}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t^i) \hat{\mu}_t(dx) \right\} < +\infty \end{aligned}$$

and

$$\begin{aligned} 0 \leq & \sup_{i \geq 1, t \in [0, T]} \sup_{g_t \in \mathbb{R}^{|\mathcal{Y}|}} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) \right. \right. \\ & \left. \left. + \int_{\mathcal{E}_y} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) \hat{m}_t^i(dy) \right\} < +\infty, \end{aligned}$$

using the bounded convergence theorem, we obtain that $I^*(\hat{\mu}, \hat{\theta}^i) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$. Thanks to Remark 7.2, this completes the proof of the lemma. \square

We now extend the conclusion of the previous lemma to the case when the mapping $[0, T] \ni t \mapsto \mu_t \in M_1(\mathcal{X})$ is not necessarily Lipschitz continuous.

Lemma 7.4. *Let $\nu \in M_1(\mathcal{X})$ and let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function such that $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. Suppose that $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_\uparrow([0, T], M(\mathcal{Y}))$ is such that $I^*(\hat{\mu}, \hat{\theta}) < +\infty$, and $\inf_{t \in [\delta, T]} \min_{x \in \mathcal{X}} \hat{\mu}_t(x) > 0$ for all $\delta > 0$. Then $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.*

Proof. Let us first suppose that the mapping $t \mapsto \hat{\mu}_t$ is locally Lipschitz continuous at $t = 0$ so that $\sup_{t \in [0, \eta]} \|\dot{\hat{\mu}}_t\| < +\infty$ for some $\eta > 0$. Define a sequence of paths $\hat{\mu}_t^i$, $i \geq 1$, by $\hat{\mu}_0^i = \hat{\mu}_0$, and

$$\dot{\hat{\mu}}_t^i = \dot{\hat{\mu}}_t 1_{\{\|\hat{\mu}_t\| \leq i\}} + \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t}^* \hat{\mu}_t^i 1_{\{\|\hat{\mu}_t\| > i\}}, \quad t \in [0, T].$$

Since $I^*(\hat{\mu}, \hat{\theta}) < +\infty$, by Lemma 5.1, it follows that the mapping $t \mapsto \hat{\mu}_t$ is absolutely continuous and by the dominated convergence theorem one easily concludes that $\hat{\mu}_t^i \rightarrow \hat{\mu}_t$ as $i \rightarrow \infty$ uniformly in $t \in [0, T]$. Thus, by the assumption $\inf_{t \in [\delta, T]} \min_{x \in \mathcal{X}} \hat{\mu}_t(x) > 0$ for all $\delta > 0$, it follows that $\hat{\mu}^i \in D([0, T], M_1(\mathcal{X}))$ for all i sufficiently large. Note that $(\hat{\mu}^i, \hat{\theta})$ satisfies the conditions of Lemma 7.3 and hence $\tilde{I}(\hat{\mu}^i, \hat{\theta}) = I^*(\hat{\mu}^i, \hat{\theta})$ for all $i \geq 1$, that $\hat{\mu}^i \rightarrow \hat{\mu}$ in $D([0, T], M_1(\mathcal{X}))$ as $i \rightarrow \infty$, and that $\hat{\mu}_t^i = \hat{\mu}_t$ for all $t \in [0, \eta]$ for all sufficiently large i .

Let us now show that $I^*(\hat{\mu}^i, \hat{\theta}) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$. By the arguments similar to those used in the proof of Lemma 7.3, using Berge's maximum theorem, for each $t \in [0, T]$, the mapping

$$u \mapsto \sup_{g_t \in B(\mathcal{Y})} - \int_{\mathcal{Y}} \left(L_{u, g_t}(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(u) \right) \hat{m}_t(dy)$$

is continuous on $M_1(\mathcal{X})$, and hence

$$\begin{aligned} & \sup_{g_t \in B(\mathcal{Y})} - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t^i} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t^i) \right) \hat{m}_t(dy) \\ & \rightarrow \sup_{g_t \in B(\mathcal{Y})} - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) \hat{m}_t(dy) \end{aligned}$$

as $i \rightarrow \infty$. Therefore, by the bounded convergence theorem, we have

$$\begin{aligned} & \int_{[0, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t^i} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t^i) \right) \hat{m}_t(dy) \right\} dt \\ & \rightarrow \int_{[0, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \right) \hat{m}_t(dy) \right\} dt \end{aligned}$$

as $i \rightarrow \infty$.

For the slow component, define

$$\begin{aligned} Z_t^i := & \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t}^* \hat{\mu}_t^i \rangle \right. \\ & \left. - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t) \hat{\mu}_t^i(dx) \right\}, \quad t \in [0, T], \end{aligned}$$

and

$$\begin{aligned} Z_t := & \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \hat{\mu}_t \rangle \right. \\ & \left. - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\}, \quad t \in [0, T]. \end{aligned}$$

Since $I^*(\hat{\mu}, \hat{\theta}) < +\infty$ it follows that $Z_t < +\infty$ for almost all $t \in [0, T]$. Thanks to the assumption $\inf_{t \in [\delta, T]} \min_{x \in \mathcal{X}} \hat{\mu}_t(x) > 0$ for all $\delta > 0$, using the Berge's maximum theorem, for almost all $t \in [\eta, T]$, we see that the mapping

$$u \mapsto \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \dot{\hat{\mu}}_t - \bar{\Lambda}_{u, \hat{m}_t}^* u \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(u, \hat{m}_t) u(dx) \right\}$$

on $M_1(\mathcal{X})$ is continuous at $\hat{\mu}_t$. Hence, noting that $Z_t^i = Z_t$ on $t \in [0, \eta]$ for all i sufficiently large, for all $t \in [0, T] \cap \{s \in [0, T] : Z_s < +\infty\}$ we have $\hat{\mu}_t^i = \hat{\mu}_t$ for all i sufficiently large, and $\hat{\mu}_t^i \rightarrow \hat{\mu}_t$ as $i \rightarrow \infty$ uniformly in $t \in [0, T]$, it follows that for all $t \in [0, T] \cap \{s \in [0, T] : Z_s < +\infty\}$ $Z_t^i \rightarrow Z_t$ as $i \rightarrow \infty$. Let us now show the convergence of the corresponding integrals. Fix $t \in (0, T]$ such that $Z_t < +\infty$ and let $\hat{\alpha}_t^i \in \mathbb{R}^{|\mathcal{X}|}$ and $\hat{\alpha}_t \in \mathbb{R}^{|\mathcal{X}|}$ attain the supremum in the definition of Z_t^i and Z_t respectively. Whenever $\|\hat{\mu}_t^i\| \leq i$, we have,

$$\begin{aligned}
0 \leq Z_t^i &= \langle \hat{\alpha}_t^i, \hat{\mu}_t^i \rangle - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} (\exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} - 1) \bar{\lambda}_{x, x'}(\hat{\mu}_t^i, \hat{m}_t) \hat{\mu}_t^i(x) \\
&= \langle \hat{\alpha}_t^i, \hat{\mu}_t \rangle - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} (\exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} - 1) \bar{\lambda}_{x, x'}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(x) \\
&\quad - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} (\exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} - 1) \times (\bar{\lambda}_{x, x'}(\hat{\mu}_t^i, \hat{m}_t) \hat{\mu}_t^i(x) - \bar{\lambda}_{x, x'}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(x)) \\
&\leq Z_t - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} (\exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} - 1) \times (\bar{\lambda}_{x, x'}(\hat{\mu}_t^i, \hat{m}_t) \hat{\mu}_t^i(x) - \bar{\lambda}_{x, x'}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(x)). \quad (7.3)
\end{aligned}$$

Since $\hat{\mu}_t^i = \hat{\mu}_t$, $t \in [0, \eta]$, for all large enough i , the second term above vanishes whenever $t \in [0, \eta]$. Since $\hat{\mu}_t^i \rightarrow \hat{\mu}_t$ as $i \rightarrow \infty$ uniformly in $t \in [0, T]$, the first order optimality condition for $(\hat{\alpha}_t^i(x), x \in \mathcal{X})$ (see (6.18)) implies that, for some constants $c_\eta > 0$, we have

$$\max_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} \leq c_\eta(1 + \|\hat{\mu}_t\|)$$

whenever $t \in [\eta, T] \cap \{s \in [0, T] : Z_s < +\infty\}$. In particular, the right hand side of (7.3) is integrable. Hence, noting that $Z_t^i = 0$ in the alternative case when $\|\hat{\mu}_t\| > i$, by an application of the dominated convergence theorem, we have that

$$\int_{[0, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t) \hat{\mu}_t^i(dx) \right\} dt$$

converges to

$$\int_{[0, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \hat{\mu}_t \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\} dt$$

as $i \rightarrow \infty$. Hence, combining the convergences for the slow and the fast components, we have $I^*(\hat{\mu}^i, \hat{\theta}) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$. Further, by Remark 7.2, it follows that $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.

In the general case when the mapping $t \mapsto \hat{\mu}_t$ is not locally Lipschitz continuous at $t = 0$, using arguments similar to those used in the proof of Lemma 7.1, one constructs a sequence $\hat{\tau}^i$, $i \geq 1$, and a sequence of elements $(\hat{\mu}^i, \hat{\theta}^i) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, $i \geq 1$, such that $\hat{\tau}^i \rightarrow 0$ as $i \rightarrow \infty$, $\sup_{t \in [0, \hat{\tau}^i]} \|\hat{\mu}_t^i\| < +\infty$ (therefore the mapping $t \mapsto \hat{\mu}_t^i$ is locally Lipschitz continuous at $t = 0$), $(\hat{\mu}^i, \hat{\theta}^i) \rightarrow (\hat{\mu}, \hat{\theta})$ in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ as $i \rightarrow \infty$, $\hat{\mu}_t^i = \hat{\mu}_{t+1/i-\hat{\tau}^i}$ and $\hat{m}_t^i = \hat{m}_{t+1/i-\hat{\tau}^i}$ for all $t \in [\hat{\tau}^i, T]$, and

$$\begin{aligned}
&\int_{[0, \hat{\tau}^i] \cup [T+1/i-\hat{\tau}^i, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} dt \\
&+ \int_{[0, \hat{\tau}^i] \cup [T+1/i-\hat{\tau}^i, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t^i} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t^i) \right) \hat{m}_t^i(dy) \right\} dt
\end{aligned}$$

converges to 0 as $i \rightarrow \infty$ (by using the small cost construction of constant velocity paths). Based on what we have already shown for paths that are locally Lipschitz continuous at $t = 0$, we see

that $\tilde{I}(\hat{\mu}^i, \hat{\theta}^i) = I^*(\hat{\mu}^i, \hat{\theta}^i)$ for all $i \geq 1$. Again, using arguments similar to those used in the proof of Lemma 7.1, we conclude that $I^*(\hat{\mu}^i, \hat{\theta}^i) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$. Once again, by Remark 7.2, we have $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$. This completes the proof of the lemma. \square

We finally show that $\tilde{I}(\mu, \theta) = I^*(\mu, \theta)$ for all $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, by allowing the path μ to hit the boundary of $M_1(\mathcal{X})$.

Theorem 7.5. *Let $\nu \in M_1(\mathcal{X})$ and let $\tilde{I} : D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y})) \rightarrow [0, +\infty]$ be a subsequential rate function such that $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. Then, for all $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, we have $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.*

Proof. Since $\tilde{I}(\mu, \theta) \geq I^*(\mu, \theta)$ for all $(\mu, \theta) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$, it suffices to focus on a $(\hat{\mu}, \hat{\theta}) \in D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ such that $I^*(\hat{\mu}, \hat{\theta}) < +\infty$ and $\hat{\mu}_0 = \nu$. By Lemma 5.1, we have that the mapping $[0, T] \ni t \mapsto \hat{\mu}_t \in M_1(\mathcal{X})$ is absolutely continuous. In particular, $\hat{\mu}_t$ exists for almost all $t \in [0, T]$ and $\hat{\mu}_t = \nu + \int_{[0, t]} \hat{\mu}_s ds$ for all $t \in [0, T]$.

We shall construct a sequence of paths $\hat{\mu}^i \in D([0, T], M_1(\mathcal{X}))$, $i \geq 1$, such that $\hat{\mu}^i \rightarrow \hat{\mu}$ in $D([0, T], M_1(\mathcal{X}))$ as $i \rightarrow \infty$, $\tilde{I}(\hat{\mu}^i, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$ for all $i \geq 1$, and $I^*(\hat{\mu}^i, \hat{\theta}) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$.

Let $\varepsilon_i(x) = \frac{\hat{\mu}_{1/i}(x) + 1/i}{1 + |\mathcal{X}|/i}$, $x \in \mathcal{X}$ and $i \geq 1$. Using arguments similar to those used in the proof of Lemma 7.1, we first construct a sequence of times $\hat{\tau}^i$, $i \geq 1$, and a sequence of piecewise constant velocity trajectories $\hat{\mu}_t^i$, $t \in [0, \hat{\tau}^i]$, with the property that $\hat{\mu}_0^i = \hat{\mu}_0$ for all $i \geq 1$, $\hat{\mu}_{\hat{\tau}^i}^i(x) = \varepsilon_i(x)$ for all $x \in \mathcal{X}$ and $i \geq 1$, $\hat{\tau}^i \rightarrow 0$ as $i \rightarrow \infty$, and

$$\int_{[0, \hat{\tau}^i]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t) \hat{\mu}_t^i(dx) \right\} dt \rightarrow 0 \quad (7.4)$$

as $i \rightarrow \infty$. We then define the path $\hat{\mu}_t^i$ on $t \in (\hat{\tau}^i, T]$ by

$$\hat{\mu}_t^i(x) = \frac{\hat{\mu}_{t+1/i-\hat{\tau}^i}(x) + 1/i}{1 + |\mathcal{X}|/i}, x \in \mathcal{X}.$$

Clearly, $\hat{\mu}_t^i \rightarrow \hat{\mu}_t$ as $i \rightarrow \infty$ uniformly in $t \in [0, T]$ and hence $\hat{\mu}^i \rightarrow \hat{\mu}$ in $D([0, T], M_1(\mathcal{X}))$ as $i \rightarrow \infty$. Note that $(\hat{\mu}^i, \hat{\theta})$ satisfies the conditions of Lemma 7.4 and hence we have $\tilde{I}(\hat{\mu}^i, \hat{\theta}) = I^*(\hat{\mu}^i, \hat{\theta})$ for all $i \geq 1$.

We now show that $I^*(\hat{\mu}^i, \hat{\theta}) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$. Using arguments similar to those used in the proof of Lemma 7.4, it is easy to show that

$$\begin{aligned} & \int_{[0, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t^i} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t^i) \hat{m}_t(dy) \right) \hat{m}_t(dy) \right\} dt \\ & \rightarrow \int_{[0, T]} \sup_{g_t \in B(\mathcal{Y})} \left\{ - \int_{\mathcal{Y}} \left(L_{\hat{\mu}_t} g_t(\cdot)(y) + \int_{\mathcal{E}_{\mathcal{Y}}} \tau(Dg_t(y, \Delta)) \gamma_{y, y+d\Delta}(\hat{\mu}_t) \hat{m}_t(dy) \right) \hat{m}_t(dy) \right\} dt \end{aligned} \quad (7.5)$$

as $i \rightarrow \infty$.

To show convergence of the integral corresponding to the slow process, define

$$\begin{aligned} Z_t^i := & \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_{t-1/i+\hat{\tau}^i}^i - \bar{\Lambda}_{\hat{\mu}_{t-1/i+\hat{\tau}^i}^i, \hat{m}_t}^* \hat{\mu}_{t-1/i+\hat{\tau}^i}^i \rangle \right. \\ & \left. - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_{t-1/i+\hat{\tau}^i}^i, \hat{m}_t) \hat{\mu}_{t-1/i+\hat{\tau}^i}^i(dx) \right\}, t \in [1/i, T + 1/i - \hat{\tau}^i], \end{aligned}$$

and

$$Z_t := \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \hat{\mu}_t \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\}, t \in [0, T].$$

Note the shift in the time index in the definition of Z_t^i to enable direct comparison between Z_t and Z_t^i . For $t \in [1/i, T]$, we then have

$$Z_t^i = \frac{1}{1 + |\mathcal{X}|/i} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t \rangle - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} (\exp\{\alpha_t(x') - \alpha_t(x)\} - 1) \bar{\lambda}_{x, x'}(\hat{\mu}_{t-1/i+\hat{\tau}^i}, \hat{m}_t) (\hat{\mu}_t(x) + 1/i) \right\}.$$

The objective function above can be simplified as

$$\begin{aligned} & \langle \alpha_t, \hat{\mu}_t \rangle - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\alpha_t(x') - \alpha_t(x)\} \bar{\lambda}_{x, x'}(\hat{\mu}_{t-1/i+\hat{\tau}^i}, \hat{m}_t) (\hat{\mu}_t(x) + 1/i) \\ &= \langle \alpha_t, \hat{\mu}_t \rangle - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\alpha_t(x') - \alpha_t(x)\} \bar{\lambda}_{x, x'}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(x) \\ & \quad - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\alpha_t(x') - \alpha_t(x)\} \left[(\bar{\lambda}_{x, x'}(\hat{\mu}_{t-1/i+\hat{\tau}^i}, \hat{m}_t) - \bar{\lambda}_{x, x'}(\hat{\mu}_t, \hat{m}_t)) \hat{\mu}_t(x) + \frac{\bar{\lambda}_{x, x'}(\hat{\mu}_{t-1/i+\hat{\tau}^i}, \hat{m}_t)}{i} \right] \\ & \leq \langle \alpha_t, \hat{\mu}_t \rangle - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\alpha_t(x') - \alpha_t(x)\} \bar{\lambda}_{x, x'}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(x) \\ & \quad - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\alpha_t(x') - \alpha_t(x)\} \left(-\frac{c_L \hat{\mu}_t(x)}{i} + \frac{c}{i} \right) \end{aligned}$$

where the last inequality follows from assumption (A2); here $c = \min_{(x, x') \in \mathcal{X}} \min_{y \in \mathcal{Y}} \min_{\xi \in M_1(\mathcal{X})} \lambda_{x, x'}(\xi, y)$ and $c_L = \max_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \max_{y \in \mathcal{Y}} c_L^{x, x', y}$ where $c_L^{x, x', y}$ is the Lipschitz constant of $\lambda_{x, x'}(\cdot, y)$, $(x, x') \in \mathcal{E}_{\mathcal{X}}, y \in \mathcal{Y}$. Fix $t \in [1/i, T + 1/i - \hat{\tau}^i]$ with $Z_t < +\infty$ and let $(\hat{\alpha}_t^i(x), x \in \mathcal{X}) \in \mathbb{R}^{|\mathcal{X}|}$ denote the optimiser in the definition of Z_t^i . Then the above computation gives us

$$\begin{aligned} Z_t^i & \leq \frac{1}{1 + |\mathcal{X}|/i} \left\{ \langle \hat{\alpha}_t^i, \hat{\mu}_t \rangle - \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} \bar{\lambda}_{x, x'}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(x) \right\} \\ & \quad - \frac{1}{1 + |\mathcal{X}|/i} \left\{ \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} \left(-\frac{c_L \hat{\mu}_t(x)}{i} + \frac{c}{i} \right) \right\} \\ & \leq \frac{1}{1 + |\mathcal{X}|/i} Z_t - \frac{1}{1 + |\mathcal{X}|/i} \left\{ \sum_{(x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} \left(-\frac{c_L \hat{\mu}_t(x)}{i} + \frac{c}{i} \right) \right\}. \end{aligned}$$

If $\hat{\mu}_t(x) < c/c_L$ for some $x \in \mathcal{X}$, we see that all the terms in the summation corresponding to the edges $(x, x') \in \mathcal{E}_{\mathcal{X}}$ are negative. On the other hand, if $\hat{\mu}_t(x) > c/c_L$, noting that $\hat{\tau}^i \rightarrow 0$ as $i \rightarrow \infty$ and the convergence of $\hat{\mu}_t^i$ to $\hat{\mu}_t$ as $i \rightarrow \infty$ uniformly in $t \in [0, T]$, the first order optimality condition for $(\hat{\alpha}_t^i(x), x \in \mathcal{X})$ implies that, for some constant $c_2 > 0$,

$$\max_{x' \in \mathcal{X}: (x, x') \in \mathcal{E}_{\mathcal{X}}} \exp\{\hat{\alpha}_t^i(x') - \hat{\alpha}_t^i(x)\} \leq c_2(1 + \|\hat{\mu}_t\|),$$

and hence for all $t \in [1/i, T + 1/i - \hat{\tau}^i]$ with $Z_t < +\infty$, we obtain that

$$Z_t^i \leq \frac{1}{1 + |\mathcal{X}|/i} \{Z_t + c_2 |\mathcal{E}_{\mathcal{X}}| (1 + \|\hat{\mu}_t\|)\}.$$

Hence by the dominated convergence theorem, we see that

$$\int_{[0, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t^i - \bar{\Lambda}_{\hat{\mu}_t^i, \hat{m}_t^i}^* \hat{\mu}_t^i \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t^i, \hat{m}_t^i) \hat{\mu}_t^i(dx) \right\} \times 1_{\{t \geq \hat{\tau}^i\}} dt$$

converges to

$$\int_{[0, T]} \sup_{\alpha_t \in \mathbb{R}^{|\mathcal{X}|}} \left\{ \langle \alpha_t, \hat{\mu}_t - \bar{\Lambda}_{\hat{\mu}_t, \hat{m}_t}^* \hat{\mu}_t \rangle - \int_{\mathcal{X} \times \mathcal{E}_{\mathcal{X}}} \tau(D\alpha_t(x, \Delta)) \bar{\lambda}_{x, x+d\Delta}(\hat{\mu}_t, \hat{m}_t) \hat{\mu}_t(dx) \right\} dt$$

as $i \rightarrow \infty$. This along with the convergences (7.4) and (7.5) implies that $I^*(\hat{\mu}^i, \hat{\theta}) \rightarrow I^*(\hat{\mu}, \hat{\theta})$ as $i \rightarrow \infty$. The procedure of Remark 7.2 then completes the proof of the theorem. \square

8 Completing the Proof of Theorem 2.2

We finally complete the proof of Theorem 2.2 by extending the conclusion of Theorem 7.5 to all subsequential rate functions \tilde{I} , i.e. we remove the restriction that, for some $\nu \in M_1(\mathcal{X})$, $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$.

Proof of Theorem 2.2. Fix $\nu \in M_1(\mathcal{X})$ and suppose that $\{\mu_N\}$ is such that $\limsup_{N \rightarrow \infty} \frac{1}{N} \log P(|\mu_N(0) - \nu| \geq \varepsilon) = -\infty$ for each $\varepsilon > 0$. By Theorem 3.3, the family $\{(\mu_N, \theta_N)\}_{N \geq 1}$ is exponentially tight in $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$. Therefore, there exists a subsequence $\{N_k\}_{k \geq 1}$ of \mathbb{N} such that $\{(\mu_{N_k}, \theta_{N_k})\}_{k \geq 1}$ satisfies the LDP with rate function \tilde{I} (see, for example, Dembo and Zeitouni [9, Lemma 4.1.23]); by the above condition on the family $\{\mu_N\}$ and by the contraction principle, we see that $\tilde{I}(\mu, \theta) = +\infty$ unless $\mu_0 = \nu$. Therefore, by Theorem 7.5, $\tilde{I} = I^*$ on $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$. Hence \tilde{I} is uniquely determined for all such subsequences, and it follows that the family $\{(\mu_N, \theta_N)\}_{N \geq 1}$ satisfies the LDP with rate function I^* (see, for example, Dembo and Zeitouni [9, Exercise 4.4.15 (b)]) defined as follows: $I^*(\mu, \theta)$ is defined by (5.1) whenever μ is such that $\mu(0) = \nu$, and $I^*(\mu, \theta) = +\infty$ otherwise.

In the general case when $\{\mu_N(0)\}$ satisfies the LDP on $M_1(\mathcal{X})$ with rate function I_0 , let $p_{\nu_N}^{(N)}$ denote the regular conditional distribution of (μ_N, θ_N) on $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ given $\mu_N(0) = \nu_N \in M_1^N(\mathcal{X})$. By the above argument, whenever $\nu_N \rightarrow \nu$ in $M_1(\mathcal{X})$, $p_{\nu_N}^{(N)}$ satisfies the LDP on $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ with rate function $I^*(\mu, \theta) + \infty 1_{\mu(0) \neq \nu}$. Therefore, the family $\{(\mu_N, \theta_N)\}_{N \geq 1}$ satisfies the LDP on $D([0, T], M_1(\mathcal{X})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$ with rate function $I_0(\mu(0)) + I^*(\mu, \theta)$ (see, for example, Chaganty [7]). This completes the proof of Theorem 2.2. \square

A Examples of two time scale mean-field models

We describe two applications that can be studied using our two time scale mean-field model – a retrial system with orbit queues and a wireless local area network (WLAN) with local interaction.

Example 1. We first describe a retrial system with orbit queues (see Figure 2). Such systems have been used to model multiple competing jobs in a carrier sense multiple access network (see Avrachenkov et al. [1] and the references therein). In this model, there is a single exponential server with service rate N , N statistically identical Poisson arrival streams (of rate λ) and N orbit queues of identical (finite) size K , one corresponding to each arrival stream. Whenever an arriving

customer finds an empty server, it occupies the server and spends a random amount of time, exponentially distributed with mean $1/N$, and then leaves the system. If the arriving customer sees a busy server, it waits in the orbit queue corresponding to that arrival stream, if the queue is not full. Whenever an orbit queue is nonempty and the server is free, the head of the line customer in that orbit queue attempts for service at a fixed positive rate α . In this setting, the state of the server (i.e. idle/busy) represents the environment, and the number of waiting customers in an orbit queue represents the state of that node. Note that the state of each orbit queue evolves slowly (i.e. $O(1)$ many transitions in a given $O(1)$ duration of time). But since there are N orbit queues and each nonempty queue attempts for service with a fixed positive rate, the environment makes $O(N)$ many transitions in a given $O(1)$ duration of time. Also, the transition rates of the number of customers in a queue depend on the state of the server and the transition rates of the environment depend on the fraction of non-empty orbit queues. Figure 3 depicts the transition rates of each orbit queue when the server state is y ($y = 0$ indicates idle state and $y = 1$ indicates busy state), and Figure 4 depicts the transition rates of the server when the empirical measure of the states of all the orbit queues is ξ . Clearly, this system falls within the framework of our two time scale mean-field model.

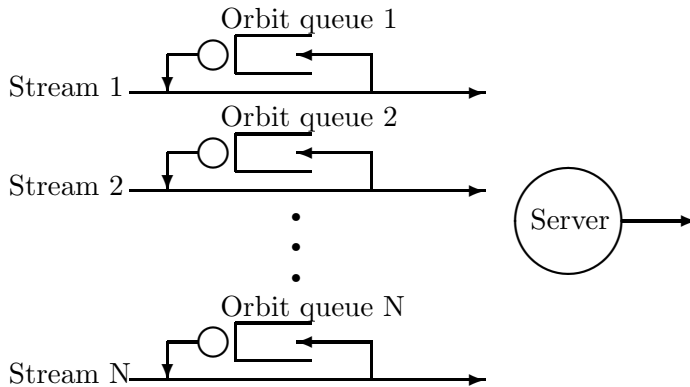


Figure 2: A retrial system with N orbit queues

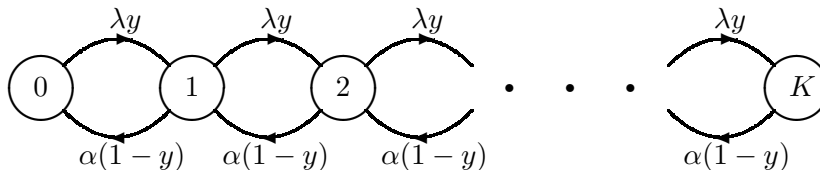


Figure 3: Transition rates of an orbit queue when the server state is y

Example 2. We now describe the setting of WLAN. Let there be N nodes. Time is divided into slots. Each node has a state associated with it, which represents the probability of attempting a packet transmission in a slot. Since the network could be spread over a large geographical area, the nodes are grouped into C classes; every node that belongs to a class can hear the transmissions of every other node in that class. Figure 5 depicts an example network with 7 nodes and 3 classes. The interaction among the nodes comes from the distributed channel access algorithm executed by the nodes. This interaction results in the evolution of the state of each node in the following

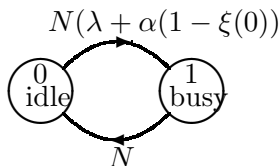


Figure 4: Transition rates of the server when the empirical measure of nodes is ξ ; $\xi(0)$ denotes the fraction of empty orbit queues

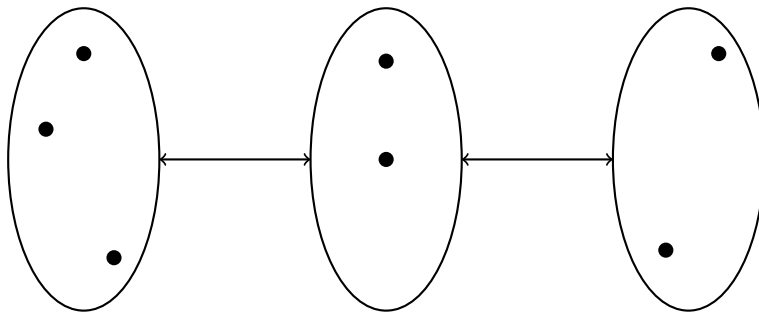


Figure 5: A wireless local area network with 3 classes and 7 users; interference among classes are indicated by arrows

fashion: a node that incurs a collision upon a packet transmission moves to a different state with a reduced probability of attempt, and upon a successful transmission moves to another state with an increased probability of attempt. Since multiple nodes could transmit at the same slot, the channel corresponding to a class of nodes could be in three different states in a given time slot: (i) an idle slot (denoted by state 0), (ii) a collision (state 2) or (iii) a successful packet transmission (state 1). We denote the channel state corresponding to each class of nodes as the environment, i.e., at each time slot, the environment is an element of $\{0, 1, 2\}^C$ with the c th coordinate representing the channel state of the c th class of nodes. Since there are $O(N)$ many nodes in each class, we see that the environment makes $O(N)$ many transitions over a given $O(1)$ time duration. Also, we see that the transition rates of the environment depend on the attempt probabilities of the nodes in that class, but only through the empirical measure of the states of the nodes in that class. On the other hand, the transition rates of the states of a node depend on the attempt probabilities of the nodes in that class (again, only through the empirical measure) as well as the environment. Hence, we have a two time scale mean-field model that describes the network, but one that operates in discrete-time. We now see how to translate this to an approximate continuous-time model.

Figure 6 depicts the set of allowed transitions of a node; in typical WLAN implementations, the most aggressive state is 0 and the least aggressive state is K . A node moves from state i to state $i + 1$ when it incurs a collision, and moves from state i to 0 when a packet is successfully transmitted. To describe the transition rates of the continuous time model, we shall consider a scaled version of the above discrete time model where each time slot is of duration $1/N$. Let p_i/N denote the attempt probability of a node in state i , and let A denote the interference matrix among

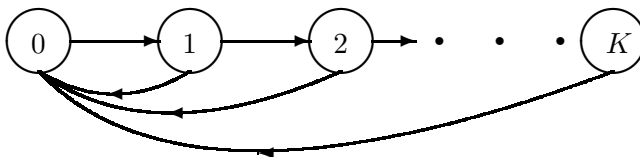


Figure 6: Set of allowed transitions for a particle in a WLAN

classes, specifically, $A_{c,d} = 1$ implies that a class c node's transmission is interfered by a class d node's transmission. Let $V_c = \{d : A_{c,d} = 1\}$ denote the classes that interfere with class c nodes' transmissions. Also, for each $i \in \{0, 1, \dots, K\}$ and $c \in \{1, 2, \dots, C\}$, let ξ_i^c denote the fraction of nodes (among the nodes in class c) in state i and let $y \in \{0, 1, 2\}^C$ denote the state of the background process. The transition probability of tagged node in class c from state i to state 0 is

$$\left(\frac{p_i}{N} \prod_{d \in V_c, d \neq c} \left[1_{\{y_d=0\}} \prod_{j \in c, j \neq i} \left(1 - \frac{p_j}{N} \right) \right] \right) \\ \times \left(\prod_{d \in V_c, d \neq c} \left[\prod_{d' \in V_d} 1_{\{y_{d'}=0\}} \left(\prod_{j \in d'} \left(1 - \frac{p_j}{N} \right) \right) + \left(1 - \prod_{d' \in V_d} 1_{\{y_{d'}=0\}} \right) \right] \right);$$

scaling the above by N and noting that $\prod_{j \in d} (1 - p_j/N) \sim \exp\{-\sum_{i=0}^K p_i \xi_i^d\}$, the corresponding transition rate of the continuous time model can be approximated as

$$p_i \left(\prod_{d \in V_c} 1_{\{y_d=0\}} \right) \times \left(\prod_{d \in V_c} \left[\prod_{d' \in V_d} 1_{\{y_{d'}=0\}} \left(\exp \left\{ -\sum_{i=0}^K p_i \xi_i^d \right\} - 1 \right) + 1 \right] \right).$$

Similarly, the transition rate of a class c node from state i to state $i + 1$ is

$$p_i \left(\prod_{d \in V_c} 1_{\{y_d=0\}} \right) \times \left(1 - \prod_{d \in V_c} \left[\prod_{d' \in V_d} 1_{\{y_{d'}=0\}} \left(\exp \left\{ -\sum_{i=0}^K p_i \xi_i^d \right\} - 1 \right) + 1 \right] \right).$$

We can also write down the transition rates of the background process; for example, a transition from the all-0 state to the state y with $y_c = 1$ and $y_d = 0$ for all $d \neq c$ (which happens when a node in class c starts a transmission) occurs with rate

$$\left(N \sum_{i=0}^K p_i \xi_i^c \right) \times \exp \left\{ -\sum_{i=0}^K p_i \xi_i^c \right\}.$$

A study of the above model in the large- N regime has been done by Bordenave et al. [3] towards understanding the average throughput obtained by a node in a given class, whereas our result in this paper provides a finer asymptotic analysis, in the realm of large deviations, which enables us to study metastability in such systems. For a continuous-time model of WLAN without a fast environment, see Boorstyn et al. [2].

References

- [1] K. Avrachenkov, P. Nain, and U. Yechiali. A retrial system with two input streams and two orbit queues. *Queueing Systems*, 77(1):1–31, 2014.
- [2] R. Boorstyn, A. Kershnerbaum, B. S. Maglaris, and V. Sahin. Throughput analysis in multihop csma packet radio networks. *IEEE Transactions on Communications*, 35:267–274, 1987.
- [3] C. Bordenave, D. McDonald, and A. Proutiere. A particle system in interaction with a rapidly varying environment: Mean field limits and applications. *Netw. Heterog. Media*, 5(1):31–62, 2010.
- [4] V. Borkar and R. Sundaresan. Asymptotics of the invariant measure in mean field models with jumps. *Stoch. Syst.*, 2(2):322–380, 2012.

- [5] A. Budhiraja, P. Dupuis, and A. Ganguly. Large deviations for small noise diffusions in a fast markovian environment. *Electronic Journal of Probability*, 23, 2018.
- [6] E. Castiel, S. Borst, L. Miclo, F. Simatos, and P. Whiting. Induced idleness leads to deterministic heavy traffic limits for queue-based random-access algorithms. *arXiv preprint arXiv:1904.03980*, 2019.
- [7] N. R. Chaganty. Large deviations for joint distributions and statistical applications. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 147–166, 1997.
- [8] D. A. Dawson and J. Gärtner. Large deviations from the mckean-vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.
- [9] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag Berlin Heidelberg, 2 edition, 2010.
- [10] B. Djehiche and I. Kaj. The rate function for some measure-valued jump processes. *Ann. Probab.*, pages 1414–1438, 1995.
- [11] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain markov process expectations for large time, i. *Communications on Pure and Applied Mathematics*, 28(1):1–47, 1975.
- [12] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*, volume 28. SIAM, 1999.
- [13] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley & Sons, 2 edition, 2005.
- [14] J. Feng and T. G. Kurtz. *Large Deviations for Stochastic Processes*. Mathematical Surveys and Monographs. American Mathematical Society, 2 edition, 2006.
- [15] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften. American Mathematical Society, 3 edition, 2012.
- [16] P. Hunt and T. Kurtz. Large loss networks. *Stochastic Processes and their Applications*, 53(2):363–378, 1994.
- [17] P. J. Hunt and C. N. Laws. Optimization via trunk reservation in single resource loss systems under heavy traffic. *The Annals of Applied Probability*, 7(4):1058–1079, 1997.
- [18] C.-R. Hwang and S.-J. Sheu. Large-time behavior of perturbed diffusion markov processes with applications to the second eigenvalue problem for fokker-planck operators and simulated annealing. *Acta Appl. Math.*, 19(3):253–295, 1990.
- [19] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288. Springer Science & Business Media, 2013.
- [20] F. P. Kelly. Loss networks. *The Annals of Applied Probability*, pages 319–378, 1991.
- [21] R. Khasminskii. On the averaging principle for stochastic differential Itô equation. *Kibernetika*, 4(3):260–279, 1968.
- [22] R. Kumar and L. Popovic. Large deviations for multi-scale jump-diffusion processes. *Stochastic Processes and their Applications*, 127(4):1297–1320, 2017.

- [23] C. Léonard. Large deviations for long range interacting particle systems with jumps. *Ann. Inst. Henri Poincaré Probab. Stat.*, 31(2):289–323, 1995.
- [24] C. Léonard. On large deviations for particle systems associated with spatially homogeneous boltzmann type equations. *Probability Theory and Related Fields*, 101(1):1–44, Mar 1995.
- [25] R. Liptser. Large deviations for two scaled diffusions. *Probability theory and Related Fields*, 106(1):71–104, 1996.
- [26] A. Puhalskii. The method of stochastic exponentials for large deviations. *Stochastic Processes and their Applications*, 54(1):45–70, 1994.
- [27] A. Puhalskii. *Large Deviations and Idempotent Probability*. Chapman and Hall/CRC, 2001.
- [28] A. A. Puhalskii. On large deviations of coupled diffusions with time scale separation. *Ann. Probab.*, 44(4):3111–3186, 07 2016.
- [29] M. M. Rao and Z. D. Ren. *Theory of Orlicz Spaces*, volume 146 of *Pure and Applied Mathematics*. Marcel Dekker, Inc., 1991.
- [30] R. K. Sundaram. *A First Course in Optimization Theory*. Cambridge University Press, 1996.
- [31] A. Y. Veretennikov. On large deviations in the averaging principle for sdes with a “full dependence”. *Annals of Probability*, 27(1):284–296, 1999.
- [32] A. Y. Veretennikov. On large deviations for SDEs with small diffusion and averaging. *Stochastic Processes and their Applications*, 89(1):69–79, 2000.
- [33] S. Yasodharan and R. Sundaresan. Large time behaviour and the second eigenvalue problem for finite state mean-field interacting particle systems. *arXiv preprint arXiv:1909.03805*, 2019.