**Radboud Repository**

Radboud University Nijmegen

# PDF hosted at the Radboud Repository of the Radboud University Nijmegen

# Large deviations for Markov-modulated diffusion processes with rapid switching

Gang Huang, Michel Mandjes, Peter Spreij

January 14, 2015

### Abstract

In this paper, we study small noise asymptotics of Markov-modulated diffusion processes in the regime that the modulating Markov chain is rapidly switching. We prove the joint sample-path large deviations principle for the Markov-modulated diffusion process and the occupation measure of the Markov chain (which evidently also yields the large deviations principle for each of them separately by applying the contraction principle). The structure of the proof is such that we first prove exponential tightness, and then establish a local large deviations principle (where the latter part is split into proving the corresponding upper bound and lower bound).

*Address.* Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands.

*Email.* {g.huang|m.r.h.mandjes|spreij}@uva.nl

## 1   Introduction

The setting studied in this paper is the following. We consider a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with a filtration $\{\mathscr{F}_t\}_{t \in \mathbb{R}_+}$, where $\mathbb{R}_+ := [0, +\infty)$. $\mathscr{F}_0$ contains all the $\mathbb{P}$-null sets of $\mathscr{F}$, and $\{\mathscr{F}_t\}_{t \in \mathbb{R}_+}$ is right continuous. Let $X_t$ be a finite-state time-homogeneous Markov chain with transition intensity matrix $Q$ and state space $\mathbb{S} := \{1, \cdots, d\}$ for some $d \in \mathbb{N}$. The Markov-modulated diffusion process is defined as the unique solution to

$$M_t = M_0 + \int_0^t b(X_s, M_s)\mathrm{d}s + \int_0^t \sigma(X_s, M_s)\mathrm{d}B_s,$$

where $B_t$ is a standard Brownian motion. We assume that there exist $i, x$ such that $\sigma(i, x) \neq 0$ throughout this paper. The concept of Markov modulation is also known as 'regime switching'; the Markov chain $X_t$ is often referred to as the 'background process', or the 'modulating Markov chain'.

The objective of this paper is to study the above stochastic differential equation under a particular parameter scaling. For a strictly positive (but typically small) $\epsilon$, we scale $Q$ to $Q/\epsilon =: Q^\epsilon$, and denote by $X_t^\epsilon$ the Markov chain with this transition intensity matrix $Q^\epsilon$. If

the expected number of jumps per unit time is $y$ for $X_t$, then the time-scaling entails that it is $y/\epsilon$ for $X_t^\epsilon$. One could therefore say that the Markov chain has been sped up by a factor $\epsilon^{-1}$, and, as a consequence, $X_t^\epsilon$ switches rapidly among its states when $\epsilon$ is small. A classical topic in large deviations theory, initiated by Freidlin and Wentzell [9], concerns small-noise large deviations. In this paper, we investigate how rapid-switching behavior of $X_t^\epsilon$ affects the small-noise asymptotics of $X_t^\epsilon$-modulated diffusion processes on the interval $[0,T]$ (for any fixed strictly positive $T$).

Let us make the scaling regime considered more concrete now. Importantly, it concerns a scaling of the function $\sigma(\cdot,\cdot)$ to $\sqrt{\epsilon}\sigma(\cdot,\cdot)$ in the Markov-modulated diffusion, but at the same time we speed up the Markovian background process in the way we described above. The resulting process $M_t^\epsilon$ is defined as the unique strong solution to

$$M_t^\epsilon = M_0^\epsilon + \int_0^t b(X_s^\epsilon, M_s^\epsilon)\mathrm{d}s + \sqrt{\epsilon}\int_0^t \sigma(X_s^\epsilon, M_s^\epsilon)\mathrm{d}B_s, \tag{1}$$

where we recall that $X_t^\epsilon$ has transition intensity matrix $Q^\epsilon$. Focusing on the regime that $\epsilon \to 0$, we call in the sequel $M_t^\epsilon$ the Markov-modulated diffusion process with rapid switching. For simplicity, we will assume throughout this paper that $M_0^\epsilon \equiv 0$, whereas $X_0^\epsilon$ starts at an arbitrary $x \in \mathbb{S}$, for all $\epsilon$. When we write e.g. $\mathbb{E}[M_t^\epsilon]$, this is to be understood as the expectation of $M_t^\epsilon$ with the above initial conditions.

Since $M_t^\epsilon$ evolves in the random environment of $X_t^\epsilon$, we need to design a coupling to separate the effects of the vanishing of the diffusion term and the fast varying of the Markov chain, but at the same time to keep track of both of them. Since the scaling $Q$ to $Q/\epsilon$ is equivalent to speeding up time by a factor $\epsilon^{-1}$, one could informally say that $X_t^\epsilon$ relates to a faster time scale than $M_t^\epsilon$, and therefore essentially exhibits stationary behavior 'around' this specific $t$. Then it is custom to consider the occupation measure of $X_t^\epsilon$, which is defined on $\Omega \times [0,T] \times \mathbb{S}$ as

$$\nu^\epsilon(\omega; t, i) = \int_0^t \mathbf{1}_{\{X_s^\epsilon(\omega)=i\}}\mathrm{d}s. \tag{2}$$

As its name suggests, $\nu^\epsilon(\cdot; T, i)$ measures the time $X_t^\epsilon$ spends in state $i$ during the time interval $[0,T]$. Moreover, we can use the derivative of $\nu^\epsilon(t)$ to gauge the infinitesimal change of the occupation measure of $X_t^\epsilon$, at any $t \in [0,T]$. We thus construct a coupling $(M^\epsilon, \nu^\epsilon)$, which is the main object studied in this paper.

A celebrated result in Donsker and Varadhan [6] concerns the large deviations principle (LDP) for $\nu^1(\omega; t, \cdot)/t$ as $t \to \infty$ (i.e., the LDP of the fraction of time spent in the individual states of the background process). The setting of the present paper, however, involves the *sample-path* LDP for $\nu^\epsilon$ on $[0,T]$ as $\epsilon \to 0$. More precisely, we define the image space $\mathbb{M}_T$ of $\nu^\epsilon$ restricted on $[0,T]$ as the space of functions $\nu$ on $[0,T] \times \mathbb{S}$ satisfying $\nu(t,i) = \int_0^t K_\nu(s,i)\mathrm{d}s$, where $\sum_{i=1}^d K_\nu(s,i) = 1$, $K_\nu(s,i) \geqslant 0$ for every $i \in \mathbb{S}, s \in [0,T]$, and $K_\nu(s,i)$ being Borel measurable with respect to $s$; $K_\nu$ is referred to as the *kernel* of $\nu$. The metric on $\mathbb{M}_T$ is defined as

$$d_T(\mu, \nu) = \sup_{0 \leqslant t \leqslant T, i \in \mathbb{S}} \left| \int_0^t K_\mu(s,i)\mathrm{d}s - \int_0^t K_\nu(s,i)\mathrm{d}s \right|.$$

We can also view $\mathbb{M}_T$ as a subset of $\mathbb{C}_{[0,T]}(\mathbb{R}^d)$ which is the space of $\mathbb{R}^d$-valued continuous functions on $[0,T]$. In addition, the metric $d_T$ on $\mathbb{M}_T$ is equivalent to the uniform metric on $\mathbb{C}_{[0,T]}(\mathbb{R}^d)$.

We also define $\mathbb{C}_T$ as the image space of $M^\epsilon$, which is the space of functions $f \in \mathbb{C}_{[0,T]}(\mathbb{R})$ and $f(0) = 0$ equipped with the uniform metric $\rho_T(f, g) := \sup_{0 \leqslant t \leqslant T} |f(t) - g(t)|$. The product metric $\rho_T \times d_T$ on $\mathbb{C}_T \times \mathbb{M}_T$ is defined by

$$(\rho_T \times d_T)((\varphi, \nu), (\varphi', \nu')) := \rho_T(\varphi, \varphi') + d_T(\nu, \nu'), \quad \forall (\varphi, \nu), (\varphi', \nu') \in \mathbb{C}_T \times \mathbb{M}_T.$$

We denote by $\mathscr{B}(\mathbb{C}_T \times \mathbb{M}_T)$ the Borel $\sigma$-algebra generated by the topology induced by $\rho_T \times d_T$.

The main result of this paper is the *joint* sample-path LDP for $(M^\epsilon, \nu^\epsilon)$ on $\mathbb{C}_T \times \mathbb{M}_T$. The associated (joint) large deviations rate function is obtained in quite an explicit form. It is actually the sum of two expressions that we introduce later in this paper, viz. (6), i.e., the rate function $I_T(\varphi, \nu)$ corresponding to $M^\epsilon$, and (5), i.e., the rate function $\tilde{I}_T(\nu)$ corresponding to $\nu^\epsilon$. Informed readers will recognize that these rate functions are variants of those for diffusion processes, as given in e.g. Freidlin and Wentzell [9], and for occupation measures of Markov processes, as given in e.g. Donsker and Varadhan [6] (where we remark again that the result in [6] relates to $\nu^1(\omega; t, \cdot)/t$ for $t$ large, whereas our statement concerns the sample paths of $\nu^\epsilon$).

One method of proving the LDP for a family of probability measures on a metric space, as was introduced in the seminal papers of Liptser and Pukhalskii [19] and Liptser [18], is to first prove exponential tightness, and then the local LDP (precise definitions of these notions will be given in the next section). Our work by and large follows this approach. Importantly, the model considered in Liptser [18] is similar to ours, in that it also studies the stochastic differential equation (1), but in the setup of Liptser [18] the process $X_t^\epsilon$ is another diffusion process (rather than a finite-state Markov chain). It means that we can roughly follow the structure of the proof presented in [18] (we also rely on the method of stochastic exponentials, for instance), but there are crucial differences at many places. For instance, as we point out below, there are several novelties that have the potential of being used in other settings, too.

One of the methodological novelties is the following. We explore a nice connection between regularity properties of the rate function $\tilde{I}_T(\nu)$ in the LDP for $(M^\epsilon, \nu^\epsilon)$ and a dense subset of the image space $\mathbb{M}_T$ of $\nu^\epsilon$. On this dense subset, the optimizer of the integrand of $\tilde{I}_T(\nu)$ is infinitely differentiable. This eliminates many difficulties in the computation and leads us to first prove the local LDP on a dense subset of $\mathbb{C}_T \times \mathbb{M}_T$. We then extend the local LDP to $\mathbb{C}_T \times \mathbb{M}_T$ by continuity properties of the rate functions $I_T(\varphi, \nu)$ and $\tilde{I}_T(\nu)$.

Let $\mathbb{U}$ denote the space of functions on $[0, T] \times \mathbb{S}$ being continuously differentiable on $[0, T]$ and $\inf_{s \in [0,T], i \in \mathbb{S}} u(s, i) > 0$. In our analysis in Section 6, we identify the following stochastic exponential which is directly related to the Markov chain $X_t^\epsilon$ and its rate function $\tilde{I}_T(\nu)$ (as given in (5)):

$$\frac{u(t, X_t^\epsilon)}{u(0, X_0^\epsilon)} \exp\left(-\int_0^t \frac{\frac{\partial}{\partial s} u(s, X_s^\epsilon) + (Q^\epsilon u)(s, X_s^\epsilon)}{u(s, X_s^\epsilon)} \mathrm{d}s\right), \quad u \in \mathbb{U},$$

which plays a key role when proving the local LDP. Here we follow the notational convention that $(Q^\epsilon u)(s, i) = \sum_{j=1}^d Q_{ij}^\epsilon u(s, j)$, for $i \in \mathbb{S}$.

As mentioned above, the main result of our paper is the *joint* sample-path LDP for $(M^\epsilon, \nu^\epsilon)$. The LDPs for each component $M^\epsilon$ and $\nu^\epsilon$ are then derived as corollaries from our main result in the standard way, i.e., by an application of the contraction principle. The small noise LDP for the Markov-modulated diffusion processes (which is $M^\epsilon$ alone) is also

3

studied in a newly published paper by He and Yin [12] in a setting of multi-dimensional processes and time-depending transition intensity matrices. In our corresponding result, which is Corollary 3.2, the rate function for $M^\epsilon$ is decomposed into two parts that allow an appealing interpretation: the first part corresponds to the rare behavior of the background process $X^\epsilon$, where the second part corresponds to the rare behavior of $M^\epsilon$ (conditional on the rare behavior of $X^\epsilon$). The rate function in He and Yin [12] is less explicit, in that it is expressed in terms of an H-functional in which the aforementioned two parts cannot be distinguished. The sample-path LDP for occupation measures of rapid switching Markov chain (which is $\nu^\epsilon$ alone) is obtained in Theorem 5.1 in He et al. [13]. The rate function, which is also expressed in terms of an H-functional, coincides with the rate function in our LDP for $\nu^\epsilon$ (Corollary 3.3) when the transition intensity matrix is time-homogeneous. However, focusing on obtaining the LDP for the Markov-modulated diffusion process together with the background process, our aim and approach in this paper are entirely different from theirs.

The large-deviations analysis for stochastic processes with Markov-modulation is a currently active research field. Besides the previously mentioned papers of He et al. [13] and He and Yin [12], we list a few more. Guillin [10] proved the averaging principle (moderate deviations) of Equation (1) where $X_t^\epsilon$ is an exponentially ergodic Markov process and $b, \sigma$ are bounded functions. He and Yin [11] studied the moderate-deviations behavior of $M_t^\epsilon$ in Equation (1), where $\sigma \equiv 0$ and $X_t^\epsilon$ is a non-homogeneous Markov chain with two time-scales. Lasry and Lions [17] and Fournié et al. [8] considered large deviations for the hitting times of Markov-modulated diffusion processes with rapid switching.

Interestingly, the present paper relates to our previous work [14]. For ease ignoring the initial position, we there considered the Markov-modulated diffusion $\check{M}_t^\epsilon$ described by

$$\check{M}_t^\epsilon = \int_0^t b(X_s^\epsilon, \check{M}_s^\epsilon)\mathrm{d}s + \int_0^t \sigma(X_s^\epsilon, \check{M}_s^\epsilon)\mathrm{d}B_s.$$

In the regime $\epsilon \to 0$ the solutions of the stochastic differential equation converge weakly to a (non-modulated) diffusion $\check{M}_t$ satisfying, with $\pi$ denoting the stationary distribution of $X_t^\epsilon$ (and hence also of $X_t$),

$$\check{M}_t = \int_0^t \sum_{i=1}^d b(i, \check{M}_s)\pi(i)\mathrm{d}s + \int_0^t \left(\sum_{i=1}^d \sigma^2(i, \check{M}_s)\pi(i)\right)^{1/2} \mathrm{d}B_s.$$

This result shows that, when the background chain switches rapidly, it is hard to distinguish from observed data a Markov-modulated diffusion process from an 'ordinary' diffusion. The work in the present paper, in contrast, indicates that no such property carries over to the large deviations. The impact of a fast switching background chain does appear in the small noise asymptotics, as shown in the LDPs in this paper.

We now describe the organization of our paper. The structure of the paper is as follows. In Section 2, we introduce some preliminary results, definitions, and notation. In Section 3, we state the paper's main result and explain the steps of its proof. In Section 4, exponential tightness of $(M^\epsilon, \nu^\epsilon)$ is verified. We identify a dense subset of $\mathbb{C}_T \times \mathbb{M}_T$ in Section 5, and explore regularity properties of the rate function on it. The upper bound and lower bound of the local LDP for $(M^\epsilon, \nu^\epsilon)$ are proved in Sections 6 and 7, respectively. We present a number of technical lemmas in the appendix.

# 2 Preliminaries

In this section we first provide the definitions of the LDP, exponential tightness and the local LDP, and state a set of related theorems that are relevant in the context of the paper. Let $\mathbb{X}$ throughout denote a Polish space with Borel $\sigma$-algebra $\mathscr{B}(\mathbb{X})$ and a metric $\rho$.

**Definition 2.1** (Varadhan [26]) *A family of probability measures $\mathbb{P}^\epsilon$ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is said to obey the LDP with a rate function $I(\cdot)$ if there exists a function $I(\cdot) : \mathbb{X} \to [0, \infty]$ satisfying:*
*(1) There exists $x \in \mathbb{X}$ such that $I(x) < \infty$; $I$ is lower semicontinuous; for every $c < \infty$ the set $\{x : I(x) \leqslant c\}$ is a compact set in $\mathbb{X}$.*
*(2) For every closed set $F \subset \mathbb{X}$, $\limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}^\epsilon(F) \leqslant -\inf_{x \in F} I(x)$.*
*(3) For every open set $O \subset \mathbb{X}$, $\liminf_{\epsilon \to 0} \epsilon \log \mathbb{P}^\epsilon(O) \geqslant -\inf_{x \in O} I(x)$.*

**Definition 2.2** (Den Hollander [5], Puhalskii [24]) *A family of probability measures $\mathbb{P}^\epsilon$ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is said to be exponentially tight, if for every $L < \infty$, there exists a compact set $K_L \subset \mathbb{X}$ such that*

$$\limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}^\epsilon(\mathbb{X} \setminus K_L) \leqslant -L.$$

**Definition 2.3** (Puhalskii [24], Liptser and Puhalskii [18]) *A family of probability measures $\mathbb{P}^\epsilon$ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is said to obey the local LDP with a rate function $I(\cdot)$ if for every $x \in \mathbb{X}$*

$$\limsup_{\delta \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}^\epsilon(\{y \in \mathbb{X} : \rho(x, y) \leqslant \delta\}) \leqslant -I(x), \tag{3}$$

$$\liminf_{\delta \to 0} \liminf_{\epsilon \to 0} \epsilon \log \mathbb{P}^\epsilon(\{y \in \mathbb{X} : \rho(x, y) \leqslant \delta\}) \geqslant -I(x). \tag{4}$$

Since $\mathbb{X}$ is a Polish space, Definition 2.1.(1) implies exponential tightness. Also, Definition 2.1.(2)–(3) guarantee that $\mathbb{P}^\epsilon$ satisfies the local LDP. Actually, the converse is also valid and is the key to prove our main result.

**Theorem 2.4** (Puhalskii [24], Liptser and Puhalskii [18]) *If a family of probability measures $\mathbb{P}^\epsilon$ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is exponentially tight and obeys the local LDP with a rate function $I$, then it obeys the LDP with the rate function $I$.*

The following lemma, which corresponds to Lemma 1.4 in Borovkov and Mogulskiĭ [2], shows that a local LDP on a dense subset of $\mathbb{X}$ is enough for the validation of the local LDP on $\mathbb{X}$, provided the rate function possesses a regularity property.

**Lemma 2.5** *(i) If (3) is fulfilled for all $\tilde{x} \in \tilde{\mathbb{X}}$, where $\tilde{\mathbb{X}}$ is dense in $\mathbb{X}$ and function $I(x)$ is lower semi-continuous, then it holds for all $x \in \mathbb{X}$.*
*(ii) If for every $x \in \mathbb{X}$ with $I(x) < \infty$ there exists a sequence $\tilde{x}_n \in \tilde{\mathbb{X}}$ converging to $x$ and $I(\tilde{x}_n) \to I(x)$, then the fullfillment of (4) for $\tilde{x} \in \tilde{\mathbb{X}}$ implies the same for all $x \in \mathbb{X}$.*

Next we impose some assumptions on the stochastic differential equation (1), as was defined in the introduction. It is noted that (A.1) ('Lipschitz continuity') implies (A.2) ('linear growth'); we chose to include (A.2) as well, however, for ease reference in later sections.

(A.1) *Lipschitz continuity*: there is a positive constant $K$ such that

$$|b(i, x) - b(i, y)| + |\sigma(i, x) - \sigma(i, y)| \leqslant K|x - y|, \quad \forall i \in \mathbb{S}, \ \ x, y \in \mathbb{R}.$$

(A.2) *Linear growth*: there exists a positive constant $K$ (which might be different from the $K$ used in (A.1)) such that

$$|b(i, x)| + |\sigma(i, x)| \leqslant K(1 + |x|), \quad \forall i \in \mathbb{S}, \ \ x \in \mathbb{R}.$$

(A.3) *Independence*: the Markov chain $X_t^\epsilon$ is independent of the Brownian motion $B_t$ for all $\epsilon$.

(A.4) *Irreducibility*: the off-diagonal entries of the transition intensity matrix $Q$ are strictly positive. Hence, the Markov chain $X_t^\epsilon$ is irreducible for all $\epsilon$ and has an invariant probability measure $\pi = (\pi(1), \cdots, \pi(d))$.

Finally, we introduce some extra notation and function spaces. For an arbitrary stochastic process or a function $Y_t$, we denote the running maximum process by $Y_t^* := \sup_{s \leqslant t} |Y_s|$. For a semimartingale $Y_t$ such that $Y_0 = 0$, its *stochastic exponential* is defined as a semimartingale $\mathscr{E}(Y)_t$ which is the unique strong solution to

$$\mathscr{E}(Y)_t = 1 + \int_0^t \mathscr{E}(Y)_{s-} \mathrm{d}Y_s.$$

We denote $\mathbb{H}_T$ the Cameron-Martin space of functions $\varphi \in \mathbb{C}_T$ such that $\varphi(t) = \int_0^t \varphi'(s)\mathrm{d}s$ and $\varphi'$ is square-integrable on $[0, T]$. We call $\varphi'$ the derivative of $\varphi$.

## 3 Main results

We first introduce the definitions of the rate functions involved in the main result. The rate function corresponding to $\nu^\epsilon$ is defined as

$$\tilde{I}_T(\nu) := \int_0^T \sup_{u \in U} \left[ -\sum_{i=1}^d \frac{(Qu)(i)}{u(i)} K_\nu(s, i) \right] \mathrm{d}s, \quad \nu \in \mathbb{M}_T, \tag{5}$$

where we recall the notation $(Qu)(i) = \sum_{j=1}^d Q_{ij} u(j)$, for $i \in \mathbb{S}$, and $U$ denotes the set of $d$-dimensional component-wise strictly positive vectors. We now define the rate function corresponding to $M^\epsilon$. For any $(\varphi, \nu) \in \mathbb{C}_T \times \mathbb{M}_T$, we define

$$I_T(\varphi, \nu) := \begin{cases} \dfrac{1}{2} \displaystyle\int_0^T \frac{[\varphi_t' - \hat{b}(\nu, \varphi_t)]^2}{\hat{\sigma}^2(\nu, \varphi_t)} \mathrm{d}t & \text{if } \varphi \in \mathbb{H}_T, \\ \infty & \text{otherwise.} \end{cases} \tag{6}$$

where

$$\hat{b}(\nu, x) := \sum_{i=1}^d b(i, x) K_\nu(t, i), \quad \hat{\sigma}(\nu, x) := \left( \sum_{i=1}^d \sigma^2(i, x) K_\nu(t, i) \right)^{1/2}.$$

In the above formulae, we follow the conventions that $0/0 = 0$ and $n/0 = \infty$, for all $n > 0$. When we fix a time $T$, $(M^\epsilon, \nu^\epsilon)$ is understood as a joint process restricted on $[0, T]$.

6

Let $\mathbb{P} \circ (M^\epsilon, \nu^\epsilon)^{-1}$ denote $\mathbb{P}((M^\epsilon, \nu^\epsilon) \in \cdot)$, which is a family of probability measures on $(\mathbb{C}_T \times \mathbb{M}_T, \mathscr{B}(\mathbb{C}_T \times \mathbb{M}_T))$. Also, $\mathbb{P} \circ (M^\epsilon)^{-1}$ and $\mathbb{P} \circ (\nu^\epsilon)^{-1}$ are families of probability measures on $(\mathbb{C}_T, \mathscr{B}(\mathbb{C}_T))$ and $(\mathbb{M}_T, \mathscr{B}(\mathbb{M}_T))$ respectively. The following theorem is our main result which states the joint sample-path LDP of $(M^\epsilon, \nu^\epsilon)$ on $[0, T]$, as $\epsilon \to 0$..

**Theorem 3.1** *For every $T > 0$, the family $\mathbb{P} \circ (M^\epsilon, \nu^\epsilon)^{-1}$ obeys the LDP in $(\mathbb{C}_T \times \mathbb{M}_T, \rho_T \times d_T)$ with the rate function*

$$L_T(\varphi, \nu) = I_T(\varphi, \nu) + \tilde{I}_T(\nu).$$

**Proof** The proof relies on applying Theorem 2.4. We first need to prove the exponential tightness of $\mathbb{P} \circ (M^\epsilon, \nu^\epsilon)^{-1}$ on $(\mathbb{C}_T \times \mathbb{M}_T, \mathscr{B}(\mathbb{C}_T \times \mathbb{M}_T))$, i.e., for every $L < \infty$, there exists a compact set $K_L \subset \mathbb{C}_T \times \mathbb{M}_T$ such that

$$\limsup_{\epsilon \to 0} \epsilon \log \mathbb{P} \left( (M^\epsilon, \nu^\epsilon) \in \mathbb{C}_T \times \mathbb{M}_T \setminus K_L \right) \leqslant -L.$$

It is obvious that $\mathbb{P} \circ (M^\epsilon, \nu^\epsilon)^{-1}$ is exponentially tight if so are $\mathbb{P} \circ (M^\epsilon)^{-1}$ and $\mathbb{P} \circ (\nu^\epsilon)^{-1}$. As we mentioned earlier, $\mathbb{M}_T$ is a subset of $\mathbb{C}_{[0,T]}(\mathbb{R}^d)$. For any $\nu \in \mathbb{M}_T$, its derivative $K_\nu(s, i)$ is bounded by 1. Then all $\nu \in \mathbb{M}_T$ have the same Lipschitz constant, and hence $\mathbb{M}_T$ is equicontinuous. It is easily seen that $\mathbb{M}_T$ is bounded and closed. Then the Arzelà-Ascoli theorem implies that $\mathbb{M}_T$ is compact. The exponential tightness of $\mathbb{P} \circ (\nu^\epsilon)^{-1}$ is satisfied since we can take $K_L = \mathbb{M}_T$. Exponential tightness of $\mathbb{P} \circ (M^\epsilon)^{-1}$ is verified in Proposition 4.3 below.

Secondly, we proceed to prove that $\mathbb{P} \circ (M^\epsilon, \nu^\epsilon)^{-1}$ obeys the local LDP with the rate function $L_T(\varphi, \nu)$. That is, for every $(\varphi, \nu) \in \mathbb{C}_T \times \mathbb{M}_T$, we need to obtain the upper bound

$$\limsup_{\delta \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) \leqslant -L_T(\varphi, \nu),$$

and the lower bound

$$\liminf_{\delta \to 0} \liminf_{\epsilon \to 0} \epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) \geqslant -L_T(\varphi, \nu).$$

The core of the proof is proving the local LDP on a dense subset of $\mathbb{C}_T \times \mathbb{M}_T$. The upper bound is validated in Proposition 6.4. The lower bound is first proved in Proposition 7.3 given the condition $\inf_{i,x} \sigma^2(i, x) > 0$. Then the condition is lifted in Proposition 7.5 by a perturbation argument. $\square$

The LDP for $\mathbb{P} \circ (M^\epsilon)^{-1}$ only (rather than for $\mathbb{P} \circ (M^\epsilon, \nu^\epsilon)^{-1}$) is then derived from Theorem 3.1 by the contraction principle in Dembo and Zeitouni [4]. We follow the convention that $\inf(\emptyset) = \infty$.

**Corollary 3.2** *The family $\mathbb{P} \circ (M^\epsilon)^{-1}$ obeys the LDP with the rate function $\inf_{\nu \in \mathbb{M}_T} L_T(\varphi, \nu)$.*

At an intuitive level, $\tilde{I}_T(\nu)$ can be interpreted as the 'cost' of forcing $\nu^\epsilon$ to behave like $\nu$ on $[0, T]$. The other term $I_T(\varphi, \nu)$, can be seen as the 'cost' of the sample paths of $M^\epsilon$ being close to $\varphi$ *conditional on $\nu^\epsilon$ behaving like $\nu$ on $[0, T]$*. Then $\inf_{\nu \in \mathbb{M}} L_T(\varphi, \nu)$ indicates the *minimal* 'cost' of the sample paths of $M^\epsilon$ being close to $\varphi$ on $[0, T]$.

Suppose $F$ is a closed or an open subset of $\mathbb{C}_T$. We can also interpret Corollary 3.2 as the concentration of the probability $\mathbb{P} \circ (M^\epsilon)^{-1}(F)$, which is the set of sample paths of $M^\epsilon$, on the

7

'most likely path' $\arg\inf_{\varphi\in F}(\inf_{\nu\in\mathbb{M}_T}[I_T(\varphi,\nu)+\tilde{I}_T(\nu)])$. So there are two sources contributing to the large deviations behavior of $M^\epsilon$: $I_T(\varphi,\nu)$ represents the contribution resulting from the small noise, and $\tilde{I}_T(\nu)$ represents the one from the rapid switching of the modulating Markov chain.

Again by the contraction principle, $\mathbb{P}\circ(\nu^\epsilon)^{-1}$ obeys the LDP in $(\mathbb{M}_T,d_T)$ with the rate function $\inf_{\varphi\in\mathbb{C}_T} I_T(\varphi,\nu)+\tilde{I}_T(\nu)$. Since there exists a $\varphi\in\mathbb{H}_T$ such that $\varphi'_t=\hat{b}(\nu,\varphi_t)$ for all $t\in[0,T]$ and all $\nu\in\mathbb{M}_T$, it immediately follows that $\inf_{\varphi\in\mathbb{C}_T} I_T(\varphi,\nu)=0$. Hence, we have the following corollary.

**Corollary 3.3** *The family* $\mathbb{P}\circ(\nu^\epsilon)^{-1}$ *obeys the LDP in* $(\mathbb{M}_T,d_T)$ *with the rate function* $\tilde{I}_T(\nu)$.

## 4 Exponential tightness

We show the exponential tightness of $\mathbb{P}\circ(M^\epsilon)^{-1}$ by Aldous-Pukhalskii-type sufficient conditions, as dealt with in e.g. Aldous [1], Liptser and Pukhalskii [19]. The following criterion for exponential tightness in $\mathbb{C}_T$, as well as an auxiliary lemma, are adapted from Theorem 3.1 and Lemma 3.1 in Liptser and Pukhaskii [19] (which consider càdlàg processes with jumps) to our setting of continuous processes. Let $\Gamma_T(\mathscr{F}_t)$ denote the family of stopping times adapted to $\mathscr{F}_t$ taking values in $[0,T]$.

**Theorem 4.1** *Let* $Y_t^\epsilon:(\Omega,\{\mathcal{F}_t\}_{t\leqslant T},\mathbb{P})\to\mathbb{C}_T$. *If*

(i)
$$\lim_{K'\to\infty}\limsup_{\epsilon\to0}\epsilon\log\mathbb{P}\left(Y_T^{\epsilon*}\geqslant K'\right)=-\infty,$$

(ii)
$$\lim_{\delta\to0}\limsup_{\epsilon\to0}\epsilon\log\sup_{\tau\in\Gamma_T(\mathcal{F}_t)}\mathbb{P}\left(\sup_{t\leqslant\delta}|Y_{\tau+t}^\epsilon-Y_\tau^\epsilon|\geqslant\eta\right)=-\infty,\quad\forall\eta>0,$$

*then* $\mathbb{P}\circ(Y^\epsilon)^{-1}$ *is exponentially tight.*

**Lemma 4.2** *Let* $Y=(Y_t)_{t\geqslant0}$ *be a continuous semimartingale with* $Y_0=0$. *Let* $D$ *denote the part corresponding to a predictable process of locally bounded variation, and* $V$ *the part corresponding to the quadratic variation of the local martingale. Assume that for* $T>0$ *there exists a convex function* $H(\lambda),\lambda\in\mathbb{R}$ *with* $H(0)=0$ *and such that for all* $\lambda\in\mathbb{R}$ *and* $t\leqslant T$

$$\lambda D_t+\lambda^2 V_t/2\leqslant tH(\lambda\xi),\quad a.s.,$$

*where* $\xi$ *is a nonnegative random variable defined on the same probability space as* $Y$. *Then, for all* $c>0$ *and* $\eta>0$,

$$\mathbb{P}(Y_T^*\geqslant\eta)\leqslant\mathbb{P}(\xi>c)+\exp\left\{-\sup_{\lambda\in R}[\lambda\eta-TH(\lambda c)]\right\}.$$

We are now ready to prove the exponential tightness claim. The technique borrows elements from Liptser [18].

**Proposition 4.3** *For every* $T>0$, *the family* $\mathbb{P}\circ(M^\epsilon)^{-1}$ *is exponentially tight on* $(\mathbb{C}_T,\mathscr{B}(\mathbb{C}_T))$.

**Proof** Firstly, we verify the condition (i) of Theorem 4.1 for the process $M_T^{\epsilon*}$. For any $T > 0$, evidently,

$$M_T^{\epsilon*} \leqslant \int_0^T |b(X_s^\epsilon, M_s^\epsilon)| \mathrm{d}s + \sup_{t \leqslant T} \left| \sqrt{\epsilon} \int_0^t \sigma(X_s^\epsilon, M_s^\epsilon) \mathrm{d}B_s \right|, \quad \text{a.s.}.$$

We denote $C_t^\epsilon := \sqrt{\epsilon} \int_0^t \sigma(X_s^\epsilon, M_s^\epsilon) \mathrm{d}B_s$. By (A.2),

$$M_T^{\epsilon*} \leqslant K \int_0^T (1 + M_s^{\epsilon*}) \mathrm{d}s + C_T^{\epsilon*} = KT + C_T^{\epsilon*} + K \int_0^T M_s^{\epsilon*} \mathrm{d}s, \quad \text{a.s.}.$$

Since $KT + C_T^{\epsilon*}$ is nonnegative and non-decreasing in $T$, Gronwall's inequality implies

$$M_T^{\epsilon*} \leqslant e^{KT} \left[ KT + C_T^{\epsilon*} \right], \quad \text{a.s.}. \tag{7}$$

Now define $j_{K'} := K' \exp(-KT) - KT$. Then (7) entails that for sufficiently large $K'$ such that $j_{K'} > 0$,

$$\mathbb{P}(M_T^{\epsilon*} \geqslant K') \leqslant \mathbb{P}(C_T^{\epsilon*} \geqslant j_{K'}) \leqslant j_{K'}^{-1/\epsilon} \mathbb{E}\left[ (C_T^{\epsilon*})^{1/\epsilon} \right],$$

using Chebyshev's inequality. We thus conclude

$$\epsilon \log \mathbb{P}(M_T^{\epsilon*} \geqslant K') \leqslant -\log j_{K'} + \epsilon \log \mathbb{E}\left[ (C_T^{\epsilon*})^{1/\epsilon} \right]. \tag{8}$$

We assume that $1/\epsilon > 2$ in the rest of the proof (justified by the fact that we consider the limit $\epsilon \to 0$). Since $C_t^\epsilon$ is a local martingale, the process $|C_t^\epsilon|^{1/\epsilon}$ has a unique Doob-Meyer decomposition; let $\check{C}_t^\epsilon$ denote the unique predictable increasing process in this decomposition. Applying a local martingale maximal inequality (see e.g. Liptser and Shiryaev [20, Thm. 1.9.2]) to $C_t^\epsilon$, we have for the running maximum process that

$$\mathbb{E}\left[ (C_T^{\epsilon*})^{1/\epsilon} \right] \leqslant \left( \frac{1}{1-\epsilon} \right)^{1/\epsilon} \mathbb{E}\left[ \check{C}_T^\epsilon \right]. \tag{9}$$

In order to obtain an explicit expression for $\check{C}_t^\epsilon$, we apply Itô's formula to $|C_t^\epsilon|^{1/\epsilon}$. This means that, for any $t \in [0, T]$,

$$|C_t^\epsilon|^{1/\epsilon} = \frac{1}{\sqrt{\epsilon}} \int_0^t |C_s^\epsilon|^{1/\epsilon - 1} \mathrm{sign}(C_s^\epsilon) \sigma(X_s^\epsilon, M_s^\epsilon) \mathrm{d}B_s + \frac{1-\epsilon}{2\epsilon} \int_0^t |C_s^\epsilon|^{1/\epsilon - 2} \sigma^2(X_s^\epsilon, M_s^\epsilon) \mathrm{d}s.$$

We notice that the first part is a local martingale and the second part is a predictable increasing process. As a consequence,

$$\check{C}_T^\epsilon = \frac{1-\epsilon}{2\epsilon} \int_0^T |C_s^\epsilon|^{1/\epsilon - 2} \sigma^2(X_s^\epsilon, M_s^\epsilon) \mathrm{d}s. \tag{10}$$

Invoking (A.2) again, we have that $\sigma^2(X_s^\epsilon, M_s^\epsilon) \leqslant K^2(1 + M_s^{\epsilon*})^2$. Since (7) remains valid when replacing $T$ by $s$, for any $s \leqslant T$, we find

$$\begin{aligned}
|C_s^\epsilon|^{1/\epsilon - 2} \sigma^2(X_s^\epsilon, M_s^\epsilon) &\leqslant (C_s^{\epsilon*})^{1/\epsilon - 2} K^2 \left[ 1 + e^{Ks}(Ks + C_s^{\epsilon*}) \right]^2 \\
&\leqslant (C_s^{\epsilon*})^{1/\epsilon - 2} K^2 \left[ 1 + e^{KT}(KT + C_s^{\epsilon*}) \right]^2 \\
&\leqslant (C_s^{\epsilon*})^{1/\epsilon - 2} K^2 [2(1 + e^{KT} KT)^2 + 2e^{2KT}(C_s^{\epsilon*})^2].
\end{aligned}$$

Let $L_{T,K} = 2K^2 \max\left\{(1 + e^{KT}KT)^2, e^{2KT}\right\}$. Then

$$|C_s^\epsilon|^{1/\epsilon-2}\sigma^2(X_s^\epsilon, M_s^\epsilon) \leqslant (C_s^{\epsilon*})^{1/\epsilon-2}L_{T,K}\left[1 + (C_s^{\epsilon*})^2\right] \leqslant L'_{T,K}\left[1 + (C_s^{\epsilon*})^{1/\epsilon}\right],$$

where $L \equiv L'_{T,K}$ is a positive constant not depending on $K'$ (nor $\epsilon$). We plug (10) and the above estimate into (9), so as to obtain

$$
\begin{aligned}
\mathbb{E}[(C_T^{\epsilon*})^{1/\epsilon}] &\leqslant \left(\frac{1}{1-\epsilon}\right)^{1/\epsilon}\mathbb{E}\left[\frac{1-\epsilon}{2\epsilon}\int_0^T |C_s^\epsilon|^{1/\epsilon-2}\sigma^2(X_s^\epsilon, M_s^\epsilon)\mathrm{d}s\right] \\
&\leqslant \left(\frac{1}{1-\epsilon}\right)^{1/\epsilon-1}\frac{L}{2\epsilon}\left(T + \int_0^T \mathbb{E}\left[(C_s^{\epsilon*})^{1/\epsilon}\right]\mathrm{d}s\right) \\
&\leqslant \left(\frac{1}{1-\epsilon}\right)^{1/\epsilon-1}\frac{LT}{2\epsilon}\exp\left[\left(\frac{1}{1-\epsilon}\right)^{1/\epsilon-1}\frac{LT}{2\epsilon}\right],
\end{aligned}
$$

the last inequality following from Gronwall's inequality. Now observe that $(1-\epsilon)^{1-1/\epsilon}$ is decreasing on $\epsilon \in [0, \frac{1}{2})$, with limiting value $e$ as $\epsilon \to 0$. As a result, we have the following upper bound on the exponential decay rate of $\mathbb{E}[(C_T^{\epsilon*})^{1/\epsilon}]$:

$$
\begin{aligned}
\limsup_{\epsilon \to 0} \epsilon \log \mathbb{E}\left[(C_T^{\epsilon*})^{1/\epsilon}\right] &\leqslant \limsup_{\epsilon \to 0}\left[\epsilon \log\left(\frac{1}{1-\epsilon}\right)^{1/\epsilon-1} + \epsilon \log \frac{LT}{2\epsilon} + \left(\frac{1}{1-\epsilon}\right)^{1/\epsilon-1}\frac{LT}{2}\right] \\
&\leqslant \frac{eLT}{2} < \infty.
\end{aligned}
$$

Hence, by (8), for all $T > 0$, condition (i) of Thm. 4.1 follows for the process $M_T^{\epsilon*}$:

$$\lim_{K' \to \infty} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}\left(M_T^{\epsilon*} \geqslant K'\right) = -\infty. \tag{11}$$

Secondly, we verify condition (ii) of Theorem 4.1. To this end, note that for arbitrary $T > 0$, $\delta \leqslant 1$, and stopping time $\tau \in \Gamma_T(\mathcal{F}_t)$,

$$\mathbb{P}\left(\sup_{t \leqslant \delta}|M_{\tau+t}^\epsilon - M_\tau^\epsilon| \geqslant \eta\right) \leqslant \mathbb{P}\left(\sup_{t \leqslant \delta}(M_{\tau+t}^\epsilon - M_\tau^\epsilon) \geqslant \eta\right) + \mathbb{P}\left(\sup_{t \leqslant \delta}(M_\tau^\epsilon - M_{\tau+t}^\epsilon) \geqslant \eta\right). \tag{12}$$

We can see that $M_{\tau+t}^\epsilon - M_\tau^\epsilon$ is a semimartingale with respect to the filtration $\{\mathcal{F}_{\tau+t}\}_{t \geqslant 0}$. For any $\tau \in \Gamma_T(\mathcal{F}_t)$, we denote

$$D_t^\epsilon := \int_\tau^{\tau+t} b(X_s^\epsilon, M_s^\epsilon)\mathrm{d}s, \quad V_t^\epsilon := \epsilon\int_\tau^{\tau+t}\sigma^2(X_s^\epsilon, M_s^\epsilon)\mathrm{d}s.$$

By (A.2), we have, for all $\lambda \in \mathbb{R}$, $t \leqslant \delta \leqslant 1$ and $\tau \leqslant T$,

$$\lambda D_t^\epsilon + \frac{\lambda^2}{2}V_t^\epsilon \leqslant |\lambda|K(1 + M_{T+1}^{\epsilon*})t + \frac{\lambda^2\epsilon}{2}K^2(1 + M_{T+1}^{\epsilon*})^2t, \quad \text{a.s..}$$

We define

$$H(\lambda) := |\lambda| + \frac{\lambda^2\epsilon}{2}, \quad \xi := K(1 + M_{T+1}^{\epsilon*}).$$

Then $M_{\tau+t}^\epsilon - M_\tau^\epsilon$ satisfies the conditions of Lemma 4.2 and, for all $c > 0, \eta > 0$,

$$\mathbb{P}\left(\sup_{t \leqslant \delta}(M_{\tau+t}^\epsilon - M_\tau^\epsilon) \geqslant \eta\right) \leqslant \mathbb{P}(\xi > c) + \exp\left\{-\sup_{\lambda \in \mathbb{R}}[\lambda\eta - \delta H(\lambda c)]\right\}.$$

10

Since $\mathbb{P}(\xi > c) = \mathbb{P}(M_{T+1}^{\epsilon*} > c/K - 1)$, it follows that

$$\mathbb{P}\left(\sup_{t \leqslant \delta}(M_{\tau+t}^{\epsilon} - M_{\tau}^{\epsilon}) \geqslant \eta\right) \leqslant 2\max\left(\mathbb{P}\left(M_{T+1}^{\epsilon*} > \frac{c}{K} - 1\right), \exp\left\{-\sup_{\lambda \in \mathbb{R}}[\lambda\eta - \delta H(\lambda c)]\right\}\right).$$

The supremum of $\lambda\eta - \delta H(\lambda c)$ can be explicitly evaluated:

$$\sup_{\lambda \in \mathbb{R}}[\lambda\eta - \delta H(\lambda c)] = \sup_{\lambda \in \mathbb{R}}\left[\lambda\eta - \delta|\lambda|c - \delta\frac{\lambda^2 c^2 \epsilon}{2}\right] = \frac{1}{\epsilon}\sup_{\lambda > 0}\left[(\eta\epsilon - \delta c\epsilon)\lambda - \frac{\delta c^2 \epsilon^2}{2}\lambda^2\right] = \frac{(\eta - \delta c)^2}{2\epsilon\delta c^2}.$$

As a consequence, for all positive $c$,

$$\lim_{\delta \to 0}\limsup_{\epsilon \to 0}\epsilon\log\exp\left\{-\sup_{\lambda \in \mathbb{R}}[\lambda\eta - \delta H(\lambda c)]\right\} = \lim_{\delta \to 0} -\frac{(\eta - \delta c)^2}{2\delta c^2} = -\infty.$$

It is concluded that for any $\tau \in \Gamma_T(\mathcal{F}_t)$ and $c > 0$,

$$\lim_{\delta \to 0}\limsup_{\epsilon \to 0}\epsilon\log\mathbb{P}\left(\sup_{t \leqslant \delta}(M_{\tau+t}^{\epsilon} - M_{\tau}^{\epsilon}) \geqslant \eta\right) \leqslant \limsup_{\epsilon \to 0}\epsilon\log\mathbb{P}\left(M_{T+1}^{\epsilon*} > \frac{c}{K} - 1\right).$$

By (11), we know

$$\lim_{c \to \infty}\limsup_{\epsilon \to 0}\epsilon\log\mathbb{P}\left(M_{T+1}^{\epsilon*} > \frac{c}{K} - 1\right) = -\infty.$$

It implies

$$\lim_{\delta \to 0}\limsup_{\epsilon \to 0}\epsilon\log\sup_{\tau \in \Gamma_T(\mathcal{F}_t)}\mathbb{P}\left(\sup_{t \leqslant \delta}(M_{\tau+t}^{\epsilon} - M_{\tau}^{\epsilon}) \geqslant \eta\right)$$
$$\leqslant \inf_{c > 0}\limsup_{\epsilon \to 0}\epsilon\log\mathbb{P}\left(M_{T+1}^{\epsilon*} > \frac{c}{K} - 1\right) = -\infty.$$

Moreover, the claim

$$\lim_{\delta \to 0}\limsup_{\epsilon \to 0}\epsilon\log\sup_{\tau \in \Gamma_T(\mathcal{F}_t)}\mathbb{P}\left(\sup_{t \leqslant \delta}(M_{\tau}^{\epsilon} - M_{\tau+t}^{\epsilon}) \geqslant \eta\right) = -\infty$$

is proved in the same way. Thus the desired claim follows from (12). $\qquad\square$

## 5 Auxiliary results

We first identify a dense subset of $\mathbb{C}_T \times \mathbb{M}_T$ which substantially simplifies the proof of the local LDP in the next two sections. Let $\mathbb{M}_T^+$ be the subset of $\mathbb{M}_T$ such that $K_\nu(s, i) > 0, \forall s \in [0, T], i \in \mathbb{S}$, and let $\mathbb{M}_T^{++}$ be the subset of $\mathbb{M}_T^+$ such that $K_\nu(\cdot, i) \in \mathbb{C}_{[0,T]}^\infty, \forall i \in \mathbb{S}$.

**Lemma 5.1** $\mathbb{M}_T^{++}$ *is dense in* $\mathbb{M}_T$.

**Proof** We prove the claim in two steps. Firstly, we show that $\mathbb{M}_T^{++}$ is dense in $\mathbb{M}_T^+$. We begin by introducing the standard *mollifier* $J(x)$ on $\mathbb{R}$, i.e.,

$$J(x) := \begin{cases} k\exp\left(\frac{1}{|x|^2 - 1}\right) & \text{if } |x| < 1, \\ 0 & \text{if } |x| \geqslant 1, \end{cases}$$

where $k > 0$ is selected so that $\int_{\mathbb{R}} J(x)\mathrm{d}x = 1$. For each $\eta > 0$, we define $J_\eta(x) := \eta^{-1}J(x/\eta)$. For any $\nu$ in $\mathbb{M}_T^+$, we extend the domain of $K_\nu(\cdot, i)$ to $(-1, T+1)$, as follows:

$$
K_\nu(s, i) := \begin{cases} K_\nu(s, i) & \text{if } s \in [0, T], \\ K_\nu(0, i) & \text{if } s \in (-1, 0), \\ K_\nu(1, i) & \text{if } s \in (T, T+1). \end{cases}
$$

Since $K_\nu(\cdot, i)$ is integrable on $(-1, T+1)$ for each $i \in \mathbb{S}$, we can define its mollification as

$$
K_\nu^\eta(s, i) = \int_{-\eta}^\eta J_\eta(y)K_\nu(s - y, i)\mathrm{d}y, \text{ for } s \in (-1 + \eta, T + 1 - \eta), \eta < 1.
$$

By Theorem C.6 in Evans [7], $K_\nu^\eta(\cdot, i)$ is smooth on $(-1+\eta, T+1-\eta)$ and $K_\nu^\eta(\cdot, i) \to K_\nu(\cdot, i)$ almost everywhere as $\eta \to 0$.

Next we proceed to show that $\nu^\eta$ with the kernel $K_\nu^\eta(s, i)$ is an element of $\mathbb{M}_T^{++}$. It is clear that $K_\nu^\eta(\cdot, i) \in \mathbb{C}_{[0,T]}^\infty$ when restricted to $[0, T]$, and in addition we have $K_\nu^\eta(s, i) > 0$, for all $s \in [0, T]$ and $i \in \mathbb{S}$. As a consequence, we only need to prove that $\sum_{i=1}^d K_\nu^\eta(s, i) = 1$, for all $s \in [0, T]$. For any $s \in [0, T]$ and $y \in [-\eta, \eta]$, it holds that $\sum_{i=1}^d K_\nu(s - y, i) = 1$ since $s - y \in (-1, T+1)$. We thus have that

$$
\sum_{i=1}^d K_\nu^\eta(s, i) = \int_{-\eta}^\eta J_\eta(y) \sum_{i=1}^d K_\nu(s - y, i)\mathrm{d}y = \int_{-\eta}^\eta J_\eta(y)\mathrm{d}y = 1.
$$

For any $t \leqslant T, i \in \mathbb{S}$, due to the fact that $K_\nu^\eta(\cdot, i) \to K_\nu(\cdot, i)$ on $(0, t)$ almost everywhere as $\eta \to 0$ and $K_\nu^\eta(s, i) \leqslant 1$, it holds that

$$
\int_0^t K_\nu^\eta(s, i)\mathrm{d}s \to \int_0^t K_\nu(s, i)\mathrm{d}s, \text{ as } \eta \to 0,
$$

appealing to the dominated convergence theorem. Since $\int_0^t K_\nu^\eta(s, i)\mathrm{d}s$ is increasing in $t$, $\int_0^t K_\nu(s, i)\mathrm{d}s$ is continuous in $t$ and $d$ is finite, we obtain the following uniform convergence:

$$
d_T(\nu^\eta, \nu) = \sup_{t \in [0,T], i \in \mathbb{S}} \left| \int_0^t K_\nu^\eta(s, i)\mathrm{d}s - \int_0^t K_\nu(s, i)\mathrm{d}s \right| \to 0, \text{ as } \eta \to 0,
$$

using Result 1.1.21 in Jacod [15].

Secondly, we prove that $\mathbb{M}_T^+$ is dense in $\mathbb{M}_T$. Noticing that $K_\nu(s, i)$ can be 0 for some $i, s$, we define (for any $\nu \in \mathbb{M}$) an $\nu^\eta \in \mathbb{M}_T^+$ through

$$
K_\nu^\eta(s, i) := \frac{K_\nu(s, i) + \eta}{1 + \eta d},
$$

$\eta > 0$, for all $i, s$. As is directly verified, $d_T(\nu^\eta, \nu) \leqslant T(\eta + \eta d)/(1 + \eta d)$. Then the desired result holds. Consequently, $\mathbb{M}_T^{++}$ is dense in $\mathbb{M}_T$ by the triangle inequality. $\qquad \square$

We then present a regularity property of the rate function $\tilde{I}_T(\nu)$ on $\mathbb{M}_T^{++}$.

**Lemma 5.2** *Fix* $s \in [0, T]$ *and* $\nu \in \mathbb{M}_T^{++}$. *Then there is an optimizer* $u^*(s, \cdot)$ *of*

$$
\inf_{u \in U} \left[ \sum_{i=1}^d \frac{(Qu)(i)}{u(i)} K_\nu(s, i) \right]
$$

*such that* $u^*(\cdot, i) \in \mathbb{C}_{[0,T]}^\infty$, *for all* $i \in \mathbb{S}$, *and* $u^* \in \mathbb{U}$.

**Proof** As obviously

$$\sum_{i=1}^{d} \frac{(Qu)(i)}{u(i)} K_\nu(s,i) = \sum_{i=1}^{d} Q_{ii} K_\nu(s,i) + \sum_{i=1}^{d} \frac{\sum_{j\neq i}^{d} Q_{ij} u(j)}{u(i)} K_\nu(s,i),$$

the optimization problem essentially reduces to

$$\inf_{u\in U} \left[ \sum_{i=1}^{d} \frac{\sum_{j\neq i}^{d} Q_{ij} u(j)}{u(i)} K_\nu(s,i) \right].$$

We let $r_{ji} := u(j)/u(i)$, for $i \neq j$. Since $r_{ji} = 1/r_{ij}$, the optimization problem can be written as a minimization over $d(d-1)/2$ variables:

$$\inf_{r_{ji}>0} \sum_{i=1}^{d} \sum_{j=1}^{i-1} \left[ Q_{ij} r_{ji} K_\nu(s,i) + Q_{ji} r_{ji}^{-1} K_\nu(s,j). \right]$$

Observe that for any $i, j, k$ the equality $r_{ij} r_{jk} = r_{ik}$ needs to hold, which corresponds to $\psi(d) := (d-1)(d-2)/2$ constraints. We then perform the change of variables $x_{ji} := \log r_{ji}$, and denote by $\boldsymbol{X} = (x_{21}, \cdots, x_{d(d-1)})^{\mathrm{T}}$ the $d(d-1)/2$ variables. Letting $\boldsymbol{K}_\nu(s) = (K_\nu(s,1), \cdots, K_\nu(s,d))^{\mathrm{T}}$, we transform the above optimization problem into

$$\inf_{\boldsymbol{X}} f(\boldsymbol{K}_\nu(s), \boldsymbol{X}), \quad \text{where } f(\boldsymbol{K}_\nu(s), \boldsymbol{X}) := \sum_{i=1}^{d} \sum_{j=1}^{i-1} \left[ Q_{ij} e^{x_{ji}} K_\nu(s,i) + Q_{ji} e^{-x_{ji}} K_\nu(s,j) \right],$$

with $(d-1)(d-2)/2$ additional constraints to be imposed.

The gradient vector of $f$ with respect to $\boldsymbol{X}$ is

$$\mathbf{D}_{\boldsymbol{X}} f = \left( \frac{\partial f}{\partial x_{21}}, \cdots, \frac{\partial f}{\partial x_{d(d-1)}} \right),$$

and the corresponding Hessian matrix $\boldsymbol{D}_{\boldsymbol{X}}^2 f$ is the diagonal matrix which has entries of the form $Q_{ij} e^{x_{ji}} K_\nu(s,i) + Q_{ji} e^{-x_{ji}} K_\nu(s,j)$ on its diagonal. The idea is now to split the vector $\boldsymbol{X}$ into $\boldsymbol{X}_0 = (x_{21}, \cdots, x_{d1})^{\mathrm{T}}$ and $\boldsymbol{X}_1$ (where the latter vector corresponds with the remaining $\psi(d)$ variables). Due to the constraints, we have $\boldsymbol{X}_1 = \boldsymbol{L}\boldsymbol{X}_0$ where $\boldsymbol{L}$ is a matrix of dimension $\psi(d) \times (d-1)$. The next step is to include the constraints into the optimization equation $f$. It yields the following new optimization problem, on which no additional constraints need to ne imposed:

$$\inf_{\boldsymbol{X}_0} \hat{f}(\boldsymbol{K}_\nu(s), \boldsymbol{X}_0), \quad \text{where } \hat{f}(\boldsymbol{K}_\nu(s), \boldsymbol{X}_0) = f(\boldsymbol{K}_\nu(s), (\boldsymbol{X}_0, \boldsymbol{L}\boldsymbol{X}_0)).$$

Observe that $f$ is a globally strictly convex function of $\boldsymbol{X}$ on a convex domain, and consequently $\hat{f}$ is a strictly convex function of $\mathbf{X}_0$. Hence, there is a unique minimizer $\boldsymbol{X}_0^*(s) = (x_{21}^*(s), \cdots, x_{d1}^*(s))^{\mathrm{T}}$ for any $s \in [0,T]$. Since we have that both $K_\nu(s,i) > 0$ and $Q_{ij} > 0$ for $i \neq j$, any entry of $\boldsymbol{X}_0^*(s)$ cannot be $-\infty$ or $\infty$. We thus conclude that $\boldsymbol{X}_0^*(s) \in \mathbb{R}^{d-1}$.

Let $\boldsymbol{I}$ be the $(d-1)$-dimensional identity matrix. Then we define

$$\tilde{f}(\boldsymbol{K}_\nu(s), \boldsymbol{X}_0) := \mathbf{D}_{\boldsymbol{X}_0} \hat{f}(\boldsymbol{K}_\nu(s), \boldsymbol{X}_0) = \mathbf{D}_{\boldsymbol{X}} f(\boldsymbol{K}_\nu(s), (\boldsymbol{X}_0, \boldsymbol{L}\boldsymbol{X}_0)) \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{L} \end{pmatrix},$$

which is a smooth function on $\mathbb{R}^d \times \mathbb{R}^{d-1}$ such that $\tilde{f}(\boldsymbol{K}_\nu(s), \boldsymbol{X}_0^*(s)) = 0$. The gradient matrix of $\tilde{f}$ with respect to $\boldsymbol{X}_0$ evaluated in $\boldsymbol{X}_0^*(s)$ is

$$\boldsymbol{G} := \mathbf{D}_{\boldsymbol{X}_0}^2 \tilde{f}(\boldsymbol{K}_\nu(s), \boldsymbol{X}_0^*(s)) = (\boldsymbol{I} \ \ \boldsymbol{L}^{\mathrm{T}}) \, \mathbf{D}_{\boldsymbol{X}}^2 f(\boldsymbol{K}_\nu(s), (\boldsymbol{X}_0^*(s), \boldsymbol{L}\boldsymbol{X}_0^*(s))) \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{L} \end{pmatrix}.$$

Let $|\boldsymbol{G}|$ denote the determinant of $\boldsymbol{G}$. Since $\boldsymbol{H}$ is a positive-definite diagonal matrix and $\boldsymbol{L}$ is of full rank, we conclude that $|\boldsymbol{G}| \neq 0$, for all $s \in (0, T)$.

Hence, the implicit function theorem (cf. Theorem C.8 in Evans [7]) implies that $\boldsymbol{X}_0^*(s)$ is a smooth function of $\boldsymbol{K}_\nu(s)$: since $K_\nu(\cdot, i) \in \mathbb{C}_{[0,T]}^\infty$ for all $i \in \mathbb{S}$, we conclude that $\boldsymbol{X}_0^*(s) \in \mathbb{C}_{[0,T]}^\infty(\mathbb{R}^{d-1})$. It also follows that the corresponding minimizer in terms of the variables $r_{ij}$, say $(r_{21}^*(s), \cdots, r_{d1}^*(s))$, is in $\mathbb{C}_{[0,T]}^\infty((0, \infty)^{d-1})$. Recalling that $r_{ji} = u(j)/u(i)$, we set

$$(u^*(s, 1), u^*(s, 2), \cdots, u^*(s, d)) \equiv (1, r_{21}^*(s), \cdots, r_{d1}^*(s))$$

on $[0, T]$. Then $(u^*(s, 1), \cdots, u^*(s, d))$ is an optimizer corresponding to

$$\inf_{u \in U} \left[ \sum_{i=1}^d \frac{(Qu)(i)}{u(i)} K_\nu(s, i) \right],$$

and $u^*(\cdot, i) \in \mathbb{C}_{[0,T]}^\infty$ for all $i \in \mathbb{S}$. It is easily seen that $u^*(s, i) > 0$. Then $\inf_{s \in [0,T], i \in \mathbb{S}} u^*(s, i) > 0$ by continuity of $u^*$ on $[0, T]$. Hence, $u^* \in \mathbb{U}$. $\qquad\square$

The following continuity property of the rate functions will be used in proving the upper and lower bounds.

**Lemma 5.3** *Let $\nu^\eta, \nu \in \mathbb{M}_T$ with kernels $K_\nu^\eta$ and $K_\nu$ such that $K_\nu^\eta(\cdot, i) \to K_\nu(\cdot, i)$ a.e. as $\eta \to 0$ on $[0, T]$ for each $i \in \mathbb{S}$. Then*

(i) $\tilde{I}_T(\nu^\eta) \to \tilde{I}_T(\nu)$ *as $\eta \to 0$;*

(ii) $I_T(\varphi, \nu^\eta) \to I_T(\varphi, \nu)$ *as $\eta \to 0$, $\forall \varphi \in \mathbb{H}_T$, if $\inf_{i,x} \sigma^2(i, x) > 0$.*

**Proof** (i) Let $\rho$ be a $d$-dimensional vector such that $\sum_{i=1}^d \rho(i) = 1$ and $\rho(i) \geqslant 0$. By Lemma 4.22 in den Hollander [5],

$$-\inf_{u \in U} \left[ \sum_{i=1}^d \frac{(Qu)(i)}{u(i)} \rho(i) \right]$$

is continuous in $\rho$ and positive. Moreover, for all $\rho$, realizing that the $Q_{ii}$ are negative,

$$-\inf_{u \in U} \left[ \sum_{i=1}^d \frac{(Qu)(i)}{u(i)} \rho(i) \right] \leqslant -\sum_{i=1}^d Q_{ii}$$

as a consequence of

$$-\inf_{u \in U} \left[ \sum_{i=1}^d \frac{(Qu)(i)}{u(i)} \rho(i) \right] = -\sum_{i=1}^d Q_{ii} \rho(i) - \inf_{u \in U} \left[ \sum_{i=1}^d \frac{\sum_{j \neq i}^d Q_{ij} u(j)}{u(i)} \rho(i) \right].$$

Hence,

$$\sup_{u \in U} \left[ -\sum_{i=1}^d \frac{(Qu)(i)}{u(i)} K_\nu^\eta(s, i) \right] \to \sup_{u \in U} \left[ -\sum_{i=1}^d \frac{(Qu)(i)}{u(i)} K_\nu(s, i) \right],$$

14

as $\eta \to 0$, almost everywhere on $[0, T]$. Also,

$$-\inf_{u \in U}\left[\sum_{i=1}^{d} \frac{(Qu)(i)}{u(i)} K_{\nu}^{\eta}(s, i)\right] \leqslant -\sum_{i=1}^{d} Q_{ii}$$

for all $s$. Then the desired result follows directly by applying the dominated convergence theorem.

(ii) When $\inf_{i,x} \sigma^2(i, x) > 0$, it is easily seen by continuity that

$$\frac{\varphi_t' - \hat{b}(\nu^\eta, \varphi_t)]^2}{\hat{\sigma}^2(\nu^\eta, \varphi_t)} \to \frac{[\varphi_t' - \hat{b}(\nu, \varphi_t)]^2}{\hat{\sigma}^2(\nu, \varphi_t)}$$

a.e. as $\eta \to 0$. Let $\sigma^2$ denote $\inf_{i,x} \sigma^2(i, x)$. For every $\nu \in \mathbb{M}_T$, we have

$$\frac{[\varphi_t' - \hat{b}(\nu, \varphi_t)]^2}{\hat{\sigma}^2(\nu, \varphi_t)} \leqslant \frac{|\varphi_t'|^2 + (\sum_{i=1}^{d} |b(i, \varphi_t)|)^2 + 2|\varphi_t'|(\sum_{i=1}^{d} |b(i, \varphi_t)|)}{\sigma^2}.$$

Since $\varphi$ is absolutely continuous and $b(i, x)$ is Lipschitz continuous in $x$, $\sum_{i=1}^{d} |b(i, \varphi_t)| < b < \infty$ on $[0, T]$. Hence, $[\varphi_t' - \hat{b}(\nu, \varphi_t)]^2/\hat{\sigma}^2(\nu, \varphi_t) \leqslant (|\varphi_t'| + b)^2/\sigma^2$. Since $\varphi'$ is square-integrable on $[0, T]$, $I_T(\varphi, \nu^\eta) \to I_T(\varphi, \nu)$ as $\eta \to 0$ by again applying the dominated convergence theorem. $\quad\square$

## 6 Upper bound for the local LDP

This section considers the upper bound in the local LDP, whereas the next section concentrates on the corresponding lower bound. Recall that our aim is to establish

$$\limsup_{\delta \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) \leqslant -L_T(\varphi, \nu),$$

with $L_T(\varphi, \nu)$ as defined in Section 2. Our approach, which has a simlar structure as the one used in Liptser in [18], finds an exponential (in $\epsilon$, that is) upper bound on the probability $\mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta)$ relying on the method of stochastic exponentials. As it turns out, this bound should contain the rate function $L_T(\varphi, \nu)$, as desired.

We start by introducing some additional notation. Let $\mathbb{S}_T$ denote the space of all step functions on $[0, T]$ of the form, for $k \in \mathbb{N}$ and real numbers $\lambda_0, \cdots, \lambda_k$,

$$\lambda(t) = \lambda_0 \mathbf{1}_{\{t=0\}}(t) + \sum_{i=0}^{k} \lambda_i \mathbf{1}_{(t_i, t_{i+1}]}(t), \quad 0 = t_0 < \cdots < t_{k+1} = T.$$

For any $\varphi \in \mathbb{C}_T$, we introduce the following notation

$$\int_0^T \lambda(s) \mathrm{d}\varphi_s := \sum_{i=0}^{k} \lambda_i [\varphi_{T \wedge t_{i+1}} - \varphi_{T \wedge t_i}].$$

In the sequel we frequently use the process

$$N_t^\epsilon := \frac{1}{\sqrt{\epsilon}} \int_0^t \lambda(s) \sigma(X_s^\epsilon, M_s^\epsilon) \mathrm{d}B_s, \quad \lambda \in \mathbb{S}_T,$$

15

which has the stochastic exponential

$$\mathscr{E}(N^\epsilon)_t = \exp\left(N_t^\epsilon - \frac{1}{2}\langle N^\epsilon\rangle_t\right), \quad \text{where} \quad \langle N^\epsilon\rangle_t = \frac{1}{\epsilon}\int_0^t \lambda^2(s)\sigma^2(X_s^\epsilon, M_s^\epsilon)\mathrm{d}s.$$

Next we introduce a stochastic exponential associated with the occupation measure $\nu^\epsilon$. For any $u(\cdot,\cdot) \in \mathbb{U}$,

$$\hat{N}_t^\epsilon = u(t, X_t^\epsilon) - u(0, X_0^\epsilon) - \int_0^t \frac{\partial}{\partial s}u(s, X_s^\epsilon)\mathrm{d}s - \int_0^t (Q^\epsilon u)(s, X_s^\epsilon)\mathrm{d}s$$

is a local martingale on $[0,T]$ due to Itô's formula. We define

$$\tilde{N}_t^\epsilon := \int_0^t \frac{1}{u(s-, X_{s-}^\epsilon)}\mathrm{d}\hat{N}_s^\epsilon.$$

Then

$$\mathscr{E}(\tilde{N}^\epsilon)_t = \frac{u(t, X_t^\epsilon)}{u(0, X_0^\epsilon)}\exp\left(-\int_0^t \frac{\frac{\partial}{\partial s}u(s, X_s^\epsilon) + (Q^\epsilon u)(s, X_s^\epsilon)}{u(s, X_s^\epsilon)}\mathrm{d}s\right) \tag{13}$$

is the stochastic exponential of $\tilde{N}_t^\epsilon$. Indeed,

$$
\begin{aligned}
\mathrm{d}\mathscr{E}(\tilde{N}^\epsilon)_t &= \frac{u(t, X_t^\epsilon)}{u(0, X_0^\epsilon)}\exp\left(-\int_0^t \frac{\frac{\partial}{\partial s}u(s, X_s^\epsilon) + (Q^\epsilon u)(s, X_s^\epsilon)}{u(s, X_s^\epsilon)}\mathrm{d}s\right) \\
&\quad \times \left(-\frac{\frac{\partial}{\partial t}u(t, X_t^\epsilon) + (Q^\epsilon u)(t, X_t^\epsilon)}{u(t, X_t^\epsilon)}\mathrm{d}t\right) \\
&\quad + \exp\left(-\int_0^t \frac{\frac{\partial}{\partial s}u(s, X_s^\epsilon) + (Q^\epsilon u)(s, X_s^\epsilon)}{u(s, X_s^\epsilon)}\mathrm{d}s\right)\frac{\mathrm{d}u(t, X_t^\epsilon)}{u(0, X_0^\epsilon)} \\
&= \frac{\mathscr{E}(\tilde{N}^\epsilon)_{t-}}{u(t-, X_{t-}^\epsilon)}\left[\mathrm{d}u(t, X_t^\epsilon) - \frac{\partial}{\partial t}u(t, X_t^\epsilon)\mathrm{d}t - (Q^\epsilon u)(t, X_t)\mathrm{d}t\right] \\
&= \frac{\mathscr{E}(\tilde{N}^\epsilon)_{t-}}{u(t-, X_{t-}^\epsilon)}\mathrm{d}\hat{N}_t^\epsilon.
\end{aligned}
$$

Since $\inf_{t\in[0,T], i\in\mathbb{S}} u(t,i) > 0$, $\tilde{N}_t^\epsilon$ is a local martingale and its stochastic exponential $\mathscr{E}(\tilde{N}^\epsilon)_t$ is also a local martingale by Theorem 1.4.61 in Jacod and Shiryaev [16]. Then $\mathscr{E}(\tilde{N}^\epsilon)_t$ is a martingale since it is bounded. We will use this martingale property when applying a change of measure in the next section. The martingale $\mathscr{E}(\tilde{N}^\epsilon)_t$ is an extension of the exponential martingale studied by Palmowski and Rolski in [22].

**Lemma 6.1** $\mathscr{E}(\tilde{N}^\epsilon)_t\mathscr{E}(N^\epsilon)_t$ is a local martingale, and $\mathbb{E}[\mathscr{E}(\tilde{N}^\epsilon)_t\mathscr{E}(N^\epsilon)_t] \leqslant 1$.

**Proof** By Protter [23, Thm. 2.38],

$$\mathscr{E}(\tilde{N}^\epsilon)_t\mathscr{E}(N^\epsilon)_t = \mathscr{E}(\tilde{N}^\epsilon + N^\epsilon + [\tilde{N}^\epsilon, N^\epsilon])_t,$$

where $[\tilde{N}^\epsilon, N^\epsilon]$ denotes the quadratic covariation process. Since $\tilde{N}_t^\epsilon$ is a pure jump local martingale and $N_t^\epsilon$ is a continuous local martingale, $[\tilde{N}^\epsilon, N^\epsilon] = 0$. Then $\mathscr{E}(\tilde{N}^\epsilon)_t\mathscr{E}(N^\epsilon)_t$ is the stochastic exponential of the local martingale $\tilde{N}_t^\epsilon + N_t^\epsilon$ and a local martingale too. Since a positive local martingale is a supermartingale, $\mathbb{E}[\mathscr{E}(\tilde{N}^\epsilon)_t\mathscr{E}(N^\epsilon)_t] \leqslant \mathbb{E}[\mathscr{E}(\tilde{N}^\epsilon)_0\mathscr{E}(N^\epsilon)_0] = 1$. $\square$

The above lemma evidently implies that

$$\mathbb{E}\left[\mathbf{1}_{\{\rho_T(M^{\epsilon,T},\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta\}}\mathscr{E}(\tilde{N}^\epsilon)_T\mathscr{E}(N^\epsilon)_T\right]\leqslant 1.$$

In order to find an exponential upper bound on $\mathbb{P}(\rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta)$, we derive non-random exponential *lower* bounds on $\mathscr{E}(\tilde{N}^\epsilon)_T$ and $\mathscr{E}(N^\epsilon)_T$ in case that both $M^\epsilon$ is close to $\varphi$ and $\nu^\epsilon$ close to $\nu$ (i.e., on the set $\{\rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta\}$). The next two lemmas present the results; Lemma 6.2 focuses on $\mathscr{E}(N^\epsilon)_T$, whereas Lemma 6.3 covers $\mathscr{E}(\tilde{N}^\epsilon)_T$.

**Lemma 6.2** *For every* $(\varphi,\nu)\in\mathbb{C}_T\times\mathbb{M}_T$ *and every* $\lambda\in\mathbb{S}_T$, $\delta>0$, *there exists a positive constant* $K_{\lambda,\varphi,T}$ *not depending on* $\epsilon$ *or* $\delta$ *such that*

$$\mathscr{E}(N^\epsilon)_T\geqslant\exp\left\{\frac{1}{\epsilon}\left[\int_0^T\lambda(s)\mathrm{d}\varphi_s-\int_0^T\lambda(s)\hat{b}(\nu,\varphi_s)\mathrm{d}s-\int_0^T\frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu,\varphi_s)\mathrm{d}s\right]-\frac{\delta}{\epsilon}K_{\lambda,\varphi,T}\right\}$$

*on the set* $\{\rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta\}$.

**Proof** It is first realized that, by (1), $N_t^\epsilon$ can be rearranged as

$$N_t^\epsilon=\frac{1}{\epsilon}\left[\int_0^t\lambda(s)\mathrm{d}M_s^\epsilon-\int_0^t\lambda(s)b(X_s^\epsilon,M_s^\epsilon)\mathrm{d}s\right].$$

Then a straightforward computation yields that

$$N_T^\epsilon-\frac{1}{2}\langle N^\epsilon\rangle_T=\frac{1}{\epsilon}\left[\int_0^T\lambda(s)\mathrm{d}M_s^\epsilon-\int_0^T\lambda(s)b(X_s^\epsilon,M_s^\epsilon)\mathrm{d}s-\int_0^T\frac{\lambda^2(s)}{2}\sigma^2(X_s^\epsilon,M_s^\epsilon)\mathrm{d}s\right]$$

$$=\frac{1}{\epsilon}\left[\int_0^T\lambda(s)\mathrm{d}M_s^\epsilon-\int_0^T\lambda(s)\mathrm{d}\varphi_s\right]-\frac{1}{\epsilon}\left[\int_0^T\lambda(s)b(X_s^\epsilon,M_s^\epsilon)\mathrm{d}s-\int_0^T\lambda(s)\hat{b}(\nu,\varphi_s)\mathrm{d}s\right]$$

$$-\frac{1}{\epsilon}\left[\int_0^T\frac{\lambda^2(s)}{2}\sigma^2(X_s^\epsilon,M_s^\epsilon)\mathrm{d}s-\int_0^T\frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu,\varphi_s)\mathrm{d}s\right]$$

$$+\frac{1}{\epsilon}\left[\int_0^T\lambda(s)\mathrm{d}\varphi_s-\int_0^T\lambda(s)\hat{b}(\nu,\varphi_s)\mathrm{d}s-\int_0^T\frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu,\varphi_s)\mathrm{d}s\right].$$

As a consequence, we evidently have

$$N_T^\epsilon-\frac{1}{2}\langle N^\epsilon\rangle_T\geqslant\frac{1}{\epsilon}\left[\int_0^T\lambda(s)\mathrm{d}\varphi_s-\int_0^T\lambda(s)\hat{b}(\nu,\varphi_s)\mathrm{d}s-\int_0^T\frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu,\varphi_s)\mathrm{d}s\right]$$

$$-\frac{1}{\epsilon}\left|\int_0^T\lambda(s)\mathrm{d}M_s^\epsilon-\int_0^T\lambda(s)\mathrm{d}\varphi_s\right|-\frac{1}{\epsilon}\left|\int_0^T\lambda(s)b(X_s^\epsilon,M_s^\epsilon)\mathrm{d}s-\int_0^T\lambda(s)\hat{b}(\nu,\varphi_s)\mathrm{d}s\right|$$

$$-\frac{1}{\epsilon}\left|\int_0^T\frac{\lambda^2(s)}{2}\sigma^2(X_s^\epsilon,M_s^\epsilon)\mathrm{d}s-\int_0^T\frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu,\varphi_s)\mathrm{d}s\right|\quad\text{a.s.}.$$

Hence, by repeated use of the triangle inequality, we find that $N_T^\epsilon-\frac{1}{2}\langle N^\epsilon\rangle_T\geqslant\epsilon^{-1}G_T^\epsilon$ a.s., where $G_T^\epsilon$ is given by

$$\left[\int_0^T\lambda(s)\mathrm{d}\varphi_s-\int_0^T\lambda(s)\hat{b}(\nu,\varphi_s)\mathrm{d}s-\int_0^T\frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu,\varphi_s)\mathrm{d}s\right]$$

$$-\left|\int_0^T\lambda(s)\mathrm{d}M_s^\epsilon-\int_0^T\lambda(s)\mathrm{d}\varphi_s\right|$$

$$-\left|\int_0^T\lambda(s)b(X_s^\epsilon,M_s^\epsilon)-\lambda(s)b(X_s^\epsilon,\varphi_s)\mathrm{d}s\right|-\left|\int_0^T\lambda(s)b(X_s^\epsilon,\varphi_s)-\lambda(s)\hat{b}(\nu,\varphi_s)\mathrm{d}s\right|$$

$$-\left|\int_0^T\frac{\lambda^2(s)}{2}\sigma^2(X_s^\epsilon,M_s^\epsilon)-\frac{\lambda^2(s)}{2}\sigma^2(X_s^\epsilon,\varphi_s)\mathrm{d}s\right|-\left|\int_0^T\frac{\lambda^2(s)}{2}\sigma^2(X_s^\epsilon,\varphi_s)-\frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu,\varphi_s)\mathrm{d}s\right|.$$

17

In the rest of the proof, all objects are considered on the set $\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}$; we analyze all absolute values in the previous display separately. Let us start with considering the first absolute value; we thus find that

$$\left| \int_0^T \lambda(s)\mathrm{d}M_s^\epsilon - \int_0^T \lambda(s)\mathrm{d}\varphi_s \right| = \left| \sum_{j=1}^k \lambda_j \left( M_{T \wedge t_{j+1}}^\epsilon - \varphi_{T \wedge t_{j+1}} - (M_{T \wedge t_j}^\epsilon - \varphi_{T \wedge t_j}) \right) \right| \leqslant 2\lambda_T^* \delta.$$

Now consider the second absolute value. The Lipschitz condition (A.1) implies that

$$\left| \int_0^T \lambda(s)b(X_s^\epsilon, M_s^\epsilon) - \lambda(s)b(X_s^\epsilon, \varphi_s)\mathrm{d}s \right| \leqslant \int_0^T |\lambda(s)|K|M_s^\epsilon - \varphi_s|\mathrm{d}s \leqslant \lambda_T^* \delta K T.$$

For the fourth one, (A.1) also entails that

$$\left| \int_0^T \frac{\lambda^2(s)}{2} \left[ \sigma^2(X_s^\epsilon, M_s^\epsilon) - \sigma^2(X_s^\epsilon, \varphi_s) \right] \mathrm{d}s \right| \leqslant \frac{\lambda_T^{2*}}{2} \int_0^T K\delta|\sigma(X_s^\epsilon, M_s^\epsilon) + \sigma(X_s^\epsilon, \varphi_s)|\mathrm{d}s$$

Since $\varphi$ is continuous on $[0,T]$, there exists a positive constant $r$ such that $\varphi_T^* \leqslant r - \delta$. It yields that $M_T^{\epsilon*} \leqslant r$ on the set $\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}$. By the linear growth condition (A.2) and the above reasoning

$$|\sigma(X_s^\epsilon, M_s^\epsilon) + \sigma(X_s^\epsilon, \varphi_s)| \leqslant K(1 + |M_s^\epsilon|) + K(1 + |\varphi_s|) \leqslant 2K(1 + r).$$

We conclude that

$$\left| \int_0^T \frac{\lambda^2(s)}{2} \left[ \sigma^2(X_s^\epsilon, M_s^\epsilon) - \sigma^2(X_s^\epsilon, \varphi_s) \right] \mathrm{d}s \right| \leqslant \lambda_T^{2*} \delta K^2(1 + r)T.$$

Then, concerning the third absolute value,

$$
\begin{aligned}
\left| \int_0^T \lambda(s)b(X_s^\epsilon, \varphi_s) - \lambda(s)\hat{b}(\nu, \varphi_s)\mathrm{d}s \right| &= \left| \sum_{j=0}^k \int_{t_j}^{t_{j+1}} \lambda_j[b(X_s^\epsilon, \varphi_s) - \hat{b}(\nu, \varphi_s)]\mathrm{d}s \right| \\
&\leqslant \sum_{j=0}^k \left| \int_{t_j}^{t_{j+1}} \lambda_j \sum_{i=1}^d b(i, \varphi_s)[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right| \\
&= \sum_{j=0}^k \left| \int_{t_j}^{t_{j+1}} \sum_{i=1}^d f_j(i,s)[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right| \\
&\leqslant \sum_{j=0}^k \sum_{i=1}^d \left| \int_{t_j}^{t_{j+1}} f_j(i,s)[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right|
\end{aligned}
$$

where $f_j(i,s) := \lambda_j b(i, \varphi_s)$. Since $b(i, \cdot)$ is Lipschitz continuous and $\varphi_s$ is absolutely continuous, $f_j(i,s)$ is of bounded variation. Then, by Lemma A.1,

$$\sup_{i \in \mathbb{S}} \left| \int_{t_j}^{t_{j+1}} f_j(i,s)[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right| \leqslant C_j \delta,$$

where $C_j$ is a constant. We thus conclude that

$$\left| \int_0^T \lambda(s)b(X_s^\epsilon, \varphi_s) - \lambda(s)\hat{b}(\nu, \varphi_s)\mathrm{d}s \right| \leqslant \sum_{j=0}^k C_j \delta d \leqslant C\delta.$$

18

A similar procedure yields for the last absolute value

$$\left| \int_0^T \frac{\lambda^2(s)}{2} \sigma^2(X_s^\epsilon, \varphi_s) - \frac{\lambda^2(s)}{2} \hat\sigma^2(\nu, \varphi_s) \mathrm{d}s \right| \leqslant C'\delta.$$

Upon collecting these inequalities, we find

$$\mathscr{E}(N^\epsilon)_T \geqslant \exp\left\{ \frac{1}{\epsilon} \left[ \int_0^T \lambda(s)\mathrm{d}\varphi_s - \int_0^T \lambda(s)\hat b(\nu, \varphi_s)\mathrm{d}s - \int_0^T \frac{\lambda^2(s)}{2}\hat\sigma^2(\nu, \varphi_s)\mathrm{d}s \right] - \frac{\delta}{\epsilon} K_{\lambda,\varphi,T} \right\},$$

where we denote

$$K_{\lambda,\varphi,T} := 2\lambda_T^* + \lambda_T^* KT + \lambda_T^{2*} K^2(1+r)T + C + C',$$

which is a positive constant not depending on $\delta$ or $\epsilon$. $\qquad\square$

**Lemma 6.3** *For every $\nu \in \mathbb{M}_T$, every $u \in \mathbb{U}$ and every $\gamma, \delta > 0$, there exist positive constants $C_u$, $C'_u$, $K_u$ and $K_{Q,u}$ not depending on $\epsilon$ or $\delta$ such that*

$$\mathscr{E}(\tilde N^\epsilon)_T \geqslant K_u \exp\left( -C_u\delta d - \gamma d - C'_u Td - \frac{1}{\epsilon}K_{Q,u}\delta d - \frac{1}{\epsilon}\int_0^T \sum_{i=1}^d \frac{Qu(s,i)}{u(s,i)} K_\nu(s,i)\mathrm{d}s \right)$$

*on the set $\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}$.*

**Proof** First observe that

$$
\begin{aligned}
\mathscr{E}(\tilde N^\epsilon)_T &= \frac{u(T, X_T^\epsilon)}{u(0, X_0^\epsilon)} \exp\left( -\int_0^T \sum_{i=1}^d \frac{\frac{\partial}{\partial s}u(s,i) + (Q^\epsilon u)(s,i)}{u(s,i)} \mathbf{1}_{\{X_s^\epsilon=i\}}\mathrm{d}s \right) \\
&= \frac{u(T, X_T^\epsilon)}{u(0, X_0^\epsilon)} \exp\left( -\sum_{i=1}^d \int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)} [\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right. \\
&\qquad\qquad\qquad\qquad \left. -\sum_{i=1}^d \int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)} K_\nu(s,i)\mathrm{d}s \right) \times \\
&\quad \exp\left( -\frac{1}{\epsilon}\sum_{i=1}^d \int_0^T \frac{Qu(s,i)}{u(s,i)} [\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s - \frac{1}{\epsilon}\sum_{i=1}^d \int_0^T \frac{Qu(s,i)}{u(s,i)} K_\nu(s,i)\mathrm{d}s \right)
\end{aligned}
$$

By the definition of $u$ and $X_0^\epsilon = x$, we have that $K_u := \min_{i,x} u(T,i)/u(0,x)$ is a positive constant. Hence $\mathscr{E}(\tilde N^\epsilon)_T$ majorizes

$$
K_u \exp\left( -\sum_{i=1}^d \left| \int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)} [\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right| - \sum_{i=1}^d \left| \int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)} K_\nu(s,i)\mathrm{d}s \right| \right) \times
$$

$$
\exp\left( -\frac{1}{\epsilon}\sum_{i=1}^d \left| \int_0^T \frac{Qu(s,i)}{u(s,i)} [\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right| - \frac{1}{\epsilon}\sum_{i=1}^d \int_0^T \frac{Qu(s,i)}{u(s,i)} K_\nu(s,i)\mathrm{d}s \right)
$$

On the set $\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}$, Lemma A.1 implies that, for any $\gamma > 0$, $i \in \mathbb{S}$,

$$\left| \int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)} [\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s \right| \leqslant C_u\delta + \gamma,$$

19

$$\left| \int_0^T \frac{Qu(s,i)}{u(s,i)} [\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s,i)] \mathrm{d}s \right| \leqslant K_{Q,u} \delta, \quad \forall i \in \mathbb{S}.$$

Since $\frac{\partial}{\partial s} u(s,i)/u(s,i)$ is continuous on $[0,T]$,

$$\left| \int_0^T \frac{\frac{\partial}{\partial s} u(s,i)}{u(s,i)} K_\nu(s,i) \mathrm{d}s \right| \leqslant C_u' T.$$

Hence,

$$\mathscr{E}(\tilde{N}^\epsilon)_T \geqslant K_u \exp\left( -C_u \delta d - \gamma d - C_u' T d - \frac{1}{\epsilon} K_{Q,u} \delta d - \frac{1}{\epsilon} \int_0^T \sum_{i=1}^d \frac{Qu(s,i)}{u(s,i)} K_\nu(s,i) \mathrm{d}s \right).$$

We have thus proven our claim. $\qquad\qquad\square$

Now we are ready to prove the upper bound in the local LDP.

**Proposition 6.4** *For every $(\varphi, \nu) \in \mathbb{C}_T \times \mathbb{M}_T$,*

$$\limsup_{\delta \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) \leqslant -L_T(\varphi, \nu).$$

**Proof** Due to Lemma 5.1, $\mathbb{C}_T \times \mathbb{M}_T^{++}$ is dense in $\mathbb{C}_T \times \mathbb{M}_T$. We first prove that the upper bound holds on $\mathbb{C}_T \times \mathbb{M}_T^{++}$. For every $\nu \in \mathbb{M}_T^{++}$, it is an immediate implication of Lemma 5.2 that there is an optimizer $u^*(\cdot, \cdot)$ of

$$\inf_{u \in U} \left[ \sum_{i=1}^d \frac{(Qu)(i)}{u(i)} K_\nu(s,i) \right]$$

such that $u^* \in \mathbb{U}$. We denote

$$\mathscr{E}_t^{u^*} = \frac{u^*(t, X_t^\epsilon)}{u^*(0, X_0^\epsilon)} \exp\left( -\int_0^t \frac{\frac{\partial}{\partial s} u^*(s, X_s^\epsilon) + (Q^\epsilon u^*)(s, X_s^\epsilon)}{u^*(s, X_s^\epsilon)} \mathrm{d}s \right).$$

Lemma 6.1 implies that

$$\mathbb{E}\left[ \mathbf{1}_{\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}} \mathscr{E}_T^{u^*} \mathscr{E}(N^\epsilon)_T \right] \leqslant 1 \tag{14}$$

for every $\lambda \in \mathbb{S}_T$. By virtue of Lemmas 6.2 and 6.3, we have a non-random lower bound for $\mathscr{E}_T^{u^*} \mathscr{E}(N^\epsilon)_T$ on the set $\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}$. Hence, (14) implies that, for all $\lambda \in \mathbb{S}_T$,

$$\mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta)$$
$$\leqslant \frac{1}{K_{u^*}} \exp\left( C_{u^*} \delta d + C_{u^*}' T d + \frac{1}{\epsilon} K_{Q,u^*} \delta d + \frac{1}{\epsilon} \int_0^T \sum_{i=1}^d \frac{Qu^*(s,i)}{u^*(s,i)} K_\nu(s,i) \mathrm{d}s \right) \times$$
$$\exp\left\{ -\frac{1}{\epsilon} \left[ \int_0^T \lambda(s) \mathrm{d}\varphi_s - \int_0^T \lambda(s) \hat{b}(\nu, \varphi_s) \mathrm{d}s - \int_0^T \frac{\lambda^2(s)}{2} \hat{\sigma}^2(\nu, \varphi_s) \mathrm{d}s \right] + \frac{\delta}{\epsilon} K_{\lambda,\varphi,T} \right\}.$$

We observe that

$$\int_0^T \sum_{i=1}^d \frac{Qu^*(s,i)}{u^*(s,i)} K_\nu(s,i) \mathrm{d}s = -\int_0^T \sup_{u \in U} \left[ -\sum_{i=1}^d \frac{(Qu)(i)}{u(i)} K_\nu(s,i) \right] \mathrm{d}s = -\tilde{I}_T(\nu).$$

20

It directly entails that, again for all $\lambda \in \mathbb{S}_T$,

$$\epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta)$$
$$\leqslant -\left[\int_0^T \lambda(s)\mathrm{d}\varphi_s - \int_0^T \lambda(s)\hat{b}(\nu, \varphi_s)\mathrm{d}s - \int_0^T \frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu, \varphi_s)\mathrm{d}s\right] + K_{\lambda,\varphi,T}\delta$$
$$-\epsilon \log K_{u^*} + \epsilon\left(C_{u^*}\delta d + C'_{u^*}Td\right) + K_{Q,u^*}\delta d - \tilde{I}_T(\nu). \tag{15}$$

It is easily seen that all the terms with $\delta$ or $\epsilon$ vanish as $\delta \to 0$, $\epsilon \to 0$. As a consequence we conclude, by minimizing the right hand-side over $\lambda$, that the decay rate

$$\limsup_{\delta \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta)$$

is majorized by

$$-\sup_{\lambda \in \mathbb{S}_T}\left[\int_0^T \lambda(s)\mathrm{d}\varphi_s - \int_0^T \lambda(s)\hat{b}(\nu, \varphi_s)\mathrm{d}s - \int_0^T \frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu, \varphi_s)\mathrm{d}s\right] - \tilde{I}_T(\nu).$$

Since $b(i,x)$ and $\sigma(i,x)$ satisfy the linear growth condition (A.2), $\hat{b}(\nu,x)$ and $\hat{\sigma}(\nu,x)$ are of linear growth as well. Then Liptser and Pukhalskii [19, Lemma 6.1] implies that

$$\sup_{\lambda \in \mathbb{S}_T}\left[\int_0^T \lambda(s)\mathrm{d}\varphi_s - \int_0^T \lambda(s)\hat{b}(\nu, \varphi_s)\mathrm{d}s - \int_0^T \frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu, \varphi_s)\mathrm{d}s\right]$$
$$= \begin{cases} \int_0^T \sup_{\lambda \in \mathbb{R}}\left[\lambda\varphi'_s - \lambda\hat{b}(\nu, \varphi_s) - \frac{\lambda^2}{2}\hat{\sigma}^2(\nu, \varphi_s)\right]\mathrm{d}s & \text{if } \varphi \in \mathbb{H}_T, \\ \infty & \text{otherwise.} \end{cases}$$

For $s \in [0,T]$ such that $\hat{\sigma}^2(\nu, \varphi_s) \neq 0$ and $\varphi \in \mathbb{H}_T$, it is well-known (cf. Fredlin and Wentzell [9], Liptser [18]) that

$$\sup_{\lambda \in \mathbb{R}}\left[\lambda\varphi'_s - \lambda\hat{b}(\nu, \varphi_s) - \frac{\lambda^2}{2}\hat{\sigma}^2(\nu, \varphi_s)\right] = \frac{[\varphi'_s - \hat{b}(\nu, \varphi_s)]^2}{2\hat{\sigma}^2(\nu, \varphi_s)}.$$

For $s \in [0,T]$ such that $\hat{\sigma}^2(\nu, \varphi_s) = 0$ and $\varphi \in \mathbb{H}_T$,

$$\sup_{\lambda \in \mathbb{R}}\left[\lambda\varphi'_s - \lambda\hat{b}(\nu, \varphi_s) - \frac{\lambda^2}{2}\hat{\sigma}^2(\nu, \varphi_s)\right] = \begin{cases} 0 & \text{if } \varphi'_s = \hat{b}(\nu, \varphi_s), \\ \infty & \text{otherwise.} \end{cases}$$

Hence, with the conventions $0/0 = 0$ and $n/0 = \infty$ (for all $n > 0$) being in force,

$$\int_0^T \sup_{\lambda \in \mathbb{R}}\left[\lambda\varphi'_s - \lambda\hat{b}(\nu, \varphi_s) - \frac{\lambda^2}{2}\hat{\sigma}^2(\nu, \varphi_s)\right]\mathrm{d}s = \frac{1}{2}\int_0^T \frac{[\varphi'_t - \hat{b}(\nu, \varphi_t)]^2}{\hat{\sigma}^2(\nu, \varphi_t)}\mathrm{d}t$$

if $\varphi \in \mathbb{H}_T$.

Hence the lower bound for the dense subset $\mathbb{C}_T \times \mathbb{M}_T^{++}$ is established. In consideration of Lemma 2.5, the upper bound is proved for $\mathbb{C}_T \times \mathbb{M}_T$ if we can show $I_T(\varphi, \nu)$ and $\tilde{I}_T(\nu)$ are lower semi-continuous on $\nu$. We denote

$$F_\lambda(\varphi, \nu) = \int_0^T \lambda(s)\mathrm{d}\varphi_s - \int_0^T \lambda(s)\hat{b}(\nu, \varphi_s)\mathrm{d}s - \int_0^T \frac{\lambda^2(s)}{2}\hat{\sigma}^2(\nu, \varphi_s)\mathrm{d}s.$$

By the above computation, we know for every $(\varphi, \nu) \in \mathbb{C}_T \times \mathbb{M}_T$, $I_T(\varphi, \nu) = \sup_{\lambda \in \mathbb{S}_T} F_\lambda(\varphi, \nu)$. For every $\lambda \in \mathbb{S}_T$, $F_\lambda(\nu, \varphi)$ is continuous on $\nu$ due to Lemma A.1. Then $I_T(\varphi, \nu)$ is lower semi-continuous on $\nu$ since it is the pointwise supremum of continuous functions. By Lemma 5.3, $\tilde{I}_T(\nu)$ also satisfies the requirement. The claim is established. $\qquad\square$

# 7 Lower bound for the local LDP

This section studies the lower bound of the local LDP. To this end, it is realized that only finite rate functions need to be investigated. The rate function $\tilde{I}_T(\nu)$ is finite for every $\nu \in \mathbb{M}_T$ since $0 \leqslant \tilde{I}_T(\nu) \leqslant -T \sum_{i=1}^d Q_{ii}$. We further observe that the rate function $I_T(\varphi, \nu)$ is finite for every $(\varphi, \nu) \in \mathbb{H}_T \times \mathbb{M}_T$ if $\inf_{i,x} \sigma^2(i,x) > 0$. Hence we consider the case of $\inf_{i,x} \sigma^2(i,x) > 0$ first. Let $(\varphi, \nu) \in \mathbb{H}_T \times \mathbb{M}_T$. We define

$$\bar{N}_t^\epsilon := \frac{1}{\sqrt{\epsilon}} \int_0^t \frac{\varphi_s' - \hat{b}(\varphi_s, \nu)}{\hat{\sigma}(\varphi_s, \nu)} dB_s. \tag{16}$$

Then its stochastic exponential is

$$\mathscr{E}(\bar{N}^\epsilon)_t = \exp\left( \bar{N}_t^\epsilon - \frac{1}{2} \langle \bar{N}^\epsilon \rangle_t \right), \quad \text{where} \quad \langle \bar{N}^\epsilon \rangle_t = \frac{1}{\epsilon} \int_0^t \left[ \frac{\varphi_s' - \hat{b}(\varphi_s, \nu)}{\hat{\sigma}(\varphi_s, \nu)} \right]^2 ds.$$

For simplicity, we denote

$$h_s := \frac{\varphi_s' - \hat{b}(\varphi_s, \nu)}{\hat{\sigma}(\varphi_s, \nu)}$$

throughout this section. Recall from (13) that, for a given $u(\cdot, \cdot) \in \mathbb{U}$,

$$\mathscr{E}(\tilde{N}^\epsilon)_t = \frac{u(t, X_t^\epsilon)}{u(0, X_0^\epsilon)} \exp\left( - \int_0^t \frac{\frac{\partial}{\partial s} u(s, X_s^\epsilon) + (Q^\epsilon u)(s, X_s^\epsilon)}{u(s, X_s^\epsilon)} ds \right).$$

In order to perform a change of measure in Proposition 7.3, we show that $\{\mathscr{E}(\tilde{N}^\epsilon)_t \mathscr{E}(\bar{N}^\epsilon)_t\}_{t \in [0,T]}$ is a true martingale. It is noted that in the first results of this section, we impose the condition $\inf_{i,x} \sigma^2(i,x) > 0$, which will be lifted later on.

**Lemma 7.1** *For every $(\varphi, \nu) \in \mathbb{H}_T \times \mathbb{M}_T$ and $u(\cdot, \cdot) \in \mathbb{U}$, $\{\mathscr{E}(\tilde{N}^\epsilon)_t \mathscr{E}(\bar{N}^\epsilon)_t\}_{t \in [0,T]}$ is a martingale if $\inf_{i,x} \sigma^2(i,x) > 0$.*

**Proof** We have shown in last section that $\mathscr{E}(\tilde{N}^\epsilon)_t$ is a martingale. Since $\varphi \in \mathbb{H}_T$ and recalling that we assumed $\inf_{i,x} \sigma^2(i,x) > 0$, it follows that $\langle \bar{N}^\epsilon \rangle_T = \frac{1}{\epsilon} \int_0^T h_s^2 ds < \infty$. Then Novikov's condition implies that $\mathscr{E}(\bar{N}^\epsilon)_t$ is also a martingale. Since $X_t^\epsilon$ is independent of the Brownian motion $B_t$, $\mathscr{E}(\tilde{N}^\epsilon)_t$ is also independent of $\mathscr{E}(\bar{N}^\epsilon)_t$. So,

$$\mathbb{E}[\mathscr{E}(\tilde{N}^\epsilon)_T \mathscr{E}(\bar{N}^\epsilon)_T] = \mathbb{E}[\mathscr{E}(\tilde{N}^\epsilon)_T] \mathbb{E}[\mathscr{E}(\bar{N}^\epsilon)_T] = \mathbb{E}[\mathscr{E}(\tilde{N}^\epsilon)_0] \mathbb{E}[\mathscr{E}(\bar{N}^\epsilon)_0] = \mathbb{E}[\mathscr{E}(\tilde{N}^\epsilon)_0 \mathscr{E}(\bar{N}^\epsilon)_0].$$

By the same reasoning as in the proof of Lemma 6.1, we know that $\mathscr{E}(\tilde{N}^\epsilon)_t \mathscr{E}(\bar{N}^\epsilon)_t$ is a supermartingale. Hence, it is a martingale by Liptser and Shiryaev [21, Lemma 6.4]. $\qquad\square$

**Lemma 7.2** *For every $\nu \in \mathbb{M}_T$, every $u \in \mathbb{U}$ and every $\gamma, \delta > 0$, there exist positive constants $C_u$, $C_u'$, $K_u'$ and $K_{Q,u}$ not depending on $\epsilon$ or $\delta$ such that*

$$[\mathscr{E}(\tilde{N}^\epsilon)_T]^{-1} \geqslant K_u' \exp\left( -C_u \delta d - \gamma d - C_u' T d - \frac{1}{\epsilon} K_{Q,u} \delta d + \frac{1}{\epsilon} \int_0^T \sum_{i=1}^d \frac{Qu(s,i)}{u(s,i)} K_\nu(s,i) ds \right)$$

*on the set $\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}$.*

**Proof** According to the computation in Lemma 6.3, we have $[\mathscr{E}(\tilde{N}^\epsilon)_T]^{-1}$ is equal to

$$\frac{u(0, X_0^\epsilon)}{u(T, X_T^\epsilon)} \exp\left(\sum_{i=1}^d \int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)}[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s + \sum_{i=1}^d \int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)}K_\nu(s,i)\mathrm{d}s\right) \times$$

$$\exp\left(\frac{1}{\epsilon}\sum_{i=1}^d \int_0^T \frac{Qu(s,i)}{u(s,i)}[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s + \frac{1}{\epsilon}\sum_{i=1}^d \int_0^T \frac{Qu(s,i)}{u(s,i)}K_\nu(s,i)\mathrm{d}s\right)$$

Defining $K_u' := \min_{i,j} u(0,j)/u(T,i)$, we have that $[\mathscr{E}(\tilde{N}^\epsilon)_T]^{-1}$ is not less than

$$K_u' \exp\left(-\sum_{i=1}^d \left|\int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)}[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s\right| - \sum_{i=1}^d \left|\int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)}K_\nu(s,i)\mathrm{d}s\right|\right) \times$$

$$\exp\left(-\frac{1}{\epsilon}\sum_{i=1}^d \left|\int_0^T \frac{Qu(s,i)}{u(s,i)}[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s\right| + \frac{1}{\epsilon}\sum_{i=1}^d \int_0^T \frac{Qu(s,i)}{u(s,i)}K_\nu(s,i)\mathrm{d}s\right)$$

On the set $\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}$, Lemma A.1 implies

$$\left|\int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)}[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s\right| \leqslant C_u\delta + \gamma, \quad \forall i \in \mathbb{S}, \forall \gamma > 0,$$

$$\left|\int_0^T \frac{Qu(s,i)}{u(s,i)}[\mathbf{1}_{\{X_s^\epsilon=i\}} - K_\nu(s,i)]\mathrm{d}s\right| \leqslant K_{Q,u}\delta, \quad \forall i \in \mathbb{S}.$$

Also,

$$\left|\int_0^T \frac{\frac{\partial}{\partial s}u(s,i)}{u(s,i)}K_\nu(s,i)\mathrm{d}s\right| \leqslant C_u'T.$$

Hence,

$$[\mathscr{E}(\tilde{N}^\epsilon)_T]^{-1} \geqslant K_u' \exp\left(-C_u\delta d - \gamma d - C_u'Td - \frac{1}{\epsilon}K_{Q,u}\delta d + \frac{1}{\epsilon}\int_0^T \sum_{i=1}^d \frac{Qu(s,i)}{u(s,i)}K_\nu(s,i)\mathrm{d}s\right).$$

We have thus derived the desired lower bound. □

We proceed to prove the lower bound of the local LDP under the condition $\inf_{i,x} \sigma^2(i,x) > 0$.

**Proposition 7.3** *For every $(\varphi, \nu) \in \mathbb{H}_T \times \mathbb{M}_T$, if $\inf_{i,x} \sigma^2(i,x) > 0$,*

$$\liminf_{\delta \to 0} \liminf_{\epsilon \to 0} \epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) \geqslant -L_T(\varphi, \nu).$$

**Proof** For any $\nu \in \mathbb{M}_T$, there is a sequence $\nu^\eta \in \mathbb{M}_T^{++}$ such that $\nu^\eta \to \nu$ by Lemma 5.1. Actually, the convergence happens in the way that $K_\nu^\eta(\cdot, i) \to K_\nu(\cdot, i)$ a.e.. Then by Lemma 5.3, the rate function $L_T(\varphi, \nu)$ satisfies the continuity property required in Lemma 2.5. Hence we only need to prove the lower bound on the dense subset $\mathbb{H}_T \times \mathbb{M}_T^{++}$. Recall that for every $\nu \in \mathbb{M}_T^{++}$, Lemma 5.2 implies that there is an optimizer $u^*(\cdot, \cdot)$ of

$$\inf_{u \in U}\left[\sum_{i=1}^d \frac{(Qu)(i)}{u(i)}K_\nu(s,i)\right]$$

23

such that $u^* \in \mathbb{U}$ and

$$\mathscr{E}_t^{u^*} = \frac{u^*(t, X_t^\epsilon)}{u^*(0, X_0^\epsilon)} \exp\left(-\int_0^t \frac{\frac{\partial}{\partial s} u^*(s, X_s^\epsilon) + (Q^\epsilon u^*)(s, X_s^\epsilon)}{u^*(s, X_s^\epsilon)} \mathrm{d}s\right).$$

By Lemma 7.1, we know $\mathbb{E}[\mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T] = 1$. On $(\Omega, \mathscr{F}_T)$, we define a new probability measure $\mathbb{P}_{u^*}$ through $\mathrm{d}\mathbb{P}_{u^*} = \mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T \mathrm{d}\mathbb{P}$. Since $\mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T$ is strictly positive, $\mathbb{P}_{u^*}$ is equivalent to $\mathbb{P}$ and $\mathrm{d}\mathbb{P} = \left[\mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T\right]^{-1} \mathrm{d}\mathbb{P}_{u^*}$. So that we can translate the probability of our interest under the original measure $\mathbb{P}$ into the mean of a certain random quantity under the alternative measure $\mathbb{P}_{u^*}$:

$$\mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) = \int_{\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta\}} \left[\mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T\right]^{-1} \mathrm{d}\mathbb{P}_{u^*}. \qquad (17)$$

By Girsanov's theorem, $\tilde{B}_t := B_t - \frac{1}{\sqrt{\epsilon}} \int_0^t h_s \mathrm{d}s$ is a $\mathbb{P}_{u^*}$-Brownian motion on $(\Omega, (\mathscr{F}_t)_{t \leqslant T})$. We substitute the above equation in (16), and obtain

$$\bar{N}_T^\epsilon - \frac{1}{2} \langle \bar{N}^\epsilon \rangle_T = \frac{1}{\sqrt{\epsilon}} \int_0^T h_s \mathrm{d}\tilde{B}_s + \frac{1}{2\epsilon} \int_0^T h_s^2 \mathrm{d}s.$$

It thus follows that $[\mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T]^{-1}$ is equal to

$$\frac{u^*(0, X_0^\epsilon)}{u^*(t, X_t^\epsilon)} \exp\left(\int_0^t \frac{\frac{\partial}{\partial s} u^*(s, X_s^\epsilon) + (Q^\epsilon u^*)(s, X_s^\epsilon)}{u^*(s, X_s^\epsilon)} \mathrm{d}s - \frac{1}{\sqrt{\epsilon}} \int_0^T h_s \mathrm{d}\tilde{B}_s - \frac{1}{2\epsilon} \int_0^T h_s^2 \mathrm{d}s\right).$$

Now let $L$ be a positive constant. We define $\Theta^\epsilon := \left\{\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta, \left|\int_0^T h_s \mathrm{d}\tilde{B}_s\right| \leqslant L\right\}$. Then (17) implies

$$\mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) \geqslant \int_{\Theta^\epsilon} \left[\mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T\right]^{-1} \mathrm{d}\mathbb{P}_{u^*}.$$

By Lemma 7.2, we obtain the following non-random lower bound of $[\mathscr{E}_T^{u^*} \mathscr{E}(\bar{N}^\epsilon)_T]^{-1}$, valid on the set $\Theta^\epsilon$:

$$K'_{u^*} \exp\left(-C_{u^*}\delta d - C'_{u^*} T d - \frac{1}{\epsilon} K_{Q,u^*} \delta d - \frac{\tilde{I}_T(\nu)}{\epsilon} - \frac{I_T(\varphi, \nu)}{\epsilon} - \frac{L}{\sqrt{\epsilon}}\right).$$

As a consequence, we have the following lower bound of the probability $\mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta)$

$$K'_{u^*} \exp\left(-C_{u^*}\delta d - C'_{u^*} T d - \frac{1}{\epsilon} K_{Q,u^*} \delta d - \frac{\tilde{I}_T(\nu)}{\epsilon} - \frac{I_T(\varphi, \nu)}{\epsilon} - \frac{L}{\sqrt{\epsilon}}\right) \times \mathbb{P}_{u^*}(\Theta^\epsilon).$$

This, in turn, leads to the following lower bound on the corresponding exponential decay rate:

$$\begin{aligned}
\epsilon \log \mathbb{P}(\rho_T(M^\epsilon, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \delta) \geqslant \quad & \epsilon \log K'_{u^*} - \epsilon(C_{u^*}\delta d + C'_{u^*} T d) - K_{Q,u^*} \delta d \\
& -\tilde{I}_T(\nu) - I_T(\varphi, \nu) - \sqrt{\epsilon} L + \epsilon \log \mathbb{P}_{u^*}(\Theta^\epsilon). \quad (18)
\end{aligned}$$

Then a sufficient condition for desired result to hold is $\lim_{\epsilon \to 0} \mathbb{P}_{u^*}(\Theta^\epsilon) > 0$. It is evident that

$$\mathbb{P}_{u^*}(\Theta^\epsilon) \geqslant 1 - \mathbb{P}_{u^*}\left(\left|\int_0^T h_s \mathrm{d}\tilde{B}_s\right| > L\right) - \mathbb{P}_{u^*}(d_T(\nu^\epsilon, \nu) > \delta) - \mathbb{P}_{u^*}(\rho_T(M^\epsilon, \varphi) > \delta). \quad (19)$$

We proceed by consecutively proving that the three probabilities in the right-hand side of (19) vanishes as $\epsilon \to 0$. We start by analyzing the first probability. By Chebyshev's inequality,

$$\tilde{\mathbb{P}}_{u^*}\left(\left|\int_0^T h_s \mathrm{d}\tilde{B}_s\right| > L\right) \leqslant \frac{\tilde{\mathbb{E}}_{u^*}\left|\int_0^T h_s \mathrm{d}\tilde{B}_s\right|^2}{L^2} = \frac{\int_0^T h_s^2 \mathrm{d}s}{L^2}.$$

Since $\int_0^T h_s^2 \mathrm{d}s < \infty$, we can make this upper bound arbitrarily small by picking $L$ sufficiently large.

Next we consider the second probability in the right-hand sider of (19). We notice that the part

$$\exp\left(\frac{1}{\sqrt{\epsilon}}\int_0^T h_s \mathrm{d}\tilde{B}_s + \frac{1}{2\epsilon}\int_0^T h_s^2 \mathrm{d}s\right)$$

in the change of measure procedure is not related to the Markov chain . Then by Proposition 11.2.3 in Bielecki and Rutkowski [3], a Markov chain $X_t$ with transition intensity matrix $Q$ under $\mathbb{P}$ becomes a Markov chain under $\mathbb{P}_{u^*}$ with transition intensity matrix $Q(u^*)(t)$ where

$$Q(u^*)(t)_{ij} = Q_{ij}\frac{u^*(t,j)}{u^*(t,i)} \;\; \text{for } i \neq j; \quad Q(u^*)(t)_{ii} = -\sum_{j \neq i} Q_{ij}\frac{u^*(t,j)}{u^*(t,i)}.$$

Hence, $Q(u^*)(t)/\epsilon$ is the transition intensity matrix of $X_t^\epsilon$ under $\mathbb{P}_{u^*}$. By Lemma A.2, for every $t \in [0,T]$, $\boldsymbol{K}_\nu(t) = (K_\nu(t,1), \ldots, K_\nu(t,d))$ is the unique solution of

$$\mu(t)Q(u^*)(t) = 0, \quad \sum_{i=1}^d \mu(t,i) = 1, \quad \mu(t,i) \geqslant 0.$$

Also, all entries of the matrix $Q(u^*)(t)$ are smooth on $[0,T]$ by Lemma 5.2. Then by Corollary 5.8 in Yin and Zhang [27],

$$\mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t \mathbf{1}_{\{X_s^\epsilon=i\}}\mathrm{d}s - \int_0^t K_\nu(i,s)\mathrm{d}s\right| > \epsilon^{1/4}\right) \leqslant K \exp\left\{-\frac{C_T}{\epsilon^{1/4}(T+1)^{3/2}}\right\},$$

where $C_T$ is a strictly positive constant. For any $\delta > 0$, the following (obvious) inequality holds for every $\epsilon$ such that $\epsilon \in (0, \delta^4)$

$$\mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t \mathbf{1}_{\{X_s^\epsilon=i\}}\mathrm{d}s - \int_0^t K_\nu(i,s)\mathrm{d}s\right| > \delta\right)$$
$$\leqslant \mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t \mathbf{1}_{\{X_s^\epsilon=i\}}\mathrm{d}s - \int_0^t K_\nu(i,s)\mathrm{d}s\right| > \epsilon^{1/4}\right).$$

Hence we have

$$\mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t \mathbf{1}_{\{X_s^\epsilon=i\}}\mathrm{d}s - \int_0^t K_\nu(i,s)\mathrm{d}s\right| > \delta\right) \to 0 \text{ as } \epsilon \to 0.$$

That is, $\mathbb{P}_{u^*}(d_T(\nu^\epsilon, \nu) > \delta) \to 0$, as $\epsilon \to 0$.

Now we proceed by showing the third probability in the right-hand side of (19) vanishes as $\epsilon \to 0$. We substitute $\tilde{B}_t$ for $B_t$ in (1), yielding

$$M_t^\epsilon = \int_0^t b(X_s^\epsilon, M_s^\epsilon) + h_s\sigma(X_s^\epsilon, M_s^\epsilon)\mathrm{d}s + \sqrt{\epsilon}\int_0^t \sigma(X_s^\epsilon, M_s^\epsilon)\mathrm{d}\tilde{B}_s.$$

By setting $\tilde{M}_t^\epsilon := M_t^\epsilon - \varphi_t$, we obtain

$$
\begin{aligned}
\tilde{M}_t^\epsilon \;=\;& \sqrt{\epsilon}\int_0^t \sigma(X_s^\epsilon, M_s^\epsilon)\mathrm{d}\tilde{B}_s + \int_0^t [b(X_s^\epsilon, M_s^\epsilon) - b(X_s^\epsilon, \varphi_s)]\mathrm{d}s + \int_0^t [b(X_s^\epsilon, \varphi_s) - \hat{b}(\nu, \varphi_s)]\mathrm{d}s \\
&+ \int_0^t h_s[\sigma(X_s^\epsilon, M_s^\epsilon) - \sigma(X_s^\epsilon, \varphi_s)]\mathrm{d}s + \int_0^t h_s[\sigma(X_s^\epsilon, \varphi_s) - \hat{\sigma}(\nu, \varphi_s)]\mathrm{d}s.
\end{aligned}
$$

Using the Lipschitz continuity featuring in (A1), we find that both

$$
\sup_{t\leqslant T}\left|\int_0^t [b(X_s^\epsilon, M_s^\epsilon) - b(X_s^\epsilon, \varphi_s)]\mathrm{d}s\right| \leqslant \int_0^T K\tilde{M}_s^{\epsilon*}\mathrm{d}s,
$$

and

$$
\sup_{t\leqslant T}\left|\int_0^t h_s[\sigma(X_s^\epsilon, M_s^\epsilon) - \sigma(X_s^\epsilon, \varphi_s)]\mathrm{d}s\right| \leqslant \int_0^T |h_s|K\tilde{M}_s^{\epsilon*}\mathrm{d}s.
$$

Recalling that we denote throughout this paper running maximum processes by adding an asterisk ('*'), it is now immediate that

$$
\tilde{M}_T^{\epsilon*} \leqslant I_T^{1*} + I_T^{2*} + I_T^{3*} + \int_0^T K(1 + |h_s|)\tilde{M}_s^{\epsilon*}\mathrm{d}s,
$$

where

$$
\begin{aligned}
I_t^1 \;&:=\; \sqrt{\epsilon}\int_0^t \sigma(X_s^\epsilon, M_s^\epsilon)\mathrm{d}\tilde{B}_s, \quad I_t^2 := \int_0^t [b(X_s^\epsilon, \varphi_s) - \hat{b}(\nu, \varphi_s)]\mathrm{d}s, \\
I_t^3 \;&:=\; \int_0^t h_s[\sigma(X_s^\epsilon, \varphi_s) - \hat{\sigma}(\nu, \varphi_s)]\mathrm{d}s.
\end{aligned}
$$

Then Gronwall's inequality implies

$$
\tilde{M}_T^{\epsilon*} \leqslant [I_T^{1*} + I_T^{2*} + I_T^{3*}]\exp\left(\int_0^T K(1 + |h_s|)\mathrm{d}s\right). \tag{20}
$$

The next step is to study the impact of $I_T^{1*}$, $I_T^{2*}$, and $I_T^{3*}$ separately. For any $\delta > 0$, it is an immediate consequence of Chebyshev's inequality that $\mathbb{P}_{u^*}(I_T^{1*} > \delta) \leqslant \delta^{-3}\tilde{\mathbb{E}}_{u^*}[(I_T^{1*})^3]$. We notice the close similarity between $I_t^1$ and $C_t^\epsilon$ in the proof of Proposition 4.3. The quantity $\mathbb{E}_{u^*}[(I_T^{1*})^3]$ can be dealt with using essentially the same procedure that was used to bound $\mathbb{E}[(C_T^{\epsilon*})^{1/\epsilon}]$: we derive an inequality similar to (9), i.e.,

$$
\mathbb{E}_{u^*}[(I_T^{1*})^3] \leqslant \frac{27}{8}\mathbb{E}_{u^*}\left[3\epsilon\int_0^T |I_s^1|\sigma^2(X_s^\epsilon, M_s^\epsilon)\mathrm{d}s\right].
$$

We thus obtain

$$
\mathbb{E}_{u^*}[(I_T^{1*})^3] \leqslant \frac{81\epsilon kT}{8}\exp\left(\frac{81\epsilon kT}{8}\right),
$$

where $k$ is a positive constant. As a consequence, $\lim_{\epsilon\to 0}\mathbb{P}_{u^*}(I_T^{1*} > \delta) = 0$.

The claim $\lim_{\epsilon\to 0}\mathbb{P}_{u^*}(I_T^{2*} > \delta) = 0$ can be established as follows. As a first stap we observe that since

$$
\begin{aligned}
\sup_{t\leqslant T}\left|\int_0^t [b(X_s^\epsilon, \varphi_s) - \hat{b}(\nu, \varphi_s)]\mathrm{d}s\right| \;&=\; \sup_{t\leqslant T}\left|\sum_{i=1}^d \int_0^t b(i, \varphi_s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right| \\
&\leqslant\; d\sup_{t\leqslant T, i\in\mathbb{S}}\left|\int_0^t b(i, \varphi_s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right|,
\end{aligned}
$$

26

the following upper bound applies:

$$\mathbb{P}_{u^*}(I_T^{2*} > \delta) = \mathbb{P}_{u^*}\left(\sup_{t \leqslant T}\left|\int_0^t [b(X_s^\epsilon, \varphi_s) - \hat{b}(\nu, \varphi_s)]\mathrm{d}s\right| > \delta\right)$$

$$\leqslant \mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t b(i, \varphi_s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right| > \delta/d\right)$$

Since $b(i, x)$ is Lipschitz continuous in $x$ and $\varphi_t$ is absolutely continuous, $b(i, \varphi_t)$ is bounded on $[0, T]$. Then by Corollary 5.8 in Yin and Zhang [27] again, $\mathbb{P}_{u^*}(I_T^{2*} > \delta) \to 0$, as $\epsilon \to 0$.

Similar to the above computation, we can obtain that

$$\mathbb{P}_{u^*}(I_T^{3*} > \delta) \leqslant \mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t h_s \sigma(i, \varphi_s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right| > \delta/d\right).$$

We know that $h_s \sigma(i, \varphi_s)$ is square-integrable for every $i \in \mathbb{S}$. Then by the method of mollification in Theorem C.6 in Evans [7], there exists a sequence of smooth functions $h^\eta(i, s)$ such that $h^\eta(i, s) \to h_s \sigma(i, \varphi_s)$ as $\eta \to 0$ in $L^2[0, T]$. By the Cauchy-Schwarz inequality,

$$\left|\int_0^t [h_s \sigma(i, \varphi_s) - h^\eta(i, s)][\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right| \leqslant \sqrt{2t}\left(\int_0^t [h_s \sigma(i, \varphi_s) - h^\eta(i, s)]^2 \mathrm{d}s\right)^{1/2}.$$

Then,

$$\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t h_s \sigma(i, \varphi_s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right|$$

$$\leqslant \sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t h^\eta(i, s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right| + \sup_{i \in \mathbb{S}}\sqrt{2T}\left(\int_0^T [h_s \sigma(i, \varphi_s) - h^\eta(i, s)]^2 \mathrm{d}s\right)^{1/2}.$$

We let $H(\eta) := \sup_{i \in \mathbb{S}}\sqrt{2T}\left(\int_0^T [h_s \sigma(i, \varphi_s) - h^\eta(i, s)]^2 \mathrm{d}s\right)^{1/2}$. It is clear that $H(\eta) \to 0$ as $\eta \to 0$. Hence,

$$\mathbb{P}_{u^*}(I_T^{3*} > \delta) \leqslant \mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t h^\eta(i, s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right| + H(\eta) > \delta/d\right).$$

For any $\delta > 0$, we can choose all $\eta > 0$ small enough such that $H(\eta) < \delta/2d$. It yields

$$\mathbb{P}_{u^*}(I_T^{3*} > \delta) \leqslant \mathbb{P}_{u^*}\left(\sup_{t \leqslant T, i \in \mathbb{S}}\left|\int_0^t h^\eta(i, s)[\mathbf{1}_{\{X_s^\epsilon = i\}} - K_\nu(s, i)]\mathrm{d}s\right| > \delta/2d\right).$$

Since $h^\eta(i, s)$ is bounded on $[0, T]$, the probability in the right-hand of above inequality vanishes as $\epsilon \to 0$ for any small enough $\eta$ by Corollary 5.8 in Yin and Zhang [27]. Hence we conclude that $\lim_{\epsilon \to 0}\mathbb{P}_{u^*}(I_T^{3*} > \delta) = 0$.

We have thus shown that $\mathbb{P}_{u^*}(\Theta^\epsilon)$ remains bounded away from 0 as $\epsilon \to 0$. Upon combining all the above, the proof of the lemma is now complete. □

So far we have focused on the case $\inf_{i,x}\sigma^2(i, x) > 0$; to complete the analysis, we next consider the situation that this condition is lifted, in which case $\hat{\sigma}(\varphi_s, \nu)$ can be singular.

Our proof uses arguments used in the method presented by Liptser [18, Lemma A.6]. Given $\gamma > 0$, we study the stochastic differential equation

$$M_t^{\epsilon,\gamma} = \int_0^t b(X_s^\epsilon, M_s^{\epsilon,\gamma}) \mathrm{d}s + \sqrt{\epsilon} \int_0^t \sigma(X_s^\epsilon, M_s^{\epsilon,\gamma}) \mathrm{d}B_s + \sqrt{\epsilon}\gamma W_t, \qquad (21)$$

where $M_0^{\epsilon,\gamma} \equiv 0$ and $W_t$ is another standard $\mathbb{P}$-Brownian motion, independent of $B_t$ and $X_t^\epsilon$. We provide an auxiliary lemma which is to be used when proving the lower bound; informally, it states that $M^{\epsilon,\gamma}$ and $M^\epsilon$ are 'superexponentially close'.

**Lemma 7.4** *For every $T > 0$ and $\eta > 0$,*

$$\lim_{\gamma \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}\left(\rho_T(M^{\epsilon,\gamma}, M^\epsilon) > \eta\right) = -\infty. \qquad (22)$$

**Proof** We define $A_t^{\epsilon,\gamma} := M_t^{\epsilon,\gamma} - M_t^\epsilon$, and

$$\alpha_t^\epsilon := \frac{b(X_t^\epsilon, M_t^{\epsilon,\gamma}) - b(X_t^\epsilon, M_t^\epsilon)}{M_t^{\epsilon,\gamma} - M_t^\epsilon}, \quad \beta_t^\epsilon := \frac{\sigma(X_t^\epsilon, M_t^{\epsilon,\gamma}) - \sigma(X_t^\epsilon, M_t^\epsilon)}{M_t^{\epsilon,\gamma} - M_t^\epsilon}.$$

By (A.1), i.e., the Lipschitz condition, we conclude

$$|\alpha_t^\epsilon| \leqslant K, \quad |\beta_t^\epsilon| \leqslant K, \quad \forall t \in [0, T]. \qquad (23)$$

As an immediate consequence of (1) and (21), we have

$$\mathrm{d}A_t^{\epsilon,\gamma} = \alpha_t^\epsilon A_t^{\epsilon,\gamma} \mathrm{d}t + \sqrt{\epsilon}\beta_t^\epsilon A_t^{\epsilon,\gamma} \mathrm{d}B_t + \sqrt{\epsilon}\gamma \mathrm{d}W_t.$$

We define

$$E_t^\epsilon := \exp\left(\int_0^t \left[\alpha_s^\epsilon - \frac{\epsilon}{2}(\beta_s^\epsilon)^2\right] \mathrm{d}s + \sqrt{\epsilon}\int_0^t \beta_s^\epsilon \mathrm{d}B_s\right).$$

We apply Itô's formula to $(E_t^\epsilon)^{-1}$, so as to obtain

$$\mathrm{d}(E_t^\epsilon)^{-1} = (E_t^\epsilon)^{-1}\left(\epsilon(\beta_t^\epsilon)^2 \mathrm{d}t - \alpha_t^\epsilon \mathrm{d}t - \sqrt{\epsilon}\beta_t^\epsilon \mathrm{d}B_t\right).$$

Since $W_t$ is independent of $B_t$, we have

$$\mathrm{d}\langle A_t^{\epsilon,\gamma}, (E_t^\epsilon)^{-1}\rangle = \langle -\sqrt{\epsilon}\beta_t^\epsilon (E_t^\epsilon)^{-1} \mathrm{d}B_t, \sqrt{\epsilon}\beta_t^\epsilon A_t^{\epsilon,\gamma} \mathrm{d}B_t + \sqrt{\epsilon}\gamma \mathrm{d}W_t\rangle = -\epsilon(\beta_t^\epsilon)^2 (E_t^\epsilon)^{-1} A_t^{\epsilon,\gamma} \mathrm{d}t.$$

By applying the integration-by-parts formula,

$$\mathrm{d}A_t^{\epsilon,\gamma}(E_t^\epsilon)^{-1} = A_t^{\epsilon,\gamma} \mathrm{d}(E_t^\epsilon)^{-1} + (E_t^\epsilon)^{-1} \mathrm{d}A_t^{\epsilon,\gamma} + \mathrm{d}\langle A_t^{\epsilon,\gamma}, (E_t^\epsilon)^{-1}\rangle = \sqrt{\epsilon}\gamma (E_t^\epsilon)^{-1} \mathrm{d}W_t,$$

and hence

$$A_t^{\epsilon,\gamma} = \sqrt{\epsilon}\,\gamma E_t^\epsilon \int_0^t (E_s^\epsilon)^{-1} \mathrm{d}W_s.$$

We define the set $\Gamma_N := \{1/N \leqslant \inf_{t \leqslant T} E_t^\epsilon \leqslant \sup_{t \leqslant T} E_t^\epsilon \leqslant N\}$, for $N \in \mathbb{N}$. Observe that it holds that $\rho_T(M^{\epsilon,\gamma}, M^\epsilon) = (A^{\epsilon,\gamma})_T^*$, and therefore

$$\begin{aligned} \mathbb{P}(\rho_T(M^{\epsilon,\gamma}, M^\epsilon) > \eta) &\leqslant \mathbb{P}((A^{\epsilon,\gamma})_T^* > \eta, \Gamma_N) + \mathbb{P}(\Omega \setminus \Gamma_N) \\ &\leqslant 2\max\{\mathbb{P}((A^{\epsilon,\gamma})_T^* > \eta, \Gamma_N), \mathbb{P}(\Omega \setminus \Gamma_N)\}. \end{aligned}$$

28

We now consider each of the probabilities $\mathbb{P}((A^{\epsilon,\gamma})_T^* > \eta, \Gamma_N)$ and $\mathbb{P}(\Omega \setminus \Gamma_N)$ separately. On the set $\Gamma_N$,

$$(A^{\epsilon,\gamma})_T^* \leqslant \sqrt{\epsilon}\gamma E_T^{\epsilon*} \sup_{t \leqslant T} \left| \int_0^t (E_s^\epsilon)^{-1} \mathrm{d}W_s \right| \leqslant \sqrt{\epsilon}\gamma N \sup_{t \leqslant T} \left| \int_0^t (E_s^\epsilon)^{-1} \mathrm{d}W_s \right|.$$

Since $\alpha_t^\epsilon$ and $\beta_t^\epsilon$ are bounded as $\epsilon \to 0$, it follows that $\int_0^t [\alpha_s^\epsilon - \frac{\epsilon}{2}(\beta_s^\epsilon)^2] \mathrm{d}s$ is bounded as well, and therefore we omit it for brevity when analyzing $E_t^\epsilon$. Based on the above, the stated holds if we can prove that (A) for all $N \in \mathbb{N}$, covering the contribution of $\mathbb{P}((A^{\epsilon,\gamma})_T^* > \eta, \Gamma_N)$,

$$\lim_{\gamma \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}\left( \sqrt{\epsilon}\gamma N \sup_{t \leqslant T} \left| \int_0^t (E_s^\epsilon)^{-1} \mathrm{d}W_s \right| > \eta, \Gamma_N \right) = -\infty, \tag{24}$$

and (B), covering the contribution of $\mathbb{P}(\Omega \setminus \Gamma_N)$,

$$\lim_{N \to \infty} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}\left( \sqrt{\epsilon} \sup_{t \leqslant T} \left| \int_0^t \beta_s^\epsilon \mathrm{d}B_s \right| > \log N \right) = -\infty. \tag{25}$$

Let us first consider contribution (A). To this end, define

$$\tau := T \wedge \inf \left\{ t \leqslant T : \left| \int_0^t (E_s^\epsilon)^{-1} \mathrm{d}W_s \right| > \frac{\eta}{\sqrt{\epsilon}\gamma N} \right\}.$$

Then (24) is equivalent to, for all $N \in \mathbb{N}$,

$$\lim_{\gamma \to 0} \limsup_{\epsilon \to 0} \epsilon \log \mathbb{P}\left( \sqrt{\epsilon}\gamma N \int_0^\tau (E_s^\epsilon)^{-1} \mathrm{d}W_s \geqslant \eta \ (\text{or} \leqslant -\eta), \Gamma_N \right) = -\infty. \tag{26}$$

For $N \in \mathbb{N}$ and $\eta > 0$, we define the process $\tilde{E}_t^\epsilon$ and its stochastic exponential $\mathscr{E}(\tilde{E}^\epsilon)_t$:

$$\tilde{E}_t^\epsilon := \frac{\eta}{\sqrt{\epsilon}\gamma N^3 T} \int_0^t (E_s^\epsilon)^{-1} \mathrm{d}W_s, \quad \mathscr{E}(\tilde{E}^\epsilon)_t = \exp\left( \tilde{E}_t^\epsilon - \frac{1}{2}\langle \tilde{E}^\epsilon \rangle_t \right).$$

Since $\mathscr{E}(\tilde{E}^\epsilon)_t$ is a supermartingale, we have

$$\mathbb{E}\left[ \mathbf{1}_{\{\sqrt{\epsilon}\gamma N \int_0^\tau (E_s^\epsilon)^{-1} \mathrm{d}W_s \geqslant \eta, \Gamma_N\}} \mathscr{E}(\tilde{E}^\epsilon)_\tau \right] \leqslant 1$$

On the set $\{\sqrt{\epsilon}\gamma N \int_0^\tau (E_s^\epsilon)^{-1} \mathrm{d}W_s \geqslant \eta, \Gamma_N\}$, we have

$$\mathscr{E}(\tilde{E}^\epsilon)_\tau \geqslant \exp\left( \frac{\eta}{\sqrt{\epsilon}\gamma N^3 T} \frac{\eta}{\sqrt{\epsilon}\gamma N} - \frac{1}{2}\left( \frac{\eta}{\sqrt{\epsilon}\gamma N^3 T} \right)^2 N^2 T \right) = \exp\left( \frac{\eta^2}{2\epsilon\gamma^2 N^4 T} \right),$$

and consequently

$$\exp\left( \frac{\eta^2}{2\epsilon\gamma^2 N^4 T} \right) \mathbb{P}\left( \sqrt{\epsilon}\gamma N \int_0^\tau (E_s^\epsilon)^{-1} \mathrm{d}W_s \geqslant \eta, \Gamma_N \right) \leqslant 1.$$

We conclude that the part corresponding to "$\geqslant \eta$" in (26) is valid, but it is immediately verified that the part corresponding to "$\leqslant -\eta$" in (26) can be addressed in the same way.

We now turn to contribution (B). The validity of (25) can be proved in a similar way by defining the stopping time

$$\tau' := \inf \left\{ t \leqslant T : \left| \int_0^t \beta_s^\epsilon \mathrm{d}B_s \right| > \frac{\log N}{\sqrt{\epsilon}} \right\}$$

and the process $\tilde{\beta}_t^\epsilon$ and its stochastic exponential $\mathscr{E}(\tilde{\beta}^\epsilon)_t$:

$$\tilde{\beta}_t^\epsilon := \frac{\log N}{\sqrt{\epsilon}K^2 T}\int_0^t \beta_s^\epsilon \mathrm{d}B_s, \quad \mathscr{E}(\tilde{\beta}^\epsilon)_t = \exp\left(\tilde{\beta}_t^\epsilon - \frac{1}{2}\langle\tilde{\beta}^\epsilon\rangle_t\right),$$

where $K$ is the constant in (23). $\qquad\square$

The following result establishes the lower bound of the local LDP.

**Proposition 7.5** *For every* $(\varphi,\nu)\in\mathbb{C}_T\times\mathbb{M}$,

$$\liminf_{\delta\to 0}\liminf_{\epsilon\to 0}\epsilon\log\mathbb{P}(\rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta)\geqslant -L_T(\varphi,\nu).$$

**Proof** As mentioned in the beginning of this section, only the case $(\varphi,\nu)\in\mathbb{H}_T\times\mathbb{M}_T$ such that

$$\int_0^T \frac{[\varphi_t' - \hat{b}(\nu,\varphi_t)]^2}{\hat{\sigma}^2(\nu,\varphi_t)}\mathrm{d}t < \infty$$

needs to be considered. If $\inf_{i,x}\sigma^2(i,x)>0$, then the result is valid due to Proposition 7.3. If $\inf_{i,x}\sigma^2(i,x)=0$, then we consider $M_t^{\epsilon,\gamma}$ as defined in (21). The idea is that we decompose the probability $\mathbb{P}\left(\rho_T(M^{\epsilon,\gamma},\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\frac{\delta}{2}\right)$ into the sum of

$$\mathbb{P}\left(\rho_T(M^{\epsilon,\gamma},\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\frac{\delta}{2}, \rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta\right) \tag{27}$$

and

$$\mathbb{P}\left(\rho_T(M^{\epsilon,\gamma},\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\frac{\delta}{2}, \rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)>\delta\right). \tag{28}$$

Obviously, (27) is majorized by $\mathbb{P}(\rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta)$. Using the triangle inequality, we find that $\rho_T(M^\epsilon,\varphi)\leqslant\rho_T(M^\epsilon,M^{\epsilon,\gamma})+\rho_T(M^{\epsilon,\gamma},\varphi)$. So that (28) is majorized by $\mathbb{P}\left(\rho_T(M^\epsilon,M^{\epsilon,\gamma})>\frac{\delta}{2}\right)$. Hence, $\mathbb{P}\left(\rho_T(M^{\epsilon,\gamma},\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\frac{\delta}{2}\right)$ is not greater than

$$2\max\left[\mathbb{P}(\rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta),\mathbb{P}\left(\rho_T(M^\epsilon,M^{\epsilon,\gamma})>\frac{\delta}{2}\right)\right].$$

By Lemma 7.4,

$$\lim_{\gamma\to 0}\limsup_{\epsilon\to 0}\epsilon\log\mathbb{P}\left(\rho_T(M^{\epsilon,\gamma},M^\epsilon)>\frac{\delta}{2}\right)=-\infty,$$

and, as a result,

$$\liminf_{\delta\to 0}\lim_{\gamma\to 0}\liminf_{\epsilon\to 0}\epsilon\log\mathbb{P}\left(\rho_T(M^{\epsilon,\gamma},\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\frac{\delta}{2}\right)$$
$$\leqslant\liminf_{\delta\to 0}\liminf_{\epsilon\to 0}\epsilon\log\mathbb{P}(\rho_T(M^\epsilon,\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\delta).$$

Next we compute the term on the left-hand side of the above inequality. Since $M^{\epsilon,\gamma}$ meets the conditions in Lemma 7.3, $(M^{\epsilon,\gamma},\nu^\epsilon)$ satisfies the inequality (18). Then for every $\gamma>0$, we obtain

$$\liminf_{\epsilon\to 0}\epsilon\log\mathbb{P}\left(\rho_T(M^{\epsilon,\gamma},\varphi)+d_T(\nu^\epsilon,\nu)\leqslant\frac{\delta}{2}\right)\geqslant -K_{Q,u^*}\frac{\delta}{2}d-\tilde{I}_T(\nu)-\frac{1}{2}\int_0^T \frac{[\varphi_s'-\hat{b}(\varphi_s,\nu)]^2}{\hat{\sigma}^2(\varphi_s,\nu)+\gamma^2}\mathrm{d}s.$$

By the monotone convergence theorem (recall the convention $0/0 = 0$),

$$\frac{1}{2}\int_0^T \frac{[\varphi'_s - \hat{b}(\varphi_s, \nu)]^2}{\hat{\sigma}^2(\varphi_s, \nu) + \gamma^2}\mathrm{d}s \to \frac{1}{2}\int_0^T \frac{[\varphi'_s - \hat{b}(\varphi_s, \nu)]^2}{\hat{\sigma}^2(\varphi_s, \nu)}\mathrm{d}s = I_T(\varphi, \nu), \text{ as } \gamma \to 0$$

which implies that

$$\liminf_{\delta \to 0} \lim_{\gamma \to 0}\liminf_{\epsilon \to 0} \epsilon \log \mathbb{P}\left(\rho_T(M^{\epsilon,\gamma}, \varphi) + d_T(\nu^\epsilon, \nu) \leqslant \frac{\delta}{2}\right) \geqslant -\tilde{I}_T(\nu) - I_T(\varphi, \nu).$$

We have proven the claim. $\qquad\square$

# A  Appendix

**Lemma A.1** *Let $f(t, i)$ be a continuous function on $[0, T]$ for every $i \in \mathbb{S}$. Let $\mu, \nu \in \mathbb{M}$ such that $d_T(\mu, \nu) \leqslant \delta$. For any $\gamma > 0$ and $[t_1, t_2] \subset [0, T]$, there exists a constant $C > 0$ such that*

$$\sup_{i \in \mathbb{S}}\left|\int_{t_1}^{t_2} f(s, i)[K_\mu(s, i) - K_\nu(s, i)]\mathrm{d}s\right| \leqslant C\delta + \gamma. \tag{29}$$

**Proof** We first look at functions $f(t, i)$ that are of bounded variation. By integration by parts, we have

$$\int_{t_1}^{t_2} f(s, i)[K_\mu(s, i) - K_\nu(s, i)]\mathrm{d}s = [\mu(s, i) - \nu(s, i)]f(s, i)|_{t_1}^{t_2} - \int_{t_1}^{t_2}[\mu(s, i) - \nu(s, i)]\mathrm{d}f(s, i).$$

Then

$$\left|\int_{t_1}^{t_2} f(s, i)[K_\mu(s, i) - K_\nu(s, i)]\mathrm{d}s\right| \leqslant |\mu(t_2, i) - \nu(t_2, i)||f(t_2, i)| + |\mu(t_1, i) - \nu(t_1, i)||f(t_1, i)|$$

$$+ \int_{t_1}^{t_2}|\mu(s, i) - \nu(s, i)||\mathrm{d}f(s, i)|$$

$$\leqslant C_1\delta + C_2\delta + TV_f[t_1, t_2]\delta,$$

where $TV_f[t_1, t_2]$ denotes the total variation of $f$ on $[t_1, t_2]$ and $C_1, C_2$ are two positive constants. Since $\mathbb{S}$ has finite elements, we can find a constant $C$ such that the claim holds. If $f(t, i)$ is only continuous, it can be uniformly approximated by a continuously differentiable function (see [25]), that is, for any $\gamma > 0$, there exists a continuously differentiable function $f^\gamma(t, i)$ such that

$$\sup_{t \in [t_1, t_2], i \in \mathbb{S}}|f(t, i) - f^\gamma(t, i)| < \gamma/2\,(t_2 - t_1).$$

Then

$$\sup_{i \in \mathbb{S}}\left|\int_{t_1}^{t_2} f(s, i)[K_\mu(s, i) - K_\nu(s, i)]\mathrm{d}s\right| \leqslant \sup_{i \in \mathbb{S}}\left|\int_{t_1}^{t_2}[f(s, i) - f^\gamma(s, i)][K_\mu(s, i) - K_\nu(s, i)]\mathrm{d}s\right|$$

$$+ \sup_{i \in \mathbb{S}}\left|\int_{t_1}^{t_2} f^\gamma(s, i)[K_\mu(s, i) - K_\nu(s, i)]\mathrm{d}s\right|$$

$$\leqslant \gamma + C\delta.$$

This finishes our proof. $\qquad\square$

31

For any $u \in \mathbb{U}$, let $Q(u)(t)$ be the transition matrix resulting from the measure change induced by the stochastic exponential $\mathscr{E}(\tilde{N}^\epsilon)$. It is known, see Proposition 11.2.3 in Bielecki and Rutkowski [3], that

$$Q(u)(t)_{ij} = Q_{ij}\frac{u(t,j)}{u(t,i)} \text{ if } i \neq j; \quad Q(u)(t)_{ii} = -\sum_{j \neq i} Q_{ij}\frac{u(t,j)}{u(t,i)}.$$

For a fixed $t$, we suppress this $t$, so as to make the notation more compact. In matrix notation, (where we throughout write $\text{diag}(u)$ to denote the diagonal matrix with entries $u_i\delta_{ij}$) we have

$$Q(u) = \text{diag}(u)^{-1}Q\text{diag}(u) - \text{diag}(u)^{-1}\text{diag}(Qu). \tag{30}$$

**Lemma A.2** *Let $\nu$ be a $d$-dimensional vector such that $\sum_{i=1}^d \nu(i) = 1$ and $\nu(i) > 0$. Let $u^* \in U$ be an optimizer of*

$$\inf_{u \in U} \sum_i \frac{(Qu)_i}{u_i}\nu_i = \inf_{u \in U} \nu^{\mathrm{T}}\text{diag}(u)^{-1}Qu.$$

*Then $\nu$ is the unique invariant vector of the transition matrix $Q(u^*)$.*

**Proof** To find a minimizing $u^* = u(\nu)$ (where it observed that minimizers are not necessarily unique) for the above problem, we first note that all $u_i^* > 0$. Hence the minimizer solves the system of first order conditions. Differentiation with respect to $u_k$ yields ($e_k$ denoting the $k$-th basis vector)

$$-\nu^{\mathrm{T}}\text{diag}(u)^{-1}e_k e_k^{\mathrm{T}}\text{diag}(u)^{-1}Qu + \nu^{\mathrm{T}}\text{diag}(u)^{-1}Qe_k = 0.$$

In vector notation these equations can be conveniently summarized as

$$\nu^{\mathrm{T}}\text{diag}(u)^{-1}\left(-\text{diag}(u)^{-1}\text{diag}(Qu) + Q\right) = 0. \tag{31}$$

From (30) we deduce by commutation of diagonal matrices the relation

$$Q(u) = \text{diag}(u)^{-1}(Q - \text{diag}(u)^{-1}\text{diag}(Qu))\text{diag}(u),$$

and hence

$$\text{diag}(u)Q(u)\text{diag}(u)^{-1} = Q - \text{diag}(u)^{-1}\text{diag}(Qu).$$

We can therefore rewrite (31) as

$$\nu^{\mathrm{T}}Q(u)\text{diag}(u)^{-1} = 0.$$

It follows that $\nu^{\mathrm{T}}Q(u^*) = 0$. Since $\nu(i) > 0$, $\nu$ is the unique invariant vector of $Q(u^*)$. $\qquad\square$

# References

[1] Aldous, D., 1978. Stopping times and tightness. Ann. Probab., 6, 335–340.

[2] Borovkov, A. A. and Mogulskiĭ, A. A., 2010. On large deviation principles in metric spaces. Sibirsk. Mat. Zh., 51, 1251–1269.

[3] Bielecki, T. R. and Rutkowski, M., 2002. Credit risk: modelling, valuation and hedging. Springer-Verlag, Berlin.

[4] Dembo, A. and Zeitouni, O., 2010. Large deviations techniques and applications. Stochastic Modelling and Applied Probability, 38, Springer-Verlag, Berlin.

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

[6] Donsker, M. D. and Varadhan, S. R. S., 1975. Asymptotic evaluation of certain Markov proces expectations for large time, I. Comm. Pure Appl. Math., 28, 1-47.

[7] Evans, L. C., 2010. Partial differential equations. American Mathematical Society, Providence, RI,

[8] Fournié, E., Lasry, J. M. and Lions, P. L., 1997. Some nonlinear methods for studying far-from-the-money contingent claims. Numerical methods in finance, 115–145, Cambridge Univ. Press, Cambridge.

[9] Freidlin, M. I. and Wentzell, A. D., 1998. Random perturbations of dynamical systems. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 260, Springer-Verlag, New York.

[10] Guillin, A., 2003. Averaging principle of SDE with small diffusion: moderate deviations. Ann. Probab., 31, 413–443.

[11] He, Q. and Yin, G., 2014. Moderate deviations for time-varying dynamic systems driven by non-homogeneous Markov chains with Two-time Scales. Stochastics, 86, 527–550.

[12] He, Q. and Yin, G., 2014. Large deviations for multi-scale Markovian switching systems with a small diffusion. Asymptot. Anal., 87, 123–145.

[13] He, Q., Yin, G. and Zhang, Q., 2011. Large deviations for two-time-scale systems driven by non-homogeneous Markov chains and associated optimal control problems. SIAM J. Control Optim., 49, 1737–1765.

[14] Huang, G., Mandjes, M. and Spreij, P., 2014. Weak convergence of Markov-modulated diffusion processes with rapid switching. Statist. Probab. Lett., 86, 74–79.

[15] Jacod, J., 1985 Théorèmes limite pour les processus. École d'été de probabilités de Saint-Flour, XIII—1983, Springer, Berlin.

[16] Jacod, J. and Shiryaev, A. N., 2003. Limit theorems for stochastic processes. Springer-Verlag, Berlin.

[17] Lasry, J. M., and Lions, P.L., 1995. Grandes déviations pour des processus de diffusion couplés par un processus de sauts. C. R. Acad. Sci. Paris Sér. I Math., 321, 849–854.

[18] Liptser, R., 1996. Large deviations for two scaled diffusions, Probab. Theory and Related Fields 106, 71–104.

[19] Liptser, R. SH. and Pukhalskii, A.A., 1992. Limit theorems on large deviations for semimartingales. Stochastics Stochastics Rep 38, 201–249.

[20] Liptser, R. Sh. and Shiryayev, A. N., 1989. Theory of martingales. Kluwer Academic Publishers Group, Dordrecht.

[21] Liptser, R. S. and Shiryaev, A. N., 2001. Statistics of random processes I. Applications of Mathematics (New York), 5, Springer-Verlag, Berlin.

[22] Palmowski, Zbigniew and Rolski, Tomasz, 2002. A technique for exponential change of measure for Markov processes, Bernoulli, 8, 767–785.

[23] Protter, P. E., 2005. Stochastic integration and differential equations. Stochastic Modelling and Applied Probability, 21, Springer-Verlag, Berlin.

[24] Puhalskii, A., 1991. On functional principle of large deviations. New trends in probability and statistics 1. VSP, Utrecht.

[25] Pursell, L. E., 1967. Uniform approximation of real continuous functions on the real line by infinitely differentiable functions. Math. Mag., 40, 263–265.

[26] Varadhan, S. R. S., 1984. Large deviations and applications. CBMS-NSF Regional Conference Series in Applied Mathematics 46, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

[27] Yin, G. and Zhang, Q., 2013. Continuous-time Markov chains and applications, 2nd edition. Springer, New York.

[28] Yin, G. and Zhu, C., 2010. Hybrid switching diffusions. Stochastic Modelling and Applied Probability, 63, Springer, New York.