Large deviations for two scale chemical kinetic processes

Tiejun Li [*] and Feng Lin[†]

LMAM and School of Mathematical Sciences, Peking University,
Beijing 100871, China

ABSTRACT

We formulate the large deviations for a class of two scale chemical kinetic processes motivated from biological applications. The result is successfully applied to treat a genetic switching model with positive feedbacks. The corresponding Hamiltonian is convex with respect to the momentum variable as a by-product of the large deviation theory. This property ensures its superiority in the rare event simulations compared with the result obtained by formal WKB asymptotics. The result is of general interest to understand the large deviations for multiscale problems.

1. INTRODUCTION

We will investigate the large deviations for a class of two scale chemical kinetic processes with the slow variable $\boldsymbol{z}_n \in \mathbb{N}^d / n$ satisfying

$$\boldsymbol{z}_n(t) = \boldsymbol{z}_n(0) + \sum_{i=1}^{S} \frac{1}{n} P_i \left(n \int_0^t \lambda_i(\boldsymbol{z}_n(s), \xi_n(s)) ds \right) \boldsymbol{u}_i \qquad (1.1)$$

subject to some fixed initial state $\boldsymbol{z}_n(0) = \boldsymbol{z}^0$, where $\{P_i(t)\}_{i=1,\dots,S}$ are independent uni-rate Poisson processes, $\lambda_i \in \mathbb{R}^+$ is called the propensity function which characterizes the reaction rate of the ith reaction and $\boldsymbol{u}_i \in \mathbb{Z}^d$ is called the state change vector. The number $n \in \mathbb{N}$ corresponds to the system volume, thus $\boldsymbol{z}_n$ has the meaning of concentration (number of molecules per volume) for the considered kinetic system. The fast variable $\xi_n \in \mathbb{Z}_D := \{1, 2, \cdots, D\}$ is a simple jump process with the time dependent rate $n q_{ij}(\boldsymbol{z}_n(t))$ from state i to j at time t. With this mathematical setup, the processes $\boldsymbol{z}_n(t)$ and $\xi_n(t)$ are fully coupled to each
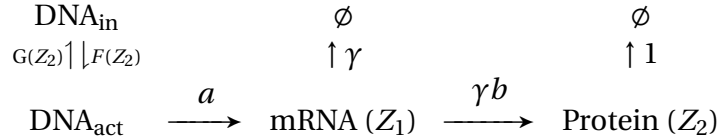
---

[*]email: *tieli@pku.edu.cn*
[†]email: *math_linfeng@pku.edu.cn*

other and the infinitesimal generator $\mathscr{L}_n$ of this system has the form

$$\mathscr{L}_n h(\boldsymbol{z}, i) = n \sum_{l=1}^{S} \lambda_l(\boldsymbol{z}, i)[h(\boldsymbol{z} + \boldsymbol{u}_l/n, i) - h(\boldsymbol{z}, i)]$$

$$+ n \sum_{\substack{j=1 \\ j \neq i}}^{D} q_{ij}(z)[h(\boldsymbol{z}, j) - h(\boldsymbol{z}, i)], \qquad (1.2)$$

where $\boldsymbol{z} \in \mathbb{N}^d/n$, $i \in \mathbb{Z}_D$ and h is any compactly supported smooth function of $\boldsymbol{z}$ for each i. For more about the notations and the backgrounds on the chemical kinetic processes, the readers may be referred to [11, 13].

The above problem is motivated by our recent rare event study in the biological applications [1, 19, 21]. In a cell, the reactions underlying gene expression usually involve low copy number of molecules, such as DNA, mRNAs and transcription factors, so the stochasticity in gene regulation process is inevitable even under constant environmental conditions [10]. When the number of the molecules for all species goes to infinity and the law of mass action holds for the propensity functions, one gets the well-known large volume limit or Kurtz's limit, which gives the deterministic reaction rate equations for the concentration of the species [17]. The convergence result can be further refined to the large deviation type [26]. Recently, the following typical biological model with positive feedbacks is utilized to investigate the robustness of the genetic switching system [1, 4, 19, 21].

$$
\begin{array}{ccc}
\text{DNA}_{\text{in}} & \varnothing & \varnothing \\
G(Z_2) \uparrow \downarrow F(Z_2) & \uparrow \gamma & \uparrow 1 \\
\text{DNA}_{\text{act}} \xrightarrow{\ a\ } & \text{mRNA } (Z_1) \xrightarrow{\ \gamma b\ } & \text{Protein } (Z_2)
\end{array}
$$

This problem is a special case of our formulation shown at the beginning of this paper for $d = 2$, $D = 2$ and $S = 4$. Denote n the system size and $\boldsymbol{z} = (z_1, z_2) = (Z_1, Z_2)/n$ the slow variables after taking large volume scaling, where Z_1 and Z_2 are the number of mRNA and protein molecules, respectively. Since there is only one molecule of DNA at active (DNA$_{\text{act}}$) or inactive state (DNA$_{\text{in}}$), for better use of notation, we take the fast variable $\xi \in \{0, 1\}$ instead of $\{1, 2\}$ to represent that the DNA is at inactive ($\xi = 0$) or active state ($\xi = 1$), respectively. By taking into account the scaling of parameters

$$a \sim nb^{-1}, \quad F(Z_2), G(Z_2) \sim O(n) \quad \text{if} \quad Z_1, Z_2 \sim O(n),$$

we further assume

$$F(Z_2) = nf(z_2), \quad G(Z_2) = ng(z_2). \qquad (1.3)$$

This assumption holds when we consider a Hill-function type jump rates with Hill coefficient 2 and large volume scaling for equilibrium constants [1]. Thus, we have the rescaled jump rates for DNA

$$q_{01}(\boldsymbol{z}) = f(z_2), \quad q_{10}(\boldsymbol{z}) = g(z_2), \tag{1.4}$$

and the following list of reactions associated with slow variables as shown in Table 1.

TABLE 1. Reaction schemes and parameters

Reaction scheme	Propensity function	State change vector
$DNA_{act} \rightarrow mRNA$	$\lambda_1(z_1, z_2, \xi) = b^{-1}\xi$	$\boldsymbol{u}_1 = (1, 0)$
$mRNA \rightarrow \emptyset$	$\lambda_2(z_1, z_2, \xi) = \gamma z_1$	$\boldsymbol{u}_2 = (-1, 0)$
$mRNA \rightarrow Protein$	$\lambda_3(z_1, z_2, \xi) = \gamma b z_1$	$\boldsymbol{u}_3 = (0, 1)$
$Protein \rightarrow \emptyset$	$\lambda_4(z_1, z_2, \xi) = z_2$	$\boldsymbol{u}_4 = (0, -1)$

The infinitesimal generator of this process has the form

$$\mathscr{L}_n h(\boldsymbol{z}, i) = n \sum_{l=1}^{4} \lambda_l(\boldsymbol{z}, i) \Big(h(\boldsymbol{z} + n^{-1}\boldsymbol{u}_l, i) - h(\boldsymbol{z}, i) \Big)$$
$$+ n\Big(f(z_2)[h(\boldsymbol{z}, 1) - h(\boldsymbol{z}, 0)] + g(z_2)[h(\boldsymbol{z}, 0) - h(\boldsymbol{z}, 1)] \Big) \tag{1.5}$$

for $i = 0, 1$. One can obtain a mean field ODE system as

$$\frac{dz_1}{dt} = \frac{b^{-1}f(z_2)}{f(z_2) + g(z_2)} - \gamma z_1, \quad \frac{dz_2}{dt} = \gamma b z_1 - z_2 \tag{1.6}$$

when n goes to infinity through the perturbation analysis for the infinitesimal generator [18, 21, 23]. With suitable choice of the functions $F(Z_2)$ and $G(Z_2)$, the final mean field ODEs have two stable stationary points and there are noise induced transitions between these two states when n is finite. To understand the robustness of the genetic switching, the biophysicists employed the WKB ansatz to the stationary distribution [1]

$$P(Z_1, Z_2) \sim \exp[-nS(z_1, z_2)] \tag{1.7}$$

and obtained a steady state Hamilton-Jacobi equation $H(z_1, z_2, \nabla S) = 0$. Mathematically the function S resembles the role of the quasi-potential of the stochastic dynamical system [12, 22, 32] but it is not sure whether it is the case in the current stage. Another related physics approach to study a similar switching system is to utilize the spin-boson path integral formalism in quantum field theory and then take the semiclassical approximation and adiabatic limit [19, 31]. Both approaches are difficult

to be rationalized in mathematical sense. So how to formulate this problem in a mathematically rigorous way? To resolve this issue, we have to answer the following two fundamental questions.

(1) Question 1. What is the large deviation principle (LDP) associated with the system (1.2)? Presumably, we can obtain the Lagrangian from the large deviation analysis, then get the Hamiltonian H through the Legendre-Fenchel transform.

(2) Question 2. What is the relation between the rigorously obtained Hamiltonian H in the above question and the Hamiltonian obtained via WKB asymptotics?

The aim of this paper is to make an exploration on these two questions. To do this, we first note that the large volume limit no longer holds in the current example. Although the mRNA and protein copy numbers scale as V, we have only one DNA, which switches between the active and inactive states. This fact excludes the direct applicability of the LDP results in [26]. However, the fast switching between the two states of the DNA ensures the averaging technique still valid as shown in (1.6) by taking the quasi-equilibrium limit [8, 16, 21]. We will show that the LDP analysis is also feasible by incorporating the Donsker-Varadhan type large deviations. Indeed, similar situation has been nicely discussed by Liptser [20] and Veretennikov [29, 30] for two-scale diffusions like

$$dX_n(t) = A(X_n(t), \xi_n(t))dt + \frac{1}{\sqrt{n}}B(X_n(t), \xi_n(t))dW_t, \qquad (1.8)$$

$$d\xi_n(t) = nb(\xi_n(t))dt + \sqrt{n}\sigma(\xi_n(t))dV_t. \qquad (1.9)$$

The main idea of this paper is to generalize the result in [20] to our two-scale chemical kinetic processes. As we will see, although the framework is similar, we have to deal with the technicalities brought by the jump processes and the full coupling between the fast and slow variables (ξ_n is independent of X_n in (1.9)).

To state the main results of this paper, let us introduce the occupation measure ν_n on $([0, T] \times \mathbb{Z}_D, \mathscr{B}([0, T]) \otimes \mathscr{B}(\mathbb{Z}_D)$ corresponding to ξ_n

$$\nu_n(\Delta \times \Gamma) = \int_0^T \mathbf{1}(t \in \Delta, \xi_n \in \Gamma)dt, \ \ \Delta \in \mathscr{B}([0, T]), \Gamma \in \mathscr{B}(\mathbb{Z}_D), \qquad (1.10)$$

where T is any fixed positive real number. Denote $\mathbb{D}^d[0, T]$ the space of d-dimensional vector functions on $[0, T]$ whose components are right continuous with left-hand limits, $\mathbb{M}_L[0, T]$ of finite measures $\nu = \nu(dt, i)$ on $([0, T] \times \mathbb{Z}_D, \mathscr{B}([0, T]) \otimes \mathscr{B}(\mathbb{Z}_D))$ which are absolutely continuous with respect to dt and have Lebesgue time marginals, i.e. we have $\nu(dt, i) = n_\nu(t, i)dt$, $n_\nu(t, i) \geq 0$ and $\sum_{i=1}^D n_\nu(t, i) = 1$. The ν_n we considered always

belongs to $\mathbb{M}_L[0, T]$. Take the metric $\rho^{(2)}$ on $\mathbb{M}_L[0, T]$ as the Lévy-Prohorov metric and $\rho^{(1)}$ on $\mathbb{D}^d[0, T]$ as the Skorohod metric defined as

$$\rho^{(1)}(\boldsymbol{r}, \tilde{\boldsymbol{r}}) = \inf_{\lambda \in \mathscr{F}} \left\{ \|\lambda\|^\circ \vee \sup_{t \in [0,T]} \|\boldsymbol{r}(t) - \tilde{\boldsymbol{r}}(\lambda(t))\| \right\}, \tag{1.11}$$

where $\|\cdot\|$ is the Euclidean norm in the corresponding space, $\mathscr{F}$ is the collection of strictly increasing functions $\lambda(t)$ such that $\lambda(0) = 0$ and $\lambda(T) = T$, and

$$\|\lambda\|^\circ := \sup_{0 \le s < t \le T} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|.$$

$\mathbb{D}^d[0, T]$ and $\mathbb{M}_L[0, T]$ are complete and separable spaces with $\rho^{(1)}$ and $\rho^{(2)}$, respectively [2]. Our task is to establish the LDP for the pair $(\boldsymbol{z}_n, \nu_n)$ in metric space $(\mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T], \rho^{(1)} \times \rho^{(2)})$.

This paper is organized as follows. In Section 2, we present the main large deviation theorem and give the rate functional of the whole system. By using the contraction principle and the Legendre-Fenchel transform we get the Hamiltonian related to the slow variable $\boldsymbol{z}_n$. As a concrete application, we then study the genetic switching model and compare the difference between the rigorously obtained Hamiltonian and that obtained by WKB ansatz. In Sections 3 and 4, we give the proof of the main theorem. Due to the technicalities to handle the non-negativity constraint for $\boldsymbol{r}$, we decompose the proof procedure into two steps. In Section 3, we prove the LDT theorem by relaxing the bounded domain condition to the whole space case. The upper bound estimate is standard in some sense. However, the proof of the lower bound is technical because of the full coupling between the fast and slow variables. The resolution is based on the approximation and change-of-measure approach. The central idea is to make a piecewise linear approximation to any given path and occupation measure $(\boldsymbol{r}, \nu)$ by $(\boldsymbol{y}, \pi)$ at first, and then construct suitable new processes $(\bar{\boldsymbol{z}}_n, \bar{\nu}_n)$ such that $\mathbb{P} - \lim_{n\to\infty} \rho^{(1)}(\bar{\boldsymbol{z}}_n, \boldsymbol{y}) = 0$ and $\mathbb{P} - \lim_{n\to\infty} \rho^{(2)}(\bar{\nu}_n, \pi) = 0$. This turns out to be technical and one key part of the whole paper. In Section 4, we strengthen the result to the half space case. Some details are left in the Appendix.

This paper should be considered as the companion of [19, 21] for studying the rare events in genetic switching system, and it is of general interest to understand the large deviations for multiscale problems [7, 8].

2. Main result and its application

2.1. **Main theorem.** We need the following technical assumptions for our main result.

Assumption 2.1. *Let* $W := \overline{\mathbb{R}_+^d}$. *Assume the following regularity conditions for the propensity functions and jump rates hold.*

(1) *(a) For each* $i \in \{1, 2, \ldots, S\}$, $j \in \mathbb{Z}_D$ *and all* $\boldsymbol{z}, \boldsymbol{x} \in W$, *the Lipschitz condition holds*

$$|\lambda_i(\boldsymbol{z}, j) - \lambda_i(\boldsymbol{x}, j)| \le L \|\boldsymbol{z} - \boldsymbol{x}\|. \tag{2.1}$$

(b) For each $i \in \{1, 2, \ldots, S\}$, $j \in \mathbb{Z}_D$ *and all* $\boldsymbol{z} \in W^\circ$, $\lambda_i(\boldsymbol{z}, j) > 0$.
(c) For each $\boldsymbol{x} \in \partial W$ *and* $\boldsymbol{y} \in \mathscr{C}\{\boldsymbol{u}_j | \lambda_j(\boldsymbol{x}) > 0\}$, *we have* $\boldsymbol{x} + s\boldsymbol{y} \in W$ *for some* $s \in (0, \infty)$, *where* $\mathscr{C}\{\boldsymbol{u}_j\}$ *is the positive cone spanned by the vectors* $\{\boldsymbol{u}_j\}$ *defined as*

$$\mathscr{C}\{\boldsymbol{u}_j\} := \left\{\boldsymbol{v} \,|\, \text{there exist } \alpha_j \ge 0 \text{ such that } \boldsymbol{v} = \sum_j \alpha_j \boldsymbol{u}_j\right\}. \tag{2.2}$$

(2) *For each* $i, j \in \mathbb{Z}_D$, $\log q_{ij}(\boldsymbol{z})$ *are bounded and Lipschitz continuous with respect to* $\boldsymbol{z} \in W$.

These assumptions hold in our application example in Section 2.2.

Theorem 2.2. *Under the Assumption 2.1, the family* $(\boldsymbol{z}_n, \nu_n)$ *defined by* (1.1) *and* (1.10) *obeys the LDP in* $(\mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T], \rho^{(1)} \times \rho^{(2)})$ *with a good rate functional* $I(\boldsymbol{r}, \nu) = I_s(\boldsymbol{r}, \nu) + I_f(\boldsymbol{r}, \nu)$, *i.e.*

(0) $I(\boldsymbol{r}, \nu)$ *values in* $[0, +\infty]$ *and its level sets are compact in* $(\mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T], \rho^{(1)} \times \rho^{(2)})$,

(1) *for every closed set* $F \in \mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T]$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}((\boldsymbol{z}_n, \nu_n) \in F) \le - \inf_{(\boldsymbol{r}, \nu) \in F} I(\boldsymbol{r}, \nu), \tag{2.3}$$

(2) *for every open set* $G \in \mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T]$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}((\boldsymbol{z}_n, \nu_n) \in G) \ge - \inf_{(\boldsymbol{r}, \nu) \in G} I(\boldsymbol{r}, \nu), \tag{2.4}$$

where the rate functional for the slow variables

$$I_s(\boldsymbol{r}, \nu) = \begin{cases} \int_0^T L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_\nu(t, \cdot)) dt, & d\boldsymbol{r}(t) = \dot{\boldsymbol{r}}(t) dt, \\ \infty, & \text{otherwise,} \end{cases} \tag{2.5}$$

$$L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) = \sup_{\boldsymbol{p} \in \mathbb{R}^d} \left(\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w})\right), \tag{2.6}$$

$$H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) = \sum_{i=1}^S \sum_{j=1}^D \lambda_i(\boldsymbol{z}, j) w_j \left(e^{\langle \boldsymbol{p}, \boldsymbol{u}_i \rangle} - 1\right), \tag{2.7}$$

and the rate functional for the fast variables

$$I_f(\boldsymbol{r}, \nu) = \int_0^T S(\boldsymbol{r}(t), n_\nu(t, \cdot)) dt, \tag{2.8}$$

$$S(\boldsymbol{z}, \boldsymbol{w}) = \sup_{\boldsymbol{\sigma} \in \mathbb{R}^D} S(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma}), \tag{2.9}$$

$$S(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma}) = - \sum_{i,j=1}^{D} w_i q_{ij}(\boldsymbol{z}) \left(e^{\langle \boldsymbol{\sigma}, \boldsymbol{e}_{ij} \rangle} - 1 \right). \tag{2.10}$$

Here we take the notation $\nu(dt, \cdot) = n_\nu(t, \cdot)dt$, thus $n_\nu(t, \cdot)$ is a probabilistic vector $(n_\nu(t, 1), n_\nu(t, 2), \ldots, n_\nu(t, D))$. $\boldsymbol{w} = (w_1, w_2, \ldots, w_D)$, and $\langle \cdot, \cdot \rangle$ is the inner product in the Euclidean space. $\boldsymbol{e}_{ij} = \boldsymbol{e}_i - \boldsymbol{e}_j$ and $\{\boldsymbol{e}_i\}_{i=1}^{D}$ are canonical basis in Euclidean space $\mathbb{R}^D$. We take the convention that $\boldsymbol{r}(t)$ is absolutely continuous with respect to time when we use the notation $d\boldsymbol{r}(t) = \dot{\boldsymbol{r}}(t)dt$, and S is a function of $(\boldsymbol{z}, \boldsymbol{w})$ (or $(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma})$) when we use $S(\boldsymbol{z}, \boldsymbol{w})$ (or $S(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma})$) by default.

The proof of Theorem 2.2 relies on first establishing a weaker statement based on the following stronger assumption on the whole space.

Assumption 2.3. *Regularity for the propensity functions and jump rates.*
   (1) *For each $i \in \{1, 2, \ldots, S\}, j \in \mathbb{Z}_D$, $\log \lambda_i(\boldsymbol{z}, j)$ is bounded and Lipschitz continuous with respect to $\boldsymbol{z} \in \mathbb{R}^d$.*
   (2) *For each $i, j \in \mathbb{Z}_D$, $\log q_{ij}(\boldsymbol{z})$ are bounded and Lipschitz continuous with respect to $\boldsymbol{z} \in \mathbb{R}^d$.*

This covers Assumption 2.1. Mathematically we express (1) as

$$\frac{1}{\Lambda} \leq \lambda_i(\boldsymbol{z}, j) \leq \Lambda, \quad \Lambda > 1 \tag{2.11}$$

for any $\boldsymbol{z} \in \mathbb{R}^d$, $i \in \{1, 2, \ldots, S\}$ and $j \in \mathbb{Z}_D$. And in this stronger set-up we simply denote the positive cone generated by $\{\boldsymbol{u}_j\}$ as

$$\mathscr{C} := \left\{ \boldsymbol{v} | \text{there exist } \alpha_j \geq 0 \text{ such that } \boldsymbol{v} = \sum_j \alpha_j \boldsymbol{u}_j \right\}. \tag{2.12}$$

Theorem 2.4. *The large deviation result in Theorem 2.2 holds for $(\boldsymbol{z}_n, \nu_n) \in \mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T]$ under the Assumption 2.3.*

As a straightforward application of the contraction principle, we have

Corollary 2.5. *The slow variables $\boldsymbol{z}_n$ obeys the LDP in $(\mathbb{D}^d[0, T], \rho^{(1)})$ with the rate functional*

$$I(\boldsymbol{r}) = \inf_{\nu \in \mathbb{M}_L[0, T]} (I_s(\boldsymbol{r}, \nu) + I_f(\boldsymbol{r}, \nu)). \tag{2.13}$$

Define the set of probabilistic transition kernels on $\mathbb{Z}_D$ as $\Delta_D = \{\boldsymbol{w} : w_1, w_2, \cdots, w_D \geq 0, \sum_{i=1}^{D} w_i = 1\}$ where $\boldsymbol{w} = (w_1, w_2, \cdots, w_D)$. We also define the reduced Lagrangian as

$$L(\boldsymbol{z}, \boldsymbol{\beta}) = \inf_{\boldsymbol{w} \in \Delta_D} \left\{ L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w}) \right\}. \tag{2.14}$$

For convenience, we will abuse the notation $n_\nu \in \mathbb{M}_L[0, T]$ and $\nu \in \mathbb{M}_L[0, T]$ in later texts.

Lemma 2.6. *For any $r(\cdot)$ which is absolutely continuous, we have*

$$\inf_{n_v \in \mathbb{M}_L[0,T]} \int_0^T L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t,\cdot)) + S(\boldsymbol{r}(t), n_v(t,\cdot)) dt$$
$$= \int_0^T L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)) dt. \tag{2.15}$$

Proof. First let us show the measurability of the integrand on the right hand side of Eq. (2.15). By Lemma A.2, $L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w})$ is convex in $\boldsymbol{w}$. So $L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w})$ is continuous with respect to $\boldsymbol{w}$ in the set $\Delta_D^\circ \subset \mathbb{R}^D$ and the interior of the low dimensional boundaries of Δ_D. Choosing a countable dense subset $\{\boldsymbol{w}^k\}_{k=1}^\infty$ in Δ_D, we have

$$L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)) = \inf_{k \geq 1} \left\{ L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), \boldsymbol{w}^k) + S(\boldsymbol{r}(t), \boldsymbol{w}^k) \right\} \tag{2.16}$$

for every $\boldsymbol{r}$ by the continuity condition. The measurability is a standard result with this formulation.

It is straightforward to have that

$$\inf_{n_v \in \mathbb{M}_L[0,T]} \int_0^T L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t,\cdot)) + S(\boldsymbol{r}(t), n_v(t,\cdot)) dt$$
$$\geq \int_0^T L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)) dt.$$

Now let us show the converse part. For any given $\epsilon > 0$, define the sets

$$A_k = \left\{ t \in [0, T] : L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)) - \left(L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), \boldsymbol{w}^k) + S(\boldsymbol{r}(t), \boldsymbol{w}^k) \right) \geq -\epsilon/T \right\}$$

for $k \geq 1$. We have that A_k are measurable sets since $L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), \boldsymbol{w}^k) + S(\boldsymbol{r}(t), \boldsymbol{w}^k)$ and $L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t))$ are both measurable functions of t. Define the measurable functions

$$F_k(t) = \begin{cases} k, & t \in A_k, \\ +\infty, & \text{otherwise} \end{cases}$$

for every $k \geq 1$ and

$$J(t) = \inf_{k \geq 1} F_k(t). \tag{2.17}$$

It is not difficult to find that $J(t) < +\infty$ for any t, $J(t)$ is measurable and takes values in positive integers. With these definitions we have

$$L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)) \geq L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), \boldsymbol{w}^{J(t)}) + S(\boldsymbol{r}(t), \boldsymbol{w}^{J(t)}) - \epsilon/T. \tag{2.18}$$

With $\boldsymbol{w}^{J(t)} := \{w_1^{J(t)}, w_2^{J(t)}, \ldots, w_D^{J(t)}\}$, define the occupation measure $\widehat{v}$

$$\widehat{v}(dt, i) = w_i^{J(t)} dt, \quad i \in \{1, 2, \cdots, D\}.$$

Then $\widehat{v} \in \mathbb{M}_L[0, T]$, $n_{\widehat{v}}(t, i) = w_i^{J(t)}$ and

$$\int_0^T L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)) dt$$

$$\geq \int_0^T L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_{\widehat{v}}(t, \cdot)) + S(\boldsymbol{r}(t), n_{\widehat{v}}(t, \cdot)) dt - \epsilon$$

$$\geq \inf_{n_v \in \mathbb{M}_L[0, T]} \int_0^T L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) + S(\boldsymbol{r}(t), n_v(t, \cdot)) dt - \epsilon$$

The proof is completed. $\qquad\square$

By Lemma 2.6, we have

$$I(\boldsymbol{r}) = \inf_{v \in \mathbb{M}_L[0, T]} (I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v))$$

$$= \inf_{n_v \in \mathbb{M}_L[0, T]} \int_0^T L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) + S(\boldsymbol{r}(t), n_v(t, \cdot)) dt$$

$$= \int_0^T L(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t)) dt. \qquad (2.19)$$

Lemma 2.7. $L(\boldsymbol{r}, \boldsymbol{\beta})$ *is convex in* $\boldsymbol{\beta}$.

Proof. By Lemma A.3,

$$L(\boldsymbol{z}, \boldsymbol{\beta}) = \inf_{\boldsymbol{w} \in \Delta_D} \big\{ L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w}) \big\}$$

$$= \inf_{\boldsymbol{w} \in \Delta_D} \sup_{\boldsymbol{p} \in \mathbb{R}^d} (\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w}))$$

$$= \sup_{\boldsymbol{p} \in \mathbb{R}^d} \inf_{\boldsymbol{w} \in \Delta_D} (\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w})).$$

It is easy to see that $\inf_{\boldsymbol{w} \in \Delta_D} (\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w}))$ is linear in $\boldsymbol{\beta}$, thus $L(\boldsymbol{r}, \boldsymbol{\beta})$ is convex in $\boldsymbol{\beta}$ according to Lemma A.2. $\qquad\square$

It is well-known that the Lagrangian L_s does not have a closed form for the standard chemical reaction kinetic system, instead it is more convenient to investigate its dual Hamiltonian H_s by Legendre-Fenchel transform. The explicit form of the Hamiltonian is important for the numerics to study the rare events in systems biology [15]. With similar idea, we

have

$$H(\boldsymbol{z}, \boldsymbol{p}) = \sup_{\boldsymbol{\beta} \in \mathbb{R}^d} \left(\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - L(\boldsymbol{z}, \boldsymbol{\beta}) \right)$$

$$= \sup_{\boldsymbol{\beta} \in \mathbb{R}^d} \left(\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - \inf_{\boldsymbol{w} \in \Delta_D} \left\{ L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) + S(\boldsymbol{z}, \boldsymbol{w}) \right\} \right)$$

$$= \sup_{\boldsymbol{\beta} \in \mathbb{R}^d} \sup_{\boldsymbol{w} \in \Delta_D} \left(\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) - S(\boldsymbol{z}, \boldsymbol{w}) \right)$$

$$= \sup_{\boldsymbol{w} \in \Delta_D} \sup_{\boldsymbol{\beta} \in \mathbb{R}^d} \left(\langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) - S(\boldsymbol{z}, \boldsymbol{w}) \right)$$

$$= \sup_{\boldsymbol{w} \in \Delta_D} \left(H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) - S(\boldsymbol{z}, \boldsymbol{w}) \right). \tag{2.20}$$

A consequence about H from its definition is that H is convex with respect to $\boldsymbol{p}$ from the convexity of L and the Legendre-Fenchel transform [9]. Furthermore if the matrix $Q = (q_{ij})_{D \times D}$ is symmetrizable, $S(\boldsymbol{z}, \boldsymbol{w})$ has an explicit expression [3]

$$S(\boldsymbol{z}, \boldsymbol{w}) = \frac{1}{2} \sum_i \sum_{j \neq i} \left[\sqrt{w_i q_{ij}(\boldsymbol{z})} - \sqrt{w_j q_{ji}(\boldsymbol{z})} \right]^2. \tag{2.21}$$

2.2. **Application to the genetic switching model.** The formula (2.20) has a nice application in the genetic switching model introduced before. In this model, we have $d = 2, D = 2$ and $S = 4$. By parameters shown in (1.4) and Table 1, we have

$$H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) = b^{-1} w_1 (e^{p_1} - 1) + A(z_1, z_2, p_1, p_2), \tag{2.22}$$

where $\boldsymbol{z} = (z_1, z_2), \boldsymbol{p} = (p_1, p_2), \boldsymbol{w} = (w_0, w_1)$ (here we utilize the notation $\boldsymbol{w} = (w_0, w_1)$ instead of $\boldsymbol{w} = (w_1, w_2)$ as mentioned in the introduction since there is only one molecule of DNA) and $A(z_1, z_2, p_1, p_2) = \gamma z_1 (e^{-p_1} - 1) + \gamma b z_1 (e^{p_2} - 1) + z_2 (e^{-p_2} - 1)$. We also have

$$S(\boldsymbol{z}, \boldsymbol{w}) = \left(\sqrt{w_0 f(z_2)} - \sqrt{w_1 g(z_2)} \right)^2.$$

Applying (2.20) with the constraints $w_0 + w_1 = 1$ and $w_0, w_1 \geq 0$, we obtain the final Hamiltonian

$$H(\boldsymbol{z}, \boldsymbol{p}) = b^{-1} s (e^{p_1} - 1) - \left(\sqrt{(1 - s) f(z_2)} - \sqrt{s g(z_2)} \right)^2 + A(z_1, z_2, p_1, p_2), \tag{2.23}$$

where

$$s = \frac{1}{2} + \frac{s_1}{2\sqrt{s_1^2 + 4}}, \quad s_1 = \frac{b^{-1}(e^{p_1} - 1) + f(z_2) - g(z_2)}{\sqrt{f(z_2) g(z_2)}}.$$

It is instructive to compare this Hamiltonian with that obtained via WKB asymptotics. In [1], another form of the Hamiltonian for this system is given via WKB asymptotics:

$$\tilde{H}(\boldsymbol{z}, \boldsymbol{p}) = A + g(z_2)^{-1}[A + b^{-1}(e^{p_1} - 1)][f(z_2) - A], \qquad (2.24)$$

where $A = A(z_1, z_2, p_1, p_2)$. The relation between the Hamiltonian $\tilde{H}$ and H is not clear so far. But one crucial difference is that H is convex with respect to the momentum variable $\boldsymbol{p}$ from the form (2.20), while $\tilde{H}$ is not. It turns out this property is crucial for the numerical computations, especially for computing the transition path in geometric minimum action method (gMAM) [15]. It is also interesting to observe that the quasi-potential $S(z_1, z_2)$ obtained from

$$H(\boldsymbol{z}, \nabla S) = 0 \quad \text{or} \quad \tilde{H}(\boldsymbol{z}, \nabla S) = 0$$

is the same even H and $\tilde{H}$ are so different [21]. It can be also verified that H is not the convex hull of $\tilde{H}$ with respect to $\boldsymbol{p}$. From the Hamilton-Jacobin theory, one may speculate that these two Hamiltonians are connected through some canonical transformation. But it is only a plausible answer which is difficult to be verified even for this concrete example.

As the large deviation results give the sharpest characterization of the considered two-scale chemical kinetic system, we can obtain the deterministic mean field ODEs and the chemical Langevin approximation for the system based on the large deviations [5], which corresponds to the law of large numbers (LLN) and the central limit theorem (CLT) for the process. Taking advantage of (2.23), we get

$$\frac{\partial H}{\partial p_1}\bigg|_{\boldsymbol{p}=\boldsymbol{0}} = \frac{b^{-1}f(z_2)}{f(z_2) + g(z_2)} - \gamma z_1, \qquad \frac{\partial H}{\partial p_2}\bigg|_{\boldsymbol{p}=\boldsymbol{0}} = \gamma b z_1 - z_2. \qquad (2.25)$$

The mean field ODEs defined by

$$\frac{dz_1}{dt} = \frac{\partial H}{\partial p_1}\bigg|_{\boldsymbol{p}=\boldsymbol{0}} \quad \text{and} \quad \frac{dz_2}{dt} = \frac{\partial H}{\partial p_2}\bigg|_{\boldsymbol{p}=\boldsymbol{0}} \qquad (2.26)$$

are exactly (1.6).

Furthermore, we have

$$\frac{\partial^2 H}{\partial p_1^2}\bigg|_{\boldsymbol{p}=\boldsymbol{0}} = \frac{b^{-1}f(z_2)}{f(z_2) + g(z_2)} + \frac{2b^{-2}f(z_2)g(z_2)}{(f(z_2) + g(z_2))^3} + \gamma z_1,$$

$$\frac{\partial^2 H}{\partial p_2^2}\bigg|_{\boldsymbol{p}=\boldsymbol{0}} = \gamma b z_1 + z_2.$$

This naturally leads to the following chemical Langevin approximation

$$\frac{dz_1}{dt} = \left[\frac{b^{-1}f}{f+g} - \gamma z_1\right]dt + \frac{1}{\sqrt{n}}\left[\sqrt{\frac{b^{-1}f}{f+g} + \frac{2b^{-2}fg}{(f+g)^3}}dB_t^1 - \sqrt{\gamma z_1}dB_t^2\right],$$

$$\frac{dz_2}{dt} = \left[\gamma b z_1 - z_2\right]dt + \frac{1}{\sqrt{n}}\left[\sqrt{\gamma b z_1}dB_t^3 - \sqrt{z_2}dB_t^4\right], \tag{2.27}$$

where f, g are abbreviations of functions $f(z_2)$ and $g(z_2)$, and B_t^i ($i = 1,\ldots,4$) are independent standard Brownian motions. It is instructive to compare (2.27) with a granted formulation by naively transplanting the Langevin approximation from the simple large volume limit [14], where the equation for z_1 reads

$$\frac{dz_1}{dt} = \left[\frac{b^{-1}f}{f+g} - \gamma z_1\right]dt + \frac{1}{\sqrt{n}}\left[\sqrt{\frac{b^{-1}f}{f+g}}dB_t^1 - \sqrt{\gamma z_1}dB_t^2\right], \tag{2.28}$$

and the equation for z_2 is the same. It is remarkable that the Eq. (2.27) has an additional term related to the noise dB_t^1. This additional fluctuation is induced by the fast switching of DNA states. Similar situation will also occur when we derive the chemical Langevin equations for enzymatic reactions, whereas we should take the fluctuation effect of the fast switching into consideration if the considered scaling is in our regime. However, this point does not seem to be paid much attention in previous research. Similar situation is further discussed in [19].

2.3. **A useful property of the Hamiltonian** H.

The Hamiltonian $H(\boldsymbol{z}, \boldsymbol{p})$ has some nice properties which can be utilized to simplify the computations in many cases. Assuming that $Q = (q_{ij})_{D\times D}$ is symmetrizable, according to (2.20), we have

$$H(\boldsymbol{z}, \boldsymbol{p}) = \sup_{\boldsymbol{w}\in\Delta_D} h(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}),$$

where

$$h(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) = H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) - S(\boldsymbol{z}, \boldsymbol{w})$$

$$= \sum_{i=1}^{d}\sum_{j=1}^{D}\lambda_i(\boldsymbol{z}, j)w_j(e^{\langle \boldsymbol{p}, \boldsymbol{u}_i\rangle} - 1)$$

$$- \frac{1}{2}\sum_i\sum_{j\neq i}\left[\sqrt{w_i q_{ij}(\boldsymbol{z})} - \sqrt{w_j q_{ji}(\boldsymbol{z})}\right]^2.$$

We will show that the supremum of h in Δ_D can be only taken in the interior Δ_D° of Δ_D. To do this, we first note that h is continuous in Δ_D

and differentiable in Δ_D°. For any $\boldsymbol{w}_b \in \partial(\Delta_D)$, define $\boldsymbol{v} = \boldsymbol{c}_0 - \boldsymbol{w}_b$ where $\boldsymbol{c}_0 = (1, 1, \ldots, 1)/D$ is the center of Δ_D. It is easy to check that

$$\lim_{t \to 0^+} \frac{1}{t} \Big(h(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}_b + t\boldsymbol{v}) - h(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}_b) \Big) = +\infty. \tag{2.29}$$

This means that the supremum of h can not be taken in $\partial(\Delta_D)$. Furthermore, since h is strictly concave in $\boldsymbol{w}$, there exists only one point $\boldsymbol{w}^*$ in Δ_D°, such that

$$\boldsymbol{w}^* = \arg\sup_{\boldsymbol{w} \in \Delta_D} h(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}).$$

An important consequence of this fact is that we can get the derivative

$$\begin{aligned}
\frac{\partial H(\boldsymbol{z}, \boldsymbol{p})}{\partial \boldsymbol{p}} &= \frac{d h(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}^*(\boldsymbol{p}))}{d \boldsymbol{p}} \\
&= \frac{\partial h(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}^*)}{\partial \boldsymbol{p}} + \frac{\partial h}{\partial \boldsymbol{w}}\Big|_{w=w^*} \frac{\partial \boldsymbol{w}^*(\boldsymbol{p})}{\partial \boldsymbol{p}} \\
&= \frac{\partial H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}^*)}{\partial \boldsymbol{p}}.
\end{aligned}$$

This is very useful to simplify the derivations when utilizing the gMAM algorithm [15] to explore the transition paths.

3. PROOF OF THEOREM 2.4

We will mainly follow the framework in [20, 26] to make the proof. First we prove the upper bound and then the lower bound.

3.1. **Upper Bound.** The proof of upper bound (2.3) is standard in some sense. It is difficult to estimate the probability of $(\boldsymbol{z}_n, \boldsymbol{v}_n) \in F$ directly. We proceed with the following steps. Firstly, we approximate $\boldsymbol{z}_n$ by $\tilde{\boldsymbol{z}}_n$, where $\tilde{\boldsymbol{z}}_n$ is an absolutely continuous path. Secondly, for a given compact set, we can get an upper bound for $(\tilde{\boldsymbol{z}}_n, \boldsymbol{v}_n)$. Thirdly, we prove that after excluding a set of exponentially small probability, $\tilde{\boldsymbol{z}}_n$ and $\boldsymbol{v}_n$ stay in compact sets, which means that $\tilde{\boldsymbol{z}}_n$ and $\boldsymbol{v}_n$ are exponentially tight sequence. And finally, we get the desired result by combing the previous steps with further estimates.

Before proceeding to the proof, let us denote $\mathbb{C}^d[0, T]$ the collection of all continuous functions of $t \in [0, T]$ with values in $\mathbb{R}^d$. Define the sup-norm for any $\boldsymbol{r}, \tilde{\boldsymbol{r}} \in \mathbb{C}^d[0, T]$

$$\rho_c^{(1)}(\boldsymbol{r}, \tilde{\boldsymbol{r}}) := \sup_{0 \le t \le T} \|\boldsymbol{r}(t) - \tilde{\boldsymbol{r}}(t)\|.$$

We have that $(\mathbb{C}^d[0, T], \rho_c^{(1)})$ is a Polish space. The metric $\rho_c^{(1)}$ is stronger than $\rho^{(1)}$ on $\mathbb{D}^d[0, T]$. As a consequence, every open set in $(\mathbb{D}^d[0, T], \rho^{(1)})$

is also open in $(\mathbb{D}^d[0,T], \rho_c^{(1)})$. And if $\mathscr{K}$ is compact in $(\mathbb{C}^d[0,T], \rho_c^{(1)})$, it is also compact in $(\mathbb{D}^d[0,T], \rho_c^{(1)})$ and in $(\mathbb{D}^d[0,T], \rho^{(1)})$.

To construct the approximation of $\boldsymbol{z}_n$, we subdivide the time interval $[0,T]$ into n pieces with nodes $t_j^n = Tj/n$, $j = 0, 1, \cdots, n$. Define the piecewise linear interpolation $\tilde{\boldsymbol{z}}_n(t)$ of $\boldsymbol{z}_n(t)$ as

$$\tilde{\boldsymbol{z}}_n(t) = (1 - \gamma_j(t))\boldsymbol{z}_n(t_j^n) + \gamma_j(t)\boldsymbol{z}_n(t_{j+1}^n), \quad t \in [t_j^n, t_{j+1}^n], \qquad (3.1)$$

where $\gamma_j(t) = (t - t_j)n/T \in [0,1]$.

We have the important characterization that $\tilde{\boldsymbol{z}}_n$ is exponentially equivalent to $\boldsymbol{z}_n$.

Lemma 3.1. *For each $\delta > 0$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\rho^{(1)}(\boldsymbol{z}_n, \tilde{\boldsymbol{z}}_n) > \delta) = -\infty. \qquad (3.2)$$

The proof of Lemma 3.1 is left in the Appendix.

For given compact sets in $\mathbb{C}^d[0,T]$, the following quasi-LDP upper bound for $(\tilde{\boldsymbol{z}}_n, \nu_n)$ holds.

Lemma 3.2. *Fix step functions $\boldsymbol{\theta}(t) \in \mathbb{R}^d$ and $\boldsymbol{\alpha}(t) \in \mathbb{R}^D$. For any $\delta > 0$ and compact sets $\mathscr{K} \in \mathbb{C}^d[0,T]$ and $\mathscr{S} \in \mathbb{M}_L[0,T]$, we have*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}((\tilde{\boldsymbol{z}}_n, \nu_n) \in \mathscr{K} \times \mathscr{S}) \le - \inf_{(\boldsymbol{r}, \nu) \in \mathscr{K} \times \mathscr{S}} \left(I_s^\delta(\boldsymbol{r}, \nu, \boldsymbol{\theta}) + I_f^\delta(\boldsymbol{r}, \nu, \boldsymbol{\alpha}) \right), \tag{3.3}$$

where

$$I_s^\delta(\boldsymbol{r}, \nu, \boldsymbol{\theta}) = \begin{cases} \int_0^T L_s^\delta(\boldsymbol{r}(t), \dot{\boldsymbol{r}}, n_\nu(t, \cdot), \boldsymbol{\theta}(t)) dt, & d\boldsymbol{r}(t) = \dot{\boldsymbol{r}}(t) dt, \\ \infty, & \text{otherwise,} \end{cases} \tag{3.4}$$

$$L_s^\delta(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p}) = \langle \boldsymbol{\beta}, \boldsymbol{p} \rangle - H_s^\delta(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}), \tag{3.5}$$

$$H_s^\delta(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w}) = \sup_{|\boldsymbol{x} - \boldsymbol{z}| < \delta} H_s(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{w}), \tag{3.6}$$

and

$$I_f^\delta(\boldsymbol{r}, \nu, \boldsymbol{\alpha}) = \int_0^T S^\delta(\boldsymbol{r}(t), n_\nu(t, \cdot), \boldsymbol{\alpha}(t)) dt, \tag{3.7}$$

$$S^\delta(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma}) = - \sup_{|\boldsymbol{x} - \boldsymbol{z}| < \delta} \sum_{i,j=1}^D w_i q_{ij}(\boldsymbol{z}) \left(e^{\langle \boldsymbol{\sigma}, \boldsymbol{e}_{ij} \rangle} - 1 \right). \tag{3.8}$$

Before the proof we remark that $H_s^\delta(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w})$ and $L_s^\delta(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p})$ are monotonically increasing and decreasing functions of δ, respectively. $S^\delta(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma})$ is a monotonically decreasing function of δ. Correspondingly, $I_s^\delta(\boldsymbol{r}, \nu, \boldsymbol{\theta})$ and $I_f^\delta(\boldsymbol{r}, \nu, \boldsymbol{\alpha})$ are decreasing functionals of δ.

Proof. We only need to consider absolutely continuous functions $\boldsymbol{r}$ on the right hand side of Eq. (3.3) since $I_s^\delta(\boldsymbol{r},v,\boldsymbol{\theta}) + I_f^\delta(\boldsymbol{r},v,\boldsymbol{\alpha}) = \infty$ otherwise. For any $\boldsymbol{r}$ and v, define the sum

$$
\begin{aligned}
J_n(\boldsymbol{r},\boldsymbol{\theta},v,\boldsymbol{\alpha}) = \sum_{j=0}^{n-1} & \left(\left\langle \boldsymbol{r}(t_{j+1}^n) - \boldsymbol{r}(t_j^n), \boldsymbol{\theta}(t_j^n) \right\rangle \right. \\
& \left. - \int_{t_j^n}^{t_{j+1}^n} H_s^\delta \left(\boldsymbol{r}(t_j^n), \boldsymbol{\theta}(t_j^n), n_v(t,\cdot) \right) dt + \int_{t_j^n}^{t_{j+1}^n} S^\delta(\boldsymbol{r}(t_j^n), n_v(t,\cdot), \boldsymbol{\alpha}(t_j^n)) dt \right).
\end{aligned}
$$
(3.9)

By Corollary A.8 in Appendix, we have

$$
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp \left(n J_n(\tilde{\boldsymbol{z}}_n, \boldsymbol{\theta}, v_n, \boldsymbol{\alpha}) \right) \le 0.
$$
(3.10)

For $(\tilde{\boldsymbol{z}}_n, v_n) \in \mathcal{K} \times \mathcal{S}$, it is obvious that

$$
J_n(\tilde{\boldsymbol{z}}_n, \boldsymbol{\theta}, v_n, \boldsymbol{\alpha}) - \inf_{(\boldsymbol{r},v)\in\mathcal{K}\times\mathcal{S}} J_n(\boldsymbol{r},\boldsymbol{\theta},v,\boldsymbol{\alpha}) \ge 0.
$$
(3.11)

So we have

$$
\exp \left(J_n(\tilde{\boldsymbol{z}}_n, \boldsymbol{\theta}, v_n, \boldsymbol{\alpha}) - \inf_{(\boldsymbol{r},v)\in\mathcal{K}\times\mathcal{S}} J_n(\boldsymbol{r},\boldsymbol{\theta},v,\boldsymbol{\alpha}) \right) \ge 1
$$

and

$$
\begin{aligned}
& \mathbb{P}((\tilde{\boldsymbol{z}}_n, v_n) \in \mathcal{K} \times \mathcal{S}) \\
\le \ & \mathbb{E} \exp \left\{ n \left[J_n(\tilde{\boldsymbol{z}}_n, \boldsymbol{\theta}, v_n, \boldsymbol{\alpha}) - \inf_{(\boldsymbol{r},v)\in\mathcal{K}\times\mathcal{S}} J_n(\boldsymbol{r},\boldsymbol{\theta},v,\boldsymbol{\alpha}) \right] \right\}.
\end{aligned}
$$

Combining this with (3.10), we get

$$
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}((\tilde{\boldsymbol{z}}_n, v_n) \in \mathcal{K} \times \mathcal{S}) \le -\liminf_{n\to\infty} \left(\inf_{(\boldsymbol{r},v)\in\mathcal{K}\times\mathcal{S}} J_n(\boldsymbol{r},\boldsymbol{\theta},v,\boldsymbol{\alpha}) \right).
$$
(3.12)

We now represent the sum on the right hand side of (3.12) as an integral. Since $\mathcal{K}$ is compact, the absolutely continuous functions $\boldsymbol{r} \in \mathcal{K}$ are thus uniformly bounded. Let V be a compact set in $\mathbb{R}^d$ such that

$$
\left\{ \boldsymbol{z} : \boldsymbol{z} = \boldsymbol{r}(t) \text{ for some } \boldsymbol{r} \in \mathcal{K} \text{ and } t \in [0, T] \right\} \subset V.
$$

For step function $\boldsymbol{\theta}$, let us investigate an interval in which $\boldsymbol{\theta}$ takes constant value $\boldsymbol{\theta}_0$, say, the interval $[0, \tau]$ without loss of generality. Then

$$
\sum_{j=0}^{n-1} \chi_{\{t_{j+1}^n \le \tau\}} \left\langle \boldsymbol{r}(t_{j+1}^n) - \boldsymbol{r}(t_j^n), \boldsymbol{\theta}_0 \right\rangle = \int_0^\tau \langle \dot{\boldsymbol{r}}, \boldsymbol{\theta}_0 \rangle dt + \epsilon_n,
$$

where the error ϵ_n takes into account the fact that τ may not match any of t_j^n. It goes to zero uniformly for $\boldsymbol{r} \in \mathcal{K}$ when n goes to infinity from the

bound

$$|\epsilon_n| \le \frac{2T}{n}|\boldsymbol{\theta}_0|\sup_{z\in V}|\boldsymbol{z}|.$$

Now $H_s^\delta(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w})$ is continuous in $\boldsymbol{z}$, $\boldsymbol{p}$ and $\boldsymbol{w}$ from the continuity of H_s on $\boldsymbol{x}, \boldsymbol{p}$ and $\boldsymbol{w}$ and the boundedness of λ_i, $\boldsymbol{\theta}$ and $n_v(t, \cdot)$ in the current setting. So we have

$$\left| H_s^\delta\left(\boldsymbol{r}(t_j^n), \boldsymbol{\theta}(t_j^n), n_v(t, \cdot)\right) - H_s^\delta\left(\boldsymbol{r}(t), \boldsymbol{\theta}(t_j^n), n_v(t, \cdot)\right)\right|, \ t_j^n \le t \le t_{j+1}^n$$

goes to zero uniformly in j for $\boldsymbol{r} \in \mathscr{K}$ and $v \in \mathscr{S}$ by equi-continuity. Therefore,

$$\sum_{j=0}^{n-1} \chi_{\{t_{j+1}^n \le \tau\}} \int_{t_j^n}^{t_{j+1}^n} H_s^\delta\left(\boldsymbol{r}(t_j^n), \boldsymbol{\theta}(t_j^n), n_v(t, \cdot)\right) dt$$

$$= \int_0^\tau H_s^\delta\left(\boldsymbol{r}(t), \boldsymbol{\theta}(t_j^n), n_v(t, \cdot)\right) dt + \epsilon_n,$$

with ϵ_n converging to zero uniformly in $(\boldsymbol{r}, v) \in \mathscr{K} \times \mathscr{S}$.

Similarly we can estimate for the part $S^\delta(\boldsymbol{r}(t_j^n), n_v(t, \cdot), \boldsymbol{\alpha}(t_j^n))$ and repeat the argument on the finite number of intervals on which $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are constants. Thanks to the uniformity in $(\boldsymbol{r}, v) \in \mathscr{K} \times \mathscr{S}$, we obtain

$$\liminf_{n\to\infty}\left(\inf_{(\boldsymbol{r},v)\in\mathscr{K}\times\mathscr{S}} J_n(\boldsymbol{r}, \boldsymbol{\theta}, v, \boldsymbol{\alpha})\right) = \inf_{(\boldsymbol{r},v)\in\mathscr{K}\times\mathscr{S}}(I_s^\delta(\boldsymbol{r}, v, \boldsymbol{\theta}) + I_f^\delta(\boldsymbol{r}, v, \boldsymbol{\alpha})).$$

Together with (3.12), the proof is completed. $\qquad\qquad\square$

Next we show the exponential tightness of the sequence $(\tilde{\boldsymbol{z}}_n, v_n)$. Define the modulus of continuity of a continuous function $\boldsymbol{z}$ as

$$V_\delta(\boldsymbol{z}) = \sup\left\{\|\boldsymbol{z}(t) - \boldsymbol{z}(s)\| : 0 \le s \le t \le T, |t - s| < \delta\right\} \qquad (3.13)$$

and the set

$$\mathscr{K}(M) = \bigcap_{m=M}^{\infty}\left\{\boldsymbol{z} \in \mathbb{C}^d[0, T] : \boldsymbol{z}(0) = \boldsymbol{z}^0, V_{2^{-m}}(\boldsymbol{z}) \le \frac{1}{\log m}\right\} \qquad (3.14)$$

for any fixed $M \in \mathbb{N}$.

Lemma 3.3 (Exponential tightness for $\tilde{\boldsymbol{z}}_n$). *For each $B > 0$, there is a compact set $\mathscr{K} \subset \mathbb{C}^d[0, T]$ such that*

$$\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}(\tilde{\boldsymbol{z}}_n \notin \mathscr{K}) \le -B.$$

Proof. For any fixed $M \in \mathbb{N}$, it is not difficult to see that the set $\mathscr{K}(M)$ is closed and the functions in $\mathscr{K}(M)$ are equicontinuous. Thus $\mathscr{K}(M)$ is compact by the Arzela-Ascoli theorem. If $2^{-m} < T/n$, we have

$$V_{2^{-m-1}}(\tilde{\boldsymbol{z}}_n) = \frac{1}{2}V_{2^{-m}}(\tilde{\boldsymbol{z}}_n)$$

since $\tilde{z}_n$ is piecewise linear. Therefore, to check whether $\tilde{z}_n$ is in $\mathcal{K}(M)$, we only need to consider a finite intersection, for values of m up to

$$M(n) = \max\left\{M, \left\lceil \frac{\log(n/T)}{\log 2} \right\rceil\right\}.$$

Using Corollary A.6 in Appendix, we have for any n with $M(n) > M$,

$$
\begin{aligned}
\mathbb{P}(\tilde{z}_n \notin \mathcal{K}(M)) &\leq \sum_{m=M}^{M(n)} \mathbb{P}\left(V_{2^{-m}}(\tilde{z}_n) > \frac{1}{\log m}\right) \\
&\leq \sum_{m=M}^{M(n)} \sum_{j=0}^{n-1} \mathbb{P}\left(\sup_{0 \leq t \leq 2^{-m}} |z_n(t_j^n + t) - z_n(t_j^n)| > \frac{1}{\log M}\right) \\
&\leq nM(n) \cdot 2d \exp\left(-n\frac{c_1}{\log M} \log\left(\frac{2^M c_2}{\log M}\right)\right)
\end{aligned}
$$

for positive constants c_1 and c_2. Thus

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\tilde{z}_n \notin \mathcal{K}(M)) \leq -c\frac{M}{\log M}$$

for some positive constant c when $M \gg 1$. $\qquad\square$

Lemma 3.4. *The measure space* $\mathbb{M}_L[0,T]$ *is compact.*

Proof. Since $[0,T] \times \{1,2,\ldots,D\}$ is compact, $\mathbb{M}_L[0,T]$ is tight. By Prohorov's theorem, $\mathbb{M}_L[0,T]$ is relatively compact. Let ν be the limit of any converging sequence $\{\nu_m\}$ in $\mathbb{M}_L[0,T]$. Since $\sum_{i=1}^{D} \nu_m(dt,i) = dt$ for all m, we have $\sum_{i=1}^{D} \nu(dt,i) = dt$ and thus $\nu(dt,i) \ll dt$. So ν also belongs to $\mathbb{M}_L[0,T]$. This proves that $\mathbb{M}_L[0,T]$ is compact. $\qquad\square$

The straightforward consequence of Lemma 3.4 is that ν_n is also exponentially tight.

Define the quasi-rate functionals for slow and fast variables corresponding to I_s and I_f in Theorem 2.2

$$I_s^\delta(\boldsymbol{r}, \nu) = \begin{cases} \int_0^T L_s^\delta(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_\nu(t,\cdot)) dt, & d\boldsymbol{r}(t) = \dot{\boldsymbol{r}}(t) dt, \\ \infty, & \text{otherwise,} \end{cases} \tag{3.15}$$

$$L_s^\delta(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) = \sup_{\boldsymbol{p} \in \mathbb{R}^d} L_s^\delta(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p}), \tag{3.16}$$

and

$$I_f^\delta(\boldsymbol{r}, \nu) = \int_0^T S^\delta(\boldsymbol{r}(t), n_\nu(t,\cdot)) dt, \tag{3.17}$$

$$S^\delta(\boldsymbol{z}, \boldsymbol{w}) = \sup_{\boldsymbol{\sigma} \in \mathbb{R}^D} S^\delta(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma}). \tag{3.18}$$

The definitions of $L_s^\delta(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p})$ and $S^\delta(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma})$ are referred to (3.5) and (3.8). We have the following approximation lemmas.

Lemma 3.5. *For any $\epsilon > 0$, the absolutely continuous function $\boldsymbol{r} \in \mathbb{C}^d[0,T]$ and $v \in \mathbb{M}_L[0,T]$, there exists neighborhood $N_{\boldsymbol{r},v} \in \mathbb{C}^d[0,T] \times \mathbb{M}_L[0,T]$ of $(\boldsymbol{r},v)$, step functions $\boldsymbol{\theta}_{\boldsymbol{r}v} \subset \mathbb{R}^d$ and $\boldsymbol{\alpha}_{\boldsymbol{r}v} \subset \mathbb{R}^D$, such that for any $\boldsymbol{q} \in N_{\boldsymbol{r}}$ and $\mu \in N_v$ which are both absolutely continuous, we have*

$$I_s^\delta(\boldsymbol{q},\mu,\boldsymbol{\theta}_{\boldsymbol{r}v}) + I_f^\delta(\boldsymbol{q},\mu,\boldsymbol{\alpha}_{\boldsymbol{r}v}) \geq I_s^\delta(\boldsymbol{r},v) + I_f^\delta(\boldsymbol{r},v) - \epsilon.$$

Lemma 3.6. *For any pair $(\boldsymbol{r},v) \in \mathbb{C}^d[0,T] \times \mathbb{M}_L[0,T]$ and $M_0 > 0$, if $\boldsymbol{r}$ is not absolutely continuous, there exists neighborhood $N_{\boldsymbol{r},v} \in \mathbb{C}^d[0,T] \times \mathbb{M}_L[0,T]$ of $(\boldsymbol{r},v)$ and step functions $\boldsymbol{\theta}_{\boldsymbol{r}v} \in \mathbb{R}^d$ and $\boldsymbol{\alpha}_{\boldsymbol{r}v} \in \mathbb{R}^D$, such that for any $(\boldsymbol{q},\mu) \in N_{\boldsymbol{r},v}$, we have*

$$I_s^\delta(\boldsymbol{q},\mu,\boldsymbol{\theta}_{\boldsymbol{r}v}) + I_f^\delta(\boldsymbol{q},\mu,\boldsymbol{\alpha}_{\boldsymbol{r}v}) \geq M_0.$$

The Lemmas 3.5 and 3.6 are direct consequences of Lemmas A.10 and A.11 in the appendix.

Simply denote the product metric $\rho^{(1)} \times \rho^{(2)}$ on $\mathbb{D}^d[0,T] \times \mathbb{M}_L[0,T]$ as $d(\cdot,\cdot)$ and define the sets

$$\Phi(K) = \left\{ (\boldsymbol{r},v) \in \mathbb{D}^d[0,T] \times \mathbb{M}_L[0,T] : I_s(\boldsymbol{r},v) + I_f(\boldsymbol{r},v) \leq K \right\} \qquad (3.19)$$

and

$$\Phi^\delta(K) = \left\{ (\boldsymbol{r},v) \in \mathbb{D}^d[0,T] \times \mathbb{M}_L[0,T] : I_s^\delta(\boldsymbol{r},v) + I_f^\delta(\boldsymbol{r},v) \leq K \right\}. \qquad (3.20)$$

We have the following characterization for $\Phi(K)$ and $\Phi^\delta(K)$.

Lemma 3.7. *For any $K > 0$, the level sets $\Phi(K)$ and $\Phi^\delta(K)$ defined in (3.19) and (3.20) are compact sets.*

Proof. By Lemma 3.4, $\mathbb{M}_L[0,T]$ is a compact set. By Lemma A.9, the functions $\boldsymbol{r} \in \Phi(K)$ are equicontinuous. Combining with the fact that $\boldsymbol{r}(0) = \boldsymbol{z}^0$, we have that $\Phi(K)$ is pre-compact. By Lemma A.10, $I_s(\boldsymbol{r},v) + I_f(\boldsymbol{r},v)$ is lower semicontinuous. Consequently, $\Phi(K)$ is closed and thus compact. The proof for $\Phi^\delta(K)$ is similar. $\qquad\square$

Proposition 3.8. *For each $K > 0$, $\delta > 0$ and $\epsilon > 0$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(d\left((\tilde{\boldsymbol{z}}_n, v_n), \Phi^\delta(K) \right) > \epsilon \right) \leq -(K - \epsilon).$$

Proof. From the exponential tightness, we can find a compact set $\mathscr{K}^N \in \mathbb{C}^d[0,T]$ for each $N > 0$ such that

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\tilde{\boldsymbol{z}}_n \notin \mathscr{K}^N) \leq -N.$$

Define the set

$$\mathscr{K}^{N,\epsilon} = \overline{\left\{ (\boldsymbol{r},v) \in \mathbb{C}^d[0,T] \times \mathbb{M}_L[0,T] : d\left((\boldsymbol{r},v), \Phi^\delta(K) \right) > \epsilon \right\}} \cap (\mathscr{K}^N \times \mathbb{M}_L[0,T]).$$

For any $(\boldsymbol{r}, v) \in \mathcal{K}^{N,\epsilon}$, we can find the neighborhood $N_{\boldsymbol{r},v}$ either satisfying Lemma 3.5 if $\boldsymbol{r}$ is absolutely continuous, or satisfying Lemma 3.6 if $\boldsymbol{r}$ is not absolutely continuous. This forms a covering of $\mathcal{K}^{N,\epsilon}$. By compactness, we can choose a finite subcover $\{N_{\boldsymbol{r}_i, v_j}\}_{i,j}$ for $\mathcal{K}^{N,\epsilon}$. Define

$$\mathcal{K}_{ij} = \overline{N_{\boldsymbol{r}_i, v_j} \cap \mathcal{K}^{N,\epsilon}}.$$

Applying Lemma 3.2, Lemma 3.5, Lemma 3.6 and letting M_0 in Lemma 3.6 larger than K, we have for any i, j,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}((\tilde{\boldsymbol{z}}_n, v_n) \in \mathcal{K}_{ij}) \le -(K - \epsilon).$$

Then we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(d\left((\tilde{\boldsymbol{z}}_n, v_n), \Phi^\delta(K) \right) > \epsilon \right)$$

$$\le \limsup_{n \to \infty} \frac{1}{n} \log \left[\mathbb{P}(\tilde{\boldsymbol{z}}_n \notin \mathcal{K}^N) + \sum_{i,j} \mathbb{P}((\tilde{\boldsymbol{z}}_n, v_n) \in \mathcal{K}_{ij}) \right]$$

$$\le -\min\{N, K - \epsilon\}.$$

Choosing N large enough, we complete the proof. $\qquad\square$

We are now ready to establish the upper bound.

Lemma 3.9. *Given $K > 0$ and $\epsilon > 0$, there exist $\delta > 0$ such that*

$$\Phi^\delta(K - \epsilon) \subset \{(\boldsymbol{r}, v) : d((\boldsymbol{r}, v), \Phi(K)) \le \epsilon\}.$$

Proof. Prove by contradiction. If the claim is false, we can choose

$$\delta_i \downarrow 0, \qquad (\boldsymbol{r}_i, v_i) \in \Phi^{\delta_i}(K - \epsilon), \quad i = 1, 2, \cdots$$

such that

$$d((\boldsymbol{r}_i, v_i), \Phi(K)) \ge \epsilon, \quad \forall i. \tag{3.21}$$

By definition of $I_s^\delta(\boldsymbol{r}, v)$ and $I_f^\delta(\boldsymbol{r}, v)$, we have the monotonicity $I_s^\delta(\boldsymbol{r}, v) \le I_s^{\delta'}(\boldsymbol{r}, v)$ and $I_f^\delta(\boldsymbol{r}, v) \le I_f^{\delta'}(\boldsymbol{r}, v)$ when $\delta \ge \delta' \ge 0$. Thus the sets $\Phi^{\delta_i}(K - \epsilon)$ are monotonically decreasing as $\delta_i \downarrow 0$, and $(\boldsymbol{r}_i, v_i)$ are contained in the set $\Phi^{\delta_1}(K - \epsilon)$ which is compact by Lemma 3.7. So there exists a subsequence converging to $(\boldsymbol{r}_0, v_0)$. With Lemma A.10 in the Appendix we have for each j

$$
\begin{aligned}
I_s^{\delta_j}(\boldsymbol{r}_0, v_0) + I_f^{\delta_j}(\boldsymbol{r}_0, v_0) &\le \liminf_{i \to \infty}\left(I_s^{\delta_j}(\boldsymbol{r}_i, v_i) + I_f^{\delta_j}(\boldsymbol{r}_i, v_i) \right) \\
&\le \liminf_{i \to \infty}\left(I_s^{\delta_i}(\boldsymbol{r}_i, v_i) + I_f^{\delta_i}(\boldsymbol{r}_i, v_i) \right) \\
&\le K - \epsilon.
\end{aligned}
$$

The monotone convergence theorem gives

$$
\begin{aligned}
I_s(\boldsymbol{r}_0, \nu_0) + I_f(\boldsymbol{r}_0, \nu_0) &= \lim_{j \to \infty} I_s^{\delta_j}(\boldsymbol{r}_0, \nu_0) + I_f^{\delta_j}(\boldsymbol{r}_0, \nu_0) \\
&\leq K - \epsilon.
\end{aligned}
$$

So $(\boldsymbol{r}_0, \nu_0) \in \Phi(K)$. For sufficiently large i, $d((\boldsymbol{r}_0, \nu_0), (\boldsymbol{r}_i, \nu_i)) \leq \epsilon$. This contradicts with (3.21). $\square$

Theorem 3.10. *For each closed set $F \subset \mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T]$,*

$$
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}((\boldsymbol{z}_n, \nu_n) \in F) \leq - \inf_{(\boldsymbol{r}, \nu) \in F} \left(I_s(\boldsymbol{r}, \nu) + I_f(\boldsymbol{r}, \nu) \right).
$$

Proof. Suppose $\inf_{(\boldsymbol{r}, \nu) \in F} \left(I_s(\boldsymbol{r}, \nu) + I_f(\boldsymbol{r}, \nu) \right) = K < \infty$. Since F and $\Phi(K-\epsilon)$ are both closed sets, we assume the distance between them is $\eta_0 > 0$. For any $\eta \leq \eta_0$,

$$
\begin{aligned}
&\mathbb{P}((\boldsymbol{z}_n, \nu_n) \in F) \\
\leq\ & \mathbb{P}\left(d((\tilde{\boldsymbol{z}}_n, \nu_n), F) \leq \frac{\eta}{2} \right) + \mathbb{P}\left(d((\tilde{\boldsymbol{z}}_n, \nu_n), (\boldsymbol{z}_n, \nu_n)) \geq \frac{\eta}{2} \right) \\
\leq\ & \mathbb{P}\left(d((\tilde{\boldsymbol{z}}_n, \nu_n), \Phi(K-\epsilon)) \geq \frac{\eta}{2} \right) + \mathbb{P}\left(\rho^{(1)}((\tilde{\boldsymbol{z}}_n, \boldsymbol{z}_n) \geq \frac{\eta}{2} \right) \quad (3.22)
\end{aligned}
$$

By Lemma 3.9, we can choose δ and η small enough such that

$$
d((\tilde{\boldsymbol{z}}_n, \nu_n), \Phi(K-\epsilon)) \geq \frac{\eta}{2} \quad \text{implies} \quad d\left((\tilde{\boldsymbol{z}}_n, \nu_n), \Phi^{\delta}(K - \epsilon - \eta/4) \right) \geq \frac{\eta}{4}.
$$

From Proposition 3.8 we have

$$
\begin{aligned}
&\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(d((\tilde{\boldsymbol{z}}_n, \nu_n), \Phi(K-\epsilon)) \geq \frac{\eta}{2} \right) \\
\leq\ & \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(d\left((\tilde{\boldsymbol{z}}_n, \nu_n), \Phi^{\delta}(K - \epsilon - \eta/4) \right) \geq \frac{\eta}{4} \right) \\
\leq\ & -(K - \epsilon - \eta/2). \quad (3.23)
\end{aligned}
$$

Combining (3.22), (3.23) and Lemma 3.1 for $\delta = \eta/4$, we obtain

$$
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}((\boldsymbol{z}_n, \nu_n) \in F) \leq -(K - \epsilon - \eta/2).
$$

The case for

$$
\inf_{(\boldsymbol{r}, \nu) \in F} \left(I_s(\boldsymbol{r}, \nu) + I_f(\boldsymbol{r}, \nu) \right) = \infty
$$

can be established similarly by choosing K arbitrarily large. $\square$

3.2. **Lower bound.** The proof of the lower bound is based on the change of measure formula. From [5], it suffices to prove that for any $(\boldsymbol{r}, v) \in \mathbb{D}^d[0, T] \times \mathbb{M}_L[0, T]$ and arbitrarily small $\epsilon > 0$ we have

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left(\boldsymbol{z}_n \in N_\epsilon(\boldsymbol{r}), v_n \in N_\epsilon(v)\right) \geq -\left(I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v)\right), \quad (3.24)$$

where $N_\epsilon(\boldsymbol{r})$ is the ϵ-neighborhood of $\boldsymbol{r}$ in $\mathbb{D}^d[0, T]$ with metric $\rho^{(1)}$, and $N_\epsilon(v)$ is the ϵ-neighborhood of v in $\mathbb{M}_L[0, T]$ with metric $\rho^{(2)}$. For given $\boldsymbol{r} \in \mathbb{D}^d[0, T]$ and $v \in \mathbb{M}_L[0, T]$, if $\boldsymbol{r}$ is not absolutely continuous, $I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v) = \infty$, thus nothing needs to be proved. Below we will exclude this case. For convenience, we further assume that $n_v(t, i)$ is continuous in t, and the case that $n_v(t, i)$ is not continuous will be discussed in Theorem 3.17 in this section. To prove the lower bound, we perform the following steps. Firstly, we approximate $\boldsymbol{r}$ by a piecewise linear path $\boldsymbol{y}$, and the occupation measure v by $\pi \in \mathbb{M}_L[0, T]$ with $n_\pi(t, \cdot)$ piecewise constant in t. Secondly, we construct new processes $\bar{\boldsymbol{z}}_n$ and $\bar{\xi}_n$ with occupation measure $\bar{v}_n$ such that

$$\mathbb{P} - \lim_{n \to \infty} \rho_c^{(1)}(\bar{\boldsymbol{z}}_n, \boldsymbol{y}) = 0, \quad \mathbb{P} - \lim_{n \to \infty} \rho^{(2)}(\bar{v}_n, \pi) = 0, \quad (3.25)$$

where the notation $\mathbb{P} - \lim$ means the convergence in probability. Moreover, we ask $\bar{\boldsymbol{z}}_n$ and the jump rates of $\bar{\xi}_n$ satisfy the conditions required by Lemmas 3.11 and 3.12. Finally, based on the change of measure formula related to $(\boldsymbol{z}_n, \xi_n)$ and $(\bar{\boldsymbol{z}}_n, \bar{\xi}_n)$, we get the limit and the proof is then finished.

As promised in the above procedure, we approximate $\boldsymbol{r}$ by a path $\boldsymbol{y}$ first. For a given J, define $\Delta = T/J$ and let $t_m = m\Delta$. On each interval $[t_m, t_{m+1}]$, define $\Delta \boldsymbol{r}_m = \boldsymbol{r}(t_{m+1}) - \boldsymbol{r}(t_m)$. Take $\boldsymbol{\mu}^m = \{\mu_i^m, i = 1, \ldots, S\}$ so as to satisfy

$$\sum_{i=1}^{S} \mu_i^m \boldsymbol{u}_i = \frac{\Delta \boldsymbol{r}_m}{\Delta} \quad \text{and} \quad \mu_i^m \geq 0. \quad (3.26)$$

If $\Delta \boldsymbol{r}_m$ are in the positive cone generated by the $\{\boldsymbol{u}_i\}$ for all m, such a choice of $\boldsymbol{\mu}^m$ is possible. If at least one of $\Delta \boldsymbol{r}_m$ is not in the positive cone generated by the $\{\boldsymbol{u}_i\}$, it is easy to check that for all $v \in \mathbb{M}[0, T]$, $I_s(\boldsymbol{r}, v) = +\infty$ (see the Remark of Lemma 5.21 in [26]) and nothing needs to be proved.

Now we construct the piecewise linear interpolation $\boldsymbol{y}$ of $\boldsymbol{r}$ such that $\boldsymbol{y}(t_0) = \boldsymbol{r}(t_0)$ and in each time interval $[t_m, t_{m+1}]$

$$\frac{d}{dt} \boldsymbol{y}(t) = \sum_{i=1}^{S} \mu_i^m \boldsymbol{u}_i. \quad (3.27)$$

Thus $\boldsymbol{y}(t_m) = \boldsymbol{r}(t_m)$ for each m. For any $\epsilon > 0$, we can choose J large enough such that

$$\rho_c^{(1)}(\boldsymbol{y}, \boldsymbol{r}) < \epsilon/4.$$

Define the sets

$$\mathscr{S} = \left\{ (\eta, \psi) \big| \eta = (\eta_{ij})_{D \times D}, \, \eta_{ij} > 0; \, \psi \in \Delta_D; \sum_{i=1}^{D} \psi_i \sum_{j=1}^{D} \eta_{ij} \boldsymbol{e}_{ij} = \boldsymbol{0} \right\} \quad (3.28)$$

and

$$\mathscr{K}_{\boldsymbol{\beta}} = \left\{ \boldsymbol{\mu} \in \mathbb{R}^S : \mu_i \geq 0, \sum_{i=1}^{S} \mu_i \boldsymbol{u}_i = \boldsymbol{\beta} \right\}. \quad (3.29)$$

We remark that the sets $\mathscr{S}$ and $\mathscr{K}_{\boldsymbol{\beta}}$ here have nothing to do with the definitions in the proof of upper bound.

Lemma 3.11. *For any $\epsilon > 0$ and large enough J, there exists a further subdivision of time interval $[t_m, t_{m+1}]$ for each $m \in \{0, 1, \cdots, J-1\}$ (i.e., $t_m = t_{m0} < t_{m1} < \cdots < t_{mK_m} = t_{m+1}$) and related $(\eta^{mk}, \psi^{mk}) \in \mathscr{S}$ ($m = 0, 1, \cdots, J-1; k = 0, 1, \cdots, K_m - 1$), such that*

$$\sum_{m=0}^{J-1} \sum_{k=0}^{K_m-1} \int_{t_{mk}}^{t_{m,k+1}} \sum_{i=1}^{D} \psi_i^{mk} \sum_{j=1}^{D} \left(\eta_{ij}^{mk} \log \frac{\eta_{ij}^{mk}}{q_{ij}(\boldsymbol{y}(t))} + q_{ij}(\boldsymbol{y}(t)) - \eta_{ij}^{mk} \right) dt$$
$$\leq \quad I_f(\boldsymbol{r}, \nu) + \epsilon.$$

and

$$\|\psi^{mk} - n_\nu(t, \cdot)\| < \epsilon/(4DT) \quad (3.30)$$

for all $t \in [t_{mk}, t_{m,k+1})$, $k = 0, 1, \cdots, K_m - 1$ and $m = 0, 1, \cdots, J-1$.

The proof of Lemma 3.11 can be found in the Appendix.

We then define the measure $\pi \in \mathbb{M}_L[0, T]$ such that $\pi(dt, i) = n_\pi(t, i) dt$ and

$$n_\pi(t, i) := \psi_i^{mk}, \quad t \in [t_{mk}, t_{m,k+1})$$

for $m = 0, 1, \cdots, J-1$ and $k = 0, 1, \cdots, K_m - 1$. With this choice $n_\pi(t, \cdot)$ is piecewise constant and

$$\rho^{(2)}(\pi, \nu) < \epsilon/4.$$

We take the frequently used notation λ_i^π in later text as the expectation of λ_i with respect to the distribution n_π

$$\lambda_i^\pi(\boldsymbol{y}(s)) = \sum_{j=1}^{D} \lambda_i(\boldsymbol{y}(s), j) n_\pi(s, j). \quad (3.31)$$

Lemma 3.12. *For any $\epsilon > 0$ and large enough J, define $\boldsymbol{\beta}_m = \Delta r_m / \Delta$, then there exists $\boldsymbol{\mu}^m \in K_{\boldsymbol{\beta}_m}$ such that*

$$\sum_{m=0}^{J-1} \int_{t_m}^{t_{m+1}} \sum_{i=1}^{S} \left(\lambda_i^{\pi}(\boldsymbol{y}(t)) - \mu_i^m + \mu_i^m \log \frac{\mu_i^m}{\lambda_i^{\pi}(\boldsymbol{y}(t))} \right) dt$$
$$\leq I_s(\boldsymbol{r}, v) + \epsilon.$$

The proof of Lemma 3.12 can be found in the Appendix.

With the constructed matrices $\{\eta^{mk}\}$ in Lemma 3.11, we define the process $\bar{\xi}_n$ with jump rate $n\eta_{ij}(t)$ where $\eta_{ij}(t) = \eta_{ij}^{mk}$, $t \in [t_{mk}, t_{m,k+1})$. Similarly, we take $\boldsymbol{\mu}^m$ constructed from Lemma 3.12 and define $\bar{\boldsymbol{z}}_n$ with jump rate

$$n\mu_i(t) \frac{\lambda_i(\bar{\boldsymbol{z}}_n(t), \bar{\xi}_n(t))}{\lambda_i^{\pi}(\boldsymbol{y}(t))}$$

for its ith component, where $\mu_i(t)$ is piecewise constant and $\mu_i(t) = \mu_i^m$ for $t \in [t_m, t_{m+1})$.

We have the following convergence result for the constructed approximations for π and $\boldsymbol{y}$.

Lemma 3.13. *Convergence of the approximation $\bar{v}_n$*

$$\mathbb{P} - \lim_{n \to \infty} \rho^{(2)}(\bar{v}^n, \pi) = 0.$$

Lemma 3.14. *Convergence of the approximation $\bar{\boldsymbol{z}}_n$*

$$\mathbb{P} - \lim_{n \to \infty} \rho_c^{(1)}(\bar{\boldsymbol{z}}^n, \boldsymbol{y}) = 0.$$

The proof of Lemmas 3.13, 3.14 will be given in the Appendix.

As we have finished the construction of $\bar{\boldsymbol{z}}_n$ and $\bar{\xi}_n$, we now perform the change of measure. Denote $\mathbb{Q}_n$ and $\bar{\mathbb{Q}}_n$ the distributions of $(\boldsymbol{z}_n(t), \xi_n(t))_{t \leq T}$ and $(\bar{\boldsymbol{z}}_n(t), \bar{\xi}_n(t))_{t \leq T}$, respectively. We have

$$\frac{d\mathbb{Q}_n}{d\bar{\mathbb{Q}}_n}(\bar{\boldsymbol{z}}_n, \bar{\xi}_n)$$

$$= \exp\left\{ -\int_0^T n \sum_{i=1}^d \left(\lambda_i(\bar{\boldsymbol{z}}_n(t), \bar{\xi}_n(t)) - \mu_i(t) \frac{\lambda_i(\bar{\boldsymbol{z}}_n(t), \bar{\xi}_n(t))}{\lambda_i^{\pi}(\boldsymbol{y}(t))} \right) dt \right.$$

$$- \int_0^T \sum_i \log \frac{\mu_i(t^-)}{\lambda_i^{\pi}(\boldsymbol{y}(t^-))} dY_t^i - \int_0^T n \sum_{i,j=1}^D \left(q_{ij}(\bar{\boldsymbol{z}}_n(t)) - \eta_{ij}(t) \right) dt$$

$$- \left. \int_0^T \sum_{i,j} \log \frac{\eta_{ij}(t^-)}{q_{ij}(\bar{\boldsymbol{z}}_n(t^-))} dM_t^{ij} \right\}$$

$$:= e^{B(\bar{\boldsymbol{z}}_n, \bar{\xi}_n)}, \tag{3.32}$$

where Y_t^i is the counting process induced by $\bar{z}_n(t)$ that will increase by one each time when a jump occurs in the u_i direction and M_t^{ij} is the counting process induced by $\bar{\xi}_n(t)$ that will increase by one each time when a jump occurs from state i to state j. The next lemma shows that the expectation of $B(\bar{z}_n, \bar{\xi}_n)$ in the exponent becomes simple in the limit $n \to \infty$.

Lemma 3.15.

$$\lim_{n \to \infty} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_0^T \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i(t) \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(y(t))} dt$$
$$= \sum_{m=0}^{J-1} \int_{t_m}^{t_{m+1}} \lambda_i^\pi(y(t)) - \mu_i^m dt. \tag{3.33}$$

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_0^T \sum_i \log \frac{\mu_i(t^-)}{\lambda_i^\pi(y(t))} dY_t^i$$
$$= \sum_{m=0}^{J-1} \int_{t_m}^{t_{m+1}} \sum_{i=1}^S \mu_i^m \log \frac{\mu_i^m}{\lambda_i^\pi(y(t))} dt. \tag{3.34}$$

$$\lim_{n \to \infty} \frac{1}{n} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_0^T \sum_{i,j} \log \frac{\eta_{ij}(t^-)}{q_{ij}(\bar{z}_n(t^-))} dM_t^{ij}$$
$$= \sum_{m=0}^{J-1} \sum_{k=0}^{K_m-1} \int_{t_{mk}}^{t_{m,k+1}} \sum_{i=1}^D n_\pi(t,i) \sum_{j=1}^D \eta_{ij}^{mk} \log \frac{\eta_{ij}^{mk}}{q_{ij}(y(t))} dt. \tag{3.35}$$

The proof of Lemma 3.15 is based on the ideas in proving Lemma 5.52 and Lemma 8.70 in [26].

Proof. Since $\mu_i(t)$ is a step function and constant in $[t_m, t_{m+1})$, to prove (3.33), we just need to prove for each m

$$\lim_{n \to \infty} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_{t_m}^{t_{m+1}} \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(y(t))} dt$$
$$= \int_{t_m}^{t_{m+1}} \lambda_i^\pi(y(t)) - \mu_i^m dt. \tag{3.36}$$

Define

$$N_\epsilon(y) := \{z \in \mathbb{D}^d[0, T] : \rho_c^{(1)}(z, y) \le \epsilon\}.$$

We have

$$\mathbb{E}_{\bar{\mathbb{Q}}_n} \int_{t_m}^{t_{m+1}} \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(y(t))} dt$$

$$=\mathbb{E}_{\bar{\mathbb{Q}}_n} \chi_{\{\bar{z}_n \in N_\epsilon(y)\}} \int_{t_m}^{t_{m+1}} \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(y(t))} dt$$

$$+\mathbb{E}_{\bar{\mathbb{Q}}_n} \chi_{\{\bar{z}_n \notin N_\epsilon(y)\}} \int_{t_m}^{t_{m+1}} \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(y(t))} dt. \quad (3.37)$$

By Lemma 3.14, the second term on the right hand side of (3.37) tends to zero as $n \to \infty$. Next let us estimate the first term.

By Assumption 2.3, we have

$$\left| \frac{\lambda_i(x', j)}{\lambda_i^\pi(x)} - \frac{\lambda_i(z, j)}{\lambda_i^\pi(z)} \right| \leq \left| \frac{\lambda_i(x', j)}{\lambda_i^\pi(x)} - \frac{\lambda_i(x', j)}{\lambda_i^\pi(z)} \right| + \left| \frac{\lambda_i(x', j) - \lambda_i(z, j)}{\lambda_i^\pi(z)} \right|$$

$$\leq \Lambda^3 L \|x - z\| + \Lambda L \|x' - z\| \quad (3.38)$$

for any $z, x, x' \in \mathbb{R}^d$, $i \in \{1, \ldots, S\}$ and $j \in \{1, 2 \ldots, D\}$.

Now take an integer N and divide $[t_m, t_{m+1}]$ into L pieces. Define $\tau_l = t_m + l(t_{m+1} - t_m)/N$ for $l = 0, \ldots, N$. Since y is continuous in $[0, T]$, we can choose N large enough such that

$$\sup_{t \in [\tau_l, \tau_{l+1}]} \|y(t) - y(\tau_l)\| \leq \epsilon \quad \text{for any } l \in \{0, \cdots, N-1\}.$$

By (2.1) and (3.38), we have

$$\chi_{\{\bar{z}_n \in N_\epsilon(y)\}} \int_{t_m}^{t_{m+1}} \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(y(t))} dt$$

$$\leq \chi_{\{\bar{z}_n \in N_\epsilon(y)\}} \sum_{l=0}^{N-1} \int_{\tau_l}^{\tau_{l+1}} \lambda_i(y(\tau_l), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(y(\tau_l), \bar{\xi}_n(t))}{\lambda_i^\pi(y(\tau_l))} + C\epsilon dt,$$

where $C = (2 + 2\Lambda + \Lambda^3)L$. So we have

$$\lim_{n \to \infty} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_{t_m}^{t_{m+1}} \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(y(t))} dt$$

$$\leq \sum_{l=0}^{N-1} \int_{\tau_l}^{\tau_{l+1}} \sum_{j=1}^{D} \lambda_i(y(\tau_l), j) n_\pi(t, j) - \mu_i^m + C\epsilon dt$$

$$\leq \sum_{l=0}^{N-1} \int_{\tau_l}^{\tau_{l+1}} \sum_{j=1}^{D} \lambda_i(y(t), j) n_\pi(t, j) - \mu_i^m + C_1 \epsilon dt$$

$$= \int_{t_m}^{t_{m+1}} \lambda_i^\pi(y(t)) - \mu_i^m dt + C_1(t_{m+1} - t_m)\epsilon$$

by ergodicity of the process $\bar{\xi}_n$, where $C_1 = (3 + 2\Lambda + \Lambda^3)L$. Similarly, we can also obtain

$$
\lim_{n\to\infty} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_{t_m}^{t_{m+1}} \lambda_i(\bar{z}_n(t), \bar{\xi}_n(t)) - \mu_i^m \frac{\lambda_i(\bar{z}_n(t), \bar{\xi}_n(t))}{\lambda_i^\pi(\boldsymbol{y}(t))} dt
$$
$$
\geq \int_{t_m}^{t_{m+1}} \lambda_i^\pi(\boldsymbol{y}(t)) - \mu_i^m dt - C_1(t_{m+1} - t_m)\epsilon.
$$

So we finish the proof for (3.33).

To prove (3.34), we first assume that $\lambda_i(\boldsymbol{x})$ are constant functions. In $[t_m, t_{m+1})$, the number of jumps $\bar{z}_n$ makes in each direction $\boldsymbol{u}_i/n$ are independent Poisson random variables with mean $n\mu_i^m(t_{m+1} - t_m)$. So

$$
\lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_{t_m}^{t_{m+1}} \sum_i \log \frac{\mu_i(t^-)}{\lambda_i^\pi(\boldsymbol{y}(t))} dY_t^i
$$
$$
= \int_{t_m}^{t_{m+1}} \sum_{i=1}^S \mu_i^m \log \frac{\mu_i^m}{\lambda_i^\pi(\boldsymbol{y}(t))} dt. \tag{3.39}
$$

For general λ_i, we can use the technique for proving (3.33) by dividing the interval $[t_m, t_{m+1}]$ into small pieces and approximating (3.34) by Riemann sums.

For (3.35), again we first assume that $q_{ij}(\boldsymbol{x})$ are constant functions. In $[t_{mk}, t_{m,k+1}]$, the number of jumps $\bar{\xi}_n$ makes in each direction $\boldsymbol{e}_{ij}$ are independent Poisson random variables with mean $n \cdot n_\pi(t_{mk}, i)\eta_{ij}^{mk}(t_{m,k+1} - t_{mk})$. So

$$
\lim_{n\to\infty} \frac{1}{n} \mathbb{E}_{\bar{\mathbb{Q}}_n} \int_{t_{mk}}^{t_{m,k+1}} \sum_{i,j} \log \frac{\eta_{ij}(t^-)}{q_{ij}(\bar{z}_n(t^-))} dM_t^{ij}
$$
$$
= \int_{t_{mk}}^{t_{m,k+1}} \sum_{i=1}^D n_\pi(t, i) \sum_{j=1}^D \eta_{ij}^{mk} \log \frac{\eta_{ij}^{mk}}{q_{ij}(\boldsymbol{y}(t))} dt. \tag{3.40}
$$

For general q_{ij}, we consider separate cases $\{\bar{z}_n \in N_\epsilon(\boldsymbol{y})\}$ and $\{\bar{z}_n \notin N_\epsilon^c(\boldsymbol{y})\}$ as in (3.37). Similar as proving (3.33), we can get the limit (3.35). $\qquad\square$

Lemma 3.16. *For given $\boldsymbol{r} \in \mathbb{D}^d[0, T]$ and $v \in \mathbb{M}_L[0, T]$, assume that $\boldsymbol{r}$ is absolutely continuous and $n_v(t, \cdot)$ is continuous in t. Then for arbitrarily small $\epsilon > 0$ we have*

$$
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\boldsymbol{z}_n \in N_\epsilon(\boldsymbol{r}), v_n \in N_\epsilon(v)) \geq -\left(I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v)\right).
$$

Proof. By Eq. (3.32) and Jensen's inequality, for any $\epsilon > 0$

$$\mathbb{P}\left(\boldsymbol{z}_n \in N_\epsilon(\boldsymbol{r}), \nu_n \in N_\epsilon(\nu)\right)$$

$$\geq \mathbb{P}\left(\boldsymbol{z}_n \in N_{\epsilon/2}(\boldsymbol{y}), \nu_n \in N_{\epsilon/2}(\pi)\right)$$

$$= \mathbb{E}_{\bar{\mathbb{Q}}_n}\left[\frac{d\mathbb{Q}_n}{d\bar{\mathbb{Q}}_n}(\bar{\boldsymbol{z}}_n(t), \bar{\xi}_n(t))\chi_{\{\bar{\boldsymbol{z}}_n \in N_{\epsilon/2}(\boldsymbol{y}), \bar{\nu}_n \in N_{\epsilon/2}(\pi)\}}\right]$$

$$= \mathbb{E}_{\bar{\mathbb{Q}}_n}\left[e^{B(\bar{\boldsymbol{z}}_n, \bar{\xi}_n)}\chi_{\{\bar{\boldsymbol{z}}_n \in N_{\epsilon/2}(\boldsymbol{y}), \bar{\nu}_n \in N_{\epsilon/2}(\pi)\}}\right]$$

$$\geq \mathbb{E}_{\bar{\mathbb{Q}}_n}\left[\chi_{\{\bar{\boldsymbol{z}}_n \in N_{\epsilon/2}(\boldsymbol{y}), \bar{\nu}_n \in N_{\epsilon/2}(\pi)\}}\right] \exp\left\{\frac{\mathbb{E}_{\bar{\mathbb{Q}}_n}\left[\chi_{\{\bar{\boldsymbol{z}}_n \in N_{\epsilon/2}(\boldsymbol{y}), \bar{\nu}_n \in N_{\epsilon/2}(\pi)\}}B(\bar{\boldsymbol{z}}_n, \bar{\xi}_n)\right]}{\mathbb{E}_{\bar{\mathbb{Q}}_n}\left[\chi_{\{\bar{\boldsymbol{z}}_n \in N_{\epsilon/2}(\boldsymbol{y}), \bar{\nu}_n \in N_{\epsilon/2}(\pi)\}}\right]}\right\}.$$
(3.41)

By Lemmas 3.13 and 3.14, we know that

$$\lim_{n \to \infty} \mathbb{E}_{\bar{\mathbb{Q}}_n}\left[\chi_{\{\bar{\boldsymbol{z}}_n \in N_{\epsilon/2}(\boldsymbol{y}), \bar{\nu}_n \in N_{\epsilon/2}(\pi)\}}\right] = 1. \tag{3.42}$$

Thus, according to Lemma 3.15, (3.41) and (3.42), we have

$$\liminf_{n \to \infty} \frac{1}{n}\log\mathbb{P}\left(\boldsymbol{z}_n \in N_\epsilon(\boldsymbol{r}), \nu_n \in N_\epsilon(\nu)\right)$$

$$\geq \liminf_{n \to \infty} \frac{1}{n}\mathbb{E}_{\bar{\mathbb{Q}}_n}\left[\chi_{\{\bar{\boldsymbol{z}}_n \in N_{\epsilon/2}(\boldsymbol{y}), \bar{\nu}_n \in N_{\epsilon/2}(\pi)\}}B(\bar{\boldsymbol{z}}_n, \bar{\xi}_n)\right]$$

$$= -\left(\sum_{m=0}^{J-1}\int_{t_m}^{t_{m+1}}\sum_{i=1}^{S}\left(\lambda_i^\pi(\boldsymbol{y}(t)) - \mu_i^m\right)dt\right.$$

$$+ \sum_{m=0}^{J-1}\int_{t_m}^{t_{m+1}}\sum_{i=1}^{S}\mu_i^m\log\frac{\mu_i^m}{\lambda_i^\pi(\boldsymbol{y}(t))}dt$$

$$+ \sum_{m=0}^{J-1}\sum_{k=0}^{K_m-1}\int_{t_{mk}}^{t_{m,k+1}}\sum_{i=1}^{D}n_\pi(t,i)\sum_{j=1}^{D}\left(\eta_{ij}^{mk}\log\frac{\eta_{ij}^{mk}}{q_{ij}(\boldsymbol{y}(t))} + q_{ij}(\boldsymbol{y}(t)) - \eta_{ij}^{mk}\right)dt\right).$$
(3.43)

Combining Lemma 3.11, Lemma 3.12 and (3.43), we finish the proof. $\quad\square$

In the final theorem, we remove the continuity assumption on $n_\nu(t, \cdot)$ to get the desired lower bound estimation.

Theorem 3.17. *For given $\boldsymbol{r} \in \mathbb{D}^d[0, T]$ and $\nu \in \mathbb{M}_L[0, T]$, assume that $\boldsymbol{r}$ is absolutely continuous, we have*

$$\liminf_{n \to \infty} \frac{1}{n}\log\mathbb{P}\left(\boldsymbol{z}_n \in N_\epsilon(\boldsymbol{r}), \nu_n \in N_\epsilon(\nu)\right) \geq -\left(I_s(\boldsymbol{r}, \nu) + I_f(\boldsymbol{r}, \nu)\right).$$

Proof. We can construct a sequence of measures $\nu^{(k)}$ ($k \geq 1$) such for any k, $n_{\nu^{(k)}}$ is continuous in t and $\rho^{(2)}(\nu, \nu^{(k)}) \to 0$. From Lemma A.10,

$I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v)$ is lower semi-continuous in v. Thus, we can choose k_0 large enough such that for any $\delta > 0$ and $\epsilon > 0$,

$$I_s(\boldsymbol{r}, v^{(k_0)}) + I_f(\boldsymbol{r}, v^{(k_0)}) \geq I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v) - \delta$$

and

$$\rho^{(2)}(v, v^{(k_0)}) < \epsilon/2.$$

Thanks to Lemma 3.16, we have

$$
\begin{aligned}
\liminf_{n\to\infty} &\frac{1}{n} \log \mathbb{P}\left(\boldsymbol{z}_n \in N_\epsilon(\boldsymbol{r}), v_n \in N_\epsilon(v)\right)\\
\geq\ & \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\boldsymbol{z}_n \in N_\epsilon(\boldsymbol{r}), v_n \in N_{\epsilon/2}(v^{(k_0)})\right)\\
\geq\ & -\left(I_s(\boldsymbol{r}, v^{(k_0)}) + I_f(\boldsymbol{r}, v^{(k_0)})\right)\\
\geq\ & -\left(I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v)\right) - \delta.
\end{aligned}
$$

The proof is completed. $\qquad\qquad\square$

3.3. **Goodness of the rate functional.** The rate functional $I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v)$ is lower semicontinuous by Lemma A.10. The goodness of the rate functional is a direct consequence of Lemma 3.7.

4. Proof of Theorem 2.2

Now we prove Theorem 2.2 under the consideration $\boldsymbol{r} \in W = \overline{(\mathbb{R}^+)^d}$ instead of the whole space. The main clue of the proof is the same as the proof of Theorem 2.4 except some technicalities to understand the behavior of jumps near the boundary of W. We will only focus on the key parts which is different from the proof of Theorem 2.4.

The difficulty in the proof of lower bound is that we can not use the change of measure formula directly, since some of the jump rates may diminish on the boundary. Mainly following [27], We overcome this issue by carefully analyzing the boundary behavior of the dynamics .

Let a d-dimensional unit vector $\boldsymbol{v} := (1, 1, \cdots, 1)/\sqrt{d}$ and define the shifting $\boldsymbol{r}_\delta(t) = \boldsymbol{r}(t) + \delta \boldsymbol{v}$ with $\delta > 0$ a sufficiently small number. With similar approach in proving Lemma 5.1 in [27], we can show that

$$\limsup_{\delta\to 0^+} \left(I_s(\boldsymbol{r}_\delta, v) + I_f(\boldsymbol{r}_\delta, v)\right) \leq I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v). \tag{4.1}$$

Next we will prove

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\boldsymbol{z}_n \in N_\delta(\boldsymbol{r}), v_n \in N_\delta(v)\right) \geq -\left(I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v)\right).$$

Denote by $V_a(\boldsymbol{r})$ the modulus of continuity of $\boldsymbol{r}$ with size a, and set $\eta(a) = \max\{V_a(\boldsymbol{r}), a\}$ so that $\eta^{-1}(a) \le a$. Now, fix δ and set $t_\delta = \eta^{-1}(\delta/3)$. Then, $t_\delta \le \delta/3$ and for $t \le t_\delta$,

$$\sup_{0 \le t \le t_\delta} \|\boldsymbol{r}(0) + t \cdot \boldsymbol{v} - \boldsymbol{r}(t)\| \le t_\delta \cdot \|\boldsymbol{v}\| + \eta(t_\delta) \le 2\delta/3.$$

Therefore, for $0 < \alpha < 1/6$,

$$\mathbb{P}(\boldsymbol{z}_n \in N_\delta(\boldsymbol{r}), v_n \in N_\delta(v)) \ge \mathbb{P}\Big(\|\boldsymbol{z}_n(t) - \boldsymbol{r}(0) - t \cdot \boldsymbol{v}\| \le \alpha\delta \text{ on } t \in [0, t_\delta],$$

$$\boldsymbol{z}_n \in N_\delta(\boldsymbol{r}; [t_\delta, T]); v_n \in N_\delta(v)\Big),$$

where $N_\delta(\boldsymbol{r}; [t_\delta, T])$ is the δ-neighborhood of $\boldsymbol{r}$ restricted on $t \in [t_\delta, T]$. Now, on this time interval

$$\sup_{t_\delta \le t \le T} \|\boldsymbol{r}(t) - \boldsymbol{r}_{t_\delta}(t)\| \le \delta/3$$

and, moreover, $d(\boldsymbol{r}_{t_\delta}(t), \partial G) \ge t_\delta/\sqrt{d}$. Therefore, for any function $\boldsymbol{u}$ on $t \in [t_\delta, T]$, $\|\boldsymbol{u} - \boldsymbol{r}_{t_\delta}\| \le t_\delta/2\sqrt{d}$ implies that $\|\boldsymbol{u} - \boldsymbol{r}\| \le 5\delta/6$ and $d(\boldsymbol{r}_{t_\delta}(t), \partial G) \ge t_\delta/2\sqrt{d}$. Now define A_δ the $\alpha\delta$-neighborhood of $\boldsymbol{r}_0 + t_\delta \boldsymbol{v}$, i.e. $A_\delta := B_{\alpha\delta}(\boldsymbol{r}_0 + t_\delta \boldsymbol{v})$ and let $\boldsymbol{r}_{t_\delta}^{\boldsymbol{y}}$ be the shift of $\boldsymbol{r}_{t_\delta}$ such that $\boldsymbol{r}_{t_\delta}^{\boldsymbol{y}}(t_\delta) = \boldsymbol{y}$. Then,

$$\mathbb{P}(\boldsymbol{z}_n \in N_\delta(\boldsymbol{r}), v_n \in N_\delta(v)) \ge$$

$$\mathbb{P}\Big(\|\boldsymbol{z}_n(t) - \boldsymbol{r}(0) - t \cdot \boldsymbol{v}\| \le \alpha\delta \text{ on } t \in [0, t_\delta]; v_n \in N_\delta(v; [0, t_\delta])\Big)$$

$$\times \inf_{\boldsymbol{y} \in A_\delta} \mathbb{P}_{\boldsymbol{y}}\Big(\boldsymbol{z}_n \in N_{\frac{t_\delta}{2\sqrt{d}}}(\boldsymbol{r}_{t_\delta}^{\boldsymbol{y}}; [t_\delta, T]); v_n \in N_\delta(v; [t_\delta, T])\Big).$$

The first term satisfies a large deviation lower bound

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\Big(\|\boldsymbol{z}_n(t) - \boldsymbol{r}(0) - t \cdot \boldsymbol{v}\| \le \alpha\delta \text{ on } t \in [0, t_\delta]; v_n \in N_\delta(v; [0, t_\delta])\Big)$$

$$\ge -C t_\delta \tag{4.2}$$

by estimating the probability of a specific path $\boldsymbol{z}_n$ lying in the $\alpha\delta$-neighborhood of the curve $\boldsymbol{r}(0) + t\boldsymbol{v}$. Because the paths in $N_{\frac{t_\delta}{2\sqrt{d}}}(\boldsymbol{r}_{t_\delta}^{\boldsymbol{y}}; [t_\delta, T])$ are bounded away from the boundary uniformly for $\boldsymbol{y} \in A_\delta$, by Theorem 3.17, we have

$$\liminf_{n \to \infty} \frac{1}{n} \log \inf_{\boldsymbol{y} \in A_\delta} \mathbb{P}_{\boldsymbol{y}}\Big(\boldsymbol{z}_n \in N_{\frac{t_\delta}{2\sqrt{d}}}(\boldsymbol{r}_{t_\delta}^{\boldsymbol{y}}; [t_\delta, T]); v_n \in N_\epsilon(v; [t_\delta, T])\Big)$$

$$\ge -\Big(I_s^{[t_\delta, T]}(\boldsymbol{r}_{t_\delta}, v) + I_f^{[t_\delta, T]}(\boldsymbol{r}_{t_\delta}, v)\Big)$$

$$\ge -\Big(I_s(\boldsymbol{r}_{t_\delta}, v) + I_f(\boldsymbol{r}_{t_\delta}, v)\Big), \tag{4.3}$$

where $I_s^{[t_\delta, T]}(\boldsymbol{r}_{t_\delta}, \nu)$ and $I_f^{[t_\delta, T]}(\boldsymbol{r}_{t_\delta}, \nu)$ are rate functionals defined on the integration interval $[t_\delta, T]$. According to (4.1), (4.2) and (4.3), we proved the lower bound.

Next let us consider the upper bound. At first we note that since the rates $\lambda_i(\boldsymbol{z}, j)$ satisfies the linear growth condition

$$\lambda_i(\boldsymbol{z}, j) \le C(1 + \|\boldsymbol{z}\|),$$

it is easy to show that

$$\lim_{K \to \infty} \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(\sup_{0 \le t \le T} \|\boldsymbol{z}_n(t)\| > K) = -\infty$$

by simple moment estimates and Doob's martingale inequality. Consequently, it suffices to prove the large deviation estimates for bounded sets and we can assume $\lambda_i(\boldsymbol{z}, j)$ are bounded.

We only need to recheck Lemma 3.9 and Lemma A.11, since the other lemmas in upper bound estimates can be verified easily under the assumption that $\lambda_i(\boldsymbol{z}, j)$ are bounded. Thanks to Corollary 4.2 and Lemma 4.6 in [27], we can obtain that Lemma 3.9 and Lemma A.11 are also correct under Assumption 2.1. Thus the upper bound is also established.

The goodness of the rate functional trivially holds under Assumption 2.1. So we complete the proof of Theorem 2.2.

APPENDIX

Lemma A.1. *Let $\{f_\alpha\}$ be a collection of lower semi-continuous functions on a metric space. Then the function f define by $f(x) = \sup_\alpha f_\alpha(x)$ is lower semicontinuous.*

Lemma A.2. *Let $\{f_\alpha\}$ be a collection of convex functions on a metric space. Then the function f define by $f(x) = \sup_\alpha f_\alpha(x)$ is convex.*

Lemma A.3. *Let $K(x, y)$ be a real-valued function, continuous in (x, y) on $\mathbb{R}^d \times \mathbb{R}^D$, convex in x for each y, and concave in y for each x. Let two non-empty closed convex sets U and V be given, at least one of which is bounded. Then*

$$\inf_{x \in U} \sup_{y \in V} K(x, y) = \sup_{y \in V} \inf_{x \in U} K(x, y).$$

The proof of Lemma A.3 may be referred to Corollary 37.3.2 of [24].

A.1. **Part 1.** Proof of lemmas related to the upper bound estimate.

Lemma A.4. *Let $\boldsymbol{z}(t) \in \mathbb{R}^d$ be any measurable process for $t \in [0, T]$. Suppose there exist numbers a and δ such that for each $\boldsymbol{p} \in \mathbb{R}^d$ with $\|\boldsymbol{p}\| = 1$,*

$$\mathbb{P}\left(\sup_{0 \le t \le T} \langle \boldsymbol{z}(t), \boldsymbol{p} \rangle \ge a\right) \le \delta.$$

Then

$$\mathbb{P}\left(\sup_{0\leq t\leq T}\|\boldsymbol{z}(t)\|\geq a\sqrt{d}\right)\leq 2d\delta.$$

Proof. It is not difficult to find that

$$\left\{\sup_{0\leq t\leq T}\|\boldsymbol{z}(t)\|\geq a\sqrt{d}\right\}\subset\bigcup_{i=1}^{2d}\left\{\sup_{0\leq t\leq T}\langle\boldsymbol{z}(t),\boldsymbol{p}_i\rangle\geq a\right\}.$$

where $\boldsymbol{p}_i:=\boldsymbol{e}_i$, $\boldsymbol{p}_{i+d}:=-\boldsymbol{e}_i$ for $i=1,\dots,d$, and $\boldsymbol{e}_i$ are chosen as the canonical orthonormal basis in Euclidean space $\mathbb{R}^d$. □

In later texts, we will take an abused notation $\xi_n(t)=\boldsymbol{e}_i\in\mathbb{R}^D$ when $\xi_n(t)=i\in\mathbb{Z}_D$. This will not bring confusion since $\xi_n(t)$ is considered as a multidimensional vector only when we take inner product with other vectors.

Lemma A.5. *There exists a function $K:\mathbb{R}^+\to\mathbb{R}^+$ with*

$$\lim_{a\to\infty}K(a)/a=+\infty,$$

such that

$$\mathbb{P}\left(\sup_{0\leq t\leq T}\|\boldsymbol{z}_n(t)-\boldsymbol{z}_n(0)\|\geq a\right)\leq 2d\exp\left(-nTK\left(\frac{a}{T}\right)\right). \qquad (4.4)$$

Proof. The inequality (4.4) holds trivially whenever $K(a/T)=0$. It suffice to prove the lemma when a is large. For $\boldsymbol{p}\in\mathbb{R}^d$, $\boldsymbol{\sigma}\in\mathbb{R}^D$ and any $\rho>0$, with the form of infinitesimal generator $\mathscr{L}_n$ (1.2), we define a mean one exponential martingale

$$M_t^{\boldsymbol{\sigma}}=\exp\left(\langle\boldsymbol{z}_n(t)-\boldsymbol{z}_n(0),\rho\boldsymbol{p}\rangle-n\int_0^t\sum_{i=1}^S\lambda_i(\boldsymbol{z}_n(s),\xi_n(s))(e^{\langle\rho\boldsymbol{p},\boldsymbol{u}_i/n\rangle}-1)ds\right.$$
$$\left.+\langle\xi_n(t)-\xi_n(0),\boldsymbol{\sigma}\rangle-n\int_0^t\sum_{i=1}^D\chi_{\{\xi_n(s)=i\}}\sum_{j=1}^D q_{ij}(\boldsymbol{z}_n(s))(e^{\langle\boldsymbol{\sigma},\boldsymbol{e}_{ij}\rangle}-1)ds\right).$$

Define $U=\max_{1\leq i\leq S}\|\boldsymbol{u}_i\|$. Fix $\|\boldsymbol{p}\|=1$, we have

$$n\int_0^t\sum_{i=1}^S\lambda_i(\boldsymbol{z}_n(s),\xi_n(s))(e^{\langle\rho\boldsymbol{p},\boldsymbol{u}_i/n\rangle}-1)ds\leq ntS\Lambda e^{U\rho/n}=:R(t,\rho)$$

by Assumption 2.3. Hence we obtain

$$
\mathbb{P}\left(\sup_{0 \le t \le T} \left\langle z_n(t) - z_n(0), p \right\rangle \ge a\right)
$$

$$
= \mathbb{P}\left(\sup_{0 \le t \le T} \exp\left(\rho \left\langle z_n(t) - z_n(0), p \right\rangle\right) \ge \exp(\rho a)\right)
$$

$$
\le \mathbb{P}\left(\sup_{0 \le t \le T} M_t^{\sigma=0} \ge \exp\left(\rho a - R(T, \rho)\right)\right)
$$

$$
\le \exp\left(nT\left[S\Lambda e^{U\rho/n} - \frac{\rho}{n}\frac{a}{T}\right]\right),
$$

where the inequality follows from Doob's martingale inequality. Take

$$
\rho = \frac{n}{U} \log \frac{a}{TS\Lambda U} > 0.
$$

Then it is not difficult to show that if we set

$$
\tilde{K}(a) = \frac{a}{U}\left(\log \frac{a}{S\Lambda U} - 1\right)
$$

for a large and $K(a) = 0$ otherwise, then

$$
\mathbb{P}\left(\sup_{0 \le t \le T} \left\langle z_n(t) - z_n(0), p \right\rangle \ge a\right) \le \exp\left(-nT\tilde{K}\left(\frac{a}{T}\right)\right).
$$

Define $K(a) = \tilde{K}(a/\sqrt{d})$, we get the desired estimate by applying Lemma A.4. $\qquad\square$

Corollary A.6. *There exist positive constants c_1 and c_2 independent of t and τ, such that for any $t, \tau \in [0, T]$ with $0 \le t + \tau \le T$,*

$$
\mathbb{P}\left(\sup_{t \le s \le t+\tau} \|z_n(s) - z_n(t)\| \ge a\right) \le 2d \exp\left(-nac_1 \log\left(\frac{ac_2}{\tau}\right)\right).
$$

Proof of Lemma 3.1. Consider a typical interval $[t_j^n, t_{j+1}^n]$. Since $z_n(t)$ and $\tilde{z}_n(t)$ agree at the endpoints of this interval, it is obvious that

$$
\|\tilde{z}_n(t_j^n) - \tilde{z}_n(t_{j+1}^n)\| > \frac{\delta}{2} \quad \text{implies} \quad \|z_n(t_{j+1}^n) - z_n(t_j^n)\| > \frac{\delta}{2}.
$$

On the other hand, we have

$$
\|z_n(t) - z_n(t_j^n)\| \ge \|z_n(t) - \tilde{z}_n(t)\| - \|\tilde{z}_n(t_{j+1}^n) - \tilde{z}_n(t_j^n)\|
$$

since $\tilde{z}_n$ is piecewise linear and $\tilde{z}_n(t_j^n) = z_n(t_j^n)$. Therefore if $\|z_n(t) - \tilde{z}_n(t)\| > \delta$ for some t in the jth interval, we must have

$$
\sup_{t_j^n \le t \le t_{j+1}^n} \|z_n(t) - z_n(t_j^n)\| \ge \delta/2.
$$

Applying Corollary A.6 with $a = \delta/2$ and $\tau = T/n$ we obtain

$$\mathbb{P}\left(\sup_{t_j^n \le t \le t_{j+1}^n} \|\mathbf{z}_n(t) - \mathbf{z}_n(t_j^n)\| \ge \delta/2\right) \le 2d \exp\left(-n\frac{\delta c_1}{2}\log\left(\frac{n\delta c_3}{2}\right)\right),$$

where $c_3 = c_2/T$. Thus,

$$\begin{aligned}
\mathbb{P}(\rho^{(1)}(\mathbf{z}_n, \tilde{\mathbf{z}}_n) > \delta) &\le \sum_{j=0}^{n-1} \mathbb{P}\left(\sup_{t_j^n \le t \le t_{j+1}^n} \|\mathbf{z}_n(t) - \tilde{\mathbf{z}}_n(t)\| > \delta\right) \\
&\le \sum_{j=0}^{n-1} \mathbb{P}\left(\sup_{t_j^n \le t \le t_{j+1}^n} \|\mathbf{z}_n(t) - \mathbf{z}_n(t_j^n)\| > \delta/2\right) \\
&\le n \cdot 2d \exp\left(-n\frac{\delta c_1}{2}\log\left(\frac{n\delta c_3}{2}\right)\right).
\end{aligned}$$

The result follows since c_1 and c_3 are positive constants. $\qquad\square$

Lemma A.7. *For any given bounded sets $A_1 \in \mathbb{R}^d$ and $A_2 \in \mathbb{R}^D$, we have that*

$$\begin{aligned}
\limsup_{n\to\infty} \mathbb{E}_{\mathbf{x},m} \exp\Big\{ &n\left\langle \tilde{\mathbf{z}}_n\left(\frac{T}{n}\right) - \tilde{\mathbf{z}}_n(0), \mathbf{p} \right\rangle - n\int_0^{T/n} H_s^\delta\left(\mathbf{x}, \mathbf{p}, n_{v_n}(t, \cdot)\right) dt \\
&+ \left\langle \xi_n\left(\frac{T}{n}\right) - \xi_n(0), \boldsymbol{\sigma} \right\rangle + n\int_0^{T/n} S^\delta(\mathbf{x}, n_{v_n}(t, \cdot), \boldsymbol{\sigma}) dt \Big\} \le 1.
\end{aligned}$$

holds uniformly in $\mathbf{x} \in \mathbb{R}^d$, $m \in \{1, 2, \cdots, D\}$, $\mathbf{p} \in A_1$ and $\boldsymbol{\sigma} \in A_2$, where $\mathbb{E}_{\mathbf{x},m}$ means the expectation with respect to the paths of $(\mathbf{z}_n, \xi_n)$ starting from $(\mathbf{x}, m)$ at $t = 0$.

Proof. For any $\mathbf{p} \in A_1$ and $\boldsymbol{\sigma} \in A_2$, define the mean one exponential martingale

$$\begin{aligned}
M_t = \exp\Big(&n\Big[\langle \mathbf{z}_n(t) - \mathbf{z}_n(0), \mathbf{p} \rangle - \int_0^t \sum_{i=1}^S \lambda_i(\mathbf{z}_n(s), \xi_n(s))(e^{\langle \mathbf{p}, \mathbf{u}_i \rangle} - 1) ds \\
&+ \langle \xi_n(t) - \xi_n(0), \boldsymbol{\sigma} \rangle - n\int_0^t \sum_{i=1}^D \chi_{\{\xi_n(s)=i\}} \sum_{j=1}^D q_{ij}(\mathbf{z}_n(s))(e^{\langle \boldsymbol{\sigma}, \mathbf{e}_{ij} \rangle} - 1) ds \Big).
\end{aligned}$$

Since $\tilde{z}_n(t_j^n) = z_n(t_j^n)$, for any $p \in A_1$ we have

$$
\begin{aligned}
1 = \mathbb{E}_{x,m} \exp \Bigg\{ n \Bigg[& \left\langle \tilde{z}_n\left(\frac{T}{n}\right) - \tilde{z}_n(0), p \right\rangle \\
& - \int_0^{T/n} \sum_{i=1}^S \sum_{j=1}^D \lambda_i(z_n(s), j)(e^{\langle p, u_i \rangle} - 1) \nu_n(ds, j) \Bigg] \\
& + \left\langle \xi_n\left(\frac{T}{n}\right) - \xi_n(0), \sigma \right\rangle - n \int_0^{T/n} \sum_{i,j=1}^D q_{ij}(z_n(s))(e^{\langle \sigma, e_{ij}\rangle} - 1)\nu_n(ds, i) \Bigg\}.
\end{aligned}
$$

By definition, the term

$$
\sum_{i=1}^S \sum_{j=1}^D \lambda_i(z_n(s), j)(e^{\langle p, u_i \rangle} - 1)\nu_n(ds, j)
$$

can be written as $H_s(z_n(s), p, n_{\nu_n}(s, \cdot)) ds$ and

$$
-\sum_{i=1}^D \sum_{j=1}^D q_{ij}(z_n(s))(e^{\langle \sigma, e_{ij} \rangle} - 1)\nu_n(ds, i)
$$

can be written as $S(z_n(s), n_{\nu_n}(s, \cdot), \sigma) ds$.

Let

$$
S_\delta = \left\{ \omega : \sup_{0 \le t \le T/n} \|z_n(t) - x\| < \frac{\delta}{2} \right\},
$$

we have

$$
\begin{aligned}
1 \;\ge\; & \mathbb{E}_{x,m} \chi_{S_\delta} \exp \Bigg\{ \left(n \left\langle \tilde{z}_n\left(\frac{T}{n}\right) - \tilde{z}_n(0), p \right\rangle - n \int_0^{T/n} H_s^\delta\left(x, p, n_{\nu_n}(t, \cdot)\right) dt \right) \\
& \quad + \left\langle \xi_n\left(\frac{T}{n}\right) - \xi_n(0), \sigma \right\rangle + n \int_0^{T/n} S^\delta(x, n_{\nu_n}(t, \cdot), \sigma) dt \Bigg\} \\
=\; & \mathbb{E}_{x,m} \exp \Bigg\{ \left(n \left\langle \tilde{z}_n\left(\frac{T}{n}\right) - \tilde{z}_n(0), p \right\rangle - n \int_0^{T/n} H_s^\delta\left(x, p, n_{\nu_n}(t, \cdot)\right) dt \right) \\
& \quad + \left\langle \xi_n\left(\frac{T}{n}\right) - \xi_n(0), \sigma \right\rangle + n \int_0^{T/n} S^\delta(x, n_{\nu_n}(t, \cdot), \sigma) dt \Bigg\} \\
& - \mathbb{E}_{x,m} \chi_{S_\delta^c} \exp \Bigg\{ \left(n \left\langle \tilde{z}_n\left(\frac{T}{n}\right) - \tilde{z}_n(0), p \right\rangle - n \int_0^{T/n} H_s^\delta\left(x, p, n_{\nu_n}(t, \cdot)\right) dt \right) \\
& \quad + \left\langle \xi_n\left(\frac{T}{n}\right) - \xi_n(0), \sigma \right\rangle + n \int_0^{T/n} S^\delta(x, n_{\nu_n}(t, \cdot), \sigma) dt \Bigg\} \qquad (4.5)
\end{aligned}
$$

Since A_1 and A_2 are bounded sets, there exist B_1 and B_2 such that $\|p\| \le B_1$ and $\|\sigma\| \le B_2$. From the Assumption 2.3 and the boundedness of p

and $\boldsymbol{\sigma}$, we have

$$
\mathbb{E}_{\boldsymbol{x},m} \chi_{S_\delta^c} \exp\left\{ \left(n \left\langle \tilde{\boldsymbol{z}}_n\left(\frac{T}{n}\right) - \tilde{\boldsymbol{z}}_n(0), \boldsymbol{p} \right\rangle - n \int_0^{T/n} H_s^\delta\left(\boldsymbol{x}, \boldsymbol{p}, n_{v_n}(t, \cdot)\right) dt \right) \right.
$$
$$
\left. + \left\langle \xi_n\left(\frac{T}{n}\right) - \xi_n(0), \boldsymbol{\sigma} \right\rangle + n \int_0^{T/n} S^\delta(\boldsymbol{z}_n(t), n_{v_n}(t, \cdot), \boldsymbol{\sigma}) dt \right\}
$$
$$
\leq \mathbb{E}_{\boldsymbol{x},m} \left(\chi_{S_\delta^c} \exp\left(n \left\langle \tilde{\boldsymbol{z}}_n\left(\frac{T}{n}\right) - \tilde{\boldsymbol{z}}_n(0), \boldsymbol{p} \right\rangle + 3K \right) \right)
$$
$$
\leq \sum_{k=1}^\infty \exp\left(n(k+1)\frac{\delta}{2}|\boldsymbol{p}| + 3K \right) \times \mathbb{P}\left(\frac{k\delta}{2} \leq \sup_{0 \leq t \leq T/n} |\boldsymbol{z}_n(t) - \boldsymbol{x}| \leq \frac{(k+1)\delta}{2} \right)
$$
$$
\leq \sum_{k=1}^\infty 2d \exp\left(n\left((k+1)\frac{\delta}{2}B_1 - \frac{k\delta c_1}{2}\log\left(\frac{k\delta c_2 n}{2T}\right) \right) \right) \times e^{3K} \to 0 \qquad (4.6)
$$

as n goes to infinity for all $\boldsymbol{x} \in \mathbb{R}^d$ with $\|\boldsymbol{p}\| \leq B_1$ and $\|\boldsymbol{\sigma}\| \leq B_2$, where K is a uniform bound depending on the bounds of $S^\delta(\cdot, \cdot, \cdot)$ and $H_s^\delta(\cdot, \cdot, \cdot)$ in the whole space, B_1, B_2 and T. Combining (4.6) and (4.5), we complete the proof. $\qquad\square$

Corollary A.8. *For any fixed step functions $\boldsymbol{\theta}(t) \in \mathbb{R}^d$ and $\boldsymbol{\alpha}(t) \in \mathbb{R}^D$, there exist constants $C > 0$ and n_0 such that*

$$
\mathbb{E} \exp\{nJ_n(\tilde{\boldsymbol{z}}_n, \boldsymbol{\theta}, v_n, \boldsymbol{\alpha})\} \leq C
$$

for all $n > n_0$, where J_n is defined in (3.9).

Proof. By definition

$$
\exp\{nJ_n(\tilde{\boldsymbol{z}}_n, \boldsymbol{\theta}, v_n, \boldsymbol{\alpha})\}
$$
$$
= \exp\left\{ \sum_{j=0}^{n-1} \left(n \left\langle \tilde{\boldsymbol{z}}_n(t_{j+1}^n) - \tilde{\boldsymbol{z}}_n(t_j^n), \boldsymbol{\theta}(t_j^n) \right\rangle \right.\right.
$$
$$
- n \int_{t_j^n}^{t_j^{n+1}} H_s^\delta\left(\tilde{\boldsymbol{z}}_n(t_j^n), \boldsymbol{\theta}(t_j^n), n_{v_n}(t, \cdot) \right) dt
$$
$$
\left.\left. + n \int_{t_j^n}^{t_j^{n+1}} S^\delta(\tilde{\boldsymbol{z}}_n(t_j^n), n_{v_n}(t, \cdot), \boldsymbol{\alpha}(t_j^n)) dt \right) \right\}
$$

$$= \exp\left\{ \sum_{j=0}^{n-1} \left(n\left\langle \tilde{z}_n(t_{j+1}^n) - \tilde{z}_n(t_j^n), \boldsymbol{\theta}(t_j^n) \right\rangle \right.\right.$$

$$\left. - n\int_{t_j^n}^{t_j^{n+1}} H_s^{\delta}\left(\tilde{z}_n(t_j^n), \boldsymbol{\theta}(t_j^n), n_{v_n}(t, \cdot) \right) dt \right.$$

$$\left. + \left\langle \xi_n(t_{j+1}^n) - \xi_n(t_j^n), \boldsymbol{\alpha}(t_j^n) \right\rangle + n\int_{t_j^n}^{t_j^{n+1}} S^{\delta}(\tilde{z}_n(t_j^n), n_{v_n}(t, \cdot), \boldsymbol{\alpha}(t_j^n)) dt \right)$$

$$\left. - \sum_{j=0}^{n-1} \left\langle \xi_n(t_{j+1}^n) - \xi_n(t_j^n), \boldsymbol{\alpha}(t_j^n) \right\rangle \right\}. \tag{4.7}$$

Now $\boldsymbol{\alpha}$ is a step function, let us first consider $\alpha(t) = \alpha_0$ on the interval $t \in [0, \tau]$. We have

$$\sum_{j=0}^{n-1} \chi_{\{t_{j+1}^n \leq \tau\}} \left\langle \xi_n(t_{j+1}^n) - \xi_n(t_j^n), \boldsymbol{\alpha}_0 \right\rangle$$

$$= \left\langle \xi_n\left(\frac{\lfloor n\tau \rfloor}{n} \right) - \xi_n(0), \boldsymbol{\alpha}_0 \right\rangle$$

where $\lfloor a \rfloor$ is the largest integer smaller than a. Since ξ_n and $\boldsymbol{\alpha}$ are bounded in [0,T], $\left| \left\langle \xi_n\left(\frac{\lfloor n\tau \rfloor}{n} \right) - \xi_n(0), \boldsymbol{\alpha}_0 \right\rangle \right|$ is uniformly bounded. Repeating this argument on the finite number of intervals on which $\boldsymbol{\alpha}$ are constants, we have that $\left| \sum_{j=0}^{n-1} \left\langle \xi_n(t_{j+1}^n) - \xi_n(t_j^n), \boldsymbol{\alpha}(t_j^n) \right\rangle \right|$ is bounded. Thus by (4.7), Lemma A.7 and the Markov property of (z_n, ξ_n),

$$\limsup_{n \to \infty} \mathbb{E} \exp\{ nJ_n(\tilde{z}_n, \boldsymbol{\theta}, v_n, \boldsymbol{\alpha}) \} \leq C$$

where C is a positive constant. □

Lemma A.9. *(Uniformly absolute continuity) Given $\boldsymbol{r} \in \mathbb{D}^d[0, T]$ and $v \in \mathbb{M}_L[0, T]$. Let $I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v) \leq K$ and fix some $\epsilon > 0$. Then there is a $\delta > 0$, independent of $\boldsymbol{r}$, such that for any collection of non-overlapping intervals in $[0, T]$ with total length δ*

$$\left\{ [t_j, s_j], j = 1, \cdots, J \right\} \quad \text{with} \quad \sum_{j=1}^{J} (s_j - t_j) = \delta,$$

we have

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

We can also find a constant B depending only on ϵ and K so that

$$\int_0^T \chi_{\{\| \dot{\boldsymbol{r}}(t) \| \geq B\}} dt \leq \epsilon.$$

Proof. For any collection of non-overlapping intervals $\{[t_j, s_j]\}_j$, define the function $k(t)$ to be equal to one if t is in some interval $[t_j, s_j]$ and zero otherwise. Since $I_s(\boldsymbol{r}, v) + I_f(\boldsymbol{r}, v) \le K$, $\boldsymbol{r}$ is absolutely continuous and $I_s(\boldsymbol{r}, v) \le K$. For any $a > 0$,

$$
\begin{aligned}
\sum_{j=1}^{J} \|\boldsymbol{r}(s_j) - \boldsymbol{r}(t_j)\| &\le \int_0^T \|\dot{\boldsymbol{r}}(t)\| k(t) dt \\
&\le \int_0^T a \cdot \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \le a\}} k(t) dt \\
&\quad + \int_0^T \frac{L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot))}{L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) / |\dot{\boldsymbol{r}}(t)|} \chi_{\{\|\dot{\boldsymbol{r}}(t)\| > a\}} k(t) dt \\
&\le a \cdot \delta + \frac{K}{f(a)}
\end{aligned}
$$

where

$$
f(a) := \inf_{\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w} \in \Delta_D} \left\{ \frac{L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})}{\|\boldsymbol{\beta}\|} : \|\boldsymbol{\beta}\| \ge a \right\}.
$$

Recalling the definition of $L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$ in (2.6), we define $U := \max_i \|\boldsymbol{u}_i\|$. For any $\boldsymbol{w} \in \Delta_D$ if we take $\boldsymbol{p} = \boldsymbol{\beta} \log \|\boldsymbol{\beta}\| / (U \|\boldsymbol{\beta}\|)$ in (2.6), we obtain

$$
L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) \ge \|\boldsymbol{\beta}\| \log \|\boldsymbol{\beta}\| / U - \|\boldsymbol{\beta}\| S\Lambda.
$$

This means that $f(a) \to \infty$ as $a \to \infty$. The choice $a = 1/\sqrt{\delta}$ and taking δ sufficiently small establishes the uniformly absolute continuity.

Now we turn to the second statement. Since

$$
\begin{aligned}
\int_0^T \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \ge B\}} dt &\le \frac{1}{B} \int_0^T \|\dot{\boldsymbol{r}}(t)\| \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \ge B\}} dt \\
&\le \frac{1}{B} \int_0^T \frac{L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot))}{L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) / \|\dot{\boldsymbol{r}}(t)\|} \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \ge B\}} dt \\
&\le \frac{1}{B} \frac{I_s(\boldsymbol{r}, v)}{f(B)},
\end{aligned}
$$

we complete the proof by choosing a sufficiently large B. $\qquad \square$

Lemma A.10. *The rate functionals are lower semicontinuous, i.e., if $(\boldsymbol{r}_n, v_n) \to (\boldsymbol{r}, v)$ as $n \to \infty$, then*

$$
\liminf_{n \to \infty} I_s(\boldsymbol{r}_n, v_n) \ge I_s(\boldsymbol{r}, v), \quad \liminf_{n \to \infty} I_f(\boldsymbol{r}_n, v_n) \ge I_f(\boldsymbol{r}, v), \tag{4.8}
$$

$$
\liminf_{n \to \infty} I_s^\delta(\boldsymbol{r}_n, v_n) \ge I_s^\delta(\boldsymbol{r}, v), \quad \liminf_{n \to \infty} I_f^\delta(\boldsymbol{r}_n, v_n) \ge I_f^\delta(\boldsymbol{r}, v) \tag{4.9}
$$

and

$$
\liminf_{n \to \infty} I_s^\delta(\boldsymbol{r}_n, v_n, \boldsymbol{\theta}) \ge I_s^\delta(\boldsymbol{r}, v, \boldsymbol{\theta}), \quad \liminf_{n \to \infty} I_f(\boldsymbol{r}_n, v_n, \boldsymbol{\alpha}) \ge I_f^\delta(\boldsymbol{r}, v, \boldsymbol{\alpha}) \tag{4.10}
$$

for any fix step functions $\boldsymbol{\theta}(t) \in R^d$ and $\boldsymbol{\alpha}(t) \in R^D$.

Proof. We only need to consider the sequences of $\boldsymbol{r}_n$ which are absolutely continuous since it will be trivial otherwise. Let $(\boldsymbol{r}_n, \nu_n) \to (\boldsymbol{r}, \nu)$ as $n \to \infty$. We may assume that $I_s(\boldsymbol{r}_n, \nu_n) + I_f(\boldsymbol{r}_n, \nu_n)$ is bounded, say by a constant K. By Lemma A.9, we know that $\boldsymbol{r}$ is also absolutely continuous.

Since $\boldsymbol{r}(t)$ is absolutely continuous in $[0, T]$, given δ, we can partition the interval $[0, T]$ into J intervals $0 = t_1 \le t_2 \le \cdots \le t_{J+1} = T$ each of length Δ such that

$$\max_j \sup_{t_j \le t \le t_{j+1}} \|\boldsymbol{r}_n(t) - \boldsymbol{r}_n(t_j)\| < \delta.$$

Denote $F_n(t, i) = \nu_n([0, t], i)$ and $F(t, i) = \nu([0, t], i)$. Recalling the definition of $L_s^\delta(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$ in (3.5), we have that $L_s^\delta(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}))$ is lower semicontinuous in δ, $\boldsymbol{z}$, $\boldsymbol{\beta}$ and $\boldsymbol{w}$ and convex in $\boldsymbol{\beta}$ and $\boldsymbol{w}$ by Lemmas A.1 and A.2. Thus for any $\epsilon > 0$ and small enough Δ, we have

$$\int_0^T L_s(\boldsymbol{r}_n(t), \dot{\boldsymbol{r}}_n(t), n_{\nu_n}(t, \cdot)) dt$$

$$\ge \sum_{j=1}^J \int_{t_j}^{t_{j+1}} L_s^\delta(\boldsymbol{r}_n(t_j), \dot{\boldsymbol{r}}_n(t), n_{\nu_n}(t, \cdot)) dt$$

$$\ge \sum_{j=1}^J \Delta \cdot L_s^\delta \left(\boldsymbol{r}_n(t_j), \frac{\int_{t_j}^{t_{j+1}} \dot{\boldsymbol{r}}_n(t) dt}{\Delta}, \frac{\int_{t_j}^{t_{j+1}} n_{\nu_n}(t, \cdot) dt}{\Delta} \right) dt$$

$$= \sum_{j=1}^J \Delta \cdot L_s^\delta \left(\boldsymbol{r}_n(t_j), \frac{\boldsymbol{r}_n(t_{j+1}) - \boldsymbol{r}_n(t_j)}{\Delta}, \frac{F_n(t_{j+1}, \cdot) - F_n(t_j, \cdot)}{\Delta} \right). \qquad (4.11)$$

Define the functions $\boldsymbol{r}_J$, F_J as

$$\boldsymbol{r}_J(t) = \boldsymbol{r}(t_j), \quad F_J(t, \cdot) = F(t_j, \cdot) \quad \text{for } t_j \le t < t_{j+1}, j = 1, \cdots, J$$

and let

$$\boldsymbol{r}^J(t) := \boldsymbol{r}_J(t + \Delta), \quad F^J(t, \cdot) := F_J(t + \Delta, \cdot) \quad \text{for } 0 \le t < T - \Delta.$$

By (4.11), we have

$$\liminf_{n \to \infty} \int_0^T L_s(\boldsymbol{r}_n(t), \dot{\boldsymbol{r}}_n(t), n_{\nu_n}(t, \cdot)) dt$$

$$\ge \sum_{j=1}^J \int_{t_j}^{t_{j+1}} \liminf_{n \to \infty} L_s^\delta \left(\boldsymbol{r}_n(t_j), \frac{\boldsymbol{r}_n(t_{j+1}) - \boldsymbol{r}_n(t_j)}{\Delta}, \frac{F_n(t_{j+1}, \cdot) - F_n(t_j, \cdot)}{\Delta} \right) dt$$

$$\ge \sum_{j=1}^{J-1} \int_{t_j}^{t_{j+1}} L_s^\delta \left(\boldsymbol{r}_J(t), \frac{\boldsymbol{r}^J(t) - \boldsymbol{r}_J(t)}{\Delta}, \frac{F^J(t, \cdot) - F_J(t, \cdot)}{\Delta} \right) dt$$

$$= \int_0^{T-\Delta} L_s^\delta \left(\boldsymbol{r}_J(t), \frac{\boldsymbol{r}^J(t) - \boldsymbol{r}_J(t)}{\Delta}, \frac{F^J(t, \cdot) - F_J(t, \cdot)}{\Delta} \right) dt.$$

Now we use the nested partitions $J_k = 2^k$, so that $\Delta_k = T/2^k$ and a corresponding sequence δ_k that converges to zero. By Fatou's Lemma,

$$\liminf_{k\to\infty} \int_0^{T-\Delta_k} L_s^{\delta_k}\left(\boldsymbol{r}_{J_k}(t), \frac{\boldsymbol{r}^{J_k}(t) - \boldsymbol{r}_{J_k}(t,\cdot)}{\Delta}, \frac{F^{J_k}(t,\cdot) - F_{J_k}(t,\cdot)}{\Delta}\right) dt$$

$$\geq \int_0^T \liminf_{k\to\infty} \chi_{\{t \leq T-\Delta_k\}} L_s^{\delta_k}\left(\boldsymbol{r}_{J_k}(t), \frac{\boldsymbol{r}^{J_k}(t) - \boldsymbol{r}_{J_k}(t,\cdot)}{\Delta}, \frac{F^{J_k}(t,\cdot) - F_{J_k}(t,\cdot)}{\Delta}\right) dt$$

$$\geq \int_0^T L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t,\cdot)) dt.$$

So we established the lower semicontinuity of $I_s(\boldsymbol{r}, v)$.

The lower semicontinuity of $I_f(\boldsymbol{r}, v)$ can be done similarly. Recall the definition of $S^\delta(\boldsymbol{z}, \boldsymbol{w})$ in (3.18), we have that $S^\delta(\boldsymbol{z}, \boldsymbol{w})$ is lower semicontinuous in δ, $\boldsymbol{z}$ and $\boldsymbol{w}$ and convex in $\boldsymbol{w}$ by Lemmas A.1 and A.2. With exactly similar procedure as proving the lower semicontinuity of $I_s(\boldsymbol{r}, v)$, we can establish

$$\liminf_{n\to\infty} \int_0^T S(\boldsymbol{r}_n(t), n_{v_n}(t,\cdot)) dt \geq \int_0^{T-\Delta} S^\delta\left(\boldsymbol{r}_J(t), \frac{F^J(t,\cdot) - F_J(t,\cdot)}{\Delta}\right) dt$$

for a fine enough partition. Again we consider the sequence of nested partition $J_k = 2^k$ and $\Delta_k = T/2^k$. By Fatou's Lemma and the lower semicontinuity of S,

$$\liminf_{k\to\infty} \int_0^{T-\Delta_k} S^{\delta_k}\left(\boldsymbol{r}_{J_k}(t), \frac{F^{J_k}(t,\cdot) - F_{J_k}(t,\cdot)}{\Delta}\right) dt$$

$$\geq \int_0^T \liminf_{k\to\infty} \chi_{\{t \leq T-\Delta_k\}} S^{\delta_k}\left(\boldsymbol{r}_{J_k}(t), \frac{F^{J_k}(t,\cdot) - F_{J_k}(t,\cdot)}{\Delta}\right) dt$$

$$\geq \int_0^T S(\boldsymbol{r}(t), n_v(t,\cdot)) dt.$$

Thus we obtain the lower semicontinuity of $I_f(\boldsymbol{r}, v)$. The proof of (4.9) and (4.10) are simlar. $\qquad\square$

Lemma A.11. *Given $\boldsymbol{r} \in \mathbb{D}^d[0, T]$, $v \in \mathbb{M}_L[0, T]$ and $\epsilon > 0$, there exist step functions $\boldsymbol{\theta}(t) \in R^d$ and $\boldsymbol{\alpha}(t) \in R^D$ such that*

$$I_s(\boldsymbol{r}, v, \boldsymbol{\theta}) \geq I_s(\boldsymbol{r}, v) - \epsilon, \tag{4.12}$$

$$I_s^\delta(\boldsymbol{r}, v, \boldsymbol{\theta}) \geq I_s^\delta(\boldsymbol{r}, v) - \epsilon, \tag{4.13}$$

and

$$I_f(\boldsymbol{r}, v, \boldsymbol{\alpha}) \geq I_f(\boldsymbol{r}, v) - \epsilon, \tag{4.14}$$

$$I_f^\delta(\boldsymbol{r}, v, \boldsymbol{\alpha}) \geq I_f^\delta(\boldsymbol{r}, v) - \epsilon. \tag{4.15}$$

The proof of (4.12) and (4.13) can be referred to Lemma 5.43 in [26] and the proof of (4.14) and (4.15) is similar. We will outline the main procedure here.

Proof. If $\boldsymbol{r}$ is not absolutely continuous, $I_f(\boldsymbol{r}, v, \boldsymbol{\alpha}) = \infty$ by definition, so nothing needs to be proved. Now let us consider the case that $\boldsymbol{r}$ is absolutely continuous. For convenience, let $L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p}) := \langle \boldsymbol{p}, \boldsymbol{\beta} \rangle - H_s(\boldsymbol{z}, \boldsymbol{p}, \boldsymbol{w})$. Since by definition $L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p}) \leq L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$ for any $\boldsymbol{p}$, we have for B large enough

$$\int_0^T \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \geq B\}} L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot), \boldsymbol{\theta}(t)) dt$$
$$\leq \int_0^T \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \geq B\}} L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) dt$$
$$\leq \epsilon/4$$

by Lemma A.9. Choose $\boldsymbol{\theta}_1(t) = \boldsymbol{0}$ whenever $\|\dot{\boldsymbol{r}}(t)\| \geq B$ or $\dot{\boldsymbol{r}}(t)$ is not in $\mathscr{C}$ as defined in (2.12). Let $R := \sup_{0 \leq t \leq T} \|\boldsymbol{r}(t)\|$. Since $\boldsymbol{r}$ is continuous, R is finite. Simply replacing λ_i by $\sum_{j=1}^D \lambda_i(\boldsymbol{z}, j) w_j$ in Lemma 5.23 of [26], we have for B_1 large enough,

$$\sup_{|\boldsymbol{p}| \leq B_1} L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p}) \geq L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) - \frac{\epsilon}{8T}$$

for all $\|\boldsymbol{z}\| \leq R$, $\|\boldsymbol{\beta}\| \leq B$ in $\mathscr{C}$ and $\boldsymbol{w}$ in Δ_D. So for any $(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$ in bounded set

$$A := \{\|\boldsymbol{z}\| \leq R, \boldsymbol{\beta} \in \mathscr{C}, \|\boldsymbol{\beta}\| \leq B, \boldsymbol{w} \in \Delta_D\},$$

there exist a $\boldsymbol{p}_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}}$ with $\|\boldsymbol{p}_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}}\| \leq B_1$ such that

$$L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p}_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}}) \geq L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) - \frac{\epsilon}{4T}.$$

On the bounded set

$$\{\|\boldsymbol{z}\| \leq R, \boldsymbol{\beta} \in \mathscr{C}, \|\boldsymbol{\beta}\| \leq B, \boldsymbol{w} \in \Delta_D, \|\boldsymbol{p}\| \leq B_1\},$$

the function $L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p})$ is uniformly continuous. What's more, by Lemmas 5.22 and 5.33 in [26], $L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$ is continuous in A. So given any $(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) \in A$, there exist a $\delta_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}} > 0$ such that

$$L_s(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{w}}, \boldsymbol{p}_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}}) \geq L_s(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{w}}) - \frac{\epsilon}{2T}$$

holds for any $(\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{\beta}}, \tilde{)} \in O_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}} \cap A$, where $O_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}}$ is the $\delta_{\boldsymbol{z}\boldsymbol{\beta}\boldsymbol{w}}$-neighborhood of $(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$. By Heine-Borel theorem, we can choose finite number of $O_{\boldsymbol{z}^i \boldsymbol{\beta}^j \boldsymbol{w}^k}$ to cover A. It means that

$$L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{p}_{\boldsymbol{z}^i \boldsymbol{\beta}^j \boldsymbol{w}^k}) \geq L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) - \frac{\epsilon}{2T}$$

whenever $\|\boldsymbol{z} - \boldsymbol{z}^i\| + \|\boldsymbol{\beta} - \boldsymbol{\beta}^j\| + \|\boldsymbol{w} - \boldsymbol{w}^k\| \le \delta_{\boldsymbol{z}^i \boldsymbol{\beta}^j \boldsymbol{w}^k}$.

Define the function $\boldsymbol{\theta}_1(t) = \boldsymbol{p}_{\boldsymbol{z}^i \boldsymbol{\beta}^j \boldsymbol{w}^k}$ whenever $\|\boldsymbol{r}(t) - \boldsymbol{z}^i\| + \|\dot{\boldsymbol{r}}(t) - \boldsymbol{\beta}^j\| + \|n_v(t, \cdot) - \boldsymbol{w}^k\| \le \delta_{\boldsymbol{z}^i \boldsymbol{\beta}^j \boldsymbol{w}^k}$ with some tie-breaking rule. The function $\boldsymbol{\theta}_1(t)$ takes finite number of values. It may not be constant on intervals of time. So we approximate $\boldsymbol{\theta}_1(t)$ by a step function. Choose η small enough such that

$$\int_0^T \chi_{\{t \in A\}} L_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) dt \le \epsilon/4$$

whenever the set A has measure less than η. Since $\boldsymbol{\theta}_1(t)$ is a simple function, we can approximate it by a step function $\boldsymbol{\theta}$ and it agrees with $\boldsymbol{\theta}_1$ outside of a set of measure η (c.f. [25]).We finish the proof for (4.12) by collecting all approximations above.

For the proof for (4.14), we take advantage of Lemma 5.23 in [26] again by replacing λ_i with $\sum_{j=1}^{D} w_i q_{ij}$. We have for C large enough

$$\sup_{\|\boldsymbol{\sigma}\| \le C} S(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma}) \ge S(\boldsymbol{z}, \boldsymbol{w}) - \frac{\epsilon}{4T}.$$

On the bounded set

$$\{\|\boldsymbol{\sigma}\| \le C, \|\boldsymbol{z}\| \le R, \boldsymbol{w} \in \Delta_D\},$$

the function $S(\boldsymbol{z}, \boldsymbol{w}, \boldsymbol{\sigma})$ is uniformly continuous. With the similar strategy for L_s we can find the desired step function $\boldsymbol{\alpha}$. So we finish the proof for (4.14). The proof for I_s^δ and I_f^δ are similar. $\qquad\square$

A.2. **Part 2.** Proof of lemmas related to the lower bound estimate.

Proof of Lemma 3.11. Since $S(\boldsymbol{z}, \boldsymbol{w})$ is bounded by $Q := \sum_{i,j=1}^{D} \sup_{\boldsymbol{z}} q_{ij}(\boldsymbol{z})$, there exists $(\eta(s), \psi(s)) \in \mathscr{S}$ for any $s \in [t_m, t_{m+1}]$ such that

$$\sum_{i=1}^{D} \psi_i(s) \sum_{j=1}^{D} \left(\eta_{ij}(s) \log \frac{\eta_{ij}(s)}{q_{ij}(\boldsymbol{y}(s))} + q_{ij}(\boldsymbol{y}(s)) - \eta_{ij}(s) \right)$$
$$\le S(\boldsymbol{y}(s), n_v(s, \cdot)) + \epsilon$$

and

$$|\psi_i(s) - n_v(s, i)| < \epsilon/(8DT)$$

by Lemma 8.61 in [26]. For each fixed $s \in [t_m, t_{m+1}]$, there exists $\delta_s > 0$ such that

$$\sum_{i=1}^{D} \psi_i(s) \sum_{j=1}^{D} \left(\eta_{ij}(s) \log \frac{\eta_{ij}(s)}{q_{ij}(\boldsymbol{y}(s))} + q_{ij}(\boldsymbol{y}(s)) - \eta_{ij}(s) \right)$$
$$\le S(\boldsymbol{y}(t), n_v(t, \cdot)) + 2\epsilon$$

and

$$|\psi_i(s) - n_v(t, i)| < \epsilon/(4DT)$$

hold for any $t \in O_s = (s - \delta_s, s + \delta_s) \cap [t_m, t_{m+1}]$. By Heine-Borel theorem, we can choose finite number of O_{s_k} in $\{O_s\}_{s \in [t_m, t_{m+1}]}$ to cover $[t_m, t_{m+1}]$. It means that there exists a further subdivision of interval $[t_m, t_{m+1}]$ (i.e., $t_m = t_{m0} < t_{m1} < \cdots < t_{mK_m} = t_{m+1}$) and related $(\psi^m(s_k), \eta^m(s_k)) \in \mathcal{S}$ such that for all $t \in [t_{mk}, t_{m,k+1}]$

$$\sum_{i=1}^{D} \psi_i^m(s_k) \sum_{j=1}^{D} \left(\eta_{ij}^m(s_k) \log \frac{\eta_{ij}^m(s_k)}{q_{ij}(\mathbf{y}(t))} + q_{ij}(\mathbf{y}(t)) - \eta_{ij}^m(s_k) \right)$$
$$\leq \quad S(\mathbf{y}(t), n_v(t, \cdot)) + 2\epsilon.$$

Since $\log q_{ij}(\mathbf{z})$ are bounded and Lipschitz continuous in $\mathbf{z}$, we can establish that $S(\mathbf{z}, \mathbf{w})$ is absolutely continuous in $\mathbf{z}$, and this absolute continuity is uniform in $\mathbf{w} \in \Delta_D$. To see this, we only need to show that the function

$$f(\mathbf{x}, \mathbf{w}) := \sup_{\boldsymbol{\sigma} \in \mathbb{R}^D} \left(-\sum_{i,j=1}^{D} w_i x_{ij} \left(e^{\langle \boldsymbol{\sigma}, e_{ij} \rangle} - 1 \right) \right)$$

is absolutely continuous in $\mathbf{x} = (x_{11}, x_{12}, \ldots, x_{DD}) \in [1/\Lambda, \Lambda]^{D^2}$ (as defined in (2.11)), uniformly in $\mathbf{w} \in \Delta_D$. For any $\mathbf{x}, \mathbf{x} + \Delta \mathbf{x} \in [1/\Lambda, \Lambda]^{D^2}$ with $\|\Delta \mathbf{x}\| \leq 1/4\Lambda$, let $h = 1/4\Lambda$, $r = \|\Delta \mathbf{x}\| / (h + \|\Delta \mathbf{x}\|)$ and define $\mathbf{q} = \mathbf{x} + \Delta \mathbf{x}/r$. With this construction, we have $\mathbf{q} \in [1/2\Lambda, M + 1/2\Lambda]^{D^2}$, $f(\mathbf{x}, \mathbf{w})$, $f(\mathbf{x} + \Delta \mathbf{x}, \mathbf{w})$, $f(\mathbf{q}, \mathbf{w}) \in [0, (\Lambda + 1/2\Lambda)D^2]$ and $\mathbf{x} + \Delta \mathbf{x} = (1 - r)\mathbf{x} + r\mathbf{q}$. From the convexity of $f(\mathbf{x}, \mathbf{w})$ in $\mathbf{x}$, we have

$$f(\mathbf{x} + \Delta \mathbf{x}, \mathbf{w}) \leq (1 - r)f(\mathbf{x}, \mathbf{w}) + rf(\mathbf{q}, \mathbf{w})$$

and thus

$$f(\mathbf{x} + \Delta \mathbf{x}, \mathbf{w}) - f(\mathbf{x}, \mathbf{w}) \leq r(f(\mathbf{q}, \mathbf{w}) - f(\mathbf{x}, \mathbf{w})) \leq 4\Lambda \left(\Lambda + \frac{1}{2\Lambda} \right) D^2 \|\Delta \mathbf{x}\|.$$

The absolute continuity in $\mathbf{z}$ and uniformity in $\mathbf{w}$ of $S(\mathbf{z}, \mathbf{w})$ ensures that the estimate

$$S(\mathbf{y}(t), n_v(t, \cdot)) \leq S(\mathbf{r}(t), n_v(t, \cdot)) + \epsilon$$

holds when J is large enough.

To simplify the notation, we will rewrite $\eta^m(s_k)$ as η^{mk} and $\psi^m(s_k)$ as ψ^{mk}. So for each m, we have

$$\sum_{k=0}^{K_m-1} \int_{t_{mk}}^{t_{m,k+1}} \sum_{i=1}^{D} \psi_i^{mk} \sum_{j=1}^{D} \left(\eta_{ij}^{mk} \log \frac{\eta_{ij}^{mk}}{q_{ij}(\mathbf{y}(t))} + q_{ij}(\mathbf{y}(t)) - \eta_{ij}^{mk} \right) dt$$
$$\leq \int_{t_m}^{t_{m+1}} S(\mathbf{r}(t), n_v(t, \cdot)) dt + 3(t_{m+1} - t_m)\epsilon.$$

The proof is completed. $\qquad \qquad \square$

Proof of Lemma 3.12. Define

$$\tilde{f}(\boldsymbol{\mu}, \lambda(\boldsymbol{z}, \cdot), \boldsymbol{w}) := \sum_{i=1}^{S} \left(\sum_{j=1}^{D} \lambda_i(\boldsymbol{z}, j) w_j - \mu_i + \mu_i \log \frac{\mu_i}{\sum_{j=1}^{D} \lambda_i(\boldsymbol{z}, j) w_j} \right)$$

and

$$\tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) = \inf_{\boldsymbol{\mu} \in \mathcal{K}_{\boldsymbol{\beta}}} \tilde{f}(\boldsymbol{\mu}, \lambda(\boldsymbol{z}, \cdot), \boldsymbol{w}).$$

Taking advantage of Theorem 5.26 of [26], we have

$$\tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) = L_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}). \tag{4.16}$$

We will show that for any B_1, $\tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$ is continuous in $\boldsymbol{z}$ and $\boldsymbol{w}$, uniformly in $\boldsymbol{\beta}$ in

$$\mathcal{V} := \{\boldsymbol{\beta} \in \mathcal{C}, \|\boldsymbol{\beta}\| \le B_1\},$$

where $\mathcal{C}$ is the cone defined in (2.12).

By Lemma 5.20 of [26], we can find a constant B_2 such that for any $\boldsymbol{\beta} \in \mathcal{V}$ there exists a $\boldsymbol{\mu} \in \mathcal{K}_{\boldsymbol{\beta}}$ with $\|\boldsymbol{\mu}\| \le B_2$. Therefore, for all $\boldsymbol{\beta} \in \mathcal{V}$ and any $\boldsymbol{\mu} \in \mathcal{K}_{\boldsymbol{\beta}}$ with $\|\boldsymbol{\mu}\| \le B_2$,

$$\tilde{L}_s(\boldsymbol{z}', \boldsymbol{\beta}, \boldsymbol{w}') - \tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$$
$$\le \tilde{f}(\boldsymbol{\mu}, \lambda(\boldsymbol{z}', \cdot), \boldsymbol{w}') - \tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$$
$$\le \tilde{f}(\boldsymbol{\mu}, \lambda(\boldsymbol{z}, \cdot), \boldsymbol{w}') - \tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) + C_1 \|\boldsymbol{z}' - \boldsymbol{z}\|$$
$$\le \tilde{f}(\boldsymbol{\mu}, \lambda(\boldsymbol{z}, \cdot), \boldsymbol{w}) - \tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) + C_1 \|\boldsymbol{z}' - \boldsymbol{z}\| + C_2 \|\boldsymbol{w}' - \boldsymbol{w}\|$$

for some positive constants C_1 and C_2. Now choose $\boldsymbol{\mu}$ to minimize $\tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w})$ to establish that

$$\tilde{L}_s(\boldsymbol{z}', \boldsymbol{\beta}, \boldsymbol{w}') - \tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) \le C_1 \|\boldsymbol{z}' - \boldsymbol{z}\| + C_2 \|\boldsymbol{w}' - \boldsymbol{w}\|. \tag{4.17}$$

By Lemma 5.17 and Lemma 5.32 of [26] (replacing $\lambda_i(\boldsymbol{x})$ with $\sum_{j=1}^{D} \lambda_i(\boldsymbol{x}, j) w_j$), we know that there exist positive constants M_1, M_2 and B so that for all $\boldsymbol{\beta} \in \mathcal{C}$ with $\|\boldsymbol{\beta}\| \ge B$, all $\boldsymbol{z} \in \mathbb{R}^d$ and all $\boldsymbol{w} \in \Delta_D$,

$$M_1 \|\boldsymbol{\beta}\| \log \|\boldsymbol{\beta}\| \le \tilde{L}_s(\boldsymbol{z}, \boldsymbol{\beta}, \boldsymbol{w}) \le M_2 \|\boldsymbol{\beta}\| \log \|\boldsymbol{\beta}\|.$$

So for any $\boldsymbol{q} \in \mathbb{D}^d[0, T]$ and any $\tilde{\nu} \in \mathbb{M}[0, T]$,

$$\int_0^T \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \ge B\}} \tilde{L}_s(\boldsymbol{q}(t), \dot{\boldsymbol{r}}(t), n_{\tilde{\nu}}(t, \cdot)) dt$$
$$\le \int_0^T \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \ge B\}} M_2 \|\dot{\boldsymbol{r}}(t)\| \log \|\dot{\boldsymbol{r}}(t)\| dt$$
$$\le \int_0^T \chi_{\{\|\dot{\boldsymbol{r}}(t)\| \ge B\}} \frac{M_2}{M_1} \tilde{L}_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_\nu(t, \cdot)) dt$$
$$:= \epsilon(B) \tag{4.18}$$

By Lemma A.9, we have $\epsilon(B) \to 0$ as $B \to \infty$. Combining (4.17) and (4.18), we have for any $\epsilon > 0$, there exist a $\delta > 0$ such that

$$\sup_{0 \le t \le T} \|\boldsymbol{q}(t) - \boldsymbol{r}(t)\| < \delta \quad \text{and} \quad \sup_{0 \le t \le T} \|n_{\tilde{v}}(t, \cdot) - n_v(t, \cdot)\| < \delta$$

implies

$$\left| \int_0^T \tilde{L}_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) dt - \int_0^T \tilde{L}_s(\boldsymbol{q}(t), \dot{\boldsymbol{r}}(t), n_{\tilde{v}}(t, \cdot)) dt \right| \le \epsilon. \quad (4.19)$$

With this continuity property, we have

$$\int_0^T \tilde{L}_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) dt$$

$$= \sum_{m=0}^{J-1} \int_{t_m}^{t_{m+1}} \tilde{L}_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) dt$$

$$\ge \sum_{m=0}^{J-1} \int_{t_m}^{t_{m+1}} \tilde{L}_s(\boldsymbol{r}(t_m), \dot{\boldsymbol{r}}(t), n_\pi(t_m, \cdot)) dt - \epsilon$$

$$\ge \sum_{m=0}^{J-1} \Delta \cdot \tilde{L}_s\left(\boldsymbol{r}(t_m), \frac{\boldsymbol{r}(t_{m+1}) - \boldsymbol{r}(t_m)}{\Delta}, n_\pi(t_m, \cdot)\right) - \epsilon.$$

By definition of $\tilde{L}_s$, for each m, we have $\boldsymbol{\mu}^m \in K_{\boldsymbol{\beta}_m}$ such that

$$\tilde{L}_s\left(\boldsymbol{r}(t_m), \frac{\boldsymbol{r}(t_{m+1}) - \boldsymbol{r}(t_m)}{\Delta}, n_\pi(t_m, \cdot)\right)$$

$$\ge \sum_{i=1}^S \left(\lambda_i^\pi(\boldsymbol{y}(t_m)) - \mu_i^m + \mu_i^m \log \frac{\mu_i^m}{\lambda_i^\pi(\boldsymbol{y}(t_m))}\right) - \epsilon/T$$

and finally we have

$$\int_0^T \tilde{L}_s(\boldsymbol{r}(t), \dot{\boldsymbol{r}}(t), n_v(t, \cdot)) dt$$

$$\ge \sum_{m=0}^{J-1} \Delta \cdot \sum_{i=1}^S \left(\lambda_i^\pi(\boldsymbol{y}(t_m)) - \mu_i^m + \mu_i^m \log \frac{\mu_i^m}{\lambda_i^\pi(\boldsymbol{y}(t_m))}\right) - 2\epsilon$$

$$\ge \int_0^T \sum_{i=1}^S \left(\lambda_i^\pi(\boldsymbol{y}(t)) - \mu_i^m + \mu_i^m \log \frac{\mu_i^m}{\lambda_i^\pi(\boldsymbol{y}(t))}\right) dt - 3\epsilon. \quad (4.20)$$

Lemma 3.12 is proved by combing (4.16) and (4.20). $\qquad \square$

Proof of Lemma 3.13. We need to prove that for any bounded continuous function $h(t, z)$,

$$\lim_{n \to \infty} \int_0^T h(t, \bar{\xi}_n(t)) dt = \int_0^T \sum_{i=1}^D h(t, i) n_\pi(t, i) dt$$

in probability. It suffices to prove that for each time interval $[t_{mk}, t_{m,k+1}]$,

$$\lim_{n\to\infty} \int_{t_{mk}}^{t_{m,k+1}} h(t, \bar{\xi}_n(t)) dt = \int_{t_{mk}}^{t_{m,k+1}} \sum_{i=1}^{D} h(t, i) n_\pi(t, i) dt.$$

Since $\bar{\xi}_n$ lives on only finite states, then for any $\epsilon > 0$, there exists $\delta > 0$ such that for $|t_k - t| < \delta$

$$|h(t, \bar{\xi}_n(t)) - h(t_k, \bar{\xi}_n(t))| < \epsilon,$$

for all $t \in [t_{mk}, t_{m,k+1}]$.

Take an integer L large enough and define $\tilde{\delta} = (t_{m,k+1} - t_{mk})/L < \delta$. Let $\tau_l = t_{mk} + l\tilde{\delta}$ for $l = 0, 1, \ldots, L$. We have

$$\begin{aligned}
&\limsup_{n\to\infty} \int_{t_{mk}}^{t_{m,k+1}} h(t, \bar{\xi}_n(t)) dt \\
&= \limsup_{n\to\infty} \sum_{l=0}^{N-1} \int_{\tau_l}^{\tau_{l+1}} h(t, \bar{\xi}_n(t)) dt \\
&\leq \limsup_{n\to\infty} \sum_{l=0}^{N-1} \int_{\tau_l}^{\tau_{l+1}} h(\tau_l, \bar{\xi}_n(t)) dt + T\epsilon \\
&= \sum_{k=0}^{N-1} \int_{\tau_l}^{\tau_{l+1}} \sum_{i=1}^{D} h(\tau_l, i) n_\pi(t, i) dt + T\epsilon \qquad (4.21) \\
&\leq \sum_{k=0}^{N-1} \int_{\tau_l}^{\tau_{l+1}} \sum_{i=1}^{D} h(t, i) n_\pi(t, i) dt + 2T\epsilon \\
&= \int_{t_{mk}}^{t_{m,k+1}} \sum_{i=1}^{D} h(t, i) n_\pi(t, i) dt + 2T\epsilon.
\end{aligned}$$

In (4.21) we utilized the ergodicity of the process $\bar{\xi}_n$ on each interval $[\tau_l, \tau_{l+1}]$. The convergence can be obtained in the almost sure and $L_{\mathbb{P}}^1$-sense rather than in probability [6]. Similarly, we can prove

$$\liminf_{n\to\infty} \int_0^T h(t, \bar{\xi}_n(t)) dt \geq \int_{t_{mk}}^{t_{m,k+1}} \sum_{i=1}^{D} h(t, i) n_\pi(t, i) dt - 2T\epsilon.$$

The proof is completed. $\qquad\square$

Proof of Lemma 3.14. The goal is to prove that for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{0 \leq t \leq T} \|\bar{z}_n(t) - y(t)\| \geq \epsilon \right) = 0.$$

For any $\boldsymbol{p} \in \mathbb{R}^d$ and $\rho > 0$, we have the martingale

$$
\begin{aligned}
M_t &= \exp\Big\{ \langle \bar{\boldsymbol{z}}_n(t) - \boldsymbol{y}(t), \rho\boldsymbol{p} \rangle \\
&\qquad - \int_0^t \sum_{i=1}^S \Big(n\mu_i(s) \frac{\lambda_i(\bar{\boldsymbol{z}}_n(s), \bar{\boldsymbol{\xi}}_n(s))}{\lambda_i^\pi(\boldsymbol{y}(s))} (e^{\langle \rho\boldsymbol{p}, \boldsymbol{u}_i/n \rangle} - 1) \\
&\qquad - \mu_i(s) \langle \rho\boldsymbol{p}, \boldsymbol{u}_i \rangle \Big) ds \Big\} \\
&= \exp\Big\{ \langle \bar{\boldsymbol{z}}_n(t) - \boldsymbol{y}(t), \rho\boldsymbol{p} \rangle \\
&\qquad - \int_0^t \sum_{i=1}^S \Big(\mu_i(s) \frac{\lambda_i(\bar{\boldsymbol{z}}_n(s), \bar{\boldsymbol{\xi}}_n(s)) - \lambda_i^\pi(\boldsymbol{y}(s))}{\lambda_i^\pi(\boldsymbol{y}(s))} \langle \rho\boldsymbol{p}, \boldsymbol{u}_i \rangle \\
&\qquad + \mu_i(s) \frac{\lambda_i(\bar{\boldsymbol{z}}_n(s), \bar{\boldsymbol{\xi}}_n(s))}{\lambda_i^\pi(\boldsymbol{y}(s))} \big(n(e^{\langle \rho\boldsymbol{p}, \boldsymbol{u}_i/n \rangle} - 1) - \langle \rho\boldsymbol{p}, \boldsymbol{u}_i \rangle \big) \Big) ds \Big\}.
\end{aligned}
$$

Recall the Assumption 2.3 and $\mu_i(t)$ is piecewise constant and bounded, we can perform similar estimate as in Lemma A.5 to obtain

$$
\begin{aligned}
&\mathbb{P}\Big(\sup_{0 \le t \le T} \Big\| \bar{\boldsymbol{z}}_n(t) - \boldsymbol{y}(t) - \int_0^t \sum_{i=1}^S \mu_i(s) \frac{\lambda_i(\bar{\boldsymbol{z}}_n(s), \bar{\boldsymbol{\xi}}_n(s)) - \lambda_i^\pi(\boldsymbol{y}(s))}{\lambda_i^\pi(\boldsymbol{y}(s))} ds\, \boldsymbol{u}_i \Big\| \ge \epsilon \Big) \\
&\le \exp\Big(- n\epsilon c_1 \log(\epsilon c_2) \Big),
\end{aligned}
\tag{4.22}
$$

where c_1 and c_2 are positive constants. By Lemma 3.13, we have

$$
\begin{aligned}
&\int_0^t \lambda_i(\bar{\boldsymbol{z}}_n(s), \bar{\boldsymbol{\xi}}_n(s)) ds - \int_0^t \lambda_i^\pi(\boldsymbol{y}(s)) ds \\
&= \Big(\int_0^t \lambda_i(\bar{\boldsymbol{z}}_n(s), \bar{\boldsymbol{\xi}}_n(s)) ds - \int_0^t \lambda_i(\boldsymbol{y}(s), \bar{\boldsymbol{\xi}}_n(s)) ds \Big) \\
&\quad + \Big(\int_0^t \sum_{j=1}^D \lambda_i(\boldsymbol{y}(s), j) n_{\bar{v}_n}(s, j) ds - \int_0^t \lambda_i^\pi(\boldsymbol{y}(s)) ds \Big) \\
&\le K \int_0^t \| \bar{\boldsymbol{z}}_n(s) - \boldsymbol{y}(s) \| ds + B_n,
\end{aligned}
\tag{4.23}
$$

where

$$
B_n = \sup_{t \in [0,T]} \Big| \sum_{j=1}^D \int_0^t \lambda_i(\boldsymbol{y}(s), j) \Big(n_{\bar{v}_n}(s, j) - n_\pi(s, j) \Big) ds \Big| \to 0
\tag{4.24}
$$

as n goes to infinity for $t \le T$.

Define $C = d\,AUT\Lambda$, where

$$
A = \max_{t \in [0,T]} \max_{i=1,\dots,S} \mu_i(t) \quad \text{and} \quad U = \max_{i=1,\cdots,S} \| \boldsymbol{u}_i \|.
$$

Combining (4.22), (4.23) and (2.11), we have

$$
\mathbb{P}\left(\sup_{0\le t\le T}\left(\left\|\bar{\boldsymbol{z}}_n(t)-\boldsymbol{y}(t)\right\|-CK\int_0^t\left\|\bar{\boldsymbol{z}}_n(s)-\boldsymbol{y}(s)\right\|ds-CB_n\right)\ge\epsilon\right)
$$

$$
\le\mathbb{P}\left(\sup_{0\le t\le T}\left\|\bar{\boldsymbol{z}}_n(t)-\boldsymbol{y}(t)-\int_0^t\sum_{i=1}^S\mu_i(t)\frac{\lambda_i(\bar{\boldsymbol{z}}_n(s),\bar{\xi}_n(s))-\lambda_i^\pi(\boldsymbol{y}(s))}{\lambda_i^\pi(\boldsymbol{y}(s))}ds\,\boldsymbol{u}_i\right\|\ge\epsilon\right)
$$

$$
\le\exp\left(-n\epsilon c_1\log(\epsilon c_2)\right). \tag{4.25}
$$

From (4.25) and Gronwall's inequality, we obtain

$$
\mathbb{P}\left(\sup_{0\le t\le T}\|\bar{\boldsymbol{z}}_n(t)-\boldsymbol{y}(t)\|\ge(\epsilon+CB_n)e^{CKT}\right)
$$

$$
\le\quad\mathbb{P}\left(\sup_{0\le t\le T}\left(\|\bar{\boldsymbol{z}}_n(t)-\boldsymbol{y}(t)\|-CK\int_0^t\|\bar{\boldsymbol{z}}_n(s)-\boldsymbol{y}(s)\|ds\right)\ge\epsilon\right)
$$

$$
\le\quad\exp\left(-n\epsilon c_1\log(\epsilon c_2)\right).
$$

Combing the condition (4.24) and the inequality

$$
\mathbb{P}\left(\sup_{0\le t\le T}\|\bar{\boldsymbol{z}}_n(t)-\boldsymbol{y}(t)\|\ge2\epsilon e^{CKT}\right)
$$

$$
\le\quad\mathbb{P}\left(\sup_{0\le t\le T}\|\bar{\boldsymbol{z}}_n(t)-\boldsymbol{y}(t)\|\ge(\epsilon+CB_n)e^{CKT}\right)+\mathbb{P}\left(B_n\ge\frac{\epsilon}{C}\right),
$$

we finish the proof. $\qquad\square$

REFERENCES

[1] M. Assaf, E. Roberts, and Z. Luthey-Schulten, *Determining the Stability of Genetic Switches: Explicitly Accounting for mRNA Noise*, Phys. Rev. Lett. 106 (2011), 2048102.

[2] P. Billingsley, *Convergence of probability measures*, John Wiley and Sons, New York, 1999.

[3] M. Chen and Y. Lu, *On evaluating the rate functions of large deviations for jump processes*, Acta. Math. Sin. 6 (1990), 206-219.

[4] Y. Chen, C. Lv, F. Li and T. Li, *Distinguishing the rates of gene activation from phenotypic variations*, BMC Syst. Biol. 9(2015), 29.

[5] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer-Verlag, New York, 2nd Edition, 1998.

[6] R. Durrett, *Probability: Theory and Examples*, 4th edition, Cambridge University Press, Cambridge, 2010.

[7] W. E, D. Liu and E. Vanden-Eijnden, Analysis of multiscale methods for stochastic differential equations, Comm. Pure App. Math. 58 (2005), 1544-1585.

[8] W. E, D. Liu and E. Vanden-Eijnden, Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales, J. Comp. Phys. 221 (2007), 158-180.

[9] R.S. Ellis, *Entropy, large deviations and statistical mechanics*, Springer-Verlag, New York, 1985.

[10] M. Elowitz, A. Levine, E. Siggia and P. Swain, *Stochastic gene expression in a single cell*, Science 297 (2002), 1183-1186.

[11] S.N. Ethier and T.G. Kurtz, *Markov processes: characterization and convergence*, John Wiley and Sons, New York, 1986.

[12] M.I. Freidlin and A.D. Wentzel, *Random perturbations of dynamical systems*, Springer-Verlag, New York, 2nd Edition, 1998.

[13] D.T. Gillespie, *Exact stochastic simulation of coupled chemical reactions*, J. Phys. Chem. 81 (1977), 2340-2361.

[14] D.T. Gillespie, *The chemical Langevin equation*, J. Chem. Phys. 113 (2000), 297-306.

[15] M. Heymann and E. Vanden-Eijnden, *The geometric minimum action method: A least action principle on the space of curves*, Comm. Pure Appl. Math. 61 (2008), 1052-1117.

[16] H. Kang and T.G. Kurtz, *Separation of time-scales and model reduction for stochastic reaction networks*, Ann. Appl. Probab. 23 (2013), 529-583.

[17] T.G. Kurtz, *The relationship between stochastic and deterministic models for chemical reactions*, J. Chem. Phys. 57 (1972) , 2976-2978.

[18] T.G. Kurtz, *A limit theorem for perturbed operator semigroups with applications to random evolutions*, J. Funct. Anal. 12 (1973), 55-67.

[19] T. Li and F. Lin, *Two-scale large deviations for chemical reaction kinetics through second quantization path integral*, J. Phys. A: Math. Theor. 49 (2016), 135204.

[20] R. Liptser, *Large deviations for two scaled diffusions* , Prob. Theory Relat. Fields 106 (1996), 71-104.

[21] C. Lv, X. Li, F. Li and T. Li, *Constructing the energy landscape for genetic switching system driven by intrinsic noise*, PLoS ONE 9 (2014), e88167.

[22] C. Lv, X. Li, F. Li and T. Li, *Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process*, PLoS Comput. Biol. 11 (2015), e1004156.

[23] G. Papanicolaou, *Introduction to the asymptotic analysis of stochastic equations*, Lectures in Mathematics, Vol. 16, American Mathematical Society, Rode Island, 1977.

[24] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[25] H.L. Royden, *Real Analysis*, Macmillan, New York, 2nd Edition, 1968.

[26] A. Shwartz and A. Weiss, *Large deivations for performance analysis: queues, communications and computing*, Chapman and Hall, London, 1995.

[27] A. Shwartz and A. Weiss, *Large deviations with diminishing rates*, Math. Oper. Res. 30 (2005), 281-310.

[28] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, Amsterdam, 3rd Edition, 2007.

[29] A.Yu. Veretennikov, *On large deviations for SDEs with small diffusion and averaging*, Stoch. Process. Appl. 89 (2000), 69-79.

[30] A.Yu. Veretennikov, *On large deviations in the averaging principle for SDE's with a "full dependence"*, Ann. Prob. 27 (1999), 284-296.

[31] K. Zhang, M. Sasai and J. Wang, *Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics*, Proc. Nat. Acad Sci. USA 110 (2013), 14930-14935.

[32] P. Zhou and T. Li, *Construction of the landscape for multi-stable systems: potential landscape, quasi-potenital, A-type integral and beyond*, J. Chem. Phys. 144 (2016), 94109.