

Article

# Moderate Averaged Deviations for a Multi-Scale System with Jumps and Memory

André de Oliveira Gomes <sup>1,\*</sup>  and Pedro Catuogno <sup>1,2,†</sup> 

<sup>1</sup> Mathematics Department, Universidade Estadual de Campinas, Campinas 13081-970, SP, Brazil

<sup>2</sup> ParisTech Applied Mathematics Department, ENSTA, 828 Boulevard des Maréchaux, 91120 Palaiseau, France

\* Correspondence: aog@m2hycon.com

† These authors contributed equally to this work.

**Abstract:** This work studies a two-time-scale functional system given by two jump diffusions under the scale separation by a small parameter  $\varepsilon \rightarrow 0$ . The coefficients of the equations that govern the dynamics of the system depend on the segment process of the slow variable (responsible for capturing delay effects on the slow component) and on the state of the fast variable. We derive a moderate deviation principle for the slow component of the system in the small noise limit using the weak convergence approach. The rate function is written in terms of the averaged dynamics associated with the multi-scale system. The core of the proof of the moderate deviation principle is the establishment of an averaging principle for the auxiliary controlled processes associated with the slow variable in the framework of the weak convergence approach. The controlled version of the averaging principle for the jump multi-scale diffusion relies on a discretization method inspired by the classical Khasminkii's averaging principle.

**Keywords:** moderate deviation principle; multi-scale stochastic differential equations with jumps and delay; segment process; stochastic averaging principle; weak convergence approach

**MSC:** 60H10; 60F10; 60J75



**Citation:** de Oliveira Gomes, A.; Catuogno, P. Moderate Averaged Deviations for a Multi-Scale System with Jumps and Memory. *Dynamics* **2023**, *3*, 171–201. <https://doi.org/10.3390/dynamics3010011>

Academic Editors: Michel L. Lapidus, Ravi P. Agarwal and Maria Alessandra Ragusa

Received: 31 January 2023

Revised: 28 February 2023

Accepted: 7 March 2023

Published: 14 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With a fixed terminal time  $T > 0$  and a certain delay  $\tau > 0$ , we consider in the small noise limit  $\varepsilon \rightarrow 0$  the two-time scale stochastic system given for any  $t \in [0, T]$  by

$$\begin{cases} dX^\varepsilon(t) &= a(X_t^\varepsilon, Y^\varepsilon(t))dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dB^1(t) + \int_{\mathbb{X}} \varepsilon c(X_{t-}^\varepsilon, z)\tilde{N}^{\frac{1}{\varepsilon}}(t, dz); \\ dY^\varepsilon(t) &= \frac{1}{\varepsilon}f(X_t^\varepsilon, Y^\varepsilon(t))dt + \frac{1}{\sqrt{\varepsilon}}g(X_t^\varepsilon, Y^\varepsilon(t))dB^2(t) + \int_{\mathbb{X}} h(X_{t-}^\varepsilon, Y^\varepsilon(t-), z)\tilde{N}^{\frac{1}{\varepsilon}}(t, dz). \end{cases} \quad (1)$$

For every  $\varepsilon > 0$ , the stochastic process  $(X^\varepsilon(t), Y^\varepsilon(t))_{t \in [0, T]}$  takes values in  $\mathbb{R}^n := \mathbb{R}^d \times \mathbb{R}^k$ . The initial datum is  $(X_0^\varepsilon, Y^\varepsilon(0)) = (\chi, y)$  where  $\chi$  is a given continuous function from  $[-\tau, 0]$  to  $\mathbb{R}^d$  (initial delay segment) and  $y \in \mathbb{R}^k$ . The processes  $X^\varepsilon$  and  $Y^\varepsilon$  are usually designated in the literature as, respectively, the slow variable and the fast variable of the multi-scale stochastic system (1). We draw the reader's attention to the use of the notation  $X_t^\varepsilon$  for the segment process, i.e.,  $X_t^\varepsilon := \{X^\varepsilon(t + \theta) \mid \theta \in [-\tau, 0]\}$  for any  $t \geq 0$ . We refer the reader to Chapters 5 and 6 of the book [1] for an introduction to the subject of stochastic functional differential equations with Brownian noise and to [2] for the study of stochastic functional differential equations with jumps. The space of the jump increments  $\mathbb{X}$  is Euclidean, the process  $B = (B^1, B^2)$  is a standard Brownian motion (BM for short) with values in  $\mathbb{R}^n$  with first component  $B^1$  being a standard BM with values in  $\mathbb{R}^d$  and the second component  $B^2$  being an independent  $\mathbb{R}^k$ -valued standard BM. For every  $\varepsilon > 0$ , the random measure  $\tilde{N}^{\frac{1}{\varepsilon}}$  is an independent compensated Poisson random measure with intensity given by  $ds \otimes \frac{1}{\varepsilon}\nu(dz)$ , where  $ds$  stands for the Lebesgue measure on the real line and  $\nu$  is a Lévy measure on  $\mathbb{X}$ . In this work, we consider  $\nu$  possibly with infinite total

mass but satisfying an exponential integrability condition that reads as the big jumps of the underlying Lévy process having exponential moments of order 2. The assumptions on the coefficients of (1) and on the measure  $\nu$  will be specified with full rigour in the following section.

Multi-scale stochastic systems such as (1) are nowadays very popular in applied mathematical and physical disciplines since they are successful models for phenomena exhibiting different levels of heterogeneity/homogeneity that can be asymptotically categorized by scaling. This technique of understanding diversity exploits the decomposition of the phase space of the model in two sets of variables, the ones with slow degrees of freedom and the ones with fast degrees of freedom through a separation scale given by an intensity parameter measuring this degree of heterogeneity/homogeneity. We refer the reader to [3] and the monograph [4] for an introduction to the subject. Typical examples are multi-factor stochastic volatility models in finance [5,6] and the dynamics of proxy-data in climatology [7] where climatic transitions are understood within the distinction between slow and fast variables that encode different factors used to build statistical parametrizations. In the description of those climatic models, short/large time scales must be taken into consideration (e.g., daily weather forecast vs climatic prediction) in order to see interesting phenomena such as metastability of the slow variable from an equilibrium state of the deterministic dynamics (cf. Appendix in [7–9]). Often in these multi-scale climatic models, the slow variable quantifies data related to large time scales (e.g., climatic data). Multi-scale stochastic systems of the type (1) offer the mathematical formalism necessary to capture more realistic attributes of the underlying stochastic climate model. The paradigmatic example in climate dynamics is the coupling of ocean temperature models (slow variable) with the atmospheric Lorentz equations (fast variable). We refer the reader for more details to [10]. The presence of an underlying Lévy process drives the stochastic dynamics of (1) in small noise models abrupt climate transitions. A typical example is given by the *Daansgard–Oeschger events* that show statistical evidence of underlying jump noise signals (cf. Chapter 10 in [10–13]). The dependence of the coefficients of (1) on the segment process of the slow variable models the memory effects exhibited by energy balance models such as the ones constructed in [14].

This type of multi-scale system is highly complex and difficult to analyze or simulate. It is desirable to approximate in a suitable sense the dynamics of the slow variable by some simpler dynamical system. The idea of the averaging principle performed first by Khasminkii in [15] is the following. Under strong dissipativity assumptions concerning the coefficients of the fast variable that ensure the existence of a unique invariant measure  $\mu^\zeta$  for the fast variable process with frozen slow variable  $\zeta$  and such that a certain ergodic property holds for the mixing coefficient  $a$  with respect to (w.r.t.) its average against  $\mu^\zeta$  (cf. Proposition 3)

$$\bar{a}(\zeta) := \int_{\mathbb{R}^k} a(\zeta, y) \mu^\zeta(dy) \tag{2}$$

we have that the (strong) averaging principle states that for any  $T > 0$  and  $\delta > 0$ , one has

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P} \left( \sup_{t \in [0, T]} |X^\varepsilon(t) - \bar{X}^0(t)| > \delta \right) = 0, \tag{3}$$

where  $\bar{X}^0$  is the unique solution of the functional averaged differential equation

$$\begin{cases} \frac{d}{dt} \bar{X}^0(t) = \bar{a}(\bar{X}_t^0), & t \in [0, T]; \\ \bar{X}_0^0 = \chi. \end{cases} \tag{4}$$

The averaging principle has applications to problems in celestial (stochastic) mechanics (cf. Chapter 7 in [16]) and climatic energy balanced models (cf. [9]) among others and has a rich and diverse history in the literature. Khasminkii’s technique was introduced in [15] and later implemented by Mark Freidlin [17] and Veretennikov in [18] in different contexts, finding huge applicability in a diverse range of problems. We refer the reader to the following exemplary but not exhaustive works on weak and strong averaging principles: ref. [19–21]

concerning multi-scale systems constituted by stochastic partial differential equations (SPDEs for short) driven by space time white noise; ref. [22–25] for multi-scale (finite and infinite dimensional) systems constituted by jump diffusions; and ref. [26,27] for stochastic dynamical systems with coefficients functionally dependent with delay. Although the averaging principle (3) yields an approximation result for small  $\varepsilon > 0$  of the slow variable process by the averaged dynamics of  $\bar{X}$ , nothing is said on the rate of convergence. Large and moderate deviation types of statements provide sharper estimates within the identification of a rate of convergence for the limit (3) in an exponentially small scale in  $\varepsilon \rightarrow 0$  and in terms of a deterministic quantity designated good rate function. We refer the reader to [28–31] for stochastic averaging under the large deviations regime and, respectively, to [32–34] for averaging under moderate deviation regimes.

The aim of this article is to derive a moderate deviation principle (MDP for short) for  $(X^\varepsilon)_{\varepsilon>0}$  as  $\varepsilon \rightarrow 0$ . More precisely, we will study deviations of  $X^\varepsilon$  from the averaged dynamical system  $\bar{X}$ ; that is,

$$Z^\varepsilon := \frac{X^\varepsilon - \bar{X}^0}{d(\varepsilon)} \quad \text{as } \varepsilon \rightarrow 0,$$

for certain families of magnitude scales  $d(\varepsilon)$  such that  $d(\varepsilon) \rightarrow 0$  and  $b(\varepsilon) := \frac{\varepsilon}{d^2(\varepsilon)} \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . We fix  $\theta \in (\frac{1}{2}, 1)$  and let  $b(\varepsilon) := \varepsilon^\theta, \varepsilon > 0$ . The restrictions on the range of  $\theta$  are due to parametric choices that are used in the course of the proof. This can be appreciated in the course of the proof of the technical but crucial Lemma 2 in Appendix A. Although we impose restrictions on the magnitudes  $d(\varepsilon)$  as stated above, the free parameter  $\theta \in (\frac{1}{2}, 1)$  still covers a big range of moderate deviation intermediary regimes. Assuming specific hypotheses on the coefficients that guarantee that  $\bar{a}$  defined in (2) exists, it is Fréchet differentiable with Lipschitz derivative, and that the Lévy measure  $\nu$  satisfies a certain exponential integrability property, we prove that the family  $(Z^\varepsilon)_{\varepsilon>0}$  satisfies a moderate deviation principle with speed  $b(\varepsilon) \rightarrow 0$  in  $\mathbb{D}([0, T]; \mathbb{R}^d)$ , the space of càdlàg functions endowed with the Skorokhod topology and the good rate function  $\mathbb{I} : \mathbb{D}([0, T]; \mathbb{R}^d) \rightarrow [0, \infty]$  given by

$$\mathbb{I}(\eta) := \inf_{(f,h) \in L^2([0,T]) \times L^2(\nu \otimes ds)} \frac{1}{2} \left( \int_0^T |f(s)|^2 ds + \int_0^T |h(s,z)|^2 \nu(dz) ds \right),$$

where for every  $(f, h) \in L^2([0, T]) \times L^2(\nu \otimes ds)$  the function  $\eta \in C([-\tau, T]; \mathbb{R}^d)$  solves uniquely the skeleton equation:

$$\begin{cases} \eta(t) &= \int_0^t D\bar{a}(\bar{X}_s^0)\eta_s ds + \int_0^t \sigma(\bar{X}_s^0)f(s)ds + \int_0^t \int_{\mathbb{X}} c(\bar{X}_s^0, z)h(s, z)\nu(dz)ds, \quad t \in [0, T]; \\ \eta_0 &= 0 \end{cases} \tag{5}$$

and the function  $\bar{X}^0 \in C([-\tau, T]; \mathbb{R}^d)$  is the unique solution of (4). Here, the coefficients  $a, \sigma$  and  $c$  are the coefficients of the stochastic Equation (1).

This means that the functional  $\mathbb{I}$  has compact sublevel sets  $\{\mathbb{I} \leq c\}$  in the Skorokhod topology for any  $c \geq 0$  and that for any open set  $G \in \mathcal{B}(\mathbb{D}([0, T]; \mathbb{R}^d))$  and closed set  $F \in \mathcal{B}(\mathbb{D}([0, T]; \mathbb{R}^d))$  the following holds:

$$\begin{aligned} \liminf_{\varepsilon \rightarrow 0} \varepsilon^\theta \ln \mathbb{P}(Z^\varepsilon \in G) &\geq - \inf_{\eta \in G} \mathbb{I}(\eta) \quad \text{and} \\ \limsup_{\varepsilon \rightarrow 0} \varepsilon^\theta \ln \mathbb{P}(Z^\varepsilon \in F) &\leq - \inf_{\eta \in F} \mathbb{I}(\eta). \end{aligned}$$

We stress that the moderate deviation regime of speed  $b(\varepsilon) = \varepsilon^\theta, \theta \in (\frac{1}{2}, 1)$ , is an intermediary regime between the central limit approximation  $d(\varepsilon) = \sqrt{\varepsilon}$  and the large deviation regime  $d(\varepsilon) = 1$ . The moderate deviation regime is a very desirable asymptotic regime for the sake of applications since the rate function involves a quadratic functional which is often easier to use in applied problems in comparison with the more involved forms

of the rate function used in large deviation statements. We refer as examples to ref. [35,36] for the application of moderate deviation principles in finance, to ref. [37] in statistics and to ref. [38] where the moderate deviation regime is used to study asymptotics of exit time results for discrete random dynamical systems.

In order to prove our result, we use the weak convergence approach of Dupuis, Ellis, Budhiraja and collaborators that rely on the equivalence in Polish spaces between the definition of the large deviation principle and the variational principle nowadays known in the literature as the Laplace–Varadhan principle. Initially, Fleming applied in [39,40] methods of stochastic control to large deviation problems. The control-theoretical approach was carried out later in order to derive variational formulas for Laplace functionals of Markov processes in different contexts (cf. [41]). In [42], the authors derive a sufficient condition for large deviation principles (LDPs for short) for Brownian diffusions and later for jump diffusions in [43,44] through the establishment of variational formulas for Laplace functionals of Markov processes. We refer the reader to the recent book [45] for an up-to-date introduction to the subject. In [46] Budhiraja, Dupuis and Ganguly derive a sufficient condition for an MDP that was successfully applied in [47,48] to the study of MDPS for SPDEs. The literature on large/moderate deviation principles for stochastic differential equations with delay is not so extensive such as in other domains of applications. We refer the reader to the works [49,50] where the authors apply Freidlin–Wentzell types of LDPs to the study of the first exit time problem in the small noise limit for Gaussian diffusions with delay. For the application of the weak convergence approach in the establishment of MDPs to stochastic differential delay equations, we mention the works [51,52].

Strategy of the proof.

The proof of the main result of this work follows from an abstract sufficient condition for moderate deviation principles stated as Theorem 9.9 in [45]. In our case, the application of this abstract condition is not straightforward due to the coupling between the slow variable  $X^\epsilon$  and the fast variable  $Y^\epsilon$  in (1) with different scaling orders in  $\epsilon \rightarrow 0$ .

More precisely, the difficult part is to prove directly the following. Fix  $\beta \in (0, 1)$ ,  $M \geq 0$ , two families of random variables  $(\zeta^\epsilon)_{\epsilon>0}$  and  $(\psi^\epsilon := \frac{\varphi^\epsilon - 1}{d(\epsilon)})_{\epsilon>0}$  such that for any  $\epsilon > 0$  one has  $\int_0^T |\zeta^\epsilon(s)|^2 ds \leq Md^2(\epsilon)$ , where  $\varphi^\epsilon \geq 0$  satisfies  $\int_0^T \int_{\mathbb{X}} (\varphi^\epsilon(s, z) \ln \varphi^\epsilon(s, z) - \varphi^\epsilon(s, z) + 1) \nu(dz) ds \leq M$   $\mathbb{P}$ -a.s. obeying the following convergences in law,  $\zeta^\epsilon \Rightarrow \zeta$  in the  $L^2$ -weak topology and  $\psi^\epsilon \mathbf{1}_{\{|\psi^\epsilon| \leq \frac{\beta}{d(\epsilon)}\}} \Rightarrow \psi$  in some ball of  $L^2(\nu \otimes ds)$  equipped with the respective  $L^2$ -weak topology. Consider the family  $\mathcal{Z}^\epsilon := \frac{\mathcal{X}^\epsilon - \bar{\mathcal{X}}^0}{d(\epsilon)}$ ,  $\epsilon > 0$ , where  $(\mathcal{X}^\epsilon)_{\epsilon>0}$  is defined for every  $\epsilon > 0$  and  $t \in [0, T]$  by

$$\begin{cases} \mathcal{X}^\epsilon(t) &= \zeta(0) + \int_0^t \left( a(\mathcal{X}_s^\epsilon, \mathcal{Y}^\epsilon(s)) + \sigma(\mathcal{X}_s^\epsilon) \bar{\zeta}_1^\epsilon(s) + \int_{\mathbb{X}} c(\mathcal{X}_s^\epsilon, z) (\varphi^\epsilon(s, z) - 1) \nu(dz) \right) ds \\ &+ \sqrt{\epsilon} \int_0^t \sigma(\mathcal{X}_s^\epsilon) dB^1(s) + \epsilon \int_0^t \int_{\mathbb{X}} c(\mathcal{X}_{s-}^\epsilon, z) \tilde{N}^{\frac{1}{\epsilon} \varphi^\epsilon}(ds, dz); \\ \mathcal{X}_0^\epsilon &= \zeta. \end{cases} \tag{6}$$

and

$$\begin{cases} \mathcal{Y}^\epsilon(t) &= y + \frac{1}{\epsilon} \int_0^t \left( f(\mathcal{X}_s^\epsilon, \mathcal{Y}^\epsilon(s)) + g(\mathcal{X}_s^\epsilon, \mathcal{Y}^\epsilon(s)) \bar{\zeta}_2^\epsilon(s) + \int_{\mathbb{X}} h(\mathcal{X}_s^\epsilon, \mathcal{Y}^\epsilon(s), z) (\varphi^\epsilon(s, z) - 1) \nu(dz) \right) ds \\ &+ \frac{1}{\sqrt{\epsilon}} \int_0^t g(\mathcal{X}_s^\epsilon, \mathcal{Y}^\epsilon(s)) dB^2(s) + \int_0^t \int_{\mathbb{X}} h(\mathcal{X}_{s-}^\epsilon, \mathcal{Y}^\epsilon(s-), z) \tilde{N}^{\frac{1}{\epsilon} \varphi^\epsilon}(ds, dz); \\ \mathcal{Y}_0^\epsilon &= y; \end{cases} \tag{7}$$

where for any  $\epsilon > 0$  the random measure  $\tilde{N}^{\frac{1}{\epsilon} \varphi^\epsilon}$  is a controlled random measure that under a change of probability measure has the same law of  $\tilde{N}^{\frac{1}{\epsilon}}$  under the original probability measure. This will be rigorously stated in Section 3.

Under the following setting, the main task in the derivation of the MDP is to prove that  $\mathcal{Z}^\epsilon \Rightarrow \bar{\mathcal{Z}}$  where  $\bar{\mathcal{Z}}$  solves (5) uniquely in  $C([-\tau, T]; \mathbb{R}^d)$  for the control  $(f, g) = (\zeta, \psi) \in$

$L^2([0, T]) \times L^2(\nu \otimes ds)$ . In order to prove that convergence in law, we show that the family  $(\mathcal{X}^\varepsilon)_{\varepsilon>0}$  satisfies a tightened averaging principle, i.e., for every  $\delta > 0$  the following holds

$$\limsup_{\varepsilon \rightarrow 0} \mathbb{P} \left( \sup_{t \in [0, T]} |\mathcal{X}^\varepsilon(t) - \bar{\mathcal{X}}^\varepsilon(t)| > \delta d(\varepsilon) \right) = 0, \tag{8}$$

where  $(\bar{\mathcal{X}}^\varepsilon)_{\varepsilon>0}$  is defined for every  $\varepsilon > 0$  and  $t \in [0, T]$  by

$$\begin{cases} \bar{\mathcal{X}}^\varepsilon(t) &= \chi(0) + \int_0^t \left( \bar{a}(\bar{\mathcal{X}}_s^\varepsilon) + \sigma(\bar{\mathcal{X}}_s^\varepsilon) \zeta_1^\varepsilon(s) + \int_{\mathbb{X}} c(\bar{\mathcal{X}}_s^\varepsilon, z) (\varphi^\varepsilon(s, z) - 1) \nu(dz) \right) ds \\ &+ \sqrt{\varepsilon} \int_0^t \sigma(\bar{\mathcal{X}}_s^\varepsilon) dB^1(s) + \varepsilon \int_0^t \int_{\mathbb{X}} c(\bar{\mathcal{X}}_{s-}^\varepsilon, z) \tilde{N}^{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz); \\ \bar{\mathcal{X}}_0^\varepsilon &= \zeta. \end{cases} \tag{9}$$

This will imply by Slutsky’s theorem (Theorem 4.1 in [53]) that  $(\bar{\mathcal{Z}}^\varepsilon)_{\varepsilon>0}$  has the same weak limit of  $(\bar{\mathcal{Z}}^\varepsilon)_{\varepsilon>0}$  where  $\bar{\mathcal{Z}}^\varepsilon := \frac{\bar{\mathcal{X}}^\varepsilon - \bar{\mathcal{X}}^0}{d(\varepsilon)}$ ,  $\varepsilon > 0$ . Therefore, we are led to the (easier) task to show that  $\bar{\mathcal{Z}}^\varepsilon \Rightarrow \bar{\mathcal{Z}}$  (since the dynamics of (9) is decoupled from the dynamics of the fast variable of the original stochastic system (1)).

The proof that  $\bar{\mathcal{Z}}^\varepsilon \Rightarrow \bar{\mathcal{Z}}$  as  $\varepsilon \rightarrow 0$  relies on classical arguments of weak convergence. We use localization techniques in order to obtain good estimates for the second moment of the processes in combination with Bernstein’s inequality for càdlàg local martingales given in the form of Theorem 3.3 of [54] implying the tightness of the respective laws. Hence, the relative compactness of the laws follows, yielding, due to Skorohod’s representation together with the well posedness of the skeleton Equation (5), the desired conclusion.

The proof of the tightened controlled averaging principle (8) is inspired on the classical Khasminkii technique introduced in [15]. In a nutshell, the procedure relies on a discretization of the time interval  $[0, T]$  and the delay initial interval  $[-\tau, 0]$  in a finite number of intervals with the same length  $\Delta(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$  satisfying some growth conditions that will interplay with the ergodic properties of the averaged dynamics via the construction of auxiliary processes  $(\hat{\mathcal{X}}^\varepsilon)_{\varepsilon>0}$  and  $(\hat{\mathcal{Y}}^\varepsilon)_{\varepsilon>0}$ . The construction of the auxiliary processes is not a straightforward generalization of the Khaminkii type of discretization used to prove the usual strong averaging principle. In our setting, we need to build stable not-straightforward discretizations  $(\hat{\mathcal{X}}^\varepsilon)_{\varepsilon>0}$  and  $(\hat{\mathcal{Y}}^\varepsilon)_{\varepsilon>0}$  in order to deal with the nonlocal integral terms that appear in the structure of the respective equations of  $(\mathcal{X}^\varepsilon)_{\varepsilon>0}$  and  $(\mathcal{Y}^\varepsilon)_{\varepsilon>0}$ . The proof of (8) builds heavily on the derivation of stable estimates for the deviations of the segment process  $(\hat{\mathcal{X}}_t^\varepsilon)_{t \in [0, T]}$  from the slow variable’s segment  $(\mathcal{X}_t^\varepsilon)_{t \in [0, T]}$  and, respectively, the deviations of the approximation  $(\hat{\mathcal{Y}}^\varepsilon(t))_{t \in [0, T]}$  from the fast variable controlled process  $(\mathcal{Y}^\varepsilon(t))_{t \in [0, T]}$ . We derive asymptotic bounds in  $\varepsilon > 0$  for the second moment of the deviations of the fast variable from its discretization in contrast with the way we estimate the respective deviations of the slow segment from its approximation. Due to dependence on the segment process given in the dynamics of  $(\mathcal{X}^\varepsilon)_{\varepsilon>0}$ , it turns out to be better to control the probability of the slow component deviations for the purpose of obtaining (8). This is a technical but major distinction of the technique for obtaining the strong controlled averaging principle (8) in comparison with the usual techniques available in the literature.

Our main result shows in particular that  $(X^\varepsilon)_{\varepsilon>0}$  obeys the same moderate deviation principle of  $(\bar{X}^\varepsilon)_{\varepsilon>0}$  where we define the averaged process  $\bar{X}^\varepsilon$  for every  $\varepsilon > 0$  and  $t \in [0, T]$  by

$$\bar{X}^\varepsilon(t) = \zeta(0) + \int_0^t \bar{a}(\bar{X}_s^\varepsilon) ds + \sqrt{\varepsilon} \int_0^t \sigma(\bar{X}_s^\varepsilon) dB^1(s) + \varepsilon \int_0^t \int_{\mathbb{X}} c(\bar{X}_{s-}^\varepsilon, z) \tilde{N}^{\frac{1}{\varepsilon}}(ds, dz).$$

One could firstly derive the moderate deviation principle for  $(\bar{X}^\varepsilon)_{\varepsilon>0}$  and secondly show that the families  $(X^\varepsilon)_{\varepsilon>0}$  and  $(\bar{X}^\varepsilon)_{\varepsilon>0}$  are exponentially equivalent, i.e., for every  $\delta > 0$  we have

$$\lim_{\varepsilon \rightarrow 0} \frac{\varepsilon}{d^2(\varepsilon)} \ln \mathbb{P} \left( \sup_{0 \leq t \leq T} \left| \frac{X^\varepsilon(t) - \bar{X}^\varepsilon(t)}{d(\varepsilon)} \right| > \delta \right) = -\infty. \tag{10}$$

This would imply that  $(X^\varepsilon)_{\varepsilon>0}$  obeys the same MDP of  $(\bar{X}^\varepsilon)_{\varepsilon>0}$  as  $\varepsilon \rightarrow 0$ . However, verifying the exponential equivalence of those families is in general hard. The reasoning employed in this work illustrates the robustness of the weak convergence approach, providing a way to reduce the proof of the MDP to the verification of properties concerning the continuity and tightness of certain auxiliary processes associated with  $(X^\varepsilon)_{\varepsilon>0}$ . Such reduction of complexity in such an endeavour can be appreciated immediately by the contrast between the 0 scale of the limit (8) with the exponential negligibility demanded in the establishment of the limit (10).

Notation.

The arrow  $\Rightarrow$  means convergence in distribution. Throughout the article, we use when convenient the shorthand notation  $A(\varepsilon) \lesssim_\varepsilon B(\varepsilon)$  to mean that there exists a constant  $c > 0$  independent of  $\varepsilon > 0$  and  $\varepsilon_0 > 0$  such that  $A(\varepsilon) \leq cB(\varepsilon)$  for every  $\varepsilon < \varepsilon_0$ . We write  $A(\varepsilon) \simeq_\varepsilon B(\varepsilon)$  as  $\varepsilon \rightarrow 0$  to mean that  $A(\varepsilon) \lesssim_\varepsilon B(\varepsilon)$  and  $B(\varepsilon) \lesssim_\varepsilon A(\varepsilon)$  as  $\varepsilon \rightarrow 0$ .

Outline of the paper.

In Section 2, we state with full detail the probabilistic framework and the hypothesis concerning the coefficients of (1) in order to state with full rigour the already announced MDP for the family  $(Z^\varepsilon)_{\varepsilon>0}$ . We finish that section with some examples. Section 3 contains the proof of the main result following the already announced strategy with full detail. Appendix A contains for the reader's convenience technical auxiliary results that can be skipped in a first reading.

## 2. Preliminaries and Statement of the Main Theorem

### 2.1. The Probabilistic and Functional Setup—The Averaged Dynamics

#### 2.1.1. The Probabilistic Setup and Notation

We follow extensively the probabilistic setup and the notation introduced by Budhiraja, Dupuis, Maroulas and collaborators in [43,44,46] and systematized in [45]. For any  $\mathbb{S}$  topological space, we denote by  $\mathcal{B}(\mathbb{S})$  its Borel  $\sigma$ -algebra. Fix  $T > 0$ ,  $n = d + k$  with  $d, k \in \mathbb{N}$  and let  $\mathbb{W} = C([0, T]; \mathbb{R}^n)$  endowed with the topology of the uniform convergence which turns out to be a Polish space. Let  $\mathbb{X} = \mathbb{R}^d \setminus \{0\}$  and  $\mathbb{M}$  be the space of locally finite measures defined on  $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ . We endow  $\mathbb{M}$  with the weakest topology such that for every  $f \in \mathcal{C}_c(\mathbb{X})$  (the space of compactly supported continuous functions) the function  $\nu \mapsto \langle \nu, f \rangle := \int_{\mathbb{X}} f(u) \nu(du)$ ,  $\nu \in \mathbb{M}$ , is continuous. This topology is known as the vague topology and can be metrized such that  $\mathbb{M}$  turns out to be a Polish space. We refer the reader to [43].

Fix a measure  $\nu \in \mathbb{M}$  and let  $\nu_T = ds \otimes \nu$  where  $ds$  is the Lebesgue measure on  $[0, T]$ . Consider the product space  $\mathbb{V} = \mathbb{W} \times \mathbb{M}$  and denote by  $\mathbb{P}$  the unique probability measure on  $(\mathbb{V}, \mathcal{B}(\mathbb{V}))$  under which the first projection  $B : \mathbb{V} \rightarrow \mathbb{W}$ ,  $B(\beta, m) = \beta$  is a standard Brownian motion with values in  $\mathbb{R}^n$  and  $N : \mathbb{V} \rightarrow \mathbb{M}$ ,  $N(\beta, m) := m$  is a Poisson random measure with intensity measure  $\nu_T$ . The corresponding expectation operator will be denoted by  $\mathbb{E}$ . We refer the reader to Theorem I.9.1 in [55].

Let  $\mathbb{Y} := \mathbb{X} \times [0, \infty)$ ,  $\mathbb{Y}_T := [0, T] \times \mathbb{Y}$ ; write  $\bar{\mathbb{M}}$  for the space of the locally finite measures defined on  $\mathbb{Y}_T$  when equipped with its Borel  $\sigma$ -algebra and  $\bar{\mathbb{V}} := \mathbb{W} \times \bar{\mathbb{M}}$ . In a slight abuse of notation and analogously to what was said for  $\mathbb{M}$ , the space  $\bar{\mathbb{M}}$  turns out to be also a Polish space and there exists a unique probability measure  $\bar{\mathbb{P}}$  defined on  $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$  such that the maps  $B : \bar{\mathbb{V}} \rightarrow \mathbb{W}$ ,  $B(\beta, \bar{m}) := \beta$  is a standard Brownian motion with values in  $\mathbb{R}^n$  and  $\bar{N} : \bar{\mathbb{V}} \rightarrow \bar{\mathbb{M}}$ ,  $\bar{N}(\beta, \bar{m}) := \bar{m}$  is a Poisson random measure with values on  $\mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d \setminus \{0\} \times [0, \infty))$  and intensity measure given by  $ds \otimes \nu \otimes dr$ , where  $dr$  stands for the Lebesgue measure on  $([0, \infty); \mathcal{B}([0, \infty))$ .

For every  $\varepsilon > 0$ , we consider  $N_\varepsilon^{\frac{1}{\varepsilon}}$  the Poisson random measure defined on the probability space  $(\mathbb{V}, \mathcal{B}(\mathbb{V}))$  with intensity measure given by  $\frac{1}{\varepsilon} ds \otimes \nu \otimes dr$  and  $\bar{N}_\varepsilon^{\frac{1}{\varepsilon}}$  for its compen-

sated counterpart. We also regard when necessary the object  $N_\varepsilon^{\frac{1}{\varepsilon}}$  as a controlled random measure on  $(\mathbb{V}, \mathcal{B}(\mathbb{V}))$  (and therefore  $\mathcal{B}(\mathbb{V})$ -measurable) under  $\mathbb{P}$  by the identity

$$N_\varepsilon^{\frac{1}{\varepsilon}}((0, t] \times U) := \int_0^t \int_U \int_0^\infty \mathbf{1}_{[0, \frac{1}{\varepsilon}]}(r) \bar{N}(ds, dx, dr), \quad t \in [0, T], U \in \mathcal{B}(\mathbb{X}). \tag{11}$$

We remark that the space  $\mathbb{Y} := \mathbb{X} \times [0, \infty)$  takes into account the jumps and the frequencies of the underlying Poisson random measure  $N$  and refer the reader to [43] for more details.

For any  $t \in [0, T]$ , define

$$\mathcal{F}_t := \sigma\{\bar{N}((0, s] \times A); B(s) \mid 0 \leq s \leq t, A \in \mathcal{B}(\mathbb{Y})\}$$

and denote by  $\bar{\mathbb{F}} := \{\bar{\mathcal{F}}_t\}_{t \in [0, T]}$  the completion of  $\mathbb{F} := \{\mathcal{F}_t\}_{t \in [0, T]}$  under  $\mathbb{P}$ . Consider  $\bar{\mathcal{P}}$  the predictable  $\sigma$ -field on  $[0, T] \times \mathbb{V}$  with the filtration  $\bar{\mathbb{F}}$  on  $(\mathbb{V}, \mathcal{B}(\mathbb{V}))$ .

We make the following assumption on  $\nu \in \mathbb{M}$ .

**Hypothesis 1.** *The measure  $\nu \in \mathbb{M}$  is a Lévy measure on  $(\mathbb{R}^d \setminus \{0\}, \mathcal{B}(\mathbb{R}^d \setminus \{0\}))$ , i.e., such that  $\int_{0 < |z| < 1} |z|^2 \nu(dz) < \infty$  and satisfying*

$$\int_{|z| \geq 1} e^{\alpha|z|^2} \nu(dz) < \infty, \quad \text{for some } \alpha > 1. \tag{12}$$

**Remark 1.** *We remark that the assumption of Gaussian tails (12) is paradigmatic within the use of weak convergence approach arguments for the derivation of moderate deviation principles for jump processes. It is used in the pioneer work [46] and further extensive follow up works that exist in the literature. The assumption of exponential tails for laws that obey large deviation principles is a classical ansatz in the literature of large deviation principles. We cite as reference the Donsker–Varadhan theorem stated as Theorem 3.34 in the monography [56]. The assumption of Gaussian tails (12) for  $\nu$  is sufficient to the proof of Lemma A1 in Appendix A which turns out to be a technical fundamental intermediary result that is fundamental in the derivation of the moderate deviation principle for  $(X^\varepsilon)_{\varepsilon > 0}$ . This restriction still captures a rich class of Lévy measures  $\nu$ , allowing the occurrence of infinitely small jumps as is exhibited in Section 2.3. We refer the reader to [57] for a discussion of the large deviation principle for symmetric stable processes that uses a very different approach than the one we use.*

### The Space of the Delays and the Segment Function

Fix now  $\tau > 0$ . Given a path  $x : [-\tau, T] \rightarrow \mathbb{R}^d$  and  $t \geq 0$ , we use the notation  $x_t$  for the segment path defined as  $x_t(\theta) := x(t + \theta)$ ,  $\theta \in [-\tau, 0]$ . Denote by  $C([-\tau, T]; \mathbb{R}^d)$  the space of continuous paths equipped with the uniform norm. We write  $\mathcal{C} := C([-\tau, 0]; \mathbb{R}^d)$ . Let  $\mathbb{D}([-\tau, T]; \mathbb{R}^d)$  be the space of the càdlàg functions equipped with the topology inherited by the  $J_1$ -metric known as the Skorokhod topology (cf. Chapter 3-p. 111 in [53]). We write  $\mathcal{D} := \mathbb{D}([-\tau, 0]; \mathbb{R}^d)$ . The space  $\mathbb{D}([-\tau, T]; \mathbb{R}^d)$  turn out to be Polish under this metric. We refer the reader to Theorems 12.1 and 12.2 in [53] for more details. For any  $x \in \mathbb{D}([-\tau, T]; \mathbb{R}^d)$ , we write  $\|x_t\|_\infty := \sup_{-\tau \leq s \leq t} |x(s)|$ ,  $t \geq 0$ .

#### 2.1.2. The Multiscale System

For every  $T > 0$ ,  $\tau > 0$  and  $\varepsilon > 0$ , we consider the following system of stochastic differential equations,

$$\begin{cases} X^\varepsilon(t) &= X^\varepsilon(0) + \int_0^t a(X_s^\varepsilon, Y^\varepsilon(s)) ds + \sqrt{\varepsilon} \int_0^t \sigma(X_s^\varepsilon) dB^1(s) + \varepsilon \int_0^t \int_{\mathbb{X}} c(X_{s-}^\varepsilon, z) \tilde{N}_\varepsilon^{\frac{1}{\varepsilon}}(ds, dz); \\ Y^\varepsilon(t) &= y + \frac{1}{\varepsilon} \int_0^t f(X_s^\varepsilon, Y^\varepsilon(s)) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^t g(X_s^\varepsilon, Y^\varepsilon(s)) dB^2(s) \\ &+ \int_0^t \int_{\mathbb{X}} h(X_{s-}^\varepsilon, Y^\varepsilon(s-), z) \tilde{N}_\varepsilon^{\frac{1}{\varepsilon}}(ds, dz), \quad t \in [0, T]; \end{cases} \tag{13}$$

subject to the initial data

$$\begin{cases} X_0^\varepsilon &= \chi \in \mathcal{C}, \\ Y^\varepsilon(0) &= y \in \mathbb{R}^k, \end{cases} \tag{14}$$

where we write  $(B(t))_{t \in [0, T]} = (B^1(t), B^2(t))_{t \in [0, T]}$  with  $(B^1(t))_{t \in [0, T]}$  and  $(B^2(t))_{t \in [0, T]}$ , two independent standard Brownian motions with values in  $\mathbb{R}^d$  and  $\mathbb{R}^k$ , respectively. We stress that the multi-scale system (13) has slow and fast components, respectively, affected by different Brownian signals in small intensity  $\varepsilon$  and by the same jump noise signal also in small intensity  $\varepsilon > 0$  but accelerated in inverse proportion. While the process  $(B^1, B^2)$  is also a BM in the space  $\mathbb{R}^{d \times k}$  due to the independence of each component the same does not hold for Poisson random measures in the respective product space of measures. For this reason, it is not clear how to use the weak convergence approach developed in [46] that builds in the derivation of a variational formula for functionals of Poisson random measures established in [43]. In order to guarantee the existence and uniqueness of the solution for (13), we assume that its coefficients are deterministic measurable functions  $a : \mathcal{D} \times \mathbb{R}^n \rightarrow \mathbb{R}^d$ ,  $\sigma : \mathcal{D} \rightarrow \mathbb{R}^{d \times d}$ ,  $c : \mathcal{D} \times \mathbb{X} \rightarrow \mathbb{R}^d$ ,  $f : \mathcal{D} \times \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ ,  $g : \mathcal{D} \times \mathbb{R}^k \rightarrow \mathbb{R}^{n \times n}$  and  $h : \mathcal{D} \times \mathbb{R}^n \times \mathbb{X} \rightarrow \mathbb{R}^n$  satisfying the following.

**Hypothesis 2.**

1. There exists  $L > 0$  such that for every  $\varphi, \tilde{\varphi} \in \mathcal{D}$  and  $y, \tilde{y} \in \mathbb{R}^n$  the following holds

$$\begin{aligned} |a(\varphi, y) - a(\tilde{\varphi}, \tilde{y})| &\leq L \left( \sup_{t \in [-\tau, 0]} |\varphi(t) - \tilde{\varphi}(t)| + |y - \tilde{y}| \right) \\ |\sigma(\varphi) - \sigma(\tilde{\varphi})| &\leq L \left( \sup_{t \in [-\tau, 0]} |\varphi(t) - \tilde{\varphi}(t)| \right) \\ \int_{\mathbb{X}} |c(\varphi, z) - c(\tilde{\varphi}, z)| \nu(dz) &\leq L \left( \sup_{t \in [-\tau, 0]} |\varphi(t) - \tilde{\varphi}(t)| \right) \\ |f(\varphi, y) - f(\tilde{\varphi}, \tilde{y})| &\leq L \left( \sup_{t \in [-\tau, 0]} |\varphi(t) - \tilde{\varphi}(t)| + |y - \tilde{y}| \right) \\ |g(\varphi, y) - g(\tilde{\varphi}, \tilde{y})| &\leq L \left( \sup_{t \in [-\tau, 0]} |\varphi(t) - \tilde{\varphi}(t)| + |y - \tilde{y}| \right) \\ \int_{\mathbb{X}} |h(\varphi, y, z) - h(\tilde{\varphi}, \tilde{y}, z)| \nu(dz) &\leq L \left( \sup_{t \in [-\tau, 0]} |\varphi(t) - \tilde{\varphi}(t)| + |y - \tilde{y}| \right). \end{aligned} \tag{15}$$

2. The functions  $c(0, z), h(0, 0, z)$  are in  $L^1(\nu)$ .

**Remark 2.** Hypothesis 2 implies that the coefficients have sublinear growth; i.e., there exists  $L_1 > 0$  such that, for any  $\varphi \in \mathcal{D}$  and  $y \in \mathbb{R}^n$ ,

$$\begin{aligned} |a(\varphi, y)| &\leq L_1 \left( 1 + \sup_{t \in [-\tau, 0]} |\varphi(t)| + |y| \right) \\ |\sigma(\varphi)| &\leq L_1 \left( 1 + \sup_{t \in [-\tau, 0]} |\varphi(t)| \right) \\ \int_{\mathbb{X}} |c(\varphi, z)| \nu(dz) &\leq L_1 \left( 1 + \sup_{t \in [-\tau, 0]} |\varphi(t)| \right) \\ |f(\varphi, y)| &\leq L_1 \left( 1 + \sup_{t \in [-\tau, 0]} |\varphi(t)| + |y| \right) \\ |g(\varphi, y)| &\leq L_1 \left( 1 + \sup_{t \in [-\tau, 0]} |\varphi(t)| + |y| \right) \\ \int_{\mathbb{X}} |h(\varphi, y, z)| \nu(dz) &\leq L_1 \left( 1 + \sup_{t \in [-\tau, 0]} |\varphi(t)| + |y| \right). \end{aligned} \tag{16}$$

The following assumption on the initial delay segment  $\zeta$  given in (14) is of great importance in the establishment of stable estimates for which we derive (8).

**Hypothesis 3.** The function  $\chi \in C$  is Lipschitz continuous with Lipschitz constant  $\lambda > 0$ , i.e.,

$$|\chi(\theta_1) - \chi(\theta_2)| \leq \lambda|\theta_1 - \theta_2|, \quad \text{for every } \theta_1, \theta_2 \in [-\tau, 0]. \tag{17}$$

**Definition 1.** Given  $T > 0, \tau > 0, \varepsilon > 0, \zeta \in C$  and  $y \in \mathbb{R}^k$  we consider the stochastic basis  $(\mathbb{V}, \mathcal{B}(\mathbb{V}), \mathbb{F}, \mathbb{P})$ . A strong solution of (13) with initial datum (14) is a stochastic process  $(X^\varepsilon, Y^\varepsilon) := \{(X^\varepsilon(t), Y^\varepsilon(t))\}_{t \in [-\tau, T]}$  such that  $X_0^\varepsilon = \chi, Y^\varepsilon(0) = y, X^\varepsilon(t)$  is  $\mathcal{F}_0$ -measurable for any  $t \in [-\tau, 0], (X^\varepsilon(t), Y^\varepsilon(t))_{t \in [0, T]}$  is  $\mathbb{F}$ -adapted and solves (13)  $\mathbb{P}$ -a.s.

We write  $\mathcal{F}_t = \mathcal{F}_0$  for any  $t \in [-\tau, 0]$ . For any  $t \in [0, T]$  and  $\varepsilon > 0$ , the random variables  $X^\varepsilon(t) \in \mathbb{R}^d$  and  $Y^\varepsilon(t) \in \mathbb{R}^k$  are called slow and fast variables, respectively, under the scale separation by the parameter  $\varepsilon > 0$  in the vanishing limit  $\varepsilon \rightarrow 0$ . We underline that the stochastic differential equation for the slow variable  $X^\varepsilon$  lifts the problem to an infinite-dimensional setting due to the dependence of the coefficients in terms of the segment path process.

Given  $T, \tau > 0, m \in \mathbb{N}$  and  $\mathbb{F} := \{\bar{\mathcal{F}}_t\}_{t \in [0, T]}$ , we define the space

$$\mathcal{S}_{\mathbb{F}}^2([-\tau, T]; \mathbb{R}^k) := \left\{ \varphi : \Omega \times [-\tau, T] \rightarrow \mathbb{R}^k \mid \varphi \text{ is } \mathbb{F} \text{-adapted with càdlàg paths such that} \right. \\ \left. \mathbb{E} \left[ \sup_{-\tau \leq u \leq T} |\varphi(u)|^2 \right] < \infty \right\}.$$

The existence and uniqueness of the solution process  $(X^\varepsilon(t), Y^\varepsilon(t))_{t \in [-\tau, T]} \in \mathcal{S}_{\mathbb{F}}^2([-\tau, T]; \mathbb{R}^d) \times \mathcal{S}_{\mathbb{F}}^2([-\tau, T]; \mathbb{R}^n)$  of (13) with initial data (14) follows from Lemma V.2 and Theorem V.7 of [58], using the convention that  $Y^\varepsilon(t) = y$  for all  $t \in [-\tau, 0]$ . This is the content of the following result.

**Theorem 1.** Fix  $T, \tau, \varepsilon > 0$  and  $y \in \mathbb{R}^k$ . Let us assume that Hypotheses 1, 2 and 3 hold for some  $\nu \in \mathbb{M}$  and  $\chi \in C$ . Then, there exists a stochastic process

$$(X^\varepsilon(t), Y^\varepsilon(t))_{t \in [-\tau, T]} \in \mathcal{S}_{\mathbb{F}}^2([-\tau, T]; \mathbb{R}^d) \times \mathcal{S}_{\mathbb{F}}^2([-\tau, T]; \mathbb{R}^n)$$

that solves uniquely (13) in the sense of Definition (1).

### 2.1.3. The Averaged Dynamics

We make the further dissipativity and boundedness assumptions on the coefficients of (13) that yield the existence and uniqueness of solution for the averaged dynamics given by (4) and some stable a priori estimates that will be crucial in the derivation of the result announced in the Introduction.

**Hypothesis 4.**

1. The function  $a$  satisfies  $a(0, y) = 0$  for any  $y \in \mathbb{R}^k$  and there exists  $\Lambda > 0$  such that

$$\begin{aligned} |g(\zeta, y)| &\leq \Lambda \\ |h(\zeta, y, z)| &\leq \Lambda|z|, \quad \text{for every } \zeta \in \mathcal{D}, y \in \mathbb{R}^k, z \in \mathbb{X}. \end{aligned} \tag{18}$$

2. There exist constants  $\beta_1, \beta_2 > 0$ , such that, for any  $\zeta, \zeta_1 \in \mathcal{D}, y, \tilde{y} \times \mathbb{R}^k$  one has

$$2\langle y, f(\zeta, y) \rangle + |g(\zeta, y)|^2 + \int_{\mathbb{X}} |h(\zeta, y, z)|^2 \nu(dz) \leq -\beta_1|y|^2 + \beta_2\|\zeta\|_\infty^2; \tag{19}$$

$$2\langle y - \tilde{y}, f(\zeta, y) - f(\zeta, \tilde{y}) \rangle + |g(\zeta, y) - g(\zeta, \tilde{y})|^2 + \int_{\mathbb{X}} |h(\varphi, y, z) - h(\varphi, \tilde{y}, z)|^2 \nu(dz) \leq -\beta_1 |y - \tilde{y}|^2 + \beta_2 \|\zeta\|_{\infty}^2 \tag{20}$$

and

$$2\langle y - \tilde{y}, f(\zeta, y) - f(\zeta_1, \tilde{y}) \rangle \leq -\beta_1 |y - \tilde{y}|^2 + \beta_2 \|\zeta - \zeta_1\|_{\infty}^2 \tag{21}$$

**Remark 3.** We do not consider a more general framework than Hypotheses 1–4 to derive the moderate deviation principle for the family of slow variables  $(X^\varepsilon)_{\varepsilon>0}$  from (13). Although it would be possible to derive the same result under the setting of locally Lipschitz coefficients and the usual weaker local versions of dissipativity conditions stated in Hypothesis 4. The reason builds on how the weak convergence approach bypasses the usual verification of exponential tightness through the verification of tightness for controlled modifications of the processes  $X^\varepsilon$  under which the use of the usual localization probabilistic techniques works well. Attaining such a degree of generality at the expense of a more technical text is beyond the scope of our work.

The following a priori estimates are straightforward and we omit their proofs.

**Proposition 1.** Fix  $T, \tau > 0$  and  $y \in \mathbb{R}^k$ . Let Hypothesis 1–4 hold for some  $\nu \in \mathbb{M}$  and  $\chi \in \mathcal{C}$ . There exists a constant  $C_1 > 0$  independent of  $\varepsilon > 0$  such that for all  $0 < \varepsilon < 1$  we have

$$\mathbb{E} \left[ \sup_{-\tau \leq t \leq T} |X^\varepsilon(t)|^2 \right] + \sup_{0 \leq t \leq T} \mathbb{E} \left[ |Y^\varepsilon(t)|^2 \right] \leq C_1. \tag{22}$$

We consider the equation for the fast variable of (13) whenever the slow component is frozen and given by  $\zeta \in \mathcal{D}$  in the regime  $\varepsilon = 1$ , i.e., fix  $y \in \mathbb{R}^k$ ; for every  $t \geq 0$ , let

$$Y^{\zeta, y}(t) = y + \int_0^t f(\zeta, Y^{\zeta, y}(s)) ds + \int_0^t g(\zeta, Y^{\zeta, y}(s)) dB^2(s) + \int_0^t \int_{\mathbb{X}} h(\zeta, Y^{\zeta, y}(s-), z) \tilde{N}^{\frac{1}{2}}(ds, dz). \tag{23}$$

We assume that Hypotheses 1–4 hold. We follow [19,20] closely in the argumentation below.

With  $\zeta \in \mathcal{D}$  fixed, we define the transition semigroup on the space  $\mathbb{B}_b(\mathbb{R}^k)$  of the bounded measurable functions associated with the jump diffusion defined by the strong solution of (23) by

$$P_t^\zeta f(y) := \mathbb{E}[f(Y^{\zeta, y}(t))], \quad t \geq 0, \quad y \in \mathbb{R}^k. \tag{24}$$

In what follows, we discuss the existence and uniqueness of an invariant measure for the family of linear operators  $(P_t^\zeta)_{t \geq 0}$ , i.e., a probability measure  $\mu^\zeta \in \mathcal{P}(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$  such that

$$\int_{\mathbb{R}^k} P_t^\zeta f(y) \mu^\zeta(dy) = \int_{\mathbb{R}^k} f(y) \mu^\zeta(dy), \quad t \geq 0, f \in \mathbb{B}_b(\mathbb{R}^k). \tag{25}$$

The dissipativity assumption given in (20) yields some  $C > 0$  such that, for any  $T_0 \geq 0$ , the following bound holds:

$$\sup_{T \geq T_0} \mathbb{E}[|Y^{\zeta, y}(T)|^2] \leq Ce^{-2\beta_1 T} (1 + \|\zeta\|_{\infty}^2 + |y|^2). \tag{26}$$

The estimate (26) implies that the family of the laws of the process  $\{\mathcal{L}(Y^{\zeta, y}(T))\}_{T \geq T_0}$  is tight in  $\mathcal{P}(\mathbb{R}^k; \mathcal{B}(\mathbb{R}^k))$  when  $T_0 \rightarrow \infty$ . Prokhorov’s theorem implies the existence of a weak limit  $\mu^\zeta$  as  $T_0 \rightarrow \infty$  and an indirect use of Krylov–Bogliobov’s theorem (Theorem 7.1 in [59]) asserts that  $\mu^\zeta$  is an invariant measure of  $(P_t^\zeta)_{t \geq 0}$ , in the sense of (25). The setting of assumptions made in Hypotheses 1–4 implies that the semigroup  $(P_t^\zeta)_{t \geq 0}$  is irreducible. We refer the reader to Proposition 2.4 in [60]. Proposition 7.5 in [59] implies that  $\mu^\zeta$  is the

unique invariant measure. Due to the estimate (26) and the definition of  $\mu^\zeta$  in (25), the simple application of monotone convergence shows, as in Lemma 3.4. in [20], that there exists  $C > 0$  such that

$$\int_{\mathbb{R}^k} |y|^2 \mu^\zeta(dy) \leq C(1 + \|\zeta\|_\infty^2 + |y|^2). \tag{27}$$

For any  $\zeta \in \mathcal{D}$ , we can define the averaged mixing coefficient

$$\bar{a}(\zeta) := \int_{\mathbb{R}^k} a(\varphi, y) \mu^\zeta(dy). \tag{28}$$

The proof of the following result concerning the Lipschitz continuity of  $\bar{a}$  is straightforward. It follows in the same way the inequality (3.4) in [61].

**Proposition 2.** Fix  $T, \tau > 0$  and  $y \in \mathbb{R}^k$ . Let Hypothesis 1–4 hold for some  $\nu \in \mathbb{M}$  and  $\chi \in \mathcal{C}$ . Then, the function  $\bar{a}$  defined by (28) is Lipschitz continuous.

Proposition 2 ensures that the averaged differential equation with initial delay data  $\chi \in \mathcal{C}$ ,

$$\begin{cases} \frac{d}{dt} \bar{X}^{0,\chi}(t) &= \bar{a}(\bar{X}_t^{0,\chi}), \\ \bar{X}_0^{0,\chi} &= \chi \end{cases} \tag{29}$$

has a unique solution  $\bar{X}^{0,\chi} \in \mathbb{C}([-\tau, T]; \mathbb{R}^d)$ .

The following proposition, which reads as a strong mixing property of the averaged coefficient  $\bar{a}$  given by (28), plays a crucial role in the establishment of the moderate deviation principle for the family  $(X^\varepsilon)_{\varepsilon>0}$  since it is a fundamental ingredient in the proof of the controlled averaging principle (8). The derivation of this ergodic property follows Lemma 5.2 of [25].

**Proposition 3.** Fix  $T, \tau > 0$  and  $y \in \mathbb{R}^k$ . Let Hypotheses 1–4 hold for some  $\nu \in \mathbb{M}$  and  $\chi \in \mathcal{C}$ . Then, there exists some function  $\alpha : [0, \infty) \rightarrow [0, \infty)$  such that  $\alpha(T) \rightarrow 0$  as  $T \rightarrow \infty$  and satisfying for any  $t \in [0, T]$

$$\mathbb{E} \left| \frac{1}{T} \int_t^{t+T} a(\zeta, Y^{\zeta,y}(s)) ds - \bar{a}(\zeta) \right|^2 \leq \alpha(T)(1 + \|\zeta\|_\infty^2 + |y|^2) \tag{30}$$

where the averaged coefficient  $\bar{a}$  is defined by (28).

### 2.2. The Main Theorem

We make the further assumption on the averaged coefficient  $\bar{a}$  defined by (28).

**Hypothesis 5.** The function  $\bar{a} : \mathcal{D} \rightarrow \mathbb{R}^d$  is Fréchet differentiable and its Fréchet derivative is a Lipschitz function, i.e., there exists some constant  $L_2 > 0$  such that

$$|D\bar{a}(\zeta) - D\bar{a}(\bar{\zeta})| \leq L_2 \left( \sup_{-\tau \leq t \leq 0} |\zeta(t) - \bar{\zeta}(t)| \right), \quad \zeta, \bar{\zeta} \in \mathcal{D}. \tag{31}$$

We define  $L^2(\nu_T) := \left\{ g : [0, T] \times \mathbb{X} \rightarrow [0, \infty) \mid \int_0^T \int_{\mathbb{X}} |g(s, z)|^2 \nu(dz) ds < \infty \right\}$ .

The main result of this work is the content of the next theorem and the reader can find its proof in the next section.

**Theorem 2.** Fix  $T, \tau > 0$  and  $y \in \mathbb{R}^k$ . Let Hypotheses 1–5 hold for some  $\nu \in \mathbb{M}$  and  $\zeta \in \mathcal{C}$ . Let

$$\mathcal{G}^0 : L^2([0, T]; \mathbb{R}^d) \times L^2(\nu_T) \rightarrow \mathbb{C}([-\tau, T]; \mathbb{R}^d)$$

such that

$$\mathcal{G}^0(f, g) = \eta,$$

where for every  $(f, g) \in L^2([0, T]; \mathbb{R}^d) \times L^2(\nu_T)$  the function  $\eta \in C([-\tau, T]; \mathbb{R}^d)$  solves uniquely the skeleton equation

$$\begin{cases} \eta(t) &= \int_0^t D\bar{a}(\bar{X}_s^{0,\xi})\eta_s ds + \int_0^t \sigma(\bar{X}_s^{0,\xi})f(s)ds + \int_0^t \int_{\mathbb{X}} c(\bar{X}_s^{0,\xi}, z)g(s, z)\nu(dz)ds, t \in [0, T] \\ \eta_0 &= 0. \end{cases} \tag{32}$$

and the function  $\bar{X}^{0,\chi} \in C([-\tau, T]; \mathbb{R}^d)$  is the unique solution of (29).

For any  $\eta \in C([-\tau, T]; \mathbb{R}^d)$  we denote

$$\mathcal{G}_\eta^0 := \left\{ (f, g) \in L^2[0, T] \times L^2(\nu_T) \mid \mathcal{G}^0(f, g) = \eta \right\}.$$

For any  $\varepsilon > 0$ , let  $d(\varepsilon) = \varepsilon^{\frac{1-\theta}{2}}$ , for some  $\theta \in (\frac{1}{2}, 1)$ .

For every  $\varepsilon > 0$ , let  $(X^{\varepsilon,\chi,y}(t), Y^{\varepsilon,\chi,y}(t))_{t \in [-\tau, T]}$  be the unique strong solution of (13) with initial condition given by (14) and

$$Z^{\varepsilon,\chi,y} := \frac{X^{\varepsilon,\chi,y} - \bar{X}^{0,\chi,y}(t)}{d(\varepsilon)}. \tag{33}$$

Then, the family  $(Z^{\varepsilon,\chi,y})_{\varepsilon>0}$  defined by (33) satisfies a large deviation principle with speed  $b(\varepsilon) = \varepsilon^\theta \rightarrow 0$  as  $\varepsilon \rightarrow 0$  for some  $\theta \in (\frac{1}{2}, 1)$  and the good rate function

$$\mathbb{I}(\eta) = \inf_{(f,g) \in \mathcal{G}_\eta^0} \frac{1}{2} \left( \int_0^T |f(s)|^2 ds + \int_0^T \int_{\mathbb{X}} |g(s, z)|^2 \nu(dz) ds \right). \tag{34}$$

with the convention that the  $\inf \emptyset = \infty$ .

### 2.3. Examples

Strongly tempered exponentially light Lévy measures.

Hypothesis 1 covers a wide class of Lévy measures and we point out the following special benchmark cases.

1. Our setting covers the simplest case of finite intensity super-exponentially light jump measures given by  $\nu(dz) = e^{-\alpha|z|^2} dz$  for some  $\alpha > 1$ . For every  $\varepsilon > 0$ , the corresponding stochastic process  $L_t^\varepsilon := \int_0^t \int_{\mathbb{X}} z \tilde{N}_\varepsilon^\frac{1}{\varepsilon}(ds, dz)$ ,  $t \geq 0$  is a compensated compound Poisson process.
2. More generally, Hypothesis 1 covers a class of Lévy measures that mimics the class of strongly tempered exponentially light measures introduced by Rosiński in [62], however, with a Gaussian damping in order to satisfy (12). For the polar coordinate  $r = |z|$  and any  $A \in \mathcal{B}(\mathbb{X})$ , we define

$$\nu(A) = \int_{\mathbb{R}^d \setminus \{0\}} \int_0^\infty \mathbf{1}_A(rz) \frac{e^{-r^2}}{r^{\alpha'+1}} dr R(dz), \quad \alpha' \in (0, 2),$$

for some measure  $R \in \mathbb{M}$  such that  $\int_{\mathbb{R}^d \setminus \{0\}} |z|^{\alpha'} R(dz) < \infty$ . We point out that, for every  $\varepsilon > 0$ , the corresponding Lévy process  $(L_t^\varepsilon)_{t \geq 0}$  differs from the compound Poisson process of the paragraph before not only from the fact that the corresponding jump measure has infinite total mass but also from the fact that although a compound Poisson process with positive jumps has almost surely nondecreasing paths, it does not have paths that are almost surely strictly increasing. Such measures and its corresponding processes were introduced in [63] for the study of dynamical fea-

tures of stochastic equations perturbed by jump accelerated noises obeying the large deviations regime.

Invariant measures for the Markov semigroup associated with the fast variable.

1. For every  $\varepsilon > 0$  and  $t \in [0, T]$ , let us consider the multiscale system

$$\begin{cases} dX^\varepsilon(t) = a(X_t^\varepsilon, Y^\varepsilon(t))dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dB^1(t), & X_0^\varepsilon = \zeta \in \mathcal{C} \\ dY^\varepsilon(t) = -\frac{1}{2\varepsilon}Y^\varepsilon(t) + \frac{1}{\varepsilon}dB^2(t), & Y^\varepsilon(0) = y \in \mathbb{R}, \end{cases}$$

where  $B^1$  and  $B^2$  are two independent standard Brownian motions with values in  $\mathbb{R}$ . We assume that the coefficients  $a$  and  $\sigma$  satisfy Hypotheses 2 and 4. For any  $\chi \in \mathcal{C}$  satisfying Hypothesis 3, the invariant measure of the fast variable (decoupled from the slow variable in this case)

$$dY(t) = -\frac{1}{2}Y(t) + dB^2(t), \quad t \geq 0,$$

is given by  $\mu(dy) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} dy$ . Hence, the averaged coefficient  $\bar{a}$  is given for any  $\zeta \in \mathcal{D}$  by

$$\bar{a}(\zeta) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} a(\zeta, y)e^{-\frac{y^2}{2}} dy.$$

The function  $\bar{a}$  satisfies Hypothesis 5 if  $a$  is  $C^1$ -Fréchet differentiable with respect to the first variable  $\zeta$ .

2. Fix  $T, \tau > 0$  and  $y \in \mathbb{R}^k$ . Let Hypotheses 1–5 hold for some  $\nu \in \mathbb{M}$  and  $\chi \in \mathcal{C}$ . For every  $\varepsilon > 0$  and  $t \geq 0$ , let us consider the multiscale system (13) with  $d = k = 1$ . We take  $f(\zeta, y) = -f_1(\zeta)y$  and  $g(\zeta, y) = g(\zeta)$  for every  $\zeta \in \mathcal{D}$  and  $y \in \mathbb{R}^k$  with  $f_1(\zeta) > 0$  and  $g(\zeta) > 0$  for any  $\zeta \in \mathcal{D}$ . Fix the Lévy measure  $\nu(dz) = e^{-|z|^2} dz$  and since this is a finite measure we consider the non-compensated Poisson random measure  $N^\varepsilon$  instead of  $\tilde{N}^\varepsilon$ . Fixed  $\zeta \in \mathcal{D}$ , the Markov semigroup of the the fast variable governed by the dynamics

$$dY^{\zeta, y}(t) = -f_1(\zeta)Y^{\zeta, y}(t) + g(\zeta)dB^2(t) + \int_{\mathbb{R} \setminus \{0\}} \left( \frac{g(\zeta)}{\sqrt{f_1(\zeta)}}z - Y^{\zeta, y}(t) \right) dN^1(ds, dz), \quad t \geq 0,$$

has a unique invariant distribution given by

$$\mu^\zeta(dy) = \sqrt{\frac{f_1(\zeta)}{\pi g^2(\zeta)}} e^{-\frac{f_1(\zeta)y^2}{g^2(\zeta)}} dy.$$

The averaged coefficient  $\bar{a}$ , given for any  $\zeta \in \mathcal{D}$  by

$$\bar{a}(\zeta) = \int_{\mathbb{R} \setminus \{0\}} a(\zeta, y)\mu^\zeta(dy),$$

satisfies Hypothesis 5 if  $a, f$  and  $g$  are  $C^1$ -Fréchet differentiable in order  $\zeta$ . This example was inspired by the examples illustrated in [30] and illustrates that the class of assumptions we make on the coefficients of (13) is not empty.

### 3. Proof of the Main Theorem

Throughout this section, let the standing assumptions made in Theorem 2 hold. Let

$$d(\varepsilon) = \varepsilon^{\frac{1-\theta}{2}}, \varepsilon > 0, \quad \text{for some } \theta \in \left(\frac{1}{2}, 1\right). \tag{35}$$

The speed of the MDP is given by  $b(\varepsilon) := \frac{\varepsilon}{d^2(\varepsilon)} = \varepsilon^\theta \rightarrow 0$ , as  $\varepsilon \rightarrow 0$ .

### 3.1. The Setup of the Weak Convergence Approach

Notation.

We follow extensively the notation introduced by Budhiraja, Dupuis and Ganguly in [46].

Let  $\bar{\mathcal{A}}_+$  (resp.  $\bar{\mathcal{A}}$ ) be the class of all  $(\mathcal{B}(\mathbb{X}) \otimes \bar{\mathcal{P}}) / \mathcal{B}([0, \infty))$  (resp.  $(\mathcal{B}(\mathbb{X}) \otimes \bar{\mathcal{P}}) / \mathcal{B}(\mathbb{R})$ )-measurable maps from  $[0, T] \times \mathbb{X} \times \bar{\mathbb{V}}$  to  $[0, \infty)$  (resp.  $\mathbb{R}$ ). For  $\varphi \in \bar{\mathcal{A}}_+$ , let us define a counting process  $N^\varphi$  on  $\bar{\mathbb{X}}_T$  by

$$N^\varphi(U \times (0, t])(\bar{\omega}) := \int_U \int_0^t \int_0^\infty \mathbf{1}_{[0, \varphi(x, s)(\bar{\omega})]}(r) \bar{N}(dx, dr, ds), \quad t \in [0, T], U \in \mathcal{B}(\mathbb{X}). \quad (36)$$

One can think of  $N^\varphi$  as a controlled random measure with  $\varphi$  selecting the intensity for the points at location  $x$  and time  $s$  in a possibly random but non-anticipating way. When  $\varphi(x, s, \bar{m}) = \theta \in (0, \infty)$ , we write  $N^\varphi = N^\theta$ . For more details, we refer the reader to [43].

Define  $\ell : [0, \infty) \rightarrow [0, \infty)$  by

$$\ell(r) = r \ln r - r + 1, \quad r \in [0, \infty).$$

For any  $\varphi \in \bar{\mathcal{A}}_+$  and  $t \in [0, T]$ , define the quantity

$$L_t(\varphi)(\bar{\omega}) := \int_0^t \int_{\mathbb{X}} \ell(\varphi(s, z, \bar{\omega})) \nu(dz) ds.$$

This is a well defined quantity as a  $[0, \infty]$ -valued random variable.

Let  $\{K_n\}_{n \in \mathbb{N}} \subset \mathbb{X}$  be an increasing sequence of compact sets such that  $\bigcup_{n=1}^\infty K_n = \mathbb{X}$ . For each  $n \in \mathbb{N}$ , let

$$\bar{\mathcal{A}}_{b,n} := \left\{ \varphi \in \bar{\mathcal{A}}_+ \mid \text{for all } (t, \bar{\omega}) \in [0, T] \times \bar{\mathbb{V}} \right. \\ \left. \varphi(t, x, \bar{m}) \in \left[ \frac{1}{n}, n \right] \text{ if } x \in K_n \text{ and } \varphi(t, x, \bar{m}) = 1 \text{ if } x \in K_n^c \right\}$$

and let  $\bar{\mathcal{A}}_b := \bigcup_{n \in \mathbb{N}} \bar{\mathcal{A}}_{b,n}$ . Considering  $\varphi$  as a control that perturbs jump rates away from 1 when  $\varphi \neq 1$ , we see that the controls in  $\bar{\mathcal{A}}_b$  are bounded and perturb only off a compact set where the bounds of the set can depend on  $\varphi$ .

Consider now the space of random variables

$$\mathcal{P}_2 := \left\{ \zeta : [0, T] \times \bar{\mathbb{V}} \rightarrow \mathbb{R}^n \mid \zeta \text{ is } \bar{\mathcal{P}} \otimes \mathcal{B}(\mathbb{R}^n) \text{ measurable such that } \int_0^T |\zeta(s, \omega)|^2 ds < \infty \quad \bar{\mathbb{P}} - \text{a.s.} \right\}$$

and set  $\mathcal{U} = \mathcal{P}_2 \times \bar{\mathcal{A}}_+$ .

For  $\zeta \in \mathcal{P}_2$  define

$$\tilde{L}_T(\zeta)(\bar{\omega}) := \frac{1}{2} \int_0^T |\zeta(\bar{\omega}, s)|^2 ds, \quad \bar{\omega} \in \bar{\mathbb{V}}.$$

For a given random control  $u = (\zeta, \varphi) \in \mathcal{U}$ , define the energy  $\bar{L}_T(u) := \tilde{L}_T(\zeta) + L_T(\varphi)$ . For any  $M > 0$ , let

$$\tilde{S}^M := \{f \in L^2([0, T]; \mathbb{R}^n) \mid \tilde{L}_T(f) \leq M\}.$$

Under the  $L^2$ -weak topology  $\tilde{S}^M$  is a compact subset of  $L^2([0, T]; \mathbb{R}^n)$ . Throughout the rest of this work we consider  $\tilde{S}^M$  to be endowed with this topology. Let

$$S^M := \{g : [0, T] \times \mathbb{X} \rightarrow [0, \infty) \mid L_T(g) \leq M\}.$$

For any  $M > 0$  and under the following identity,

$$S^M \simeq \left\{ \nu_T^g \in \mathbb{M} \mid \nu_T^g(A) := \int_A g(s, z) \nu(dz) ds, \quad A \in \mathcal{B}([0, T] \times \mathbb{X}) \right\},$$

when considering the vague topology in  $\mathbb{M}$  the space  $S^M$  turns out to be compact. For more details, we refer the reader to Lemma 5.1 in [44].

For any  $\varepsilon > 0$  and  $M > 0$ , let us consider the following tightened sublevel sets

$$S_{+, \varepsilon}^M := \left\{ g : [0, T] \times \mathbb{X} \rightarrow [0, \infty) \mid L_T(g) \leq Md^2(\varepsilon) \right\},$$

$$S_\varepsilon^M := \left\{ h : [0, T] \times \mathbb{X} \rightarrow \mathbb{R} \mid h := \frac{g-1}{d(\varepsilon)}, \varphi \in S_{+, \varepsilon}^M \right\}$$

and  $\tilde{S}_\varepsilon^M := \left\{ f : [0, T] \rightarrow \mathbb{R}^n \mid \tilde{L}_T(f) \leq Md^2(\varepsilon) \right\}.$

Define also the random sublevel sets

$$\mathcal{U}_{+, \varepsilon}^M := \left\{ \varphi \in \bar{\mathcal{A}}_b \mid \varphi(\cdot, \cdot, \omega) \in S_{+, \varepsilon}^M \quad \bar{\mathbb{P}} - a.s. \right\},$$

$$\mathcal{U}_\varepsilon^M := \left\{ \psi \in \bar{\mathcal{A}} \mid \psi(\cdot, \cdot, \omega) \in S_\varepsilon^M \quad \bar{\mathbb{P}} - a.s. \right\} \tag{37}$$

and  $\tilde{\mathcal{U}}_\varepsilon^M := \left\{ \xi \in \mathcal{P}_2 \mid \xi(\cdot, \omega) \in \tilde{S}_\varepsilon^M \quad \bar{\mathbb{P}} - a.s. \right\}.$

We reserve the notation  $B_2(R)$  for the closed ball of radius  $R > 0$  in  $L^2(\nu_T)$  and  $\tilde{B}_2(R)$  for the closed ball in  $L^2([0, T]; \mathbb{R}^n)$ .

Fix a given Polish space  $\mathbb{U}$ . Given a measurable map  $\mathcal{G}^0 : \mathbb{W} \times L^2(\nu_T) \rightarrow \mathbb{U}$ , let us write the set of fixed points of  $\eta$  under  $\mathcal{G}^0$ ,

$$\mathbb{S}[\eta] := \left\{ (f, g) \in \mathbb{W} \times L^2(\nu_T) \mid \eta = \mathcal{G}^0(f, g) \right\}$$

and define the quadratic form

$$\mathbb{I}(\eta) := \inf_{(f, g) \in \mathbb{S}[\eta]} \frac{1}{2} \left( \int_0^T |f(s)|^2 ds + \int_0^T \int_{\mathbb{X}} |g(s, z)|^2 \nu_T(ds, dz) \right), \quad \eta \in \mathbb{U}. \tag{38}$$

**Remark 4.** We note that a collection  $\{\psi^\varepsilon\}_{\varepsilon>0} \subset \bar{\mathcal{A}}$  with the property that  $\sup_{\varepsilon>0} \|\psi^\varepsilon\|_2 \leq M$   $\mathbb{P}$ -a.s. for some  $M < \infty$  is regarded as a collection of  $B_2(M)$ -valued random variables where  $B_2(M)$  is equipped with the weak topology on the Hilbert space  $L^2(\nu_T)$ . Since  $B_2(M)$  is weakly compact, such a collection of random variables is automatically tight. Suppose  $\varphi \in S_{+, \varepsilon}^M$ , which, we recall, means that  $L_T(\varphi) \leq Md^2(\varepsilon)$ . Due to Lemma 3.2. in [46] there exists  $\kappa_2(1) \in (0, \infty)$  independent of  $\varepsilon > 0$  and such that  $\psi \mathbf{1}_{\left\{ |\psi| \leq \frac{1}{d(\varepsilon)} \right\}} \in B_2(\sqrt{M\kappa_2(1)})$ , where  $\psi := \frac{\varphi-1}{d(\varepsilon)}$ .

The following set of conditions imply the moderate deviation regime.

**Hypothesis 6.** Let  $\mathbb{U}$  be a Polish space. For any  $\varepsilon > 0$ , let  $\mathcal{G}^\varepsilon : \mathbb{V} \rightarrow \mathbb{U}$  and  $\mathcal{G}^0 : \mathbb{W} \times L^2(\nu_T) \rightarrow \mathbb{U}$  be measurable maps satisfying the following two conditions.

1. **Continuity of the limiting map on the controls.** Suppose  $(f_n, g_n), (f, g) \in \tilde{S}^M \times B_2(M)$  such that  $(f_n, g_n) \rightarrow (f, g)$  as  $n \rightarrow \infty$ . Then

$$\mathcal{G}^0(f_n, g_n) \rightarrow \mathcal{G}^0(f, g) \quad \text{as } n \rightarrow \infty.$$

2. **Weak law for the map under shifts by random tightened controls.** For every  $M < \infty$ , let  $u^\varepsilon := (\xi^\varepsilon, \varphi^\varepsilon) \in \tilde{\mathcal{U}}_\varepsilon^M \times \mathcal{U}_{+, \varepsilon}^M$ . For some  $\beta \in (0, 1)$ , let us assume that  $\psi^\varepsilon \mathbf{1}_{\left\{ |\psi^\varepsilon| < \frac{\beta}{d(\varepsilon)} \right\}} \Rightarrow \psi$  in  $B_2(\sqrt{M\kappa_2(1)})$  where  $\psi^\varepsilon := \frac{\varphi^\varepsilon-1}{d(\varepsilon)}$  and  $\frac{1}{d(\varepsilon)} \xi^\varepsilon \Rightarrow \xi$  as  $\varepsilon \rightarrow 0$  in the weak topology of  $L^2([0, T]; \mathbb{R}^n)$ . Then,

$$\mathcal{G}^\varepsilon \left( \sqrt{\varepsilon} B + \int_0^\cdot \xi^\varepsilon(s) ds, \varepsilon N^{\frac{1}{\varepsilon}} \varphi^\varepsilon \right) \Rightarrow \mathcal{G}^0(\xi, \psi), \quad \text{as } \varepsilon \rightarrow 0.$$

The following theorem is the moderate deviation principle stated in an abstract manner and that will be applied to prove our main result.

**Theorem 3.** *Suppose that for every  $\varepsilon > 0$  the maps  $\mathcal{G}^\varepsilon : \mathbb{V} \rightarrow \mathbb{U}$  and  $\mathcal{G}^0 : \mathbb{W} \times L^2(\nu_T) \rightarrow \mathbb{U}$  satisfy the conditions of Hypothesis 6. Then, the family  $\{Z^\varepsilon\}_{\varepsilon>0}$  defined by*

$$Z^\varepsilon := \mathcal{G}^\varepsilon\left(\sqrt{\varepsilon}B, \varepsilon N^{\frac{1}{\varepsilon}}\right), \quad \varepsilon > 0, \quad \varepsilon > 0, \tag{39}$$

*satisfies a large deviation principle with speed  $b(\varepsilon) \rightarrow 0$  in  $\mathbb{U}$  with good rate function  $\mathbb{I}$  given by (38).*

Theorem 3 is a particular case of Theorem 9.9 in [45]. In what follows, we apply Theorem 3 to our setting.

Let us fix  $T > 0, \tau > 0, (\zeta, y) \in \mathcal{C} \times \mathbb{R}^k$  and for every  $\varepsilon > 0$  let  $(X^{\varepsilon, \zeta, y}(t), Y^{\varepsilon, \zeta, y}(t))_{t \in [-\tau, T]}$  be the unique strong solution of (13) with initial datum (14). For every  $\varepsilon > 0$ , consider  $(Z^\varepsilon)_{\varepsilon>0}$  given by (33). Under the standing assumptions made at the beginning of this section, for any  $\varepsilon > 0$ , Yamada–Watanabe’s theorem ensures the existence of a measurable map  $\mathcal{G}^\varepsilon : \mathbb{V} \rightarrow \mathbb{D}([-\tau, T]; \mathbb{R}^d)$  such that

$$Z^\varepsilon := \mathcal{G}^\varepsilon(\sqrt{\varepsilon}B, \varepsilon N^{\frac{1}{\varepsilon}}). \tag{40}$$

We recall that  $B = (B^1, B^2)$  is a Brownian motion in  $\mathbb{R}^{d \times k}$  due to the independence of  $B^1$  and  $B^2$  and for any  $\varepsilon > 0$  the Poisson random measure  $N^{\frac{1}{\varepsilon}}$  is independent of  $B^1$  and  $B^2$  and hence of  $B$ , which justifies the existence of the Ito map  $\mathcal{G}^\varepsilon$ . The proof of Theorem 2 consists in checking the conditions (1) and (2) of Hypothesis 6 for  $(\mathcal{G}^\varepsilon)_{\varepsilon>0}$  and  $\mathcal{G}^0 : \mathbb{W} \times L^2(\nu_T) \rightarrow C([-\tau, T]; \mathbb{R}^d), \mathcal{G}^0(f, g) = \eta$ , with  $\eta \in C([-\tau, T]; \mathbb{R}^d)$  defined by the skeleton Equation (32). Hence, Theorem 3 allows us to conclude.

### 3.2. The Skeleton Equations and the Compactness Condition

For any  $\chi \in \mathcal{C}$  and  $u = (f, g) \in L^2([0, T]; \mathbb{R}^d) \times L^2(\nu_T)$ , let us denote by  $\bar{Z}^u \in C([-\tau, T]; \mathbb{R}^d)$  the unique solution of (32). By definition, we have

$$\mathcal{G}^0(f, g) = \bar{Z}^u.$$

**Proposition 4.** *For every  $M < \infty$ , one has that the set*

$$\mathbb{K}_M := \left\{ \mathcal{G}^0(f, g) \mid (f, g) \in \tilde{B}_2(M) \times B_2(M) \right\}$$

*is compact in  $C([-\tau, T]; \mathbb{R}^d)$ .*

**Remark 5.** *Proposition 4 is implied by the following. Fix  $0 \leq M < \infty$ . Let  $(f_n, g_n)_{n \in \mathbb{N}} \subset \tilde{B}_2(M) \times B_2(M)$  such that  $(f_n, g_n) \rightarrow (f, g)$  as  $n \rightarrow \infty$  weakly. Therefore,*

$$\mathcal{G}^0(f_n, g_n) \rightarrow \mathcal{G}^0(f, g) \quad \text{as } n \rightarrow \infty.$$

The proof of the sentence of Remark 5 that implies Proposition 4 is standard. We refer the reader to Lemma 4.1 in the seminal work [46].

### 3.3. The Weak Limit of the Controlled Auxiliary Processes

#### 3.3.1. The Equations for the Controlled Auxiliary Processes

This section serves the purpose of verifying the second condition in Hypothesis 6 for  $\mathcal{G}^0$  and the family  $\{\mathcal{G}^\varepsilon : \mathbb{V} \rightarrow \mathbb{D}([-\tau, T]; \mathbb{R}^d)\}_{\varepsilon>0}$ . For every  $\varepsilon > 0$ , recall the random sublevel sets  $\mathcal{U}_\varepsilon^M$  and  $\tilde{\mathcal{U}}_{+\varepsilon}^M$  given by (37) and let  $u^\varepsilon := (\zeta^\varepsilon, \varphi^\varepsilon) \in \mathcal{U}_\varepsilon^M \times \tilde{\mathcal{U}}_{+\varepsilon}^M$ . Set  $\tilde{\varphi}^\varepsilon = \frac{1}{\varphi^\varepsilon}$ . The definition of  $\tilde{\varphi}^\varepsilon$  makes sense since one has  $\varphi^\varepsilon \in \mathcal{A}_b$   $\mathbb{P}$ -a.s. For any  $t \in [0, T]$ , we define the  $\mathbb{F}$ -martingales

$$\begin{aligned} \mathcal{E}(\zeta^\varepsilon)(t) &:= \exp\left(\int_0^t \zeta^\varepsilon(s) dB(s) - \frac{1}{2} \int_0^t |\zeta^\varepsilon(s)|^2 ds\right) \quad \text{and} \\ \mathcal{E}(\tilde{\varphi}^\varepsilon)(t) &:= \exp\left(\int_0^t \int_{\mathbb{X}} \int_0^{\frac{1}{\varepsilon}} \ln \tilde{\varphi}^\varepsilon(s, z) \bar{N}(ds, dz, dr) \right. \\ &\quad \left. + \int_0^t \int_{\mathbb{X}} \int_0^{\frac{1}{\varepsilon}} (-\tilde{\varphi}^\varepsilon(s, z) + 1) ds v(dz) dr\right). \end{aligned}$$

For every  $t \in [0, T]$ , let  $\bar{\mathcal{E}}(u^\varepsilon)(t) := \bar{\mathcal{E}}(\zeta^\varepsilon)(t)\mathcal{E}(\tilde{\varphi}^\varepsilon)(t)$ . Girsanov’s theorem stated in the form of Theorem III.3.24 of [64] ensures that  $(\bar{\mathcal{E}}(u^\varepsilon)(t))_{t \in [0, T]}$  is an  $\bar{\mathbb{F}}$ -martingale. Hence, the probability measures defined on  $(\bar{\mathbb{V}}, \mathcal{B}(\bar{\mathbb{V}}))$  by

$$\mathbb{Q}_T^\varepsilon(G) := \int_G \bar{\mathcal{E}}(u^\varepsilon)(T) d\bar{\mathbb{P}}, \quad \text{for all } G \in \mathcal{B}(\bar{\mathbb{V}})$$

are absolutely continuous with respect to  $\bar{\mathbb{P}}$ . Under  $\mathbb{Q}_T^\varepsilon$ , the stochastic process

$$\tilde{B}^\varepsilon(t) := B(s) - \int_0^t \zeta^\varepsilon(s) ds, \quad t \in [0, T],$$

is a standard Brownian motion and  $\varepsilon N^{\frac{1}{\varepsilon} \varphi^\varepsilon}$  is an independent random measure with the same law of  $\varepsilon N^{\frac{1}{\varepsilon}}$  under  $\bar{\mathbb{P}}$ . We recall that

$$N^{\frac{1}{\varepsilon} \varphi^\varepsilon}((0, t] \times U) := \int_0^t \int_U \int_0^\infty \mathbf{1}_{[0, \frac{1}{\varepsilon} \varphi^\varepsilon]}(r) \bar{N}(ds, dz, dr).$$

For every  $\varepsilon > 0$  and  $t \in [0, T]$ , we write  $\zeta^\varepsilon(t) = (\zeta_1^\varepsilon, \zeta_2^\varepsilon)(t) \in \mathbb{R}^d \times \mathbb{R}^k$ . For any  $(\chi, y) \in \mathcal{C} \times \mathbb{R}^k$ , we define the slow controlled process  $(\mathcal{X}^\varepsilon(t))_{t \in [0, T]}$  and the fast controlled process  $(\mathcal{Y}^\varepsilon(t))_{t \in [0, T]}$  given as the strong solutions of (6) and, respectively, (7) with respect to  $\bar{\mathbb{P}}$  (since  $\mathbb{Q}_T^\varepsilon \ll \bar{\mathbb{P}}$ ).

For every  $\varepsilon > 0$ , we define  $(\bar{\mathcal{X}}^\varepsilon(t))_{t \in [0, T]}$  the fast averaged controlled process as the strong solution under  $\bar{\mathbb{P}}$  of the controlled stochastic differential Equation (9).

For every  $\varepsilon > 0$ , let

$$\mathcal{Z}^\varepsilon := \frac{\mathcal{X}^\varepsilon - \bar{X}^0}{d(\varepsilon)} = \mathcal{G}^\varepsilon\left(\sqrt{\varepsilon} B + \int_0^\cdot \zeta^\varepsilon(s) ds, \varepsilon N^{\frac{1}{\varepsilon} \varphi^\varepsilon}\right) \tag{41}$$

and, respectively,

$$\bar{\mathcal{Z}}^\varepsilon := \frac{\bar{\mathcal{X}}^\varepsilon - \bar{X}^0}{d(\varepsilon)}. \tag{42}$$

The weak limit for the maps under shifts by random tightened controls.

Let  $M < \infty$  and  $\beta \in (0, 1)$ . Let  $(\tilde{\zeta}^\varepsilon, \varphi^\varepsilon) \in \tilde{\mathcal{U}}_\varepsilon^M \times \mathcal{U}_{+, \varepsilon}^M$  such that  $\psi^\varepsilon \mathbf{1}_{|\psi^\varepsilon| < \frac{\beta}{d(\varepsilon)}} \Rightarrow \psi$  in  $B_2(\sqrt{M\kappa_2(1)})$  where  $\psi^\varepsilon := \frac{\varphi^\varepsilon - 1}{d(\varepsilon)}$  and  $\frac{1}{d(\varepsilon)} \tilde{\zeta}^\varepsilon \Rightarrow \zeta$  in  $\tilde{B}_2(M)$ . The conclusion in the second statement in Hypothesis 6 for  $(\mathcal{G}^\varepsilon)_{\varepsilon > 0}$  and  $\mathcal{G}^0$  reads as  $\mathcal{Z}^\varepsilon \Rightarrow \bar{Z}$ , as  $\varepsilon \rightarrow 0$ , where  $\bar{Z} \in \mathcal{C}([-\tau, T]; \mathbb{R}^d)$  solves uniquely

$$\begin{cases} \bar{Z}(t) = \int_0^t D\bar{a}(\bar{X}_s^{0, \chi}) \bar{Z}(s) ds + \int_0^t \sigma(\bar{X}_s^{0, \chi}) \zeta(s) ds + \int_0^t \int_{\mathbb{X}} c(\bar{X}_s^{0, \chi}, z) \psi(s, z) v(dz) ds, & t \in [0, T] \\ \bar{Z}(t) = 0, & t \in [-\tau, 0]. \end{cases} \tag{43}$$

In order to prove that  $\mathcal{Z}^\varepsilon \Rightarrow \bar{Z}$ , as  $\varepsilon \rightarrow 0$ , we proceed as follows.

1. This step passes through two intermediary tasks. Firstly, one shows that the laws of  $(\bar{\mathcal{Z}}^\varepsilon)_{\varepsilon > 0}$  are tight in  $\mathcal{P}(C([-\tau, T]; \mathbb{R}^d))$  (since compact sets in the topology generated by the uniform convergence are also compact sets in the Skorokhod topology). Then, it follows that there exists  $\bar{Z} \in C([-\tau, T]; \mathbb{R}^d)$  such that  $\bar{\mathcal{Z}}^\varepsilon \Rightarrow \bar{Z}$  as  $\varepsilon \rightarrow 0$ . Passing to

the pointwise limit in the equation satisfied by  $\bar{Z}^\varepsilon$  and due to the uniqueness of the solution of (43), we conclude that  $\bar{Z} = \bar{Z}$ .

2. We prove the following strong (controlled) averaging principle:

$$\lim_{\varepsilon \rightarrow 0} \bar{\mathbb{P}} \left( \sup_{t \in [0, T]} |\mathcal{Z}^\varepsilon(t) - \bar{Z}^\varepsilon(t)| > \delta \right) = 0, \quad \text{for any } \delta > 0.$$

From the limit above and Theorem 4.1. in [53], commonly known as Slutsky’s theorem, we can identify  $\bar{Z}$  as the weak limit of  $(\mathcal{Z}^\varepsilon)_\varepsilon$  as  $\varepsilon \rightarrow 0$ .

### 3.3.2. A Priori Estimates and a Localization Procedure

For every  $\varepsilon > 0$ , let  $\mathcal{R}(\varepsilon) > 0$  such that  $\mathcal{R}(\varepsilon) \rightarrow \infty$  and  $d(\varepsilon)\mathcal{R}^2(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . For example,  $\mathcal{R}(\varepsilon) := \frac{1}{\sqrt[4]{d(\varepsilon)}}$ ,  $\varepsilon > 0$  does the job. Consequently,  $\sqrt{\varepsilon}\mathcal{R}^2(\varepsilon) \rightarrow 0$  and therefore  $\varepsilon\mathcal{R}^2(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . For every  $\varepsilon > 0$  and this choice of  $\mathcal{R}(\varepsilon)$ , we define the  $\bar{\mathbb{P}}$ -stopping times

$$\bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon := \inf\{t \in [0, T] \mid \mathcal{X}^\varepsilon(t) \notin B_{\mathcal{R}(\varepsilon)}(0)\}. \tag{44}$$

and

$$\bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon := \inf\{t \in [0, T] \mid \bar{\mathcal{X}}^\varepsilon(t) \notin B_{\mathcal{R}(\varepsilon)}(0)\} \tag{45}$$

The following list of propositions and lemmas are fundamental estimates used in the strategy described above to obtain the conclusion that  $\mathcal{Z}^\varepsilon \Rightarrow \bar{Z}$  as  $\varepsilon \rightarrow 0$ .

**Proposition 5.** *Let the standing assumptions of Theorem 2 hold. For any  $0 < M < \infty$ ,  $(\bar{\xi}^\varepsilon, \varphi^\varepsilon)_{\varepsilon > 0} \subset \bar{\mathcal{U}}_{+, \varepsilon}^M \times \mathcal{U}_{+, \varepsilon}^M$ ,  $\mathcal{R} : (0, 1] \rightarrow (0, \infty)$  such that  $\mathcal{R}(\varepsilon) \rightarrow \infty$  and  $d(\varepsilon)\mathcal{R}^2(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$  and  $T, \tau > 0$ , we have the following. Given  $(\mathcal{X}^\varepsilon(s))_{s \in [-\tau, T]}$  defined by (6) and  $(\bar{\mathcal{X}}^\varepsilon(s))_{s \in [-\tau, T]}$  by (9), there exists  $0 < \varepsilon_0 < 1$  and a constant  $\mathcal{C} > 0$  such that for every  $0 < \varepsilon < \varepsilon_0$  the following estimates hold:*

$$\bar{\mathbb{P}} \left( \sup_{0 \leq t \leq T} |\mathcal{X}^\varepsilon(s)| > \mathcal{R}(\varepsilon) \right) \leq 2e^{-\frac{1}{2}\mathcal{R}(\varepsilon)} + \mathcal{C}\varepsilon\mathcal{R}(\varepsilon) \tag{46}$$

and

$$\bar{\mathbb{P}} \left( \sup_{0 \leq t \leq T} |\bar{\mathcal{X}}^\varepsilon(s)| > \mathcal{R}(\varepsilon) \right) \leq 2e^{-\frac{1}{2}\mathcal{R}(\varepsilon)} + \mathcal{C}\varepsilon\mathcal{R}(\varepsilon) \tag{47}$$

The proof follows the same reasoning employed in Lemma 2.1 of [63].

**Proposition 6.** *Let  $M > 0$ . Fix a function  $\mathcal{R} : (0, \infty) \rightarrow (0, \infty)$  satisfying the assumptions of Proposition 5 and for every  $\varepsilon > 0$  let  $\bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon$  defined by (44). Under the assumptions of Hypotheses 1–5, there exists some  $\varepsilon_0 > 0$  such that the following bound holds:*

$$\Gamma_1(M) := \sup_{0 < \varepsilon < \varepsilon_0} \sup_{(\bar{\xi}, \varphi) \in \bar{\mathcal{U}}_\varepsilon^M \times \mathcal{U}_{+, \varepsilon}^M} \left( \bar{\mathbb{E}} \left[ \sup_{-\tau \leq t \leq \bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\mathcal{X}^\varepsilon(t)|^2 \right] + \sup_{-\tau \leq t \leq T} \bar{\mathbb{E}} \left[ |\mathcal{Y}^\varepsilon(t)|^2 \mathbf{1}_{\{\bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon > T\}} \right] \right) < \infty. \tag{48}$$

The proof follows from applying successfully Ito’s formula, BDG inequalities and Lemma A1 presented in Section A.1.1 of Appendix A.

**Proposition 7.** *Fix  $M > 0$ ,  $\mathcal{R} : (0, \infty) \rightarrow (0, \infty)$  satisfying the hypotheses of Proposition 5 and for every  $\varepsilon > 0$  let  $\bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon$  be defined by (45). Under Hypotheses 1–5, there exists some  $\varepsilon_0 > 0$  such that the following holds:*

$$\Gamma_2(M) := \sup_{0 < \varepsilon < \varepsilon_0} \sup_{(\bar{\xi}, \varphi) \in \bar{\mathcal{U}}_\varepsilon^M \times \mathcal{U}_{+, \varepsilon}^M} \bar{\mathbb{E}} \left[ \sup_{-\tau \leq t \leq \bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\bar{\mathcal{X}}^\varepsilon(t)|^2 \right] < \infty. \tag{49}$$

The proof of Proposition 7 follows analogously from the proof of (48). For this reason, we omit it.

**Lemma 1.** Fix  $M > 0$ ,  $\mathcal{R} : (0, \infty) \rightarrow (0, \infty)$  under the hypotheses of Proposition 5 and for every  $\varepsilon > 0$  let  $\bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon$  defined by (45). Under the assumptions of Hypotheses 1–5, there exists some  $\varepsilon_0 > 0$  such that the following holds:

$$\Gamma_3(M) := \sup_{0 < \varepsilon < \varepsilon_0} \sup_{(\zeta, \varphi) \in \tilde{\mathcal{U}}_\varepsilon^M \times \mathcal{U}_{+, \varepsilon}^M} \mathbb{E} \left[ \sup_{-\tau \leq t \leq \bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\bar{\mathcal{Z}}^\varepsilon(t)|^2 \right] < \infty. \tag{50}$$

The proof of (50) is straightforward and we omit it.

### 3.3.3. Identification of the Weak Limit

Given  $M < \infty$  and  $\varepsilon > 0$ , let  $\zeta^\varepsilon \in \tilde{\mathcal{U}}_\varepsilon^M$ ,  $\varphi^\varepsilon \in \mathcal{U}_{+, \varepsilon}^M$  and write  $\psi^\varepsilon := \frac{\varphi^\varepsilon - 1}{d(\varepsilon)}$ . Assume that for some  $\beta \in (0, 1)$  the following convergences (in law) are satisfied

$$\psi^\varepsilon \mathbf{1}_{\{|\psi^\varepsilon| \leq \frac{\beta}{d(\varepsilon)}\}} \Rightarrow \psi \quad \text{and} \quad \frac{1}{d(\varepsilon)} \zeta^\varepsilon \Rightarrow \zeta, \quad \text{as } \varepsilon \rightarrow 0.$$

Then, the following result holds.

**Proposition 8.** Let the standing assumptions of Theorem 2 hold for some  $\nu \in \mathbb{M}$  and  $\zeta \in \mathcal{C}$ . For every  $\varepsilon > 0$ , let  $(\bar{\mathcal{Z}}^\varepsilon(t))_{t \in [-\tau, T]}$  be defined by (42). Then, the family  $(\bar{\mathcal{Z}}^\varepsilon, \frac{1}{d(\varepsilon)} \zeta^\varepsilon, \psi^\varepsilon \mathbf{1}_{\{|\psi^\varepsilon| \leq \frac{\beta}{d(\varepsilon)}\}})_{\varepsilon > 0}$  is tight in  $\mathbb{D}([-\tau, T]; \mathbb{R}^d) \times \tilde{\mathcal{B}}_2(M) \times B_2(\sqrt{M\kappa_2(1)})$  for some  $\beta \in (0, 1)$  and  $\kappa_2(1)$  given in the Remark 4. Furthermore, any limit point in law  $(\bar{\mathcal{Z}}, \zeta, \psi)$  satisfies (43).

The proof follows with standard arguments used by the weak convergence approach to moderate deviation principles for stochastic differential equations with jumps. We refer the reader to Lemma 4.9 in the seminal work [46].

### 3.4. The Controlled Averaging Principle

The main result of this section allows us to identify the weak limit of  $(\mathcal{Z}^\varepsilon)_{\varepsilon > 0}$  with the weak limit of the family  $(\bar{\mathcal{Z}}^\varepsilon)_{\varepsilon > 0}$  as  $\varepsilon \rightarrow 0$ .

**Theorem 4.** Let the hypotheses of Theorem 2 hold. Then, given the families  $(\mathcal{Z}^\varepsilon)_{\varepsilon > 0}$  and  $(\bar{\mathcal{Z}}^\varepsilon)_{\varepsilon > 0}$ , defined, respectively, by (41) and (42), we have for any  $\delta > 0$

$$\lim_{\varepsilon \rightarrow 0} \mathbb{P} \left( \sup_{0 \leq t \leq \bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\mathcal{Z}^\varepsilon(t) - \bar{\mathcal{Z}}^\varepsilon(t)| > \delta \right) = 0. \tag{51}$$

The reader can find the proof in Section 3.4.4. km

#### 3.4.1. Khasminkii’s Auxiliary Processes

We follow the technique introduced in [15] with the required modifications to our settings in order to deal with the nonlocal components of the auxiliary processes  $(\mathcal{X}^\varepsilon)_{\varepsilon > 0}$  and  $(\mathcal{Y}^\varepsilon)_{\varepsilon > 0}$  given, respectively, by (6) and (7).

Let  $[-\tau, T]$  be divided into intervals of the same length parametrized for every  $\varepsilon > 0$

$$\Delta = \Delta(\varepsilon) := \varepsilon^\gamma d^2(\varepsilon) |\ln \varepsilon|^p, \quad \text{for some } \gamma \in \left(0, \theta - \frac{1}{2}\right) \quad \text{and } p > 0. \tag{52}$$

where the scale  $d(\varepsilon)$  is given by (35).

We note the following convergences that follow directly from the choice of  $\Delta = \Delta(\varepsilon)$  in (52):

$$\Delta(\varepsilon) \rightarrow 0; \quad \frac{\Delta(\varepsilon)}{d^2(\varepsilon)} \rightarrow 0; \quad \text{and} \quad \frac{\Delta(\varepsilon)}{\varepsilon} \rightarrow \infty \quad \text{as } \varepsilon \rightarrow 0. \tag{53}$$

For any  $t \in [-\tau, T]$ , we denote  $t_\Delta := \lfloor \frac{t}{\Delta} \rfloor \Delta$ .

We construct the auxiliary processes  $(\hat{\mathcal{Y}}^\varepsilon(t))_{t \in [0, T]}$  and  $(\hat{\mathcal{X}}^\varepsilon(t))_{t \in [0, T]}$  by means of the following equations: for any  $t \in [0, T]$ , let

$$\begin{aligned} \hat{\mathcal{Y}}^\varepsilon(t) &= \mathcal{Y}^\varepsilon(t_\Delta) + \frac{1}{\varepsilon} \int_{t_\Delta}^t \left( f(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) + g(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) \zeta_2^\varepsilon(s) \right) ds + \int_{\mathbb{X}} h(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s), z) (\varphi^\varepsilon(s, z) - 1) \nu(dz) ds \\ &+ \frac{1}{\sqrt{\varepsilon}} \int_{t_\Delta}^t g(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) dB^2(s) + \int_{t_\Delta}^t \int_{\mathbb{X}} h(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s-), z) \tilde{N}^{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz) \end{aligned} \tag{54}$$

and

$$\begin{aligned} \hat{\mathcal{X}}^\varepsilon(t) &= \zeta(0) + \int_0^t \left( a(\mathcal{X}_{s_\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(s)) + \sigma(\mathcal{X}_s^\varepsilon) \bar{\zeta}_1^\varepsilon(s) + \int_{\mathbb{X}} c(\mathcal{X}_s^\varepsilon, z) (\varphi^\varepsilon(s, z) - 1) \nu(dz) \right) ds \\ &+ \sqrt{\varepsilon} \int_0^t \sigma(\mathcal{X}_s^\varepsilon) dB^1(s) + \varepsilon \int_0^t \int_{\mathbb{X}} c(\mathcal{X}_{s-}^\varepsilon, z) \tilde{N}^{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz). \end{aligned} \tag{55}$$

### 3.4.2. Auxiliary Estimates

For every  $\varepsilon > 0$ , let us recall the  $\bar{\mathbb{F}}$ -stopping time  $\bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon$  given by (44) for the fixed parametrization  $\mathcal{R}$  given in Proposition 5. The following lemmas are essential a priori bounds that we use in the proof of the controlled averaging principle stated in Theorem 4.

**Lemma 2.** For every  $\varepsilon > 0$ , let  $\mathcal{R}(\varepsilon) > 0$ ,  $b(\varepsilon) := \frac{\varepsilon}{d^2(\varepsilon)}$  and  $\Delta(\varepsilon) > 0$  be fixed as above. Then, for any  $(g_\varepsilon)_{\varepsilon > 0}$  such that  $g(\varepsilon) \simeq_\varepsilon d(\varepsilon)$  as  $\varepsilon \rightarrow 0$ , the following asymptotic regime holds:

$$\bar{\mathbb{P}} \left( \sup_{0 \leq t \leq \bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} \|\mathcal{X}_t^\varepsilon - \mathcal{X}_{t_\Delta}^\varepsilon\|_\infty > g_\varepsilon \right) \lesssim_\varepsilon \Xi(\varepsilon) \rightarrow 0, \quad \text{as } \varepsilon \rightarrow 0, \tag{56}$$

where

$$\Xi(\varepsilon) := b^2(\varepsilon) \frac{\varepsilon}{|\ln \varepsilon|^{2p}} + \frac{\varepsilon^{2\theta-1-2\gamma}}{|\ln \varepsilon|^{2p-q}} + \frac{\varepsilon^{2\theta-1-\gamma}}{|\ln \varepsilon|^{p-q}} + \varepsilon^\gamma \varepsilon^{1-\theta} |\ln \varepsilon|^{2p}, \quad q > 2\gamma + 3, \quad \varepsilon > 0. \tag{57}$$

The proof is given in Section A.2.1 of Appendix A.

**Lemma 3.** For every  $\varepsilon > 0$ , let  $\mathcal{R}(\varepsilon)$  be fixed as in Proposition 5 and  $\Delta(\varepsilon)$  given by (52). Then, the following convergence holds,

$$\sup_{0 \leq t \leq T} \bar{\mathbb{E}} \left[ |\mathcal{Y}^\varepsilon(t) - \hat{\mathcal{Y}}^\varepsilon(t)| \mathbf{1}_{\{T < \bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} \right] \lesssim_\varepsilon \frac{C(\varepsilon)}{\Delta(\varepsilon)} e^{-\frac{2\Delta(\varepsilon)}{2\varepsilon} + 1} \text{as } \varepsilon \rightarrow 0. \tag{58}$$

for some  $C(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$  uniformly in the initial condition  $(\chi, y) \in \mathcal{C} \times \mathbb{R}^k$ .

The proof is given in Section A.2.2 of Appendix A.

### 3.4.3. Khasminkii’s Technique

**Proposition 9.** For any  $\delta > 0$ , we have

$$\limsup_{\varepsilon \rightarrow 0} \bar{\mathbb{P}} \left( \sup_{0 \leq t \leq \bar{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\mathcal{X}^\varepsilon(t) - \hat{\mathcal{X}}^\varepsilon(t)| > \frac{\delta d(\varepsilon)}{2} \right) = 0. \tag{59}$$

**Proof.** The definitions of  $(\mathcal{Y}^\varepsilon(t))_{t \in [0, T]}$  and  $(\hat{\mathcal{Y}}^\varepsilon(t))_{t \in [0, T]}$  given in (7) and (54), respectively, combined with Hypothesis 2 yield for every  $\varepsilon > 0$  and  $t \in [0, T]$  that

$$\begin{aligned} \hat{\mathcal{X}}^\varepsilon(t) - \mathcal{X}^\varepsilon(t) &= \int_0^t \left( a(\mathcal{X}_{s_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - a(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s)) \right) ds \\ &\leq L \int_0^t \|\mathcal{X}_{s_\Delta}^\varepsilon - \mathcal{X}_s^\varepsilon\|_\infty ds + L \int_0^t |\hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)| ds. \end{aligned}$$

The asymptotic behaviour (53) of  $\Delta(\varepsilon) > 0$  fixed in (52) combined with Lemma 2, (A14) and (A15) of Lemma 3 yield some  $C = C(L, T) > 0$  such that

$$\begin{aligned} \mathbb{P}\left(\sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\hat{\mathcal{X}}^\varepsilon(t) - \mathcal{X}^\varepsilon(t)| > \frac{d(\varepsilon)}{2}\right) &\leq \mathbb{P}\left(\int_0^{T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |a(\mathcal{X}_{s\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - a(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s))| ds > \frac{d(\varepsilon)}{2}\right) \\ &\leq \mathbb{P}\left(\sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} \|\mathcal{X}_{t\Delta}^\varepsilon - \mathcal{X}_t^\varepsilon\|_\infty > Cd(\varepsilon)\right) \\ &\quad + \mathbb{P}\left(\int_0^T |\hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} ds > Cd^2(\varepsilon)\right) \\ &\lesssim_\varepsilon \Xi(\varepsilon) + \frac{1}{d^2(\varepsilon)} \int_0^T \mathbb{E}\left[|\hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} ds\right] \\ &\lesssim_\varepsilon \Xi(\varepsilon) + \frac{C_2(\varepsilon)}{\Delta(\varepsilon)d^2(\varepsilon)} e^{-\frac{\Delta(\varepsilon)}{2\varepsilon} + 1} \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \end{aligned}$$

This finishes the proof of (59).  $\square$

**Proposition 10.** For any  $\delta > 0$ , we have

$$\limsup_{\varepsilon \rightarrow 0} \mathbb{P}\left(\sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\hat{\mathcal{X}}^\varepsilon(t) - \bar{\mathcal{X}}^\varepsilon(t)| > \frac{\delta d(\varepsilon)}{2}\right) = 0. \tag{60}$$

**Proof.** For every  $\varepsilon > 0, t \in [0, T], \zeta \in \mathcal{D}, \xi \in \tilde{\mathcal{U}}_{+,\varepsilon}^M$  and  $\varphi^\varepsilon \in \mathcal{U}_{+,\varepsilon}^M$ , we define the function

$$b^\varepsilon(\zeta)(t) := \int_0^t \left(\sigma(\zeta)\xi_1^\varepsilon(s) + \int_{\mathbb{X}} c(\zeta, z)(\varphi^\varepsilon(s, z) - 1)\nu(dz)\right).$$

The definitions of  $(\mathcal{X}^\varepsilon(t))_{t \in [0, T]}$  and  $(\hat{\mathcal{X}}^\varepsilon(t))_{t \in [0, T]}$  given, respectively, in (6) and (55) combined with the definition of  $b^\varepsilon$  given above imply for every  $t \in [0, T]$  and  $\varepsilon > 0$  the following identity  $\mathbb{P}$ -a.s. on the event  $\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}$ :

$$\begin{aligned} \hat{\mathcal{X}}^\varepsilon(t) - \bar{\mathcal{X}}^\varepsilon(t) &= \int_0^t \left(b^\varepsilon(\hat{\mathcal{X}}_s^\varepsilon) - b^\varepsilon(\bar{\mathcal{X}}_s^\varepsilon)\right) ds \\ &\quad + \int_0^t \left(a(\mathcal{X}_{s\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - \bar{a}(\mathcal{X}_s^\varepsilon)\right) ds \\ &\quad + \int_0^t \left(\bar{a}(\mathcal{X}_s^\varepsilon) - \bar{a}(\hat{\mathcal{X}}_s^\varepsilon)\right) ds + \int_0^t \left(\bar{a}(\hat{\mathcal{X}}_s^\varepsilon) - \bar{a}(\bar{\mathcal{X}}_s^\varepsilon)\right) ds \\ &\quad + \sqrt{\varepsilon} \int_0^t \left(\sigma(\mathcal{X}_s^\varepsilon) - \sigma(\bar{\mathcal{X}}_s^\varepsilon)\right) dB^1(s) \\ &\quad + \varepsilon \int_0^t \int_{\mathbb{X}} \left(c(\mathcal{X}_{s-}^\varepsilon, z) - c(\bar{\mathcal{X}}_{s-}^\varepsilon, z)\right) \tilde{N}^{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz). \end{aligned} \tag{61}$$

Hypothesis 2, Proposition 2 and (61) yield some constant  $C = C(L, T) > 0$  such that on the event  $\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}$  we have  $\mathbb{P}$ -a.s.

$$\begin{aligned} \sup_{0 \leq s \leq t} |\hat{\mathcal{X}}^\varepsilon(s) - \bar{\mathcal{X}}^\varepsilon(s)|^2 &\leq C \left(\int_0^t \sup_{0 \leq u \leq s} |\hat{\mathcal{X}}^\varepsilon(u) - \bar{\mathcal{X}}^\varepsilon(u)|^2 ds + \sup_{0 \leq s \leq t} \left|\int_0^s \left(a(\mathcal{X}_{u\Delta}^\varepsilon, \mathcal{Y}_u^\varepsilon) - \bar{a}(\mathcal{X}_u^\varepsilon)\right) du\right|^2\right) \\ &\quad + \sup_{t \in [0, T]} |J_1^\varepsilon(t)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} + \sup_{t \in [0, T]} |J_2^\varepsilon(t)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}}, \end{aligned}$$

where for any  $\varepsilon > 0$  we write

$$\begin{cases} J_1^\varepsilon(t) := \sqrt{\varepsilon} \int_0^t \left(\sigma(\mathcal{X}_s^\varepsilon) - \sigma(\bar{\mathcal{X}}_s^\varepsilon)\right) dB_1(s) \text{ and} \\ J_2^\varepsilon(t) := \varepsilon \int_0^t \int_{\mathbb{X}} \left(c(\mathcal{X}_{s-}^\varepsilon, z) - c(\bar{\mathcal{X}}_{s-}^\varepsilon, z)\right) \tilde{N}^{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz). \end{cases}$$

Gronwall’s lemma implies for any  $\varepsilon > 0$  that

$$\begin{aligned} \sup_{-\tau \leq t \leq T} |\mathcal{X}^\varepsilon(t) - \bar{\mathcal{X}}^\varepsilon(t)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} &\leq e^{CT} \left( \sup_{0 \leq s \leq t} \left| \int_0^s \left( a(\mathcal{X}_{u_\Delta}^\varepsilon, \mathcal{Y}_u^\varepsilon) - \bar{a}(\mathcal{X}_u^\varepsilon) \right) du \right|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} \right. \\ &\quad \left. + \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_1^\varepsilon(t)|^2 + \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_2^\varepsilon(t)|^2 \right). \end{aligned} \tag{62}$$

The estimate (62) yields for any  $\delta > 0$

$$\begin{aligned} \mathbb{P} \left( \sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\hat{\mathcal{X}}^\varepsilon(t) - \bar{\mathcal{X}}^\varepsilon(t)| > \frac{d(\varepsilon)\delta}{2} \right) &\leq \mathbb{P} \left( \sup_{0 \leq s \leq t} \left| \int_0^s \left( a(\mathcal{X}_{u_\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(u)) - \bar{a}(\mathcal{X}_u^\varepsilon) \right) du \right|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta^2 d^2(\varepsilon) e^{-2CT}}{12} \right) \\ &\quad + \mathbb{P} \left( \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_1^\varepsilon(t)|^2 > \frac{\delta^2 d^2(\varepsilon) e^{-2CT}}{12} \right) \\ &\quad + \mathbb{P} \left( \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_2^\varepsilon(t)|^2 > \frac{\delta^2 d^2(\varepsilon) e^{-2CT}}{12} \right). \end{aligned} \tag{63}$$

Burkholder–Davis–Gundy’s inequalities and the sublinear growth of  $\sigma$  given by (16) in Remark 2 yield some constant  $C_2 = C_2(\delta, C_1, L_1, \Gamma_1, \Gamma_2) > 0$ , where  $\Gamma_1, \Gamma_2$  are given, respectively, by (48) in Proposition 6 and (49) in Proposition 7, such that

$$\begin{aligned} \mathbb{P} \left( \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_1^\varepsilon(t)|^2 > \frac{\delta^2 d^2(\varepsilon) e^{-2CT}}{12} \right) &\leq \frac{12e^{2CT}}{\delta^2 d^2(\varepsilon)} \mathbb{E} \left[ \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_1^\varepsilon(t)|^2 \right] \\ &\leq C_2 \frac{\varepsilon}{d^2(\varepsilon)} \rightarrow 0, \quad \text{as } \varepsilon \rightarrow 0. \end{aligned} \tag{64}$$

Analogously, due to Burkholder–Davis–Gundy’s inequalities and (A1) given in Lemma A1 of Appendix A—Section A.1, there exists some constant  $C_3 = C_3(\delta, C_1, L_1, \Gamma_1, \Gamma_2, M) > 0$  that may change from line to line, such that

$$\begin{aligned} \mathbb{P} \left( \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_2^\varepsilon(t)|^2 > \frac{\delta^2 d^2(\varepsilon) e^{-2CT}}{12} \right) &\leq \frac{12e^{2CT}}{\delta^2 d^2(\varepsilon)} \mathbb{E} \left[ \sup_{0 \leq t \leq T \wedge \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |J_2^\varepsilon(t)|^2 \right] \\ &\leq \frac{\varepsilon}{d^2(\varepsilon)} C_3 \sup_{g \in \mathcal{S}_{+\varepsilon}^M} \int_0^T \int_{\mathbb{X}} |z|^2 g(s, z) \nu(dz) ds \\ &\leq C_3 b(\varepsilon) (T + d^2(\varepsilon)) \rightarrow 0. \end{aligned} \tag{65}$$

We estimate now the first term on the right-hand side of (63). For every  $\varepsilon > 0$  and  $t \in [0, T]$ , we write  $\mathbb{P}$ -a.s. on the event  $\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}$

$$\begin{aligned} \int_0^t \left( a(\mathcal{X}_{s_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - \bar{a}(\mathcal{X}_s^\varepsilon) \right) ds &= \sum_{k=0}^{\lfloor \frac{t}{\Delta} \rfloor - 1} \int_{k\Delta}^{(k+1)\Delta} \left( a(\mathcal{X}_{k\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(s)) - \bar{a}(\mathcal{X}_{k\Delta}^\varepsilon) \right) ds \\ &\quad + \sum_{k=0}^{\lfloor \frac{t}{\Delta} \rfloor - 1} \int_{k\Delta}^{(k+1)\Delta} \left( \bar{a}(\mathcal{X}_{k\Delta}^\varepsilon) - \bar{a}(\mathcal{X}_s^\varepsilon) \right) ds \\ &\quad + \int_{t_\Delta}^t \left( a(\mathcal{X}_{s_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - \bar{a}(\mathcal{X}_s^\varepsilon) \right) ds \\ &:= I_1^\varepsilon + I_2^\varepsilon + I_3^\varepsilon. \end{aligned} \tag{66}$$

It follows from (66) that

$$\begin{aligned} \mathbb{P}\left(\sup_{0 \leq s \leq t} \left| \int_0^s (a(\mathcal{X}_{u_\Delta}^\varepsilon, \mathcal{Y}_u^\varepsilon) - \bar{a}(\mathcal{X}_u^\varepsilon)) du \right|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta^2 d^2(\varepsilon) e^{-2CT}}{12}\right) &\leq \mathbb{P}\left(\sup_{0 \leq t \leq T} |I_1^\varepsilon(t)| \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta d(\varepsilon) e^{-CT}}{6\sqrt{3}}\right) \\ &+ \mathbb{P}\left(\sup_{0 \leq t \leq T} |I_2^\varepsilon(t)| \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta d(\varepsilon) e^{-CT}}{6\sqrt{3}}\right) \\ &+ \mathbb{P}\left(\sup_{0 \leq t \leq T} |I_3^\varepsilon(t)| \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta d(\varepsilon) e^{-CT}}{6\sqrt{3}}\right). \end{aligned} \tag{67}$$

We estimate  $I_2^\varepsilon$ .

We observe that for any  $\varepsilon > 0$

$$I_2^\varepsilon = \int_0^{t_\Delta} (\bar{a}(\mathcal{X}_{s_\Delta}^\varepsilon) - \bar{a}(\mathcal{X}_s^\varepsilon)) ds.$$

Proposition 2 and Lemma 2 imply for some  $C_4 = C(T) > 0$ , any  $\delta > 0$  and  $\varepsilon > 0$  small enough that

$$\mathbb{P}\left(\sup_{t \in [0, T]} |I_2^\varepsilon(t)| \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta d(\varepsilon) e^{-CT}}{6\sqrt{3}}\right) \leq \mathbb{P}\left(\sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} \int_0^{t_\Delta} |\mathcal{X}_{s_\Delta}^\varepsilon - \mathcal{X}_s^\varepsilon| > C_4 d(\varepsilon)\right) \lesssim_\varepsilon \Xi(\varepsilon) \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \tag{68}$$

We estimate  $I_3^\varepsilon$ .

Hypothesis 2, Proposition 2 and Proposition 6 yield some constant  $C_5 = C_5(L, \Gamma_1(M)) > 0$  that may change from line to line such that, for every  $\varepsilon > 0$  small enough and any  $\delta > 0$ , one has

$$\begin{aligned} \mathbb{P}\left(\sup_{t \in [0, T]} |I_3^\varepsilon(t)| \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta d(\varepsilon) e^{-CT}}{6\sqrt{3}}\right) &\leq \frac{C_5}{d^2(\varepsilon)} \mathbb{E}\left[\sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} \left| \int_{t_\Delta}^t (a(\mathcal{X}_{s_\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(s)) - \bar{a}(\mathcal{X}_s^\varepsilon)) ds \right|^2\right] \\ &\leq \frac{C_5 \Delta(\varepsilon)}{d^2(\varepsilon)} \mathbb{E}\left[\int_0^T (1 + \|\mathcal{X}_s^\varepsilon\|_\infty^2 + \|\mathcal{X}_{s_\Delta}^\varepsilon\|_\infty^2 + \|\mathcal{Y}^\varepsilon(s)\|^2) \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} ds\right] \\ &\lesssim_\varepsilon \frac{\Delta(\varepsilon)}{d^2(\varepsilon)} \rightarrow 0, \text{ as } \varepsilon \rightarrow 0, \end{aligned} \tag{69}$$

due to (53).

We estimate  $I_1^\varepsilon$ . We construct a new process  $Z := \mathcal{Y}^\varepsilon(\mathcal{X}_{k\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(k\Delta))$  where the notation that is displayed here stresses that the process is the fast variable process  $\mathcal{Y}^\varepsilon$  with frozen slow component  $\mathcal{X}_{k\Delta}^\varepsilon$  and initial condition  $\mathcal{Y}^\varepsilon(k\Delta)$ . It is a classical fact in the course of the Khasminkii technique employed in [15] for the proof of the strong averaging principle that for every  $s \in [0, \Delta]$  we have

$$(\mathcal{X}_{k\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(s + k\Delta)) =^d \left(\mathcal{X}_{k\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(\mathcal{X}_{k\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(k\Delta))\right) \left(\frac{s}{\varepsilon}\right).$$

We may assume in addition that the fabricated noises above are independent of  $\mathcal{X}_{k\Delta}^\varepsilon$  and  $\mathcal{Y}^\varepsilon(k\Delta)$ . For the proof of the statements above, we refer the reader to Section 5 in [25]. Hence, Proposition 3 together with the Markov property of  $(X_t^\varepsilon, Y^\varepsilon(t))_{t \in [0, T]}$  implies for every  $k = 0, \dots, \lfloor \frac{t}{\Delta} \rfloor$  the following:

$$\begin{aligned} \mathbb{E}\left[\left| \int_{k\Delta}^{(k+1)\Delta} (a(\mathcal{X}_{k\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(s)) - \bar{a}(\mathcal{X}_{k\Delta}^\varepsilon)) ds \right|\right] &\leq \Delta \mathbb{E}\left[\left| \frac{\varepsilon}{\Delta} \int_0^\Delta (a(\mathcal{X}_{k\Delta}^\varepsilon, Z(s)) - \bar{a}(\mathcal{X}_{k\Delta}^\varepsilon)) ds \right|\right] \\ &= \Delta \mathbb{E}\left[\mathbb{E}\left[\left| \frac{\varepsilon}{\Delta} \int_0^\Delta a(\zeta, Z^{\zeta, \mathcal{Y}}) - \bar{a}(\zeta) \right| \middle| (\zeta, \mathcal{Y}) = (\mathcal{X}_{k\Delta}^\varepsilon, \mathcal{Y}^\varepsilon(k\Delta))\right]\right] \\ &\leq \Delta \alpha \left(\frac{\Delta}{\varepsilon}\right) \left(1 + \mathbb{E}\|\mathcal{X}_{k\Delta}^\varepsilon\| + \mathbb{E}\|\mathcal{Y}^\varepsilon(k\Delta)\|\right). \end{aligned} \tag{70}$$

Proposition 3, Proposition 6, (53) and (70) yield, for any  $\delta > 0$  and  $\varepsilon > 0$  sufficiently small, that

$$\begin{aligned} \mathbb{P}\left(\sup_{0 \leq t \leq T} |I_1^\varepsilon| \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} > \frac{\delta d(\varepsilon) e^{-CT}}{6\sqrt{3}}\right) &\lesssim_\varepsilon \frac{1}{d^2(\varepsilon)} \mathbb{E}\left[\sup_{0 \leq t \leq T} |I_1^\varepsilon(t)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}}\right] \\ &\lesssim_\varepsilon \frac{1}{d^2(\varepsilon)} \sum_{k=0}^{\lfloor \frac{T}{\Delta(\varepsilon)} \rfloor} \left(\mathbb{E}\left|\int_{k\Delta}^{(k+1)\Delta} (a(\mathcal{X}_{k\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - \bar{a}(\mathcal{X}_{k\Delta}^\varepsilon)) \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} ds\right|^2\right) \\ &\lesssim_\varepsilon \frac{\Delta(\varepsilon)}{d^2(\varepsilon)} \alpha\left(\frac{\Delta}{\varepsilon}\right) \rightarrow 0 \text{ as } \varepsilon \rightarrow 0. \end{aligned} \tag{71}$$

The convergence above follows from the choice of the parametrization  $\Delta = \Delta(\varepsilon)$  fixed in (52) and  $\alpha$  constructed in Proposition 3.  $\square$

### 3.4.4. Proof of Theorem 3.2

For any  $\varepsilon > 0$ , fix  $\mathcal{R}(\varepsilon) > 0$  such as in Proposition 5 and recall the definition of  $\tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon$  in (44).

For any  $\delta > 0$ , we have

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \mathbb{P}\left(\sup_{0 \leq t \leq T} |\mathcal{Z}^\varepsilon(t) - \tilde{\mathcal{Z}}^\varepsilon(t)| > \delta\right) &\leq \limsup_{\varepsilon \rightarrow 0} \mathbb{P}\left(\sup_{0 \leq t \leq T} |\mathcal{X}^\varepsilon(t) - \tilde{\mathcal{X}}^\varepsilon(t)| > \delta d(\varepsilon)\right) \\ &\leq \limsup_{\varepsilon \rightarrow 0} \mathbb{P}\left(\sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\mathcal{X}^\varepsilon(t) - \hat{\mathcal{X}}^\varepsilon(t)| > \frac{\delta d(\varepsilon)}{2}\right) \\ &\quad + \limsup_{\varepsilon \rightarrow 0} \mathbb{P}\left(\sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} |\hat{\mathcal{X}}^\varepsilon(t) - \tilde{\mathcal{X}}^\varepsilon(t)| > \frac{\delta d(\varepsilon)}{2}\right) \\ &\quad + \limsup_{\varepsilon \rightarrow 0} \mathbb{P}\left(\tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon \leq T\right) \\ &= 0, \end{aligned} \tag{72}$$

due to Propositions 5, 9 and 10.

$\square$

### 3.5. Conclusions

#### Conclusion-Proof of Theorem 2

We recall the collection of measurable maps  $(\mathcal{G}^\varepsilon)_{\varepsilon > 0}$  introduced in (40) and  $\mathcal{G}^0$  defined by means of the skeleton Equation (32). We note that Proposition 4 reads as Condition 1 of Hypothesis 6 for  $(\mathcal{G}^\varepsilon)_{\varepsilon > 0}$  and  $\mathcal{G}^0$ . Proposition 8 combined with Theorem 4 yield, due to Slutsky’s theorem, that Condition 2 of Hypothesis 6 is verified for  $(\mathcal{G}^\varepsilon)_{\varepsilon > 0}$  and  $\mathcal{G}^0$ . Hence, the result follows from Theorem 3.

$\square$

#### Conclusion from the main result.

The work presented in this article shows how robust the use of the weak convergence approach is in the proof of a moderate deviation principle for a slow–fast system of stochastic equations given by (13). More precisely, the work presented here reduces the usual proof of exponential tightness for the family  $(X^\varepsilon)_{\varepsilon > 0}$  to the proof of a controlled stochastic averaging principle (Theorem 4) that follows from the Markov property of the system and easier weak compactness arguments.

**Author Contributions:** The authors contributed equally to the work shown in this article. Conceptualization, A.d.O.G. and P.C.; methodology, A.d.O.G.; validation, A.d.O.G. and P.C.; formal analysis, A.d.O.G.; investigation, A.d.O.G. and P.C.; writing—original draft preparation, A.d.O.G.; writing—review and editing, A.d.O.G. and P.C.; supervision, P.C.; project administration, P.C.; funding acquisition, P.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors acknowledge and thank the financial support from the FAPESP grant number 2018/06531-1 at the University of Campinas (UNICAMP), SP-Brazil.

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

### Appendix A.1. Auxiliary Results for the Derivation of the Moderate Deviation Principle

#### Appendix A.1.1. Integrability Properties of the Controls

The following lemma is heavily used in the derivation of the moderate deviation principle stated in Theorem 2. We refer the reader to Section 3.1 for notation.

**Lemma A1.** Fix  $M > 0$  and  $\nu \in \mathbb{M}$  a measure satisfying the Hypothesis 1. The following holds.

1. There exists  $\tau > 0$  such that for all  $\varepsilon > 0$  we have

$$\sup_{g \in S_{+, \varepsilon}^M} \int_I \int_{\mathbb{X}} |z|^2 g(s, z) \nu(dz) ds < \tau(d^2(\varepsilon) + |I|), \tag{A1}$$

$$\sup_{g \in S_{+, \varepsilon}^M} \int_I \int_{\mathbb{X}} |z| |g(s, z) - 1| \nu(dz) ds < \tau(d(\varepsilon) + |I|) \tag{A2}$$

and there exists  $\tilde{\tau} > 0$  yielding for all  $\varepsilon, \beta > 0$  some  $c(\beta) \rightarrow 0$  as  $\beta \rightarrow \infty$  and such that

$$\sup_{h \in S_{\varepsilon}^M} \int_I \int_{\mathbb{X}} |z| |h(s, z)| \nu(dz) ds < \tilde{\tau}(\sqrt{|I|} + |I| + d(\varepsilon) + c(\beta)), \tag{A3}$$

for any Borel measurable  $I \subset [0, T]$ .

2. For every  $\varepsilon > 0$ , let  $\psi^\varepsilon \in \mathcal{U}_{\varepsilon}^M$ . We assume that for some  $\beta \in (0, 1)$  the following convergence in law holds,  $\psi^\varepsilon \mathbf{1}_{\{|\psi^\varepsilon| < \frac{\beta}{d(\varepsilon)}\}} \Rightarrow \psi$  in the compact ball  $B_2(\sqrt{M\kappa_2(1)})$ , where  $\kappa_2(1)$  is given by Remark 4. Then, the following convergence in distribution holds, for every  $t \in [0, T]$ ,

$$\int_0^t \int_{\mathbb{X}} |z| \psi^\varepsilon(s, z) \nu(dz) ds \rightarrow \int_0^t \int_{\mathbb{X}} |z|^r \psi(s, z) \nu(dz) ds. \tag{A4}$$

For the proof of the first statement we refer the reader to Lemma 2.1 in [63]. The conclusion of the second statement is proved as in Lemma 4.8 of [46].

### Appendix A.2. Auxiliary Estimates for the Controlled Averaging Principle

#### Appendix A.2.1. Proof of Lemma 3.2

For any  $\varepsilon > 0$  we fix  $\Delta := \Delta(\varepsilon)$  given by (52),  $d(\varepsilon)$  given in (35) and  $\mathcal{R}(\varepsilon) > 0$  such as in Proposition 5. We recall that due to Proposition 6 we have for any  $\varepsilon > 0$  small enough that

$$\sup_{0 < \varepsilon < \varepsilon_0} \mathbb{E} \left[ \sup_{0 \leq t \leq \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon} \|\mathcal{X}_t^\varepsilon\|_\infty \right] < \infty, \tag{A5}$$

where  $\tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon$  is the  $\mathbb{F}$ -stopping time defined by (44).

Let us work on the event  $\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}$ . Fix  $\varepsilon > 0$ ,  $t \in [0, T]$  and  $t_\Delta := \lfloor \frac{t}{\Delta} \rfloor \Delta$ . For every  $\varepsilon > 0$ , let  $K_\varepsilon := \lfloor \frac{T}{\Delta(\varepsilon)} \rfloor \in \mathbb{N}$  and  $N_\varepsilon := \lfloor \frac{\tau}{\Delta(\varepsilon)} \rfloor \in \mathbb{N}$ . For any  $k = 0, \dots, K_\varepsilon - 1$  and  $m = 0, \dots, N_\varepsilon - 1$  we label  $I_k^\varepsilon := [k\Delta, (k+1)\Delta]$  and  $J_m^\varepsilon := [-(m+1)\Delta, -m\Delta]$ .

Given  $t \in [0, T]$  and  $\theta \in [-\tau, 0]$ , let  $k, m \geq 0$  such that  $t \in [k\Delta, (k+1)\Delta]$  and  $\theta \in [-(m+1)\Delta, -m\Delta]$ . It is immediate that

$$t + \theta \in [(k - m - 1)\Delta, (k + 1 - m)\Delta] \quad \text{and} \quad t_\Delta + \theta \in [(k - m - 1)\Delta, (k - m)\Delta].$$

We have to distinguish three possible cases:

- (i)  $m \leq k - 1$ ;
- (ii)  $m \geq k + 1$  and
- (iii)  $m = k$ .

It follows that

$$\begin{aligned} \sup_{0 \leq t \leq T} \|\mathcal{X}_t^\varepsilon - \mathcal{X}_{t_\Delta}^\varepsilon\| &= \sup_{0 \leq t \leq T} \sup_{-\tau \leq \theta \leq 0} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| \\ &= \sup_{t \in \cup_{k=0}^{K_\varepsilon-1} I_k^\varepsilon} \sup_{\theta \in \cup_{m=0}^{N_\varepsilon-1} J_m^\varepsilon} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)|. \end{aligned}$$

Let us fix  $(g_\varepsilon)_{\varepsilon>0}$  such that  $g_\varepsilon \simeq_\varepsilon d(\varepsilon)$  as  $\varepsilon \rightarrow 0$ . It follows that

$$\begin{aligned} \mathbb{P}\left(\sup_{0 \leq t \leq T} |\mathcal{X}_t^\varepsilon - \mathcal{X}_{t_\Delta}^\varepsilon| > g_\varepsilon; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) &\leq N_\varepsilon K_\varepsilon \max_{\substack{k=0, \dots, K_\varepsilon-1 \\ m=0, \dots, N_\varepsilon-1}} \mathbb{P}\left(\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(m+1)\Delta \leq \theta \leq -m\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| > g_\varepsilon; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) \\ &:= K_\varepsilon N_\varepsilon (p_1^\varepsilon + p_2^\varepsilon + p_3^\varepsilon), \end{aligned}$$

where

$$\begin{cases} p_1^\varepsilon := \mathbb{P}\left(\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(m+1)\Delta \leq \theta \leq -m\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| > g_\varepsilon; m \leq k - 1; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) \\ p_2^\varepsilon := \mathbb{P}\left(\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(m+1)\Delta \leq \theta \leq -m\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| > g_\varepsilon; m \geq k + 1; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) \text{ and} \\ p_3^\varepsilon := \mathbb{P}\left(\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(m+1)\Delta \leq \theta \leq -m\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| > g_\varepsilon; m = k; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right). \end{cases}$$

Case (i):  $m \leq k - 1$ .

In this case we have that  $t + \theta > 0$  and  $t_\Delta + \theta > 0$ . Then, we have that

$$\begin{aligned} \mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta) &= \int_{t_\Delta + \theta}^{t + \theta} \left( a(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s)) + \sigma(\mathcal{X}_s^\varepsilon) \tilde{\zeta}_1^\varepsilon(s) + \int_{\mathbb{X}} c(\mathcal{X}_s^\varepsilon, z) (\varphi^\varepsilon(s, z) - 1) \nu(dz) \right) ds \\ &\quad + \sqrt{\varepsilon} \int_{t_\Delta + \theta}^{t + \theta} \sigma(\mathcal{X}_s^\varepsilon) dB_1(s) + \varepsilon \int_{t_\Delta + \theta}^{t + \theta} c(\mathcal{X}_{s-}^\varepsilon, z) \tilde{N}_{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz). \end{aligned}$$

Let us fix the parametrization  $L = L_\varepsilon > 0$ , given by

$$L = L_\varepsilon > 0 := \frac{d^2(\varepsilon)}{|\ln \varepsilon|^q} \text{ for some } q > 2\gamma + 3, \quad \varepsilon > 0. \tag{A6}$$

The Bernstein inequality given in the form of Theorem 3.3. in [54] implies for every  $\varepsilon > 0$  that

$$p_1^\varepsilon \lesssim_\varepsilon e^{-\frac{g_\varepsilon^2}{L_\varepsilon}} + \mathbb{P}\left([\mathcal{X}^\varepsilon - \mathcal{X}^\varepsilon(t_\Delta + \theta)]_{(k+1)\Delta - m\Delta} > L_\varepsilon; m \leq k - 1; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right).$$

Due to (A5) it follows for any  $\varepsilon > 0$  on the event  $\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}$  that

$$\begin{aligned} [\mathcal{X}^\varepsilon - \mathcal{X}^\varepsilon(t_\Delta + \theta)]_{(k+1)\Delta - m\Delta} &\lesssim_\varepsilon \varepsilon \Delta (1 + \mathcal{R}^2(\varepsilon)) + \varepsilon^2 \int_{t_\Delta - (m+1)\Delta}^{(k+1)\Delta - m\Delta} |z|^2 N_{\frac{1}{\varepsilon}}(ds, dz) \\ &:= \varepsilon (1 + \mathcal{R}^2(\varepsilon)) \Delta + \varepsilon^2 I_{(k+1)\Delta - m\Delta}^\varepsilon. \end{aligned}$$

Due to the choice of  $L_\varepsilon$  in (A6) and  $\Delta(\varepsilon)$  in (52) let  $\varepsilon_0 > 0$  sufficiently small such that for any  $\varepsilon < \varepsilon_0$  we have  $\varepsilon (1 + \mathcal{R}^2(\varepsilon)) \varepsilon^\gamma |\ln \varepsilon|^{p-q} < \frac{1}{2}$ . Then, it follows that

$$\varepsilon(1 + \mathcal{R}^2(\varepsilon))\Delta = \varepsilon(1 + \mathcal{R}^2(\varepsilon))\varepsilon^\gamma |\ln \varepsilon|^{p-q} \frac{d^2(\varepsilon)}{|\ln \varepsilon|^q} < \frac{L_\varepsilon}{2} \tag{A7}$$

for every  $\varepsilon < \varepsilon_0$ .

The estimate (A1) in Lemma A1 (Section A.1 of the Appendix A) implies for any  $\varepsilon > 0$  small enough such that (A7) holds that

$$\begin{aligned} \mathbb{P}\left([\mathcal{X}^\varepsilon - \mathcal{X}^\varepsilon(t_\Delta + \theta)]_{(k+1)\Delta - m\Delta} > L_\varepsilon; m \leq k - 1; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) &\leq \bar{\mathbb{P}}\left(\varepsilon^2 I_{(k+1)\Delta - m\Delta}^\varepsilon > \frac{L_\varepsilon}{2}\right) \\ &\lesssim_\varepsilon \frac{\varepsilon^2}{2L_\varepsilon} \bar{\mathbb{E}}\left[I_{(k+1)\Delta - m\Delta}^\varepsilon\right] \\ &\lesssim_\varepsilon \frac{\varepsilon}{L_\varepsilon} \int_{t_\Delta - (m+1)\Delta}^{(k+1)\Delta - m\Delta} |z|^2 \varphi^\varepsilon(s, z) \nu(dz) ds \\ &\lesssim_\varepsilon \frac{\varepsilon}{L_\varepsilon} (d^2(\varepsilon) + \Delta). \end{aligned}$$

Due to (53), (A6) and  $d(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$  we conclude for every  $\varepsilon > 0$  small enough that

$$p_1^\varepsilon \lesssim_\varepsilon e^{-\frac{\varepsilon^2}{L_\varepsilon}} + \frac{\varepsilon}{L_\varepsilon} (d^2(\varepsilon) + \Delta(\varepsilon)) \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0. \tag{A8}$$

The case  $m \geq k + 1$ .

In this case we have that  $t + \theta < 0$  and  $t_\Delta + \theta < 0$ . Since the initial delay  $\chi$  is Lipschitz continuous (cf. (17)) it follows that

$$\begin{aligned} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| &= |\chi(t + \theta) - \chi(t_\Delta + \theta)| \\ &\leq \lambda |t - t_\Delta|. \end{aligned}$$

Then, for any  $\varepsilon > 0$  we have

$$\begin{aligned} p_2^\varepsilon &= \bar{\mathbb{P}}\left(\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(m+1)\Delta \leq \theta \leq -m\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)|^4 > (g_\varepsilon)^4; m \geq k + 1; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) \\ &\lesssim_\varepsilon \frac{1}{(g_\varepsilon)^4} \bar{\mathbb{E}}\left[\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(m+1)\Delta \leq \theta \leq -m\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)|^4\right] \lesssim_\varepsilon \left(\frac{\Delta(\varepsilon)}{g_\varepsilon}\right)^4 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0, \end{aligned} \tag{A9}$$

due to the definition of  $\Delta(\varepsilon)$  in (52) and  $g_\varepsilon \simeq_\varepsilon d(\varepsilon)$  as  $\varepsilon \rightarrow 0$ .

The case  $m = k$ .

In this case we have  $t + \theta \in [-\Delta, \Delta]$  and  $t_\Delta + \theta \in [-\Delta, 0]$ . It is immediate that

$$|\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| = |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| \mathbf{1}_{\{t+\theta > 0\}} + |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| \mathbf{1}_{\{t+\theta < 0\}}.$$

Due to the two previous cases already analyzed we have, for any  $\varepsilon > 0$  small enough, that

$$\begin{aligned} p_3^\varepsilon &\leq \bar{\mathbb{P}}\left(\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(k+1)\Delta \leq \theta \leq -k\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| > g_\varepsilon; \mathbf{1}_{\{t+\theta > 0\}}; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) \\ &\quad + \bar{\mathbb{P}}\left(\sup_{\substack{k\Delta \leq t \leq (k+1)\Delta \\ -(k+1)\Delta \leq \theta \leq -k\Delta}} |\mathcal{X}^\varepsilon(t + \theta) - \mathcal{X}^\varepsilon(t_\Delta + \theta)| > g_\varepsilon; \mathbf{1}_{\{t+\theta < 0\}}; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) \\ &\lesssim_\varepsilon e^{-\frac{\varepsilon^2}{L_\varepsilon}} + \frac{\varepsilon}{L_\varepsilon} (d^2(\varepsilon) + \Delta) + \left(\frac{\Delta}{g_\varepsilon}\right)^4 \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0. \end{aligned} \tag{A10}$$

Combining (A8)–(A10) it follows, for  $\Delta(\varepsilon)$ ,  $L_\varepsilon$  given by (52) and respectively (A6) and any  $\varepsilon > 0$  small enough, that

$$\begin{aligned} & \bar{\mathbb{P}}\left(\sup_{0 \leq t \leq T} |\mathcal{X}_t^\varepsilon - \mathcal{X}_{t_\Delta}^\varepsilon| > g\varepsilon; T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\right) \\ & \lesssim_\varepsilon N_\varepsilon K_\varepsilon \left( e^{-\frac{g\varepsilon}{L_\varepsilon}} + \frac{\varepsilon}{L_\varepsilon} (d^2(\varepsilon) + \Delta) + \left(\frac{\Delta}{g\varepsilon}\right)^4 \right) \\ & \lesssim_\varepsilon \frac{1}{(\Delta(\varepsilon))^2} \left( e^{-|\ln \varepsilon|^q} + \varepsilon |\ln \varepsilon|^q + b(\varepsilon) |\ln \varepsilon|^q \Delta + \left(\frac{\Delta(\varepsilon)}{d(\varepsilon)}\right)^4 \right) \\ & \lesssim_\varepsilon \frac{\varepsilon^q}{\varepsilon^{2\gamma} a^4(\varepsilon) |\ln \varepsilon|^{2p}} + \frac{\varepsilon}{\varepsilon^{2\gamma} a^4(\varepsilon) |\ln \varepsilon|^{2p-q}} + \frac{\varepsilon}{\varepsilon^\gamma a^4(\varepsilon) |\ln \varepsilon|^{p-q}} + d^2(\varepsilon) \varepsilon^{2\gamma} |\ln \varepsilon|^{2p} \\ & \lesssim_\varepsilon b^2(\varepsilon) \frac{\varepsilon}{|\ln \varepsilon|^{2p}} + \frac{\varepsilon^{2\theta-1-2\gamma}}{|\ln \varepsilon|^{2p-q}} + \frac{\varepsilon^{2\theta-1-\gamma}}{|\ln \varepsilon|^{p-q}} + \varepsilon^\gamma \varepsilon^{1-\theta} |\ln \varepsilon|^{2p} =: \Xi(\varepsilon). \end{aligned}$$

Since  $\gamma \in (0, \theta - \frac{1}{2})$ ,  $d(\varepsilon) = \varepsilon^{\frac{1-\theta}{2}}$ ,  $\theta \in (\frac{1}{2}, 1)$ ,  $b(\varepsilon) = \frac{\varepsilon}{d^2(\varepsilon)}$  we conclude that  $\Xi(\varepsilon) \rightarrow 0$  as  $\varepsilon \rightarrow 0$ . This finishes the proof. □

Appendix A.2.2. Proof of Lemma 3

Ito’s formula yields for any  $t \in [t_\Delta, t_\Delta + 1]$  and  $\bar{\mathbb{P}}$ -a.s.

$$\begin{aligned} & |\hat{\mathcal{Y}}^\varepsilon(t) - \mathcal{Y}^\varepsilon(t)|^2 \\ & = \frac{2}{\varepsilon} \int_{t_\Delta}^t \langle f(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - f(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s)), \hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s) \rangle ds \\ & + \frac{2}{\varepsilon} \int_{t_\Delta}^t \langle (g(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - g(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s))) \zeta_2^\varepsilon(s), \hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s) \rangle ds \\ & + \frac{2}{\varepsilon} \int_{t_\Delta}^t \int_{\mathbb{X}} \langle h(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s), z) - h(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s), z), \hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s) \rangle (\varphi^\varepsilon(s, z) - 1) \nu(dz) ds \\ & + \frac{2}{\sqrt{\varepsilon}} \int_{t_\Delta}^t \langle g(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - g(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s)), (\hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)) dB^2(s) \rangle \\ & + \frac{1}{\varepsilon} \int_{t_\Delta}^t |g(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - g(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s))|^2 ds \\ & + \int_{t_\Delta}^t \int_{\mathbb{X}} 2 \langle h(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}_{s-}^\varepsilon, z) - h(\mathcal{X}_{s-}^\varepsilon, \mathcal{Y}_{s-}^\varepsilon, z), \mathcal{Y}_{s-}^\varepsilon - \hat{\mathcal{Y}}_{s-}^\varepsilon \rangle \tilde{N}^{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz) \\ & + \int_{t_\Delta}^t \int_{\mathbb{X}} |h(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}_{s-}^\varepsilon, z) - h(\mathcal{X}_{s-}^\varepsilon, \mathcal{Y}_{s-}^\varepsilon, z), \hat{\mathcal{Y}}_{s-}^\varepsilon - \mathcal{Y}_{s-}^\varepsilon|^2 \tilde{N}^{\frac{1}{\varepsilon} \varphi^\varepsilon}(ds, dz) \\ & + \frac{1}{\varepsilon} \int_{t_\Delta}^t \int_{\mathbb{X}} |h(\mathcal{X}_{t_\Delta}^\varepsilon, \hat{\mathcal{Y}}^\varepsilon(s)) - h(\mathcal{X}_s^\varepsilon, \mathcal{Y}^\varepsilon(s))|^2 \varphi^\varepsilon(s, z) \nu(dz) ds \\ & = \sum_{i=1}^8 I_i^\varepsilon(t). \end{aligned}$$

Using (21) in Hypothesis 4 yields for any  $\varepsilon > 0$  and  $t \in [t_\Delta, t_\Delta + 1]$

$$I_1^\varepsilon(t) \leq -\frac{2\beta_1}{\varepsilon} \int_{t_\Delta}^t |\hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)|^2 ds + \frac{2\beta_2 \Delta}{\varepsilon} \|\mathcal{X}_{t_\Delta}^\varepsilon - \mathcal{X}_t^\varepsilon\|_\infty^2. \tag{A11}$$

The boundedness of  $g$  given by (18) in Hypothesis 4, the fact that  $\zeta^\varepsilon \in \tilde{\mathcal{U}}_\varepsilon^M$  and Cauchy-Schwartz’s inequality imply for any  $\varepsilon > 0$  and  $t \in [t_\Delta, t_\Delta + 1]$  that

$$I_2^\varepsilon(t) \leq \frac{4\Lambda \sqrt{Md}(\varepsilon)}{\varepsilon} \left( 1 + \int_{t_\Delta}^t |\hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)|^2 ds \right). \tag{A12}$$

Analogously, (21) in Hypothesis 4 together with (A1), (A2) given in Lemma A1 of Section A.1 of the Appendix A combined with the numeric fact  $x\lambda \leq x^2 + \frac{1}{\lambda}$ ,  $x, \lambda \geq 0$  yield some  $C_1 = C_1(M, \Lambda) > 0$  such that for any  $\varepsilon > 0$  and  $t \in [t_\Delta, t_\Delta + 1]$  we have

$$I_3^\varepsilon(t) \leq \frac{C_1}{\varepsilon\lambda}(d(\varepsilon) + \Delta) + \frac{C_1\lambda}{\varepsilon} \int_{t_\Delta}^t |\mathcal{Y}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)|^2 \Theta^\varepsilon(s) ds \quad \text{and}$$

$$I_5^\varepsilon(t) + I_8^\varepsilon(t) \leq \frac{C}{\varepsilon}(d^2(\varepsilon) + \Delta) \tag{A13}$$

where  $\Theta^\varepsilon(t) := \int_0^t |z| |\varphi^\varepsilon(s, z) - 1| \nu(dz)$ ,  $t \in [0, T]$ .

The estimates (A11)–(A13) imply for  $t \in [t_\Delta, t_\Delta + 1]$ ,  $\varepsilon > 0$  and  $\lambda = \lambda(\varepsilon) > 0$  fixed below the following  $\mathbb{P}$ -a.s. bound on the event  $\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}$ :

$$|\hat{\mathcal{Y}}^\varepsilon(t) - \mathcal{Y}^\varepsilon(t)|^2 \lesssim_\varepsilon \int_{t_\Delta}^t \frac{1}{\varepsilon} \left( -1 + d(\varepsilon) + \lambda(\varepsilon)\Theta^\varepsilon(s) \right) |\hat{\mathcal{Y}}^\varepsilon(s) - \mathcal{Y}^\varepsilon(s)|^2 ds + C_2(\varepsilon) + I_4^\varepsilon(t) + I_6^\varepsilon(t) + I_7^\varepsilon(t),$$

where

$$C_2(\varepsilon) \simeq_\varepsilon \frac{1}{\varepsilon} \left( \Delta(\varepsilon)\mathcal{R}^2(\varepsilon) + d(\varepsilon) + \frac{d(\varepsilon)}{\lambda(\varepsilon)}(1 + \Delta(\varepsilon)) + d^2(\varepsilon) + \Delta(\varepsilon) \right) \quad \text{as } \varepsilon \rightarrow 0. \tag{A14}$$

Due to Gronwall’s lemma, the estimate (A2) in Lemma A1 (Section A.1 of the Appendix A) and the fact that  $\mathbb{E}[I_4^\varepsilon] = \mathbb{E}[I_6^\varepsilon] = \mathbb{E}[I_7^\varepsilon] = 0$  it follows, for any  $\varepsilon > 0$ ,  $\lambda = \lambda(\varepsilon) = \varepsilon$  and  $t \in [t_\Delta, t_\Delta + 1]$  that

$$\mathbb{E} \left[ |\hat{\mathcal{Y}}^\varepsilon(t) - \mathcal{Y}^\varepsilon(t)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} \right] \lesssim_\varepsilon C_2(\varepsilon) \exp \left( \frac{-\Delta(\varepsilon)}{\varepsilon} (1 - d(\varepsilon)) + d(\varepsilon) + \Delta(\varepsilon) \right).$$

Let  $\varepsilon_0 > 0$  small enough such that  $1 - d(\varepsilon) > \frac{1}{2}$  and  $d(\varepsilon) + \Delta(\varepsilon) < 1$  for any  $\varepsilon < \varepsilon_0$ . Therefore we have for any  $\varepsilon > 0$  small enough and  $t \in [0, T]$  that

$$\mathbb{E} \left[ |\hat{\mathcal{Y}}^\varepsilon(t) - \mathcal{Y}^\varepsilon(t)|^2 \mathbf{1}_{\{T < \tilde{\tau}_{\mathcal{R}(\varepsilon)}^\varepsilon\}} \right] \lesssim_\varepsilon \frac{C_2(\varepsilon)}{\Delta(\varepsilon)} e^{-\frac{\Delta(\varepsilon)}{2\varepsilon} + 1} \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0, \tag{A15}$$

due to the choice of  $\Delta(\varepsilon)$  fixed in (52). □

### References

1. Mao, X. *Stochastic Differential Equations and Applications*, 2nd ed.; UK Horwood Publishing Limited: Chichester, UK, 2008.
2. Baños, D.R.; Cordoni, F.; Di Nunno, G.; Di Persio, L.; Rose, E.E. Stochastic systems with memory and jumps. *J. Diff. Eq.* **2019**, *226*, 5772–5820. [\[CrossRef\]](#)
3. Weinan, E.; Liu, S.; Vanden-Eijnden, E. Analysis of multiscale methods of stochastic differential equations. *Comm. Pure Appl. Math.* **2005**, *LVIII*, 1544–1585. [\[CrossRef\]](#)
4. Pavliotis, G.; Stuart, A. *Multiscale Methods: Averaging and Homogenization*; Texts in Applied Mathematics; Springer: Berlin/Heidelberg, Germany, 2008; Volume 53.
5. Fouque, J.-P.; Papanicolaou, G.; Sircar, K.R. *Derivatives in Financial Markets with Stochastic Volatility*; Cambridge University Press: Cambridge, UK, 2000.
6. Fouque, J.-P.; Papanicolaou, G.; Sircar, K.R.; Solna, K. Multiscale stochastic volatility asymptotics. *Multiscale Model. Simul.* **2003**, *2*, 22–42. [\[CrossRef\]](#)
7. Kifer, Y. Averaging and climate models. In *Stochastic Climate Models*; Imkeller, P., Storch, J.-S.G., Eds.; Progress in Probability; Birkhäuser Verlag: Basel, Switzerland, 2001; Volume 49.
8. Debussche, A.; Högele, M.; Imkeller, P. *The Dynamics of Nonlinear Reaction-Diffusion Equations with Small Lévy Noise*; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 2013; Volume 2085.
9. Arnold, L. Hasselmann’s program revisited: The analysis of stochasticity in deterministic climate models In *Stochastic Climate Models*; Imkeller, P., Storch, J.-S.G., Eds.; Progress in Probability; Birkhäuser Verlag: Basel, Switzerland, 2001; Volume 49.
10. Dijkstra, H.A. *Nonlinear Climate Dynamics*; Cambridge University Press: New York, NY, USA, 2013.
11. Ditlevsen, P.D. Observation of a stable noise induced millennial climate changes from an ice-core record. *Geophys. Res. Lett.* **1999**, *26*, 1441–1444. [\[CrossRef\]](#)
12. Gairing, J.; Högele, M.; Kosenkova, T.; Kulik, A. *On the Calibration of Lévy Driven Time Series with Coupling Distances with an Application in Paleoclimate*; Springer: Berlin/Heidelberg, Germany, 2016.

13. Hein, C.; Imkeller, P.; Pavlyukevich, I. Limit theorems for p-variations of solutions of SDEs driven by additive Stable Levy noise and model selection for paleo-climatic data. *Interdiscip. Math. Sci.* **2009**, *8*, 137–150.
14. Dijkstra, H.A. Derivation of delay climate models using the Mori-Zwanzig formalism. *Proc. R. Soc. A* **2019**, *475*, 2227.
15. Khasminskii, R.Z. On the principle of averaging the Ito's stochastic differential equations. *Kybernetika* **1968**, *4*, 260–279.
16. Freidlin, M.I.; Wentzell, A.D. *Random Perturbations of Dynamical Systems*, 2nd ed.; Grundlehren der Mathematischen Wissenschaften 260; Springer: New York, NY, USA, 1998.
17. Freidlin, M. The Averaging Principle and Theorems on Large Deviations. *Russian Math. Surveys* **1978**, *33*, 117–176. [[CrossRef](#)]
18. Veretennikov, A.Y. On the Averaging Principle for Systems of Stochastic Differential Equations. *Math. USSR-Sbornik* **1991**, *69*, 271–284. [[CrossRef](#)]
19. Cerrai, S. A Khasminskii type of Averaging Principle for Stochastic Reaction Diffusion Equations. *Ann. Appl. Prob.* **2009**, *19*, 899–948. [[CrossRef](#)]
20. Cerrai, S.; Freidlin, M. Averaging principle for a class of stochastic reaction-diffusion equations. *Probab. Theory Relat. Fields* **2009**, *144*, 137–177. [[CrossRef](#)]
21. Cerrai, S. Normal deviations from the averaged motion for some reaction-diffusion equations with fast oscillating perturbation. *J. Math. Pures Appl.* **2009**, *91*, 614–647. [[CrossRef](#)]
22. Givon, D. Strong Convergence Rate for Two-Time-Scale Jump Diffusion Stochastic Differential Systems. *Multiscale Model Simul.* **2007**, *6*, 577–594. [[CrossRef](#)]
23. Liu, D. Strong convergence rate of principle of averaging for jump diffusion processes. *Front. Math. China* **2012**, *7*, 305–320. [[CrossRef](#)]
24. Xu, J.; Miao, Y.; Liu, J. Strong averaging principle for slow-fast SPDEs with Poisson random measures. *Discret. Contin. Dyn. Syst. Ser. B* **2015**, *20*, 2233–2256. [[CrossRef](#)]
25. Xu, J.  $L^p$ -strong convergence of the averaging principle for slow-fast SPDEs with jumps. *J. Math. Analysis Appl.* **2017**, *445*, 342–373. [[CrossRef](#)]
26. Bao, J.; Song, Q.; Yin, G.; Yuan, C. Ergodicity and strong limit results for two-time-scale functional stochastic differential equations. *Stoch. Anal. Appl.* **2017**, *35*, 1030–1060. [[CrossRef](#)]
27. Mao, W.; You, S.; Wu, X.; Mao, X. On the averaging principle for stochastic delay differential equations with jumps. *Adv. Differ. Eq.* **2015**, *2015*, 70. [[CrossRef](#)]
28. Budhiraja, A.; Dupuis, P.; Ganguly, A. Large deviations for small noise diffusions in a fast Markovian environment. *Electron. J. Probab.* **2018**, *23*, 1–33. [[CrossRef](#)]
29. Duan, J.; Wang, W.; Roberts, A.J. Large deviations and approximations for slow-fast stochastic reaction-diffusion equations. *J. Diff. Eqs.* **2012**, *253*, 3501–3522.
30. Kumar, R.; Popovic, L. Large deviations for multi-scale jump diffusion processes. *Stoch. Proc. Their Appl.* **2017**, *127*, 1297–1320. [[CrossRef](#)]
31. Veretennikov, A.Y.; Yu, A. On large deviations for SDEs with small diffusion and averaging. *Stoch. Process. Their Appl.* **2000**, *89*, 69–79. [[CrossRef](#)]
32. Feng, J.; Fouque, J.-P.; Kumar, R. Small-Time Asymptotics for Fast Mean-Reverting Stochastic Volatility Models. *Ann. Appl. Prob.* **2012**, *22*, 1541–1575. [[CrossRef](#)]
33. Guillin, A. Moderate deviations of inhomogeneous functionals of Markov processes and application to averaging. *Stoch. Proc. Appl.* **2001**, *92*, 287–313. [[CrossRef](#)]
34. Guillin, A. Averaging principle of SDE with small diffusion: Moderate deviations. *Ann. Prob.* **2003**, *31*, 413–443. [[CrossRef](#)]
35. Friz, P.; Geshold, S.; Pinter, A. Option Pricing in the Moderate Deviations Regime. *Math. Fin.* **2018**, *28*, 962–988. [[CrossRef](#)]
36. Jacquier, A.; Spiliopoulos, J.K. Pathwise Moderate Deviations in Option Pricing. *Math. Financ.* **2019**, *30*, 1–38. [[CrossRef](#)]
37. Dejjellout, H.; Guillin, A.; Wu, L. Large and Moderate Deviations for Estimators of Quadratic Variational Processes of Diffusions. *Stat. Inference Stoch. Proc.* **2000**, *2*, 195–225. [[CrossRef](#)]
38. Keblaner, F.C.; Lipster, R. Moderate deviations for randomly perturbed dynamical systems. *Stoch. Proc. Their Appl.* **1999**, *180*, 157–176.
39. Fleming, W.H. A stochastic control approach to some large deviations problems. In *Recent Mathematical Methods in Dynamic Programming*; Dolcetta, C., Fleming, W.H., Zoletti, T., Eds.; Springer Lecture notes in Math.; Springer: Berlin/Heidelberg, Germany, 1985; Volume 1119, pp. 52–66.
40. Fleming, W.H. Stochastic control and large deviations. In *Future Tendencies in Computer Science, Control and Applied Mathematics*; Bensoussan, A., Verjus, J.P., Eds.; Springer: Berlin/Heidelberg, Germany, 1992; Volume 653.
41. Dupuis, P.; Ellis, R.S. *A Weak Convergence Approach to the Theory of Large Deviations*; Wiley Series in Probability and Statistics; Wiley and Sons: New York, NY, USA, 1997.
42. Budhiraja, A.; Dupuis, P. A variational representation for positive functionals of infinite Brownian motion. *Probab. Math. Stat.* **2000**, *20*, 39–61.
43. Budhiraja, A.; Dupuis, P.; Maroulas, V. Variational representations for continuous time processes. *Ann. de l'Inst. Henr. Poinc. B Probab. Stat.* **2011**, *47*, 725–747. [[CrossRef](#)]
44. Budhiraja, A.; Chen, J.; Dupuis, P. Large deviations for stochastic partial differential equations driven by a Poisson random measure. *Stochastic Process. Appl.* **2013**, *123*, 523–560. [[CrossRef](#)]
45. Budhiraja, A.; Dupuis, P. *Analysis and Approximation of Rare Events. Representations and Weak Convergence Methods*; Series Prob. Theory and Stoch. Modelling; Springer: Berlin/Heidelberg, Germany, 2019; Volume 94.

46. Budhiraja, A.; Dupuis, P.; Ganguly, A. Moderate deviation principles for stochastic differential equations with jumps. *Ann. Probab.* **2016**, *44*, 1723–1775. [[CrossRef](#)]
47. Budhiraja, A.; Wu, R. Moderate Deviation Principles for Weakly Interacting Particle Systems. *Probab. Theory Relat. Fields* **2017**, *168*, 721–771. [[CrossRef](#)]
48. Zheng, W.; Zhai, J.; Zhang, T. Moderate deviations for stochastic models of two-dimensional second-grade fluids driven by Lévy noises. *Comm. Math. Stat.* **2018**, *6*, 583–612. [[CrossRef](#)]
49. Azencott RGeiger, B.; Ott, W. Large deviations for Gaussian diffusions with delay. *J. Stat. Phys* **2018**, *170*, 254–285. [[CrossRef](#)]
50. Lipshutz, D. Exit time asymptotics for small noise stochastic delay differential equations. *Discret. Contin. Dyn. Syst. A* **2018**, *38*, 3099–3138. [[CrossRef](#)]
51. Ma, X.; Xi, F. Moderate deviations for neutral stochastic differential delay equations. *Stat. Prob. Lett.* **2016**, *126*, 97–107. [[CrossRef](#)]
52. Suo, Y.; Tao, J.; Zhang, W. Moderate deviations and central limit theorem for stochastic differential delay equations with polynomial growth. *Front. Math. China* **2018**, *13*, 913–933. [[CrossRef](#)]
53. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; Wiley-Interscience: Hoboken, NJ, USA, 1999.
54. Dzhaparidze, K.; van Zantem, J.H. On Bernstein-type inequalities for martingales. *Stoch. Proc. Appl.* **2001**, *93*, 109–117. [[CrossRef](#)]
55. Ikeda, N.; Watanabe, S. *Stochastic Differential Equations and Diffusion Processes*; North-Holland Publishing Co.: Amsterdam, The Netherlands, 1981.
56. Stroock, D. *An Introduction to the Theory of Large Deviations*; Springer: Berlin/Heidelberg, Germany, 1984.
57. Nishimori, Y. Large deviations for symmetric stable processes with Feynman-Kac functionals and its applications to pinned polymers. *Tohoku Math. J.* **2013**, *65*, 467–494. [[CrossRef](#)]
58. Protter, P.E. *Stochastic Integration and Differential Equations*; Stochastic Modelling and Applied Probability; Springer: Berlin/Heidelberg, Germany, 2004; p. 21.
59. DaPrato, G. *An Introduction to Infinite Dimensional Analysis*; Springer: Berlin/Heidelberg, Germany, 2006.
60. Quiao, H. Exponential Ergodicity for SDEs with Jumps and non-Lipschitz coefficients. *J. Theor. Prob.* **2014**, *27*, 137–152. [[CrossRef](#)]
61. Xu, J.; Liu, J.; Miao, Y. Strong Averaging Principle for Two-Time Scale SDEs with nonLipschitz coefficients. *J. Math. Anal. Appl.* **2018**, *468*, 116–140. [[CrossRef](#)]
62. Rosiński, J. Tempering stable processes. *Stoch. Proc. Appl.* **2007**, *177*, 677–707. [[CrossRef](#)]
63. Oliveira Gomes, A.D.; Högele, M.A. The Kramers problem driven by small accelerated Lévy noise with exponentially light jumps. *Stochastics Dyn.* **2021**, *32*, 2150019. [[CrossRef](#)]
64. Jacod, J.; Shiryaev, A.N. *Limit Theorems for Stochastic Processes*; Springer: Berlin/Heidelberg, Germany, 1987.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.