Large Deviations and Metastability Analysis for Heavy-Tailed Dynamical Systems

Xingyu Wang and Chang-Han Rhee

Industrial Engineering and Management Sciences, Northwestern University
Evanston, IL, 60613, USA

September 16, 2024

Abstract

This paper introduces a novel framework that connects large deviations and metastability analysis in heavy-tailed stochastic dynamical systems. Employing this framework in the context of stochastic difference equations $X^{\eta}_{j+1}(x) = X^{\eta}_j(x) + \eta a\big(X^{\eta}_j(x)\big) + \eta\sigma\big(X^{\eta}_j(x)\big)Z_{j+1}$ and its variation with truncated dynamics, $X^{\eta|b}_{j+1}(x) = X^{\eta|b}_j(x) + \varphi_b\big(\eta a\big(X^{\eta|b}_j(x)\big) + \eta\sigma\big(X^{\eta|b}_j(x)\big)Z_{j+1}\big)$, where $\phi_b(x) = (x/|x|)\max\{|x|, b\}$, we first establish locally uniform sample path large deviations and then translate such asymptotics into a precise characterization of the joint distribution of the first exit time and exit location. As a result, we obtain the heavy-tailed counterparts of the classical Freidlin-Wentzell and Eyring-Kramers theorems. Our large deviations asymptotics are sharp enough to identify how rare events arise in heavy-tailed dynamical systems and characterize *the catastrophe principle*. Moreover, it also unveils the discrete hierarchy of phase transitions in the asymptotics of the first exit times and locations under truncated heavy-tailed noises. Our results in this paper open up the possibility of systematic analysis of the global dynamics of heavy-tailed stochastic processes. In the appendix, we also present the corresponding results for the Lévy driven SDEs.

Contents

1 Introduction

The analysis of large deviations and metastability in stochastic dynamical systems has a rich history in probability theory and continues to be a vibrant field of research. For instance, the classical Freidlin-Wentzell theorem (see [62]) analyzed sample-path large deviations of Itô diffusions, and over the past few decades, the theory has seen numerous extensions, including the discrete-time version of Freidlin–Wentzell theorem (see, e.g., [45, 38]), large deviations for finite dimensional processes under relaxed assumptions (see, e.g., [20, 24, 23, 1, 25]), Freidlin–Wentzell-type bounds for infinite dimensional processes (see, e.g., [12, 13, 37]), and large deviations for stochastic partial differential equations (see, e.g., [61, 15, 57, 44]), to name a few. On the other hand, the exponential scaling and the pre-exponents in the asymptotics of first exit times under Brownian perturbations were characterized in the Eyring–Kramers law (see [27, 40]). There have been various theoretical advancements since these seminal works, such as the asymptotic characterization of the most likely exit path and the exit times for Brownian particles under more sophisticated gradient fields (see [43]), results for discrete-time processes (see, e.g., [39, 14]), and applications in queueing systems (see, e.g., [60]). For an alternative perspective on metastability based on potential theory, which diverges from the Freidlin-Wentzell theory, we refer the readers to [9].

While such developments provide powerful means to understand rare events and metastability of classical light-tailed systems, they often fail to provide useful bounds when it comes to the heavy-tailed systems. As shown in [32, 34, 35, 33], when the stochastic processes are driven by heavy-tailed noises, the exit events are typically caused by large perturbations of a small number of components. This is in sharp contrast to the light-tailed counterparts where rare events typically arise via smooth tilting of the nominal dynamics. Due to such a stark difference in the mechanism through which rare events arise, heavy-tailed systems exhibit a fundamentally different large deviations and metastability behaviors and call for a different set of technical tools for successful analysis.

In this paper, we build a general framework for asymptotic analysis of heavy-tailed dynamical systems by developing a set of machinery that uncovers the interconnection between the large deviations and the metastability of stochastic processes. Building upon this framework, we characterize the sample-path large deviations and metastability of heavy-tailed stochastic difference equations (and stochastic differential equations in the appendix), thus offering the heavy-tailed counterparts of Freidlin–Wentzell and Eyring–Kramers theory. More precisely, the main contributions of this article can be summarized as follows:

- **Heavy-tailed Large Deviations:** We establish sample-path large deviations for heavy-tailed dynamical systems. We propose a new heavy-tailed large deviations formulation that is locally uniform w.r.t. the initial values. We accomplish this by formulating a uniform version of $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence [42, 56]. Our large deviations characterize the *catastrophe principle* (also known as the *principle of big jumps*), which reveals a discrete hierarchy governing the causes and probabilities of a wide variety of rare events associated with heavy-tailed stochastic difference equations. Moreover, this new formulation of the heavy-tailed large deviations paves the way to the analysis of local stability and global dynamics.

- **Metastability Analysis:** We establish a scaling limit of the exit-time and exit-location for stochastic difference equations. We accomplish this by developing a machinery for local stability analysis of general (heavy-tailed) Markov processes. Central to the development is the concept of asymptotic atoms, where the process recurrently enters and asymptotically regenerates. Leveraging the locally uniform version of sample-path large deviations over such asymptotic atoms, we obtain sharp asymptotics of the joint distribution of the (scaled) exit-times and exit-locations for heavy-tailed processes. Notably, this complements the investigation of the exit times under the truncated dynamics, which was first analyzed in [35] in the context of Weibull tails.

In a companion paper [64], we show that the above framework is powerful enough to identify a scaling limit and characterize the global behavior of the heavy-tailed dynamical systems over a multi-well potential at the process level. In particular, the scaling limit is a Markov jump process whose state space consists of the local minima of the potential; under the truncated dynamics, the state space consists of *only the widest minima*. This demonstrates that the truncation changes the global dynamics of the dynamical systems qualitatively compared to the untruncated counterparts; see [21, 29]. These findings systematically characterize a curious phenomena that the truncated heavy-tailed processes avoid narrow local minima altogether in the limit. As a result, it can be shown that the fraction of time such processes spend in the narrow attraction field converges to zero as the step-size tends to zero. Precise characterization of such phenomena is of fundamental importance in understanding and further improving the curious effectiveness of the stochastic gradient descent (SGD) algorithms in training deep neural networks.

In this paper, we focus on the class of heavy-tailed phenomena captured by the notion of regular variation. To be specific, let $(\boldsymbol{Z}_i)_{i\geq 1}$ be a sequence of iid random vectors in $\mathbb{R}^m$ such that $\mathbf{E}\boldsymbol{Z}_1 = \mathbf{0}$ and $\mathbf{P}(\|\boldsymbol{Z}_i\| > x)$ is regularly varying with index $-\alpha$ as $x \to \infty$ for some $\alpha > 1$. That is, there exists some slowly varying function ϕ such that $\mathbf{P}(\|\boldsymbol{Z}_1\| > x) = \phi(x)x^{-\alpha}$. For any $\eta > 0$ and $\boldsymbol{x} \in \mathbb{R}^m$, let $\big(\boldsymbol{X}^\eta_j(\boldsymbol{x})\big)_{j\geq 0}$ be the solution of the following stochastic difference equation

$$\boldsymbol{X}^\eta_0(\boldsymbol{x}) = \boldsymbol{x}; \qquad \boldsymbol{X}^\eta_{j+1}(\boldsymbol{x}) = \boldsymbol{X}^\eta_j(\boldsymbol{x}) + \eta\boldsymbol{a}\big(\boldsymbol{X}^\eta_j(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^\eta_j(\boldsymbol{x})\big)\boldsymbol{Z}_{j+1} \quad \forall j \geq 0. \tag{1.1}$$

Throughout this paper, we adopt the convention that the subscript denotes the time, and the superscript η denotes the scaling parameter that tends to zero. Furthermore, we also consider a truncated variation of $\boldsymbol{X}^\eta_{j+1}(\boldsymbol{x})$ which is arguably more relevant when $\boldsymbol{Z}_i$'s are heavy-tailed. Specifically, let $\varphi_b(\cdot)$ be the projection operator from $\mathbb{R}^m$ onto the closed ball centered at the origin with radius b. Define $\big(\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})\big)_{j\geq 0}$ with the following recursion:

$$\boldsymbol{X}^{\eta|b}_0(\boldsymbol{x}) = \boldsymbol{x}; \qquad \boldsymbol{X}^{\eta|b}_{j+1}(\boldsymbol{x}) = \boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) + \varphi_b\Big(\eta\boldsymbol{a}\big(\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})\big)\boldsymbol{Z}_{j+1}\Big) \quad \forall j \geq 0. \tag{1.2}$$

In other words, $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ is a modulated version of $\boldsymbol{X}^\eta_j(\boldsymbol{x})$ where the distance traveled at each step is truncated at b. Such dynamical systems arise in the training algorithms for deep neural networks, and their global dynamics has a close connection to the curious ability of SGDs to regularize the deep neural networks algorithmically. See, for example, [63] and the references therein for more details. Note that (1.1) and (1.2) can be viewed as discretizations of small noise SDEs driven by Lévy processes. All the results we establish for (1.1) and (1.2) in this paper can also be established for the stochastic differential equations driven by regularly-varying Lévy processes through a straightforward adaptation of the machinery we develop in this paper. Note that although (1.1) and (1.2) are probably the most natural scaling regime, more general scaling can be considered. In Appendix A, we present the corresponding results under more general scaling regimes, i.e.,

$$\begin{aligned} &\boldsymbol{X}^{\eta}_0(\boldsymbol{x}) = \boldsymbol{x}, \qquad \boldsymbol{X}^{\eta}_j(\boldsymbol{x}) = \boldsymbol{X}^{\eta}_{j-1}(\boldsymbol{x}) + \eta \boldsymbol{a}\big(\boldsymbol{X}^{\eta}_{j-1}(\boldsymbol{x})\big) + \eta^\gamma \boldsymbol{\sigma}\big(\boldsymbol{X}^{\eta}_{j-1}(\boldsymbol{x})\big)\boldsymbol{Z}_j \quad \forall j \geq 1; \\ &\boldsymbol{X}^{\eta|b}_0(\boldsymbol{x}) = \boldsymbol{x}, \qquad \boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) = \boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x}) + \varphi_b\Big(\eta \boldsymbol{a}\big(\boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x})\big) + \eta^\gamma \boldsymbol{\sigma}\big(\boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x})\big)\boldsymbol{Z}_j\Big) \quad \forall j \geq 1. \end{aligned} \tag{1.3}$$

with some $\gamma > 0$. Similar results hold for Lévy-driven SDEs, which are summarized in Appendix B.

At the crux of this study is a fundamental difference between light-tailed and heavy-tailed stochastic dynamical systems. This difference lies in the mechanism through which system-wide rare events arise. In light-tailed systems, the system-wide rare events are characterized by the *conspiracy principle*: the system deviates from its nominal behavior because the entire system behaves subtly differently from the norm, as if it has conspired. In contrast, *the catastrophe principle* governs the rare events in heavy-tailed systems: catastrophic failures (i.e., extremely large deviations from the average behavior) in a small number of components drive the system-wide rare events, and the behavior of the rest of the system is indistinguishable from the nominal behavior.

The principle of a single big jump, a special case of the catastrophe principle, has been discussed in the heavy-tail and extreme value theory literature for a long time. That is, in many heavy-tailed systems, the system-wide rare events arise due to exactly one catastrophe. This line of investigation was initiated in the classical works [46, 47]. The summary of the subsequent developments in the context of processes with independent increments can be found in, for example, [7, 22, 26, 28]. The principle of a single big jump has been rigorously confirmed for random walks in the form of heavy-tailed large deviations at the sample-path level in [31]. More recently, [56] established a fully general catastrophe principle, which goes beyond the principle of a single big jump and characterizes the rare events driven by any number of catastrophes for regularly varying Lévy processes and random walks. For example, let $\mathbb{D}$ denote the space of càdlàg functions over $[0,1]$, let $S_j \triangleq Z_1 + \cdots + Z_j$ denote a mean-zero random walk, and let $\boldsymbol{S}^n \triangleq \{S^n_{\lfloor nt \rfloor}/n : t \in [0,1]\}$ denote a scaled version of S_j. Suppose that Z_i's have a regularly varying tail with index α as above. Then, the sample path large deviations established in [56] takes the following form: for "general" $B \in \mathbb{D}$,

$$\begin{aligned} 0 < \mathbf{C}_k(B^\circ) &\leq \liminf_{n\to\infty} \frac{\mathbf{P}(\boldsymbol{S}^n \in B)}{(n\mathbf{P}(|Z_1| > n))^k} \\ &\leq \limsup_{n\to\infty} \frac{\mathbf{P}(\boldsymbol{S}^n \in B)}{(n\mathbf{P}(|Z_1| > n))^k} \leq \mathbf{C}_k(B^-) < \infty, \end{aligned} \tag{1.4}$$

where k is the minimal number of jumps that a step function must possess in order to belong to B, $\mathbf{C}_k(\cdot)$ is a measure on $\mathbb{D}$ supported on the set of step functions with k or less jumps, and B° and B^- are the interior and closure of B, respectively. Here, k, as a function of B, plays the role of the infimum of rate function over B in the classical light-tailed large deviation principle (LDP) formulation. See also [5] where asymptotic bounds similar to (1.4) were obtained for random walks under more general scaling.

Note that in contrast to the standard log-asymptotics in the classical LDP framework, (1.4) provides exact asymptotics. This formulation provides a powerful framework in heavy-tailed contexts; for instance, this formulation has enabled the design and analysis of strongly efficient rare-event simulation algorithms for a wide variety of rare events associated with $\boldsymbol{S}^n$, as demonstrated in [18]. Moreover, [56, Section 4.4] proves that it is impossible to establish the classical LDP w.r.t. J_1 topology at the sample-path level for regularly varying Lévy processes. On a related note, it is worth mentioning that by relaxing the upper bound of the standard LDP, an alternative formulation known as "extended LDP" was proposed in [8], and such a formulation is also feasible for heavy-tailed processes; see, for example, [6, 2, 3]. However, the extended LDP only provides log-asymptotics. For regularly varying processes, it is often desirable and possible to obtain exact asymptotics; for example, the extended LDP wouldn't suffice for analyzing the strong efficiency of the aforementioned rare-event simulation algorithm in [18]. We will also see that exact asymptotics are crucial in Section 2.3 and Section 4 for sharp exit time and exit location analysis. In fact, it demands an even stronger version of (1.4), which we will introduce in (1.5) shortly. Below, we describe the main contributions of this paper in greater detail.

Large Deviations for Heavy-Tailed Dynamical Systems. The first contribution of this paper is to characterize the catastrophe principle for a general class of heavy-tailed stochastic dynamical systems in the form of a "locally uniform" heavy-tailed large deviations at the sample-path level. This turns out to be the right large deviations formulation for the purpose of the subsequent metastability analysis. To be specific, let $\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x}) \triangleq \{\boldsymbol{X}^{\eta}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) : \ t \in [0,1]\}$ be the time-scaled version of the sample path of $\boldsymbol{X}^{\eta}_{j}(\boldsymbol{x})$ embedded in the continuous time, and note that $\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x})$ is a random element in $\mathbb{D}$. As η decreases, $\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x})$ converges to a deterministic limit $\{\boldsymbol{y}_t(\boldsymbol{x}) : t \in [0,1]\}$, where $d\boldsymbol{y}_t(\boldsymbol{x})/dt = \boldsymbol{a}(\boldsymbol{y}_t(\boldsymbol{x}))$ with initial value $\boldsymbol{y}_0(\boldsymbol{x}) = \boldsymbol{x}$. Let $B \subseteq \mathbb{D}$ be a Borel set w.r.t. the J_1 topology and $A \subset \mathbb{R}^m$ be a compact set. We establish the following asymptotic bound for each k:

$$\begin{aligned} \inf_{\boldsymbol{x}\in A} \mathbf{C}^{(k)}(B^{\circ};\boldsymbol{x}) &\leq \liminf_{\eta\downarrow 0} \frac{\inf_{\boldsymbol{x}\in A} \mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,1]}(x) \in B\big)}{\big(\eta^{-1}\mathbf{P}(\|\boldsymbol{Z}_1\| > \eta^{-1})\big)^k} \\ &\leq \limsup_{\eta\downarrow 0} \frac{\sup_{\boldsymbol{x}\in A} \mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x}) \in B\big)}{\big(\eta^{-1}\mathbf{P}(\|\boldsymbol{Z}_1\| > \eta^{-1})\big)^k} \leq \sup_{\boldsymbol{x}\in A} \mathbf{C}^{(k)}(B^{-};\boldsymbol{x}). \end{aligned} \tag{1.5}$$

The precise statement and the definition of $\mathbf{C}^{(k)}$ can be found in Theorem 2.3 and Section 2.2.1, but here we just point out that the index k that leads to non-degenerate upper and lower bounds in (1.5) is the minimum number of jumps that needs to be added to the path of $\boldsymbol{y}_t(\boldsymbol{x})$ for it to enter the set B given $\boldsymbol{x} \in A$. Such a k dictates the precise polynomial decay rate of the rare-event probability and corresponds to the infimum of rate function of the classical large deviations framework. Note also that as the set A shrinks to an atom, the upper and lower bounds in (1.5) become tighter, and hence, (1.5) is a *locally uniform* version of the large deviations formulation in (1.4). An important implication of (1.5) is a sharp characterization of the catastrophe principle. Specifically, Section 2.2.2 proves that the conditional distribution of $\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x})$ given the rare event of interest converges to the distribution of a piecewise deterministic random function $\boldsymbol{X}^{*}_{|B}(\boldsymbol{x})$ with precisely k random jumps whose sizes are bounded from below:

$$\mathscr{L}\big(\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x})\big|\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x}) \in B\big) \to \mathscr{L}\big(\boldsymbol{X}^{*}_{|B}(\boldsymbol{x})\big).$$

Note that the perturbation associated with $\boldsymbol{Z}_i$ is modulated by $\eta\boldsymbol{\sigma}(\boldsymbol{X}^{\eta}_{i-1}(\boldsymbol{x}))$. Hence, the jump size associated with $\boldsymbol{Z}_i$ being bounded from below implies that $\boldsymbol{Z}_i$ is of order $1/\eta$. This confirms that the rare event $\{\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x}) \in B\}$ arises almost always because of k catastrophically large $\boldsymbol{Z}_i$'s, whereas the rest of the system is indistinguishable from its nominal behavior.

The notion of $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$-convergence, introduced in [42] and further developed in [56], was a key technical tool behind (1.4). In this paper, we introduce a uniform version of the $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence to establish the uniform asymptotics in (1.5) and prove an associated Portmanteau theorem (Theorem 2.2) in Section 2.1. These developments form the backbone that supports our proofs of the uniform sample-path large deviations in (1.5). Furthermore, we also establish the locally uniform asymptotics for the truncated dynamics $\boldsymbol{X}^{\eta|b}_{j}(x)$ in Theorem 2.4. As Section 2.3 elaborates, such large deviations of $\boldsymbol{X}^{\eta|b}_{j}(x)$ leads to exit times and locations with structurally different asymptotic limits compared to those associated with $\boldsymbol{X}^{\eta}_{j}(x)$.

Metastability Analysis. The second contribution of this paper is the first exit-time analysis for heavy-tailed systems. The first exit time problem finds applications in numerous contexts, including chemical reactions [40], physics [16, 17], extreme climate events [52], mathematical finance [59], and queueing systems [60]. A classical result in this literature is the Eyring-Kramers law [27, 40], which characterizes the exit time of Brownian particles; see also [43]. In the light-tailed context, a rich set of systematic tools for exit-time analysis are available [49, 10, 11, 9].

Unlike in the light-tailed context where dynamical systems are driven by Brownian noise, the exit times of the heavy-tailed Lévy-driven SDEs exhibit fundamentally different characteristics, and their successful analysis is a relatively recent development [32, 33]. These results were extended to the higher dimensional settings in [34, 50, 21]. More recently, motivated by the discovery of heavy tails in the stochastic gradient descent algorithms in machine learning literature, stochastic difference equations driven by α-stable noises are investigated extensively; see, for example, [48, 4]. The exit times characterized in this line of research is a manifestation of the principle of a *single* big jump in the context of the exit times of the stochastic dynamical systems. In contrast, our focus in this paper is to build a systematic tool that facilitates the analysis of the exit times even when they are driven by multiple big jump events as in the case of $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$. Indeed, we characterize the asymptotics of the joint law of the first exit time and the exit location for heavy-tailed processes.

We consider (1.1) with drift coefficients $\boldsymbol{a}(\cdot) = -\nabla U(\cdot)$ for some potential function $U \in \mathcal{C}^1(\mathbb{R}^m)$. Specifically, let $I \subseteq \mathbb{R}^d$ be some open and bounded set containing the origin. Suppose that the entire domain I falls within the attraction field of the origin in the following sense: for the ODE path $d\boldsymbol{y}_t(\boldsymbol{x})/dt = -\nabla U(\boldsymbol{y}_t(\boldsymbol{x}))$ with initial condition $\boldsymbol{y}_0(\boldsymbol{x}) = \boldsymbol{x}$, it holds that $\lim_{t\to\infty} \boldsymbol{y}_t(\boldsymbol{x}) = \boldsymbol{0}$ for all $\boldsymbol{x} \in I$. As a result, when initialized within I, the deterministic process will be attracted to and be trapped around the origin. In contrast, under the presence of random perturbations, although $\boldsymbol{X}_j^{\eta}(\boldsymbol{x})$ and $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ are attracted to the origin most of the times, they will eventually escape from I if one waits long enough. Of particular interest are the asymptotics of the first exit time as $\eta \to \infty$. Theorem 2.6 establishes that the joint law of the first exit time $\tau^{\eta|b}(\boldsymbol{x}) = \min\{j \geq 0 : \ \boldsymbol{X}_j^{\eta|b}(\boldsymbol{x}) \notin I\}$ and the exit location $\boldsymbol{X}_{\tau}^{\eta|b}(\boldsymbol{x}) \triangleq \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x})$ admits the following limit (for all $\boldsymbol{x} \in I$):

$$\Big(\lambda_b^I(\eta)\cdot\tau^{\eta|b}(\boldsymbol{x}),\ \boldsymbol{X}_{\tau}^{\eta|b}(\boldsymbol{x})\Big) \Rightarrow (E, V_b) \qquad \text{as } \eta \downarrow 0 \tag{1.6}$$

with some (deterministic) time-scaling function $\lambda_b^I(\eta)$. Here, E is an exponential random variable with the rate parameter 1, and V_b is some random element independent of E and supported on I^c. The exact law of V_b and the definition of $\lambda_b^I(\eta)$ are provided in Section 2.3.1. Here, we note that $\lambda_b^I(\eta)$ is regularly varying with index $-[1 + \mathcal{J}_b^I(\alpha - 1)]$, where $\mathcal{J}_b^I$ is the "discretized width" of domain I relative to the truncation threshold b; see (2.33) for the precise definition. Intuitively speaking, $\mathcal{J}_b^I$ is the minimal number of jumps of size b to escape from I, and hence, the wider the domain I is, the longer the exit time $\tau^{\eta|b}(\boldsymbol{x})$ will be asymptotically. Theorem 2.6 also obtains the first exit time analysis for $\boldsymbol{X}_j^{\eta}(\boldsymbol{x})$ by considering an arbitrarily large truncation threshold $b \approx \infty$.

Our approach hinges on a general machinery we develop in Section 2.3.2. At the core of this development lies the concept of asymptotic atoms, namely, nested regions of recurrence at which the process asymptotically regenerates upon each visit. Our locally uniform sample-path large deviations then prove to be the right tool in this framework, empowering us to simultaneously characterize the behavior of the stochastic processes under all the initial values over the asymptotic atoms.

It should be noted that [35] also investigated the exit events driven by multiple jumps. However, the mechanism through which multiple jumps arise in their context is due to a different tail behavior of the increment distribution that is lighter than any polynomial rate—more precisely, a Weibull tail—and it is fundamentally different from that of the regularly varying case. Along with the aforementioned results [32, 33, 34] for regularly varying SDEs, [35] paints interesting picture of the hierarchy in the asymptotics of the first exit times. See [36] for the summary of such hierarchy. Our results complement the picture and provide a missing piece of the puzzle by unveiling the precise effect of truncation in the regularly varying cases. In particular, we characterize a discrete structure of phase transitions in (1.6), where we find that the first exit time $\tau^{\eta|b}(\boldsymbol{x})$ is (roughly) of order $1/\eta^{1+\mathcal{J}_b^I\cdot(\alpha-1)}$ for small η. This means that the order of the first exit time $\tau^{\eta|b}(\boldsymbol{x})$ does not vary continuously with b; rather, it exhibits a discrete dependence on b through $\mathcal{J}_b^I$.

Some of the results in Section 2.3 of this paper have been presented in a preliminary form at a conference [63]. The main focus of [63] was the connection between the metastability analysis of stochastic gradient descent (SGD) and its generalization performance in the context of training deep neural networks. Compared to the brute force approach in [63], the current paper provides a systematic framework to characterize the global dynamics for significantly more general class of heavy-tailed dynamical systems.

The rest of the paper is organized as follows. Section 2 presents the main results of this paper. Section 3 and Section 4 provide the proofs of Sections 2.1, 2.2, and 2.3. Results for SDEs driven by Lévy processes with regularly varying increments are collected in Appendix B. Results for stochastic difference equations under more general scaling regimes are presented in Appendix A.

2 Main Results

This section presents the main results of this paper and discusses their implications. Section 2.1 introduces the uniform version of $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence and presents an associated portmanteau theorem. Section 2.2 develops the sample-path large deviations, and Section 2.3 carries out the metastability analysis. All the proofs are deferred to the later sections.

Before presenting the main results, we set frequently used notations. Let $[n] \triangleq \{1, 2, \cdots, n\}$ for any positive integer n. Let $\mathbb{N} = \{0, 1, 2, \cdots\}$ be the set of non-negative integers. Let $(\mathbb{S}, \boldsymbol{d})$ be a metric space with $\mathscr{S}_{\mathbb{S}}$ being the corresponding Borel σ-algebra. For any $E \subseteq \mathbb{S}$, let E° and E^- be the interior and closure of E, respectively. For any $r > 0$, let $E^r \triangleq \{y \in \mathbb{S} : \boldsymbol{d}(E, y) \leq r\}$ be the r-enlargement of a set E. Here for any set $A \subseteq \mathbb{S}$ and any $x \in \mathbb{S}$, we define $\boldsymbol{d}(A, x) \triangleq \inf\{\boldsymbol{d}(y, x) : y \in A\}$. Also, let $E_r \triangleq ((E^c)^r)^c$ be the r-shrinkage of E. Note that for any E, the enlargement E^r of E is closed, and the shrinkage E_r of E is open. We say that set $A \subseteq \mathbb{S}$ is bounded away from another set $B \subseteq \mathbb{S}$ if $\inf_{x \in A, y \in B} \boldsymbol{d}(x, y) > 0$. For any Borel measure μ on $(\mathbb{S}, \mathscr{S}_{\mathbb{S}})$, let the support of μ (denoted as $\text{supp}(\mu)$) be the smallest closed set C such that $\mu(\mathbb{S} \setminus C) = 0$. For any function $g : \mathbb{S} \to \mathbb{R}$, let $\text{supp}(g) \triangleq \big(\{x \in \mathbb{S} : g(x) \neq 0\}\big)^-$. Given two sequences of positive real numbers $(x_n)_{n\geq 1}$ and $(y_n)_{n\geq 1}$, we say that $x_n = \boldsymbol{O}(y_n)$ (as $n \to \infty$) if there exists some $C \in [0, \infty)$ such that $x_n \leq C y_n$ $\forall n \geq 1$. Besides, we say that $x_n = \boldsymbol{o}(y_n)$ if $\lim_{n\to\infty} x_n/y_n = 0$.

2.1 Uniform $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-Convergence

This section extends the notion of $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence [42, 56] to a uniform version and prove an associated portmanteau theorem. Such developments pave the way to the locally uniform heavy-tailed sample-path large deviations.

Specifically, in this section we consider some metric space $(\mathbb{S}, \boldsymbol{d})$ that is complete and separable. Given any Borel measurable subset $\mathbb{C} \subseteq \mathbb{S}$, let $\mathbb{S}\setminus\mathbb{C}$ be a subspace of $\mathbb{S}$ equipped with the relative topology with σ-algebra $\mathscr{S}_{\mathbb{S}\setminus\mathbb{C}} \triangleq \{A \in \mathscr{S}_{\mathbb{S}} : A \subseteq \mathbb{S}\setminus\mathbb{C}\}$. Let

$$\mathbb{M}(\mathbb{S}\setminus\mathbb{C}) \triangleq \{\nu(\cdot) \text{ is a Borel measure on } \mathbb{S}\setminus\mathbb{C} : \nu(\mathbb{S}\setminus\mathbb{C}^r) < \infty\ \forall r > 0\}.$$

$\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ can be topologized by the sub-basis constructed using sets of form $\{\nu \in \mathbb{M}(\mathbb{S}\setminus\mathbb{C}) : \nu(f) \in G\}$, where $G \subseteq [0, \infty)$ is open, $f \in \mathcal{C}(\mathbb{S}\setminus\mathbb{C})$, and $\mathcal{C}(\mathbb{S}\setminus\mathbb{C})$ is the set of all real-valued, non-negative, bounded and continuous functions with support bounded away from $\mathbb{C}$ (i.e., $f(x) = 0$ $\forall x \in \mathbb{C}^r$ for some $r > 0$). Given a sequence $\mu_n \in \mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ and some $\mu \in \mathbb{M}(\mathbb{S}\setminus\mathbb{C})$, we say that μ_n converges to μ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ as $n \to \infty$ if $\lim_{n\to\infty} |\mu_n(f) - \mu(f)| = 0$ for all $f \in \mathcal{C}(\mathbb{S}\setminus\mathbb{C})$. See [42] for alternative definitions in the form of a Portmanteau Theorem. When the choice of $\mathbb{S}$ and $\mathbb{C}$ is clear from the context, we simply refer to it as $\mathbb{M}$-convergence. As demonstrated in [56], the sample path large deviations for heavy-tailed stochastic processes can be formulated in terms of $\mathbb{M}$-convergence of the scaled process in the Skorokhod space. In this paper, we introduce a stronger version of $\mathbb{M}$-convergence, which facilitates the analysis of the local stability and global dynamics in the later sections.

Definition 2.1 (Uniform $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence)**.** *Let Θ be a set of indices. Let μ^η_θ, $\mu_\theta \in \mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ for each $\eta > 0$ and $\theta \in \Theta$. We say that μ^η_θ converges to μ_θ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ uniformly in θ on Θ as $\eta \to 0$*

if

$$\lim_{\eta\downarrow 0}\sup_{\theta\in\Theta}|\mu^\eta_\theta(f)-\mu_\theta(f)|=0\qquad \forall f\in\mathcal{C}(\mathbb{S}\setminus\mathbb{C}).$$

If $\{\mu_\theta:\ \theta\in\Theta\}$ is sequentially compact, a Portmanteau-type theorem holds. The proof is provided in Section 3.1.

Theorem 2.2 (Portmanteau theorem for uniform $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence)**.** *Let Θ be a set of indices. Let $\mu^\eta_\theta,\ \mu_\theta\in\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ for each $\eta>0$ and $\theta\in\Theta$. If, for any sequence of measures $(\mu_{\theta_n})_{n\geq 1}$, there exist a sub-sequence $(\mu_{\theta_{n_k}})_{k\geq 1}$ and some $\theta^*\in\Theta$ such that*

$$\lim_{k\to\infty}\mu_{\theta_{n_k}}(f)=\mu_{\theta^*}(f)\qquad \forall f\in\mathcal{C}(\mathbb{S}\setminus\mathbb{C}),\tag{2.1}$$

then the next three statements are equivalent:

(i) μ^η_θ *converges to* μ_θ *in* $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ *uniformly in* θ *on* Θ *as* $\eta\downarrow 0$*;*

(ii) $\lim_{\eta\downarrow 0}\sup_{\theta\in\Theta}|\mu^\eta_\theta(f)-\mu_\theta(f)|=0$ *for each* $f\in\mathcal{C}(\mathbb{S}\setminus\mathbb{C})$ *that is also uniformly continuous on* $\mathbb{S}$*;*

(iii) $\limsup_{\eta\downarrow 0}\sup_{\theta\in\Theta}\mu^\eta_\theta(F)-\mu_\theta(F^\epsilon)\leq 0$ *and* $\liminf_{\eta\downarrow 0}\inf_{\theta\in\Theta}\mu^\eta_\theta(G)-\mu_\theta(G_\epsilon)\geq 0$ *for all* $\epsilon>0$*, all closed* $F\subseteq\mathbb{S}$ *that is bounded away from* $\mathbb{C}$*, and all open* $G\subseteq\mathbb{S}$ *that is bounded away from* $\mathbb{C}$*.*

Furthermore, any of the claims (i)–(iii) implies the following.

(iv) $\limsup_{\eta\downarrow 0}\sup_{\theta\in\Theta}\mu^\eta_\theta(F)\leq\sup_{\theta\in\Theta}\mu_\theta(F)$ *and* $\liminf_{\eta\downarrow 0}\inf_{\theta\in\Theta}\mu^\eta_\theta(G)\geq\inf_{\theta\in\Theta}\mu_\theta(G)$ *for all closed* $F\subseteq\mathbb{S}$ *that is bounded away from* $\mathbb{C}$ *and all open* $G\subseteq\mathbb{S}$ *that is bounded away from* $\mathbb{C}$*.*

Remark 1. *To conclude, we provide two additional remarks regarding Theorem 2.2. First, it is not possible to strengthen statement (iii) and assert that*

$$\limsup_{\eta\downarrow 0}\sup_{\theta\in\Theta}\mu^\eta_\theta(F)-\mu_\theta(F)\leq 0,\qquad \liminf_{\eta\downarrow 0}\inf_{\theta\in\Theta}\mu^\eta_\theta(G)-\mu_\theta(G)\geq 0\tag{2.2}$$

for all closed $F\subseteq\mathbb{S}$ *bounded away from* $\mathbb{C}$ *and all open* $G\subseteq\mathbb{S}$ *bounded away from* $\mathbb{C}$*. In other words, in statement (iii) the* ϵ*-fattening in* F^ϵ *and* ϵ*-shrinking in* G_ϵ *are indispensable. Indeed, we demonstrate through a counterexample that, due to the infinite cardinality of the collections of measures* $\{\mu^\eta_\theta:\theta\in\Theta\}$ *and* $\{\mu_\theta:\ \theta\in\Theta\}$*, the claims in (2.2) can easily fall apart while statements (i)–(iii) hold true. Specifically, by setting* $\mathbb{C}=\emptyset$ *and* $\mathbb{S}=\mathbb{R}$*, the* $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$*-convergence degenerates to the weak convergence of Borel measures on* $\mathbb{R}$*. Set* $\Theta=[-1,1]$ *and*

$$\mu^\eta_\theta\triangleq\boldsymbol{\delta}_{\theta-\eta},\qquad \mu_\theta\triangleq\boldsymbol{\delta}_\theta,$$

where $\boldsymbol{\delta}_x$ *is the Dirac measure at* x*. For closed set* $F=[-1,0]$ *and any* $\eta\in(0,2)$*,*

$$\begin{aligned}\sup_{\theta\in\Theta}\mu^\eta_\theta(F)-\mu_\theta(F)&\geq\boldsymbol{\delta}_{-\eta/2}\big([-1,0]\big)-\boldsymbol{\delta}_{\eta/2}\big([-1,0]\big)\qquad \textit{by picking } \theta=\eta/2\\ &=\mathbb{I}\bigg\{\frac{-\eta}{2}\in[-1,0]\bigg\}-\mathbb{I}\bigg\{\frac{\eta}{2}\in[-1,0]\bigg\}=1,\end{aligned}$$

thus implying $\limsup_{\eta\downarrow 0}\sup_{\theta\in\Theta}\mu^\eta_\theta(F)-\mu_\theta(F)\geq 1$*.*

Secondly, while statement (iv) holds as the key component when establishing the sample-path large deviation results, it is indeed strictly weaker than the other claims for one obvious reason: unlike statements (i)–(iii), the content of statement (iv) does not require μ^η_θ *to converge to* μ_θ *for any* $\theta\in\Theta$*. To illustrate that (iv) does not imply (i)–(iii), it suffices to examine the following case where* $\mathbb{C}=\emptyset$*,* $\mathbb{S}=\mathbb{R}$*,* $\Theta=[-1,1]$*,* $\mu^\eta_\theta=\boldsymbol{\delta}_{-\theta}$*, and* $\mu_\theta=\boldsymbol{\delta}_\theta$*.*

2.2 Heavy-Tailed Large Deviations

In Section 2.2.1, we study the sample-path large deviations for stochastic difference equations with heavy-tailed increments. Section 2.2.2 then characterizes the catastrophe principle of heavy-tailed systems by presenting the conditional limit theorems that reveal a discrete hierarchy of the most likely scenarios and probabilities of rare events in heavy-tailed stochastic difference equations.

2.2.1 Sample-Path Large Deviations

Let $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$ be iid copies of some random vector $\boldsymbol{Z}$ taking values in $\mathbb{R}^d$, and let $\mathcal{F}$ be the σ-algebra generated by $(\boldsymbol{Z}_j)_{j\geq 1}$. Henceforth in this paper, all vectors in Euclidean spaces are understood as column vectors. Let $\mathcal{F}_j$ be the σ-algebra generated by $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_j$ and $\mathcal{F}_0 \triangleq \{\emptyset, \Omega\}$. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$ be a filtered probability space with filtration $\mathbb{F} = (\mathcal{F}_j)_{j\geq 0}$. The goal of this section is to study the sample-path large deviations for the discrete-time process $\big\{\boldsymbol{X}^\eta_t(\boldsymbol{x}) : t \in \mathbb{N}\big\}$ in $\mathbb{R}^m$, which is driven by the recursion

$$\boldsymbol{X}^\eta_0(\boldsymbol{x}) = \boldsymbol{x}; \qquad \boldsymbol{X}^\eta_t(\boldsymbol{x}) = \boldsymbol{X}^\eta_{t-1}(\boldsymbol{x}) + \eta\boldsymbol{a}\big(\boldsymbol{X}^\eta_{t-1}(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^\eta_{t-1}(\boldsymbol{x})\big)\boldsymbol{Z}_t, \quad \forall t \geq 1 \tag{2.3}$$

as $\eta \downarrow 0$. In particular, we are interested in the case where $\boldsymbol{Z}_i$'s are heavy-tailed. Heavy-tails are typically captured with the notion of regular variation. For any measurable function $\phi : (0,\infty) \to (0,\infty)$, we say that ϕ is regularly varying as $x \to \infty$ with index β (denoted as $\phi(x) \in \mathcal{RV}_\beta(x)$ as $x \to \infty$) if $\lim_{x\to\infty} \phi(tx)/\phi(x) = t^\beta$ for all $t > 0$. For details of the definition and properties of regularly varying functions, see, for example, Chapter 2 of [55]. Throughout this paper, we say that a measurable function $\phi(\eta)$ is regularly varying as $\eta \downarrow 0$ with index β if $\lim_{\eta\downarrow 0} \phi(t\eta)/\phi(\eta) = t^\beta$ for any $t > 0$. We denote this as $\phi(\eta) \in \mathcal{RV}_\beta(\eta)$ as $\eta \downarrow 0$. Besides, we adopt the L_2 norm $\|(x_1, \cdots, x_k)\| = \sqrt{\sum_{j=1}^k x_k^2}$ on Euclidean spaces. Let

$$H(x) \triangleq \mathbf{P}(\|\boldsymbol{Z}\| > x). \tag{2.4}$$

For any $\alpha > 0$, let ν_α be the (Borel) measure on $\mathbb{R}$ with

$$\nu_\alpha[x, \infty) = x^{-\alpha}. \tag{2.5}$$

Let $\mathfrak{N}_d \triangleq \{\boldsymbol{x} \in \mathbb{R}^d : \ \|\boldsymbol{x}\| = 1\}$ be the unit sphere of $\mathbb{R}^d$. Let $\Phi : \mathbb{R}^d \to [0,\infty) \times \mathfrak{N}_d$ be

$$\Phi(\boldsymbol{x}) \triangleq \begin{cases} \Big(\|\boldsymbol{x}\|, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\Big) & \text{if } \boldsymbol{x} \neq 0, \\ \big(0, (1, 0, 0, \cdots, 0)\big) & \text{otherwise.} \end{cases} \tag{2.6}$$

Since $\Phi(\boldsymbol{x})$ will not be applied at $\boldsymbol{x} = \mathbf{0}$ in our proof, Φ can essentially be interpreted as the polar transform on $\mathbb{R}^d$. We impose the following multivariate regular variation assumption regarding the law of $\boldsymbol{Z}$.

Assumption 1 (Regularly Varying Noises)**.** $\mathbf{E}\boldsymbol{Z} = \mathbf{0}$. *Besides, there exist some $\alpha > 1$ and a probability measure $\mathbf{S}(\cdot)$ on the unit sphere of $\mathbb{R}^d$ such that*

- *$H(x) \in \mathcal{RV}_{-\alpha}(x)$ as $x \to \infty$,*
- *for the polar coordinates $(R, \boldsymbol{\Theta}) \triangleq \Phi(\boldsymbol{Z})$, we have (as $x \to \infty$)*

$$\frac{\mathbf{P}\Big((x^{-1}R, \boldsymbol{\Theta}) \in \ \cdot\ \Big)}{H(x)} \to \nu_\alpha \times \mathbf{S} \qquad \text{in } \mathbb{M}\Big(\big([0,\infty) \times \mathfrak{N}_d\big) \setminus \big(\{0\} \times \mathfrak{N}_d\big)\Big). \tag{2.7}$$

Remark 2. *It is worth noticing that the multivariate regular variation condition* (2.7) *is typically stated in terms of vague convergence; see, e.g., [54, 30]. While vague convergence is generally weaker than* $\mathbb{M}$*-convergence (see Lemma 2.1 of [42]), due to* $\alpha > 1$ *we have* $(\nu_\alpha \times \mathbf{S})(A) < \infty$ *for any set* $A \subseteq (0,\infty) \times \mathfrak{N}_d$ *bounded away from* $\{0\} \times \mathfrak{N}_d$*. Therefore, it is easy to verify that the* $\mathbb{M}$*-convergence stated in* (2.7) *is equivalent to vague convergence. Furthermore, by equivalent statements of multivariate regular variation (see [54, 30]), Assumption 1 is equivalent to that* $H^{-1}(x)\mathbf{P}(x^{-1}\boldsymbol{Z} \in \ \cdot\ )$ *converges to some Borel measure* $\mu(\cdot)$ *in* $\mathbb{M}(\mathbb{R}^d \setminus \{\mathbf{0}\})$ *and* $\mu(\cdot)$ *admits self-similarity of form* $\mu(\lambda A) = \lambda^{-\alpha}\mu(A)$ *for any Borel set* $A \subseteq \mathbb{R}^d$ *that is bounded away from the origin.*

Next, we introduce the following assumptions on the drift coefficient $\boldsymbol{a} : \mathbb{R}^m \to \mathbb{R}^m$ and the diffusion coefficient $\boldsymbol{\sigma} : \mathbb{R}^m \to \mathbb{R}^{m\times d}$. To be clear, the exact meaning of the vecotrized version of the stochastic difference equation in (2.3) is as follows. By writing $\boldsymbol{a}(\cdot) = \big(a_1(\cdot), \cdots, a_m(\cdot)\big)$, $\boldsymbol{\sigma}(\cdot) = \big(\sigma_{i,j}(\cdot)\big)_{i\in[m],j\in[d]}$, $\boldsymbol{X}^\eta_t(x) = \big(X^\eta_{t,1}(x), \cdots, X^\eta_{t,m}(x)\big)$, and $\boldsymbol{Z}_t = (Z_{t,1}, \cdots, Z_{t,d})$, we have

$$X^\eta_{t,i}(x) = X^\eta_{t-1,i}(x) + \eta a_i\big(\boldsymbol{X}^\eta_{t-1}(x)\big) + \eta \sum_{j\in[d]} \sigma_{i,j}\big(\boldsymbol{X}^\eta_{t-1}(x)\big)Z_{t,j} \qquad \forall t \geq 1,\ i \in [m] \tag{2.8}$$

as a scalar version of the recursion. Henceforth, we adopt the L_2 vector norm induced matrix norm $\|\mathbf{A}\| = \sup_{\boldsymbol{x}\in\mathbb{R}^q:\ \|\boldsymbol{x}\|=1} \|\mathbf{A}\boldsymbol{x}\|$ for any $\mathbf{A} \in \mathbb{R}^{p\times q}$. Note that the lower bounds for C and D in Assumption 2 and 3 are obviously not necessary. However, we assume that $C \geq 1$ and $D \geq 1$ w.l.o.g. for the notational simplicity.

Assumption 2 (Lipschitz Continuity)**.** *There exists some* $D \in [1,\infty)$ *such that*

$$\|\boldsymbol{\sigma}(\boldsymbol{x}) - \boldsymbol{\sigma}(\boldsymbol{y})\| \vee \|\boldsymbol{a}(\boldsymbol{x}) - \boldsymbol{a}(\boldsymbol{y})\| \leq D\,\|\boldsymbol{x} - \boldsymbol{y}\| \quad \forall \boldsymbol{x},\ \boldsymbol{y} \in \mathbb{R}^m.$$

Assumption 3 (Boundedness)**.** *There exists some* $C \in [1,\infty)$ *such that*

$$\|\boldsymbol{a}(\boldsymbol{x})\| \vee \|\boldsymbol{\sigma}(\boldsymbol{x})\| \leq C \qquad \forall \boldsymbol{x} \in \mathbb{R}^m.$$

To present the main results, we set a few notations. Let $(\mathbb{D}[0,T], \boldsymbol{d}^{[0,T]}_{J_1})$ be the metric space where $\mathbb{D}[0,T] = \mathbb{D}_m[0,T]$ is the space of all càdlàg functions with domain $[0,T]$ and codomain $\mathbb{R}^m$, and $\boldsymbol{d}^{[0,T]}_{J_1} = \boldsymbol{d}^{[0,T]}_{J_1,m}$ is the Skorodkhod J_1 metric

$$\boldsymbol{d}^{[0,T]}_{J_1}(x,y) \triangleq \inf_{\lambda\in\Lambda_T} \sup_{t\in[0,T]} |\lambda(t) - t| \vee \|x(\lambda(t)) - y(t)\|\,. \tag{2.9}$$

Here, Λ_T is the set of all homeomorphism on $[0,T]$. Throughout this paper, we fix some m and d and consider $\boldsymbol{X}^\eta_t(\boldsymbol{x})$ taking values in $\mathbb{R}^m$ driven by $\boldsymbol{Z}_t$'s in $\mathbb{R}^d$, so we omit the subscript m in $\mathbb{D}[0,T]$ and $\boldsymbol{d}^{[0,T]}_{J_1}$. Given any $A \subseteq \mathbb{R}$, let $A^{k\uparrow} \triangleq \{(t_1,\cdots,t_k) \in A^k :\ t_1 < t_2 < \cdots < t_k\}$ be the set of sequences of increasing real numbers with length k on A. For any $k \in \mathbb{N}$ and $T > 0$, define mapping $\bar{h}^{(k)}_{[0,T]} : \mathbb{R}^m \times \mathbb{R}^{d\times k} \times \mathbb{R}^{m\times k} \times (0,T]^{k\uparrow} \to \mathbb{D}[0,T]$ as follows. Given any $\boldsymbol{x} \in \mathbb{R}^m$, $\mathbf{W} = (\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\mathbf{V} = (\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$, and $\boldsymbol{t} = (t_1,\cdots,t_k) \in (0,T]^{k\uparrow}$, let $\xi = \bar{h}^{(k)}_{[0,T]}(\boldsymbol{x},\mathbf{W},\mathbf{V},\boldsymbol{t}) \in \mathbb{D}[0,T]$ be the solution to

$$\xi_0 = \boldsymbol{x} \tag{2.10}$$

$$\frac{d\xi_s}{ds} = \boldsymbol{a}(\xi_s) \qquad \forall s \in [0,T],\ s \neq t_1,\cdots,t_k \tag{2.11}$$

$$\xi_s = \xi_{s-} + \boldsymbol{v}_j + \boldsymbol{\sigma}(\xi_{s-} + \boldsymbol{v}_j)\boldsymbol{w}_j \qquad \text{if } s = t_j \text{ for some } j \in [k]. \tag{2.12}$$

Here, for any $\xi \in \mathbb{D}[0,T]$ and $t \in (0,T]$, we use $\xi_{t-} = \lim_{s\uparrow t} \xi_s$ to denote the left limit of ξ at t, and we set $\xi_{0-} = \xi_0$. Also, define a mapping $h^{(k)}_{[0,T]} : \mathbb{R}^m \times \mathbb{R}^{d\times k} \times (0,T]^{k\uparrow} \to \mathbb{D}[0,T]$ as

$$h^{(k)}_{[0,T]}(\boldsymbol{x},\boldsymbol{W},\boldsymbol{t}) \triangleq \bar{h}^{(k)}_{[0,T]}\big(\boldsymbol{x},\boldsymbol{W},(\mathbf{0},\cdots,\mathbf{0}),\boldsymbol{t}\big). \tag{2.13}$$

In essence, the mapping $h^{(k)}_{[0,T]}(\boldsymbol{x},\mathbf{W},\boldsymbol{t})$ produces the ODE path perturbed by jumps $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k$ (with sizes modulated by the drift coefficient $\boldsymbol{\sigma}(\cdot)$) at times $t_1,\cdots,t_k$, and the mapping $\bar{h}^{(k)}_{[0,T]}$ further includes perturbations $\boldsymbol{v}_j$ right before each jump. We adopt the convention that $\xi = \bar{h}^{(0)}_{[0,T]}(\boldsymbol{x})$ is the solution to the ODE $d\xi_s/ds = \boldsymbol{a}(\xi_s)\ \forall s \in [0,T]$ under the initial condition $\xi_0 = \boldsymbol{x}$.

For any $t > 0$, let $\mathcal{L}_t$ be the Lebesgue measure restricted on $(0,t)$ and $\mathcal{L}^{k\uparrow}_t$ be the Lebesgue measure restricted on $(0,t)^{k\uparrow}$. Given any $T>0$, $x \in \mathbb{R}$, and $k \geq 0$, let

$$\mathbf{C}^{(k)}_{[0,T]}(\ \cdot\ ;\boldsymbol{x}) \triangleq \int \mathbb{I}\Big\{h^{(k)}_{[0,T]}\big(\boldsymbol{x},\mathbf{W},\boldsymbol{t}\big) \in\ \cdot\ \Big\}\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^k(d\mathbf{W}) \times \mathcal{L}^{k\uparrow}_T(d\boldsymbol{t}). \tag{2.14}$$

Here, $\mathbf{S}$ is the Borel (in fact, probability) measure on the unit sphere $\mathfrak{N}_d$ characterized in Assumption 1, ν_α is specified in (2.5), $(\nu_\alpha \times \mathbf{S})\circ\Phi$ is the composition of the product measure $\nu_\alpha \times \mathbf{S}$ with the polar transform Φ, i.e.,

$$\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)(B) \triangleq (\nu_\alpha \times \mathbf{S})\big(\Phi(B)\big) \qquad \forall B \subseteq \mathbb{R}^d \text{ that is Borel}, \tag{2.15}$$

and $\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^k$ is the k-fold of $(\nu_\alpha \times \mathbf{S})\circ\Phi$. In other words, for $\mathbf{W} = (\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k) \in \mathbb{R}^{p\times k}$, we have $\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^k(d\mathbf{W}) = \bigtimes_{j\in[k]}\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)(d\boldsymbol{w}_j)$. For $\{\boldsymbol{X}^\eta_j(\boldsymbol{x}) : j \geq 0\}$, we define the time-scaled version of the sample path as

$$\boldsymbol{X}^\eta_{[0,T]}(\boldsymbol{x}) \triangleq \big\{\boldsymbol{X}^\eta_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) :\ t \in [0,T]\big\}, \quad \forall T > 0 \tag{2.16}$$

with $\lfloor t \rfloor \triangleq \max\{n \in \mathbb{Z} :\ n \leq t\}$ and $\lceil t \rceil \triangleq \min\{n \in \mathbb{Z} :\ n \geq t\}$. Note that $\boldsymbol{X}^\eta_{[0,T]}(\boldsymbol{x})$ is a $\mathbb{D}[0,T]$-valued random element.

For each $r > 0$ and $\boldsymbol{x} \in \mathbb{R}^m$, let $\bar{B}_r(\boldsymbol{x}) \triangleq \{\boldsymbol{y} \in \mathbb{R}^m :\ \|\boldsymbol{y}-\boldsymbol{x}\| \leq r\}$ be the closed ball with radius r centered at $\boldsymbol{x}$. For any $k \in \mathbb{N}$, $T > 0$, $\epsilon \geq 0$, and $A \subseteq \mathbb{R}^m$, let

$$\mathbb{D}^{(k)}_A[0,T](\epsilon) \triangleq \bar{h}^{(k)}_{[0,T]}\Big(A \times \mathbb{R}^{m\times k} \times \big(\bar{B}_\epsilon(\mathbf{0})\big)^k \times (0,T]^{k\uparrow}\Big) \tag{2.17}$$

be the set that contains all the ODE path with k jumps by time T, i.e., images of the mapping $\bar{h}^{(k)}_{[0,T]}$ defined in (2.10)–(2.12), with small perturbations $\|\boldsymbol{v}_j\| \leq \epsilon$ for all j. We adopt the convention that $\mathbb{D}^{(-1)}_A[0,T](\epsilon) \triangleq \emptyset$. Also, it is easy to see that $\mathbb{D}^{(k)}_A[0,T](\epsilon) \subseteq \mathbb{D}^{(k)}_A[0,T](\epsilon')$ for any $0 \leq \epsilon < \epsilon'$. Next, define a scale function

$$\lambda(\eta) \triangleq \eta^{-1}H(\eta^{-1})$$

with $H(x) = \mathbf{P}(\|\boldsymbol{Z}\| > x)$ defined in (2.4). From Assumption 1, one can see that $\lambda(\eta) \in \mathcal{RV}_{\alpha-1}(\eta)$ as $\eta \downarrow 0$. For any $k \geq 1$ we write $\lambda^k(\eta) = \big(\lambda(\eta)\big)^k$. In case $T = 1$, we suppress the time horizon $[0,1]$ and write $\mathbb{D}$, $\boldsymbol{d}_{J_1}$, $h^{(k)}$, $\mathbf{C}^{(k)}$, $\mathbb{D}^{(k)}_A(\epsilon)$, and $\boldsymbol{X}^\eta(x)$ to denote $\mathbb{D}[0,1]$, $\boldsymbol{d}^{[0,1]}_{J_1}$, $h^{(k)}_{[0,1]}$, $\mathbf{C}^{(k)}_{[0,1]}$, $\mathbb{D}^{(k)}_A[0,1](\epsilon)$, and $\boldsymbol{X}^\eta_{[0,1]}(x)$, respectively.

Now, we are ready to state Theorem 2.3, which establishes the uniform $\mathbb{M}$-convergence of (the law of) $\boldsymbol{X}^\eta_{[0,T]}(x)$ to $\mathbf{C}^{(k)}(\ \cdot\ ;x)$ and a uniform version of the sample-path large deviations for $\boldsymbol{X}^\eta_{[0,T]}(x)$.

Theorem 2.3. *Under Assumptions 1, 2, and 3, it holds for any $k \in \mathbb{N}$, $T > 0$, $\epsilon > 0$, and any compact $A \subseteq \mathbb{R}^m$ that*

$$\lambda^{-k}(\eta)\mathbf{P}\big(\boldsymbol{X}^\eta_{[0,T]}(\boldsymbol{x}) \in\ \cdot\ \big) \to \mathbf{C}^{(k)}_{[0,T]}(\ \cdot\ ;\boldsymbol{x}) \quad \text{in } \mathbb{M}\Big(\mathbb{D}[0,T] \setminus \mathbb{D}^{(k-1)}_A[0,T](\epsilon)\Big) \text{ uniformly in } \boldsymbol{x} \text{ on } A$$

as $\eta \downarrow 0$. Furthermore, for any $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ that is bounded away from $\mathbb{D}^{(k-1)}_A[0,T](\epsilon)$,

$$\begin{aligned}\inf_{\boldsymbol{x}\in A}\mathbf{C}^{(k)}_{[0,T]}(B^\circ;\boldsymbol{x}) &\leq \liminf_{\eta\downarrow 0}\frac{\inf_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^\eta_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta)} \\ &\leq \limsup_{\eta\downarrow 0}\frac{\sup_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^\eta_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta)} \leq \sup_{\boldsymbol{x}\in A}\mathbf{C}^{(k)}_{[0,T]}(B^-;\boldsymbol{x}) < \infty.\end{aligned} \tag{2.18}$$

Remark 3. *We add a remark on the connection between (2.18) and the classical LDP framework. Given any measurable $B \subseteq \mathbb{D}[0,T]$, there is a particular k that plays the role of the rate function. Specifically, let $\mathcal{J}_A(B) \triangleq \min\{k \in \mathbb{N} : \ B \cap \mathbb{D}_A^{(k)}[0,T](\epsilon) \neq \emptyset\}$. In great generality, this coincides with the smallest possible value of $k \in \mathbb{N}$ for which the lower bound $\inf_{\boldsymbol{x}\in A} \mathbf{C}^{(k)}_{[0,T]}(B^\circ; \boldsymbol{x})$ in (2.18) can be strictly positive, and $\lambda^{\mathcal{J}_A(B)}(\eta)$ characterizes the exact rate of decay for both $\inf_{\boldsymbol{x}\in A} \mathbf{P}(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in B)$ and $\sup_{\boldsymbol{x}\in A} \mathbf{P}(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in B)$ as $\eta \downarrow 0$. It should be noted these results are exact asymptotics as opposed to the log asymptotics in classical LDP framework. In case that the set A is a singleton, $T = 1$, $\boldsymbol{a} \equiv 0$, and $\boldsymbol{\sigma} \equiv \boldsymbol{I}_m$ (i.e., the identity matrix in $\mathbb{R}^m$), the process $\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x})$ will degenerate to a Lévy process, and $\mathcal{J}_A(\cdot)$ will reduce to $\mathcal{J}(\cdot)$ defined in equation (3.3) of [56]. This confirms that Theorem 2.3 is a proper generalization of the heavy-tailed large deviations for Lévy processes and random walks in [56].*

The proof of Theorem 2.3 will be given in Section 3.3. Interestingly enough, the results are obtained by first studying its truncated counterpart. Specifically, for any $\boldsymbol{x} \in \mathbb{R}^m$, $b > 0$, and $\eta > 0$, on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$, we define

$$\boldsymbol{X}^{\eta|b}_0(\boldsymbol{x}) = \boldsymbol{x}, \qquad \boldsymbol{X}^{\eta|b}_t(\boldsymbol{x}) = \boldsymbol{X}^{\eta|b}_{t-1}(\boldsymbol{x}) + \varphi_b\Big(\eta\boldsymbol{a}\big(\boldsymbol{X}^{\eta|b}_{t-1}(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^{\eta|b}_{t-1}(\boldsymbol{x})\big)\boldsymbol{Z}_t\Big) \ \ \forall t \geq 1, \tag{2.19}$$

where the truncation operator $\varphi_\cdot(\cdot)$ is defined as

$$\varphi_c(\boldsymbol{w}) \triangleq \Big(\frac{c}{\|\boldsymbol{w}\|} \wedge 1\Big) \cdot \boldsymbol{w} \quad \forall \boldsymbol{w} \in \mathbb{R}^m, c > 0. \tag{2.20}$$

Here, $u \wedge v = \min\{u, v\}$ and $u \vee v = \max\{u, v\}$. Note that for any $\boldsymbol{w} \neq \boldsymbol{0}$, we have $\varphi_c(\boldsymbol{w}) = (c \wedge \|\boldsymbol{w}\|) \cdot \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$. In other words, the truncation operator $\varphi_b(\boldsymbol{w})$ in (2.19) maintains the direction of the vector $\boldsymbol{w}$ but rescales it to ensure that the norm would not exceed the threshold value b. For any $T,\ \eta,\ b > 0$, and $x \in \mathbb{R}$, let $\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \triangleq \{\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) :\ t \in [0,T]\}$ be the time-scaled version of $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ embedded in $\mathbb{D}[0,T]$.

For any $b,\ T \in (0,\infty)$, and $k \in \mathbb{N}$, define a mapping $\bar{h}^{(k)|b}_{[0,T]} : \mathbb{R}^m \times \mathbb{R}^{d\times k} \times \mathbb{R}^{m \times k} \times (0,T]^{k\uparrow} \to \mathbb{D}[0,T]$ as follows. Given any $\boldsymbol{x} \in \mathbb{R}^m$, $\mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\mathbf{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$, and $\boldsymbol{t} = (t_1, \cdots, t_k) \in (0,T]^{k\uparrow}$, let $\xi = \bar{h}^{(k)|b}_{[0,T]}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t})$ be the solution to

$$\xi_0 = \boldsymbol{x}; \tag{2.21}$$

$$\frac{d\xi_s}{ds} = \boldsymbol{a}(\xi_s) \quad \forall s \in [0,T],\ s \neq t_1, t_2, \cdots, t_k; \tag{2.22}$$

$$\xi_s = \xi_{s-} + \boldsymbol{v}_j + \varphi_b\big(\boldsymbol{\sigma}(\xi_{s-} + \boldsymbol{v}_j)\boldsymbol{w}_j\big) \quad \text{ if } s = t_j \text{ for some } j \in [k] \tag{2.23}$$

The mapping $\bar{h}^{(k)|b}_{[0,T]}$ can be interpreted as a truncated analog of the mapping $\bar{h}^{(k)}_{[0,T]}$ defined in (2.10)–(2.12). In other words, $\bar{h}^{(k)|b}_{[0,T]}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t})$ returns a perturbed ODE path where the impact of the jumps $\boldsymbol{w}_j$ are modulated by $\boldsymbol{\sigma}(\cdot)$ and truncated under b. Similarly, define a mapping $h^{(k)|b}_{[0,T]} : \mathbb{R}^m \times \mathbb{R}^{d\times k} \times (0,T]^{k\uparrow} \to \mathbb{D}[0,T]$ as

$$h^{(k)|b}_{[0,T]}(\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{t}) \triangleq \bar{h}^{(k)|b}_{[0,T]}\big(\boldsymbol{x}, \boldsymbol{W}, (\boldsymbol{0}, \cdots, \boldsymbol{0}), \boldsymbol{t}\big). \tag{2.24}$$

For any $b, T > 0$, $\epsilon \geq 0$, $A \subseteq \mathbb{R}^m$ and $k \in \mathbb{N}$, let

$$\mathbb{D}^{(k)|b}_A[0,T](\epsilon) \triangleq \bar{h}^{(k)|b}_{[0,T]}\Big(A \times \mathbb{R}^{m\times k} \times \big(\bar{B}_\epsilon(\boldsymbol{0})\big)^k \times (0,T]^{k\uparrow}\Big) \tag{2.25}$$

be a truncated analog of $\mathbb{D}^{(k)}_A[0,T](\epsilon)$. We adopt the convention that $\mathbb{D}^{(-1)|b}_A[0,T](\epsilon) \triangleq \emptyset$. We collect and establish useful properties of mappings $h^{(k)}_{[0,T]}$, $h^{(k)|b}_{[0,T]}$ and sets $\mathbb{D}^{(k)}_A[0,T](\epsilon)$, $\mathbb{D}^{(k)|b}_A[0,T](\epsilon)$ in Section C.

Given any $\boldsymbol{x} \in \mathbb{R}^m$, $k \in \mathbb{N}$, $b > 0$, and $T > 0$, define a Borel measure by

$$\mathbf{C}^{(k)|b}_{[0,T]}(\ \cdot\ ; \boldsymbol{x}) \triangleq \int \mathbb{I}\Big\{ h^{(k)|b}_{[0,T]}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t}) \in \ \cdot\ \Big\} \big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^k (d\mathbf{W}) \times \mathcal{L}^{k\uparrow}_T(d\boldsymbol{t}), \tag{2.26}$$

where $\mathbf{S}$ is the probability measure on the unit sphere $\mathfrak{N}_d$ characterized in Assumption 1 and ν_α is specified in (2.5). Note that given $\boldsymbol{x} \in A$, the measure $\mathbf{C}^{(k)|b}_{[0,T]}(\ \cdot\ ; \boldsymbol{x})$ is supported on $\mathbb{D}^{(k)|b}_A[0,T](0) \subseteq \mathbb{D}^{(k)}_A[0,T](0)$. Again, in case that $T = 1$, we set $\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \triangleq \boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})$, $h^{(k)|b} \triangleq h^{(k)|b}_{[0,1]}$, $\mathbb{D}^{(k)|b}_A(\epsilon) \triangleq \mathbb{D}^{(k)|b}_A[0,1](\epsilon)$, and $\mathbf{C}^{(k)|b} \triangleq \mathbf{C}^{(k)|b}_{[0,1]}$. Now, we are ready to state the main result. It is worth noticing that Assumption 3 (i.e., the boundedness of $\boldsymbol{\sigma}(\cdot)$ and $\boldsymbol{a}(\cdot)$) is not required in the truncated case. See Section 3.3 for the proof.

Theorem 2.4. *Under Assumptions 1 and 2, it holds for any $k \in \mathbb{N}$, any $b, T, \epsilon > 0$, and any compact $A \subseteq \mathbb{R}^m$ that*

$$\lambda^{-k}(\eta)\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in\ \cdot\ \big) \to \mathbf{C}^{(k)|b}_{[0,T]}(\ \cdot\ ; \boldsymbol{x}) \quad \text{in } \mathbb{M}\Big(\mathbb{D}[0,T] \setminus \mathbb{D}^{(k-1)|b}_A[0,T](\epsilon)\Big) \text{ uniformly in } \boldsymbol{x} \text{ on } A$$

as $\eta \downarrow 0$. Furthermore, for any $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ that is bounded away from $\mathbb{D}^{(k-1)|b}_A[0,T](\epsilon)$,

$$\begin{aligned} \inf_{\boldsymbol{x} \in A} \mathbf{C}^{(k)|b}_{[0,T]}\big(B^\circ; \boldsymbol{x}\big) &\leq \liminf_{\eta \downarrow 0} \frac{\inf_{\boldsymbol{x} \in A} \mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta)} \\ &\leq \limsup_{\eta \downarrow 0} \frac{\sup_{\boldsymbol{x} \in A} \mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta)} \leq \sup_{\boldsymbol{x} \in A} \mathbf{C}^{(k)|b}_{[0,T]}\big(B^-; \boldsymbol{x}\big) < \infty. \end{aligned} \tag{2.27}$$

Here, we provide a high-level description of the proof strategy for Theorems 2.3 and 2.4. Specifically, the proof of Theorem 2.4 and Theorem 2.3 consists of the following steps.

- First, we establish the asymptotic equivalence between $\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x})$ and an ODE perturbed by the k "largest" noises in $(\boldsymbol{Z}_j)_{j \leq T/\eta}$, in the sense that they admit the same limit in terms of $\mathbb{M}$-convergence as $\eta \downarrow 0$. The key technical tools are the concentration inequalities we developed in Lemma 3.3 that tightly control the fluctuations in $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ between any two "large" noises.
- Then, to complete the proof of Theorem 2.4, it suffices to study the $\mathbb{M}$-convergence of this perturbed ODE. The foundation of this analysis is the asymptotic law of the top-k largest noises in $(\boldsymbol{Z}_j)_{j \leq T/\eta}$ studied in Lemma 3.4.
- Regarding Theorem 2.3, note that for any b sufficiently large, it is highly likely that $\boldsymbol{X}^\eta_j(x)$ coincides with $\boldsymbol{X}^{\eta|b}_j(x)$ for the entire period of $j \leq T/\eta$ (that is, the truncation operator φ_b did not come into effect for a long period due to the truncation threshold $b > 0$ being large). By sending $b \to \infty$ and carefully analyzing the limits involved, we recover the sample-path large deviations for $\boldsymbol{X}^\eta_j(\boldsymbol{x})$ and prove Theorem 2.3.

See Section 3.3 for the detailed proof and the rigorous definitions of the concepts involved.

2.2.2 Catastrophe Principle

Perhaps the most important implication of the large deviations bounds is the identification of conditional distributions of the stochastic processes given the rare events of interest. This section precisely identifies the distributional limits of the conditional laws of $\boldsymbol{X}^\eta_{[0,T]}(\boldsymbol{x})$ and $\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x})$, respectively, given the rare events.

More precisely, the conditional limit theorem below follows immediately from the sample-path large deviations established above, i.e., (2.18) and (2.27), and Portmanteau Theorem. While all the results in Section 2.2.2 can be easily extended to $\mathbb{D}[0,T]$ with arbitrary $T \in (0, \infty)$, we focus on $\mathbb{D} = \mathbb{D}[0,1]$ for the sake of clarity of the presentation.

Corollary 2.5. *Let Assumptions 1 and 2 hold.*

(i) For some $b, \epsilon > 0$, $k \in \mathbb{N}$, $\boldsymbol{x} \in \mathbb{R}^m$, and measurable $B \subseteq \mathbb{D}$, suppose that B is bounded away from $\mathbb{D}^{(k-1)|b}_{\{\boldsymbol{x}\}}(\epsilon)$, and $\mathbf{C}^{(k)|b}(B^\circ;\boldsymbol{x}) = \mathbf{C}^{(k)|b}(B^-;\boldsymbol{x}) > 0$. Then

$$\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}) \in \cdot \mid \boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}) \in B\big) \Rightarrow \frac{\mathbf{C}^{(k)|b}(\,\cdot \cap B;\boldsymbol{x})}{\mathbf{C}^{(k)|b}(B;\boldsymbol{x})} \qquad \text{as } \eta \downarrow 0.$$

(ii) Furthermore, suppose that Assumption 3 holds. For some $k \in \mathbb{N}$, $\epsilon > 0$, $\boldsymbol{x} \in \mathbb{R}^m$, and measurable $B \subseteq \mathbb{D}$, suppose that B is bounded away from $\mathbb{D}^{(k-1)}_{\{\boldsymbol{x}\}}(\epsilon)$ and $\mathbf{C}^{(k)}(B^\circ;\boldsymbol{x}) = \mathbf{C}^{(k)}(B^-;\boldsymbol{x}) > 0$. Then

$$\mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x}) \in \cdot \,\big|\, \boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x}) \in B\big) \Rightarrow \frac{\mathbf{C}^{(k)}(\,\cdot \cap B;\boldsymbol{x})}{\mathbf{C}^{(k)}(B;\boldsymbol{x})} \qquad \text{as } \eta \downarrow 0.$$

Remark 4. *Note that Corollary 2.5 is a sharp characterization of* catastrophe principle *for $\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})$ and $\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x})$. By definition of measures $\mathbf{C}^{(k)|b}$, its support is on the set of paths of the form*

$$h^{(k)|b}\big(\boldsymbol{x}, (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k), (u_1, \cdots, u_k)\big),$$

where the mapping $h^{(k)|b}$ is defined in (2.21)–(2.23), and the norms $\|\boldsymbol{w}_j\|$'s are bounded from below; see, for instance, Lemma 3.5 and 3.6. This is a clear manifestation of the catastrophe principle: whenever the rare event arises, the conditional distribution resembles the nominal path (i.e., the solution of the associated ODE) perturbed by precisely k jumps. In fact, the definition of $\mathbf{C}^{(k)|b}$ also implies that the the jump sizes are Pareto (modulated by $\boldsymbol{\sigma}(\cdot)$) and the jump times are uniform, conditional on the perturbed path belonging to B. Similar interpretation applies to $\boldsymbol{X}^{\eta}_{[0,1]}(\boldsymbol{x})$ in part (ii).

2.3 Metastability Analysis

This section analyzes the metastability of $\boldsymbol{X}^{\eta}_j(\boldsymbol{x})$ and $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$. Section 2.3.1 establishes the scaling limits of their exit times. Section 2.3.2 introduces a framework that facilitates such analysis for general Markov chains.

2.3.1 First Exit Times and Locations

In this section, we analyze the first exit times and locations of $\boldsymbol{X}^{\eta}_j(\boldsymbol{x})$ and $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ from an attraction field of some potential with a unique local minimum at the origin. Specifically, throughout Section 2.3.1, we fix an open set $I \subset \mathbb{R}^m$ that is bounded and contains the origin, i.e., $\sup_{\boldsymbol{x} \in I}\|\boldsymbol{x}\| < \infty$ and $\mathbf{0} \in I$. Let $\boldsymbol{y}_t(\boldsymbol{x})$ be the solution of ODE

$$\boldsymbol{y}_0(\boldsymbol{x}) = \boldsymbol{x}, \qquad \frac{d\boldsymbol{y}_t(\boldsymbol{x})}{dt} = \boldsymbol{a}\big(\boldsymbol{y}_t(\boldsymbol{x})\big) \ \ \forall t \geq 0. \tag{2.28}$$

We impose the following assumption on the gradient field $\boldsymbol{a}: \mathbb{R}^m \to \mathbb{R}^m$.

Assumption 4. $\boldsymbol{a}(\mathbf{0}) = \mathbf{0}$. *For all $\boldsymbol{x} \in I \setminus \{\mathbf{0}\}$,*

$$\boldsymbol{y}_t(\boldsymbol{x}) \in I \ \forall t \geq 0, \qquad \lim_{t\to\infty} \boldsymbol{y}_t(\boldsymbol{x}) = \mathbf{0}.$$

Besides, it holds for all $\epsilon > 0$ small enough that $\boldsymbol{a}(\boldsymbol{x})\boldsymbol{x} < 0 \ \forall \boldsymbol{x} \in \bar{B}_\epsilon(\mathbf{0}) \setminus \{\mathbf{0}\}$.

An immediate consequence of the condition $\lim_{t\to\infty}\boldsymbol{y}_t(\boldsymbol{x}) = \mathbf{0} \ \forall \boldsymbol{x} \in I \setminus \{\mathbf{0}\}$ is that $\boldsymbol{a}(\boldsymbol{x}) \neq \mathbf{0}$ for all $\boldsymbol{x} \in I \setminus \{\mathbf{0}\}$. Of particular interest is the case where $\boldsymbol{a}(\cdot) = -\nabla U(\cdot)$ for some potential $U \in \mathcal{C}^1(\mathbb{R}^m)$ that has a unique local minimum at $\boldsymbol{x} = 0$ over the domain I. In particular, Assumption 4 holds if

U is also locally $\mathcal{C}^2$ at the origin, and the Hessian of $U(\cdot)$ at the origin $\boldsymbol{x} = \mathbf{0}$ is positive definite. We note that Assumption 4 is a standard one in existing literature; see e.g. [51, 34].

Define

$$\tau^{\eta}(\boldsymbol{x}) \triangleq \min\big\{j \geq 0 : \ \boldsymbol{X}_j^{\eta}(\boldsymbol{x}) \notin I\big\}, \qquad \tau^{\eta|b}(\boldsymbol{x}) \triangleq \min\big\{j \geq 0 : \ \boldsymbol{X}_j^{\eta|b}(\boldsymbol{x}) \notin I\big\}, \tag{2.29}$$

as the first exit time of $\boldsymbol{X}_j^{\eta}(\boldsymbol{x})$ and $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ from I, respectively. To facilitate the presentation of the main results, we introduce a few concepts. Define the mapping $\bar{g}^{(k)|b} : \mathbb{R}^m \times \mathbb{R}^{d\times k} \times \mathbb{R}^{m \times k} \times (0,\infty)^{k\uparrow} \to \mathbb{R}^m$ as the location of the (perturbed) ODE with k jumps at the last jump time:

$$\bar{g}^{(k)|b}\big(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, (t_1, \cdots, t_k)\big) \triangleq \bar{h}^{(k)|b}_{[0,t_k+1]}\Big(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, (t_1, \cdots, t_k)\Big)(t_k), \tag{2.30}$$

where $\bar{h}^{(k)|b}_{[0,T]}$ is the perturbed ODE mapping defined in (2.21)–(2.23). Besides, define $\breve{g}^{(k)|b} : \mathbb{R}^m \times \mathbb{R}^{d\times k} \times (0,\infty)^{k\uparrow} \to \mathbb{R}^m$ by

$$\breve{g}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t}) \triangleq \bar{g}^{(k)|b}\big(\boldsymbol{x}, \mathbf{W}, (\mathbf{0}, \cdots, \mathbf{0}), \boldsymbol{t}\big) = h^{(k)|b}_{[0,t_k+1]}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t})(t_k), \tag{2.31}$$

where $\boldsymbol{t} = (t_1, \ldots, t_k) \in (0,\infty)^{k\uparrow}$, and the mapping $h^{(k)|b}_{[0,T]}$ is defined in (2.24). For $k = 0$, we adopt the convention that $\bar{g}^{(0)|b}(\boldsymbol{x}) = \boldsymbol{x}$.

With the mapping $\bar{g}^{(k)|b}$ defined, we are able to introduce (for any $k \geq 1$, $b > 0$, and $\epsilon \geq 0$)

$$\mathcal{G}^{(k)|b}(\epsilon) \triangleq \Bigg\{\bar{g}^{(k-1)|b}\Big(\boldsymbol{v}_1 + \varphi_b\big(\boldsymbol{\sigma}(\boldsymbol{v}_1)\boldsymbol{w}_1\big), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_k), (\boldsymbol{v}_2, \cdots, \boldsymbol{v}_k), \boldsymbol{t}\Big) : \\ \mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}, \mathbf{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \Big(\bar{B}_\epsilon(\mathbf{0})\Big)^k, \boldsymbol{t} \in (0,\infty)^{k\uparrow}\Bigg\} \tag{2.32}$$

as the set covered by the k^{th} jump of along ODE path (with ϵ perburbation before each jump) initialized at the origin, with each jump modulated by $\boldsymbol{\sigma}(\cdot)$ and truncated under b. Here, the truncation operator φ_b is defined in (2.20), and $\bar{B}_r(\mathbf{0})$ is the closed ball with radius r centered at the origin. For $\epsilon = 0$, we write

$$\mathcal{G}^{(k)|b} \triangleq \mathcal{G}^{(k)|b}(0) = \Big\{\breve{g}^{(k-1)|b}\Big(\varphi_b\big(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1\big), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_k), \boldsymbol{t}\Big) : \ \mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}, \boldsymbol{t} \in (0,\infty)^{k\uparrow}\Big\}.$$

Furthermore, as a convention for the case with $k = 0$, we set

$$\mathcal{G}^{(0)|b}(\epsilon) \triangleq \bar{B}_\epsilon(\mathbf{0}).$$

We note that $\mathcal{G}^{(k)|b}(\epsilon)$ is monotone in ϵ, k, and b, in the sense that $\mathcal{G}^{(k)|b}(\epsilon) \subseteq \mathcal{G}^{(k)|b}(\epsilon')$ for all $0 \leq \epsilon \leq \epsilon'$, $\mathcal{G}^{(k)|b}(\epsilon) \subseteq \mathcal{G}^{(k+1)|b}(\epsilon)$, and $\mathcal{G}^{(k)|b}(\epsilon) \subseteq \mathcal{G}^{(k)|b'}(\epsilon)$ for all $0 < b \leq b'$.

The intuition behind our metastability analysis (in particular, Theorem 2.6) is as follows. The characterization of the k-jump-coverage sets of form $\mathcal{G}^{(k)|b}$ reveals that, due to the truncation of $\varphi_b(\cdot)$, the reachable space expands as more jumps are added to the ODE path. Regarding the asymptotics of the first exit times $\tau^{\eta|b}(\boldsymbol{x})$, this results in an intriguing phase transition for the law $\tau^{\eta|b}(\boldsymbol{x})$ (as $\eta \downarrow 0$) in terms of the minimum number of jumps required for exit. More precisely, let

$$\mathcal{J}_b^I \triangleq \min\big\{k \geq 1 : \ \mathcal{G}^{(k)|b} \cap I \neq \emptyset\big\} \tag{2.33}$$

be the smallest k such that, under truncation at level b, the k-jump-coverage sets can reach outside the attraction field I. Theorem 2.6 reveals a discrete hierarchy that the asymptotics of $\tau^{\eta|b}(\boldsymbol{x})$ does not vary with the truncation level b in a continuous fashion; instead, the order of the first exit time $\tau^{\eta|b}(\boldsymbol{x})$ and the limiting law of the exit location $\boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \notin I$ are dictated by this "discretized width" metric $\mathcal{J}_b^I$ of the domain I, relative to the truncation threshold b. Here, the limiting laws will be characterized through measures

$$\check{\mathbf{C}}^{(k)|b}(\ \cdot\ ) \triangleq \int \mathbb{I}\Big\{\check{g}^{(k-1)|b}\Big(\varphi_b\big(\boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{w}_1\big), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_k), \boldsymbol{t}\Big) \in\ \cdot\ \Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^k (d\mathbf{W}) \times \mathcal{L}_\infty^{k-1\uparrow}(d\boldsymbol{t}), \tag{2.34}$$

where $\alpha > 1$ is the heavy-tail index in Assumption 1, $\mathbf{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^k$ is the k-fold of $(\nu_\alpha \times \mathbf{S})\circ\Phi$ defined in (2.15), and $\mathcal{L}_\infty^{k\uparrow}$ is the Lebesgue measure restricted on $\{(t_1, \cdots, t_k) \in (0,\infty)^k :\ 0 < t_1 < t_2 < \cdots < t_k\}$. Section D collects useful properties of the mapping $\check{g}^{(k)|b}$ and the measure $\check{\mathbf{C}}^{(k)|b}$.

Recall that $H(\cdot) = \mathbf{P}(\|\boldsymbol{Z}_1\| > \cdot)$, $\lambda(\eta) = \eta^{-1}H(\eta^{-1})$, and for any $k \geq 1$ we write $\lambda^k(\eta) = \big(\lambda(\eta)\big)^k$. Recall that $I_\epsilon = \{\boldsymbol{y} :\ \|\boldsymbol{x} - \boldsymbol{y}\| < \epsilon \implies \boldsymbol{x} \in I\}$ is the ϵ-shrinkage of I. As the main result of this section, Theorem 2.6 provides sharp asymptotics for the joint law of first exit times and exit locations of $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ and $\boldsymbol{X}_j^{\eta}(\boldsymbol{x})$. The results are obtained through a general machinery we develop in Section 2.3.2, and here we provide a brief outline of the proof. In Section 2.3.2, we introduce the notion of asymptotic atoms, where a Markov process recurrently visits and almost regenerates upon each visit. This recurrence and almost regeneration is characterized through asymptotic limits that are uniform over the entire region of the asymptotic atom. First, in Theorem 2.9 we show that once the existence of asymptotic atoms is verified, the precise asymptotic limits for the joint law of the exit times and exit locations, such as those stated in Theorem 2.6, follow immediately. More importantly, the uniform $\mathbb{M}$-convergence and uniform sample path large deviations developed earlier prove to be powerful tools for the identification of asymptotic atoms, particularly in the truncated cases of $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$, where the truncation of heavy-tailed noises results in a much more complex description of exit paths. The detailed proof of Theorem 2.6 is provided in Section 4.2.

Theorem 2.6. ***(First Exit Times and Locations: Truncated Case)*** *Let Assumptions 1, 2, and 4 hold. Let $b > 0$ such that $\mathcal{J}_b^I < \infty$. Suppose that I^c is bounded away from $\mathcal{G}^{(\mathcal{J}_b^I - 1)|b}(\epsilon)$ for some (and hence all) $\epsilon > 0$ small enough, and $\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\partial I) = 0$. Then $C_b^I \triangleq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(I^c) < \infty$. Furthermore, if $C_b^I \in (0,\infty)$, then for any $\epsilon > 0$, $t \geq 0$, and measurable set $B \subseteq I^c$,*

$$\limsup_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in I_\epsilon} \mathbf{P}\bigg(C_b^I \eta \cdot \lambda^{\mathcal{J}_b^I}(\eta)\tau^{\eta|b}(\boldsymbol{x}) > t;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg) \leq \frac{\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(B^-)}{C_b^I} \cdot \exp(-t),$$

$$\liminf_{\eta\downarrow 0} \inf_{\boldsymbol{x}\in I_\epsilon} \mathbf{P}\bigg(C_b^I \eta \cdot \lambda^{\mathcal{J}_b^I}(\eta)\tau^{\eta|b}(\boldsymbol{x}) > t;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg) \geq \frac{\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(B^\circ)}{C_b^I} \cdot \exp(-t).$$

Otherwise, we must have $C_b^I = 0$, and

$$\limsup_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in I_\epsilon} \mathbf{P}\bigg(\eta \cdot \lambda^{\mathcal{J}_b^I}(\eta)\tau^{\eta|b}(\boldsymbol{x}) \leq t\bigg) = 0 \qquad \forall \epsilon > 0,\ t \geq 0.$$

In summary, by developing the machinery of uniform $\mathbb{M}$-convergence and asymptotic atoms, we provide a general framework that connects large deviations and first exit analysis. Applying this framework for the truncated heavy-tailed dynamics $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$, we reveal an intriguing phase transition in terms of the truncation threshold b, where the discretized width $\mathcal{J}_b^I$ dictates the order of the first exit times and the limiting law of first exit locations. To conclude this section, we note that the first exit analysis for untruncated heavy-tailed dynamics (see e.g. [32, 33, 51, 34] for similar results in the existing literature) would follow immediately from Theorem 2.6. Specifically, let

$$\check{\mathbf{C}}(\ \cdot\ ) \triangleq \int \mathbb{I}\Big\{\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w} \in\ \cdot\ \Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)(d\boldsymbol{w}). \tag{2.35}$$

The exit times and locations for the untruncated dynamics $\boldsymbol{X}^{\eta}_j(\boldsymbol{x})$ then follows from the result for $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ by sending b to ∞, and the limiting laws of the exit location $\boldsymbol{X}^{\eta}_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x})$ is characterized by $\check{\mathbf{C}}(\ \cdot\ )$, as presented in Corollary 2.7. The proof is straightforward and we collect it in Section D for the sake of completeness.

Corollary 2.7. ***(First Exit Times and Locations: Untruncated Case)*** *Let Assumptions 1, 2, and 4 hold. Suppose that $\check{\mathbf{C}}(\partial I) = 0$. Then $C^I_\infty \triangleq \check{\mathbf{C}}(I^c) < \infty$. Furthermore, if $C^I_\infty > 0$, then for any $t \geq 0$ and measurable set $B \subseteq I^c$,*

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(C^I_\infty\eta\cdot\lambda(\eta)\tau^\eta(\boldsymbol{x}) > t;\ \boldsymbol{X}^\eta_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x})\in B\bigg) \leq \frac{\check{\mathbf{C}}(B^-)}{C^I_\infty}\cdot\exp(-t),$$
$$\liminf_{\eta\downarrow 0}\inf_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(C^I_\infty\eta\cdot\lambda(\eta)\tau^\eta(\boldsymbol{x}) > t;\ \boldsymbol{X}^\eta_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x})\in B\bigg) \geq \frac{\check{\mathbf{C}}(B^\circ)}{C^I_\infty}\cdot\exp(-t).$$

Otherwise, we must have $C^I_\infty = 0$, and

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(\eta\cdot\lambda(\eta)\tau^\eta(\boldsymbol{x}) \leq t\bigg) = 0 \qquad \forall\epsilon > 0,\ t\geq 0.$$

2.3.2 General Framework: Asymptotic Atoms

This section proposes a general framework that enables sharp characterization of exit times and exit locations of Markov chains. The new heavy-tailed large deviations formulation introduced in Section 2.2 is conducive to this framework.

Consider a general metric space $(\mathbb{S},\boldsymbol{d})$ and a family of $\mathbb{S}$-valued Markov chains $\big\{\{V^\eta_j(x): j\geq 0\}: \eta > 0\big\}$ parameterized by η, where $x\in\mathbb{S}$ denotes the initial state and j denotes the time index. We use $\boldsymbol{V}^\eta_{[0,T]}(x) \triangleq \{V^\eta_{\lfloor t/\eta\rfloor}(x): t\in[0,T]\}$ to denote the scaled version of $\{V^\eta_j(x): j\geq 0\}$ as a $\mathbb{D}[0,T]$-valued random element. For a given set E, let $\tau^\eta_E(x) \triangleq \min\{j\geq 0: V^\eta_j(x)\in E\}$ denote $\{V^\eta_j(s): j\geq 0\}$'s first hitting time of E. We consider an asymptotic domain of attraction $I\subseteq\mathbb{S}$, within which $\boldsymbol{V}^\eta_{[0,T]}(x)$ typically (i.e., as $\eta\downarrow 0$) stays within I throughout any fixed time horizon $[0,T]$ as far as the initial state x is in I. However, if one considers an infinite time horizon, $V^\eta_\cdot(x)$ is typically bound to escape I eventually due to the stochasticity. The goal of this section is to establish an asymptotic limit of the joint distribution of the exit time $\tau^\eta_{I^c}(x)$ and the exit location $V^\eta_{\tau^\eta_{I^c}(x)}(x)$. Throughout this section, we will denote $V^\eta_{\tau^\eta_{I(\epsilon)^c}(x)}(x)$ and $V^\eta_{\tau^\eta_{I^c}(x)}(x)$ with $V^\eta_{\tau_\epsilon}(x)$ and $V^\eta_\tau(x)$, respectively, for notation simplicity.

We introduce the notion of asymptotic atoms to facilitate the analyses. Let $\{I(\epsilon)\subseteq I: \epsilon > 0\}$ and $\{A(\epsilon)\subseteq\mathbb{S}: \epsilon > 0\}$ be collections of subsets of I such that $\bigcup_{\epsilon>0} I(\epsilon) = I$ and $\bigcap_{\epsilon>0} A(\epsilon)\neq\emptyset$. Let $C(\cdot)$ is a Borel measure on $\mathbb{S}\setminus I$ satisfying $C(\partial I) = 0$ that characterizes the (asymptotics limit of the) exit location of $V^\eta_\cdot(x)$. Specifically, we consider two different cases for the location measure $C(\cdot)$:

- *(i)* $C(I^c)\in(0,\infty)$: by incorporating the normalizing constant $C(I^c)$ into the scale function $\gamma(\eta)$, we can assume w.l.o.g. that $C(\cdot)$ **is a probability measure**, and $C(B)$ dictates the limiting probability that $\mathbf{P}(V^\eta_\tau(x)\in B)$ as shown in Theorem 2.9;
- *(ii)* $C(I^c) = 0$: as a result, $C(B) = 0$ for any Borel set $B\subseteq I^c$, and it is equivalent to stating that $C(\cdot)$ **is trivially zero**.

Definition 2.8. *$\big\{\{V^\eta_j(x): j\geq 0\}: \eta > 0\big\}$ possesses an asymptotic atom $\{A(\epsilon)\subseteq\mathbb{S}: \epsilon > 0\}$ associated with the domain I, location measure $C(\cdot)$, scale $\gamma: (0,\infty)\to(0,\infty)$, and covering $\{I(\epsilon)\subseteq I: \epsilon > 0\}$ if the following holds: For each measurable set $B\subseteq\mathbb{S}$, there exist $\delta_B: (0,\infty)\times(0,\infty)\to(0,\infty)$, $\epsilon_B > 0$, and $T_B: (0,\infty)\to(0,\infty)$ such that*

$$C(B^\circ) - \delta_B(\epsilon,T) \leq \liminf_{\eta\downarrow 0}\frac{\inf_{x\in A(\epsilon)}\mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(x)\leq T/\eta;\ V^\eta_{\tau_\epsilon}(x)\in B\big)}{\gamma(\eta)T/\eta} \tag{2.36}$$

$$\leq \limsup_{\eta \downarrow 0} \frac{\sup_{x \in A(\epsilon)} \mathbf{P}\big(\tau^{\eta}_{I(\epsilon)^c}(x) \leq T/\eta;\ V^{\eta}_{\tau_\epsilon}(x) \in B\big)}{\gamma(\eta)T/\eta} \leq C(B^-) + \delta_B(\epsilon, T) \quad (2.37)$$

$$\limsup_{\eta \downarrow 0} \frac{\sup_{x \in I(\epsilon)} \mathbf{P}\big(\tau^{\eta}_{(I(\epsilon)\setminus A(\epsilon))^c}(x) > T/\eta\big)}{\gamma(\eta)T/\eta} = 0 \quad (2.38)$$

$$\liminf_{\eta \downarrow 0} \inf_{x \in I(\epsilon)} \mathbf{P}\big(\tau^{\eta}_{A(\epsilon)}(x) \leq T/\eta\big) = 1 \quad (2.39)$$

for any $\epsilon \leq \epsilon_B$ *and* $T \geq T_B(\epsilon)$*, where* $\gamma(\eta)/\eta \to 0$ *as* $\eta \downarrow 0$ *and* δ_B*'s are such that*

$$\lim_{\epsilon \downarrow 0} \lim_{T \to \infty} \delta_B(\epsilon, T) = 0.$$

To see how Definition 2.8 asymptotically characterize the atoms in $V^{\eta}_{\cdot}(x)$ for the first exit analysis from domain I, note that the condition (2.39) requires the process to efficiently return to the asymptotic atoms $A(\epsilon)$. The conditions (2.36) and (2.37) then state that, upon hitting the asymptotic atoms $A(\epsilon)$, the process almost regenerates in terms of the law of the exit time $\tau^{\eta}_{I(\epsilon)^c}(x)$ and exit locations $V^{\eta}_{\tau_\epsilon}(x)$. Furthermore, the condition (2.38) prevents the process $V^{\eta}_{\cdot}(x)$ from spending a long time without either returning to the asymptotic atoms $A(\epsilon)$ or exiting from $I(\epsilon)$, which covers the domain I as ϵ tends to 0.

The existence of an asymptotic atom is a sufficient condition for characterization of exit time and location asymptotics as in Theorem 2.6. To minimize repetition, we refer to the existence of an asymptotic atom—with specific domain, location measure, scale, and covering—Condition 1 throughout the paper.

Condition 1. *A family* $\big\{\{V^{\eta}_j(x) : j \geq 0\} : \eta > 0\big\}$ *of Markov chains possesses an asymptotic atom* $\{A(\epsilon) \subseteq \mathbb{S} : \epsilon > 0\}$ *associated with the domain* I*, location measure* $C(\cdot)$*, scale* $\gamma : (0,\infty) \to (0,\infty)$*, and covering* $\{I(\epsilon) \subseteq I : \epsilon > 0\}$*.*

Recall that, right before Definition 2.8, we state that for the location measure $C(\cdot)$ we consider two cases that (i) $C(I^c) = 1$ (more generally, $C(\cdot)$ is a finite measure), and (ii) $C(I^c) = 0$. The following theorem is the key result of this section. See Section 4.1 for the proof of the theorem.

Theorem 2.9. *If Condition 1 holds, then the first exit time* $\tau^{\eta}_{I^c}(x)$ *scales as* $1/\gamma(\eta)$*, and the distribution of the location* $V^{\eta}_{\tau}(x)$ *at the first exit time converges to* $C(\cdot)$*. Moreover, the convergence is uniform over* $I(\epsilon)$ *for any* $\epsilon > 0$*. That is,*

(i) If $C(I^c) = 1$*, then for each* $\epsilon > 0$*, measurable* $B \subseteq I^c$*, and* $t \geq 0$*,*

$$\begin{aligned} C(B^\circ) \cdot e^{-t} &\leq \liminf_{\eta \downarrow 0} \inf_{x \in I(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^{\eta}_{I^c}(x) > t,\, V^{\eta}_{\tau}(x) \in B\big) \\ &\leq \limsup_{\eta \downarrow 0} \sup_{x \in I(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^{\eta}_{I^c}(x) > t,\, V^{\eta}_{\tau}(x) \in B\big) \leq C(B^-) \cdot e^{-t}; \end{aligned}$$

(ii) If $C(I^c) = 0$*, then for each* $\epsilon, t > 0$*,*

$$\lim_{\eta \downarrow 0} \sup_{x \in I(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^{\eta}_{I^c}(x) > t\big) = 0.$$

3 Uniform $\mathbb{M}$-Convergence and Sample Path Large Deviations

Here, we collect the proofs for Sections 2.1 and 2.2. Specifically, Section 3.1 provides the proof of Theorem 2.2, i.e., the Portmanteau theorem for the uniform $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$-convergence. Section 3.2 further develops a set of technical tools, which will then be applied to establish the sample-path large deviations results (i.e., Theorems 2.3 and 2.4) in Section 3.3.

3.1 Proof of Theorem 2.2

Proof of Theorem 2.2. **Proof of** $(i) \Rightarrow (ii)$. It follows directly from Definition 2.1.

Proof of $(ii) \Rightarrow (iii)$. We consider a proof by contradiction. Suppose that the upper bound $\limsup_{\eta \downarrow 0} \sup_{\theta \in \Theta} \mu_\theta^\eta(F) - \mu_\theta(F^\epsilon) \leq 0$ does not hold for some closed F bounded away from $\mathbb{C}$ and some $\epsilon > 0$. Then there exist a sequence $\eta_n \downarrow 0$, a sequence $\theta_n \in \Theta$, and some $\delta > 0$ such that $\mu_{\theta_n}^{\eta_n}(F) - \mu_{\theta_n}(F^\epsilon) > \delta \ \forall n \geq 1$. Now, we make two observations. First, using Urysohn's lemma (see, e.g., lemma 2.3 of [42]), one can identify some $f \in \mathcal{C}(\mathbb{S} \setminus \mathbb{C})$, which is also uniformly continuous on $\mathbb{S}$, such that $\mathbb{I}_F \leq f \leq \mathbb{I}_{F^\epsilon}$. This leads to the bound $\mu_{\theta_n}^{\eta_n}(F) - \mu_{\theta_n}(F^\epsilon) \leq \mu_{\theta_n}^{\eta_n}(f) - \mu_{\theta_n}(f)$ for each n. Secondly, from statement (ii) we get $\lim_{n\to\infty} \left|\mu_{\theta_n}^{\eta_n}(f) - \mu_{\theta_n}(f)\right| = 0$. In summary, we yield the contradiction

$$\limsup_{n\to\infty} \mu_{\theta_n}^{\eta_n}(F) - \mu_{\theta_n}(F^\epsilon) \leq \limsup_{n\to\infty} \mu_{\theta_n}^{\eta_n}(f) - \mu_{\theta_n}(f) \leq \lim_{n\to\infty} \left|\mu_{\theta_n}^{\eta_n}(f) - \mu_{\theta_n}(f)\right| = 0.$$

Analogously, if the claim $\liminf_{\eta\downarrow 0} \inf_{\theta\in\Theta} \mu_\theta^\eta(G) - \mu_\theta(G^\epsilon) \geq 0$, supposedly, does not hold for some open G bounded away from $\mathbb{C}$ and some $\epsilon > 0$, then we can yield a similar contradiction by applying Urysohn's lemma and constructing some uniformly continuous $g \in \mathcal{C}(\mathbb{S} \setminus \mathbb{C})$ such that $\mathbb{I}_{G_\epsilon} \leq g \leq \mathbb{I}_G$. This concludes the proof of $(ii) \Rightarrow (iii)$.

Proof of $(iii) \Rightarrow (i)$**.** Again, we proceed with a proof by contradiction. Suppose that the claim $\lim_{\eta\downarrow 0} \sup_{\theta\in\Theta} \left|\mu_\theta^\eta(g) - \mu_\theta(g)\right| = 0$ does not hold for some $g \in \mathcal{C}(\mathbb{S} \setminus \mathbb{C})$. Then, there exist some sequences $\eta_n \downarrow 0$, $\theta_n \in \Theta$ and some $\delta > 0$ such that

$$|\mu_{\theta_n}^{\eta_n}(g) - \mu_{\theta_n}(g)| > \delta \qquad \forall n \geq 1. \tag{3.1}$$

To proceed, we arbitrarily pick some closed $F \subseteq \mathbb{S}$ that is bounded away from $\mathbb{C}$ and some open $G \subseteq \mathbb{S}$ that is bounded away from $\mathbb{C}$. First, using claims in (iii), we get $\limsup_{n\to\infty} \mu_{\theta_n}^{\eta_n}(F) - \mu_{\theta_n}(F^\epsilon) \leq 0$ and $\liminf_{n\to\infty} \mu_{\theta_n}^{\eta_n}(G) - \mu_{\theta_n}(G_\epsilon) \geq 0$ for any $\epsilon > 0$. Next, due to condition (2.1), by picking a sub-sequence of θ_n if necessary we can find some μ_{θ^*} such that $\lim_{n\to\infty} \left|\mu_{\theta_n}(f) - \mu_{\theta^*}(f)\right| = 0$ for all $f \in \mathcal{C}(\mathbb{S} \setminus \mathbb{C})$. By Portmanteau theorem for standard $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$-convergence (see theorem 2.1 of [42]), we yield $\limsup_{n\to\infty} \mu_{\theta_n}(F^\epsilon) \leq \mu_{\theta^*}(F^\epsilon)$ and $\liminf_{n\to\infty} \mu_{\theta_n}(G_\epsilon) \geq \mu_{\theta^*}(G_\epsilon)$. In summary, for any $\epsilon > 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \mu_{\theta_n}^{\eta_n}(F) &\leq \limsup_{n\to\infty} \mu_{\theta_n}(F^\epsilon) + \limsup_{n\to\infty} \mu_{\theta_n}^{\eta_n}(F) - \mu_{\theta_n}(F^\epsilon) \leq \mu_{\theta^*}(F^\epsilon), \\
\liminf_{n\to\infty} \mu_{\theta_n}^{\eta_n}(G) &\geq \liminf_{n\to\infty} \mu_{\theta_n}(G_\epsilon) + \liminf_{n\to\infty} \mu_{\theta_n}^{\eta_n}(G) - \mu_{\theta_n}(G_\epsilon) \geq \mu_{\theta^*}(G_\epsilon).
\end{aligned}$$

Lastly, note that $\lim_{\epsilon\downarrow 0} \mu_{\theta^*}(F^\epsilon) = \mu_{\theta^*}(F)$ and $\lim_{\epsilon\downarrow 0} \mu_{\theta^*}(G_\epsilon) = \mu_{\theta^*}(G)$ due to continuity of measures and $\bigcap_{\epsilon>0} F^\epsilon = F$, $\bigcup_{\epsilon>0} G_\epsilon = G$. This allows us to apply Portmanteau theorem for standard $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence again and obtain $\lim_{n\to\infty} \left|\mu_{\theta_n}^{\eta_n}(g) - \mu_{\theta^*}(g)\right| = 0$ for the $g \in \mathcal{C}(\mathbb{S}\setminus\mathbb{C})$ fixed in (3.1). However, recall that we have already obtained $\lim_{n\to\infty} \left|\mu_{\theta_n}(g) - \mu_{\theta^*}(g)\right| = 0$ using assumption (2.1). We now arrive at the contradiction

$$\lim_{n\to\infty} \left|\mu_{\theta_n}^{\eta_n}(g) - \mu_{\theta_n}(g)\right| \leq \lim_{n\to\infty} \left|\mu_{\theta_n}^{\eta_n}(g) - \mu_{\theta^*}(g)\right| + \lim_{n\to\infty} \left|\mu_{\theta^*}(g) - \mu_{\theta_n}(g)\right| = 0$$

and conclude the proof of $(iv) \Rightarrow (i)$.

Proof of $(i) \Rightarrow (iv)$**.** Due to the equivalence of (i), (ii), and (iii), it only remains to show that $(i) \Rightarrow (iv)$. Suppose, for the sake of contradiction, that the claim $\limsup_{\eta\downarrow 0} \sup_{\theta\in\Theta} \mu_\theta^\eta(F) \leq \sup_{\theta\in\Theta} \mu_\theta(F)$ in (iv) does not hold for some closed F bounded away from $\mathbb{C}$. Then we can find sequences $\eta_n \downarrow 0$, $\theta_n \in \Theta$ and some $\delta > 0$ such that $\mu_{\theta_n}^{\eta_n}(F) > \sup_{\theta\in\Theta} \mu_\theta(F) + \delta \ \forall n \geq 1$. Next, due to the assumption (2.1), by picking a sub-sequence of θ_n if necessary we can find some μ_{θ^*} such that $\lim_{n\to\infty} \left|\mu_{\theta_n}(f) - \mu_{\theta^*}(f)\right| = 0$

for all $f \in \mathcal{C}(\mathbb{S} \setminus \mathbb{C})$. Meanwhile, (i) implies that $\lim_{n\to\infty} \left|\mu^{\eta_n}_{\theta_n}(f) - \mu_{\theta_n}(f)\right| = 0$ for all $f \in \mathcal{C}(\mathbb{S} \setminus \mathbb{C})$. Therefore,

$$\lim_{n\to\infty} \left|\mu^{\eta_n}_{\theta_n}(f) - \mu_{\theta^*}(f)\right| \leq \lim_{n\to\infty} \left|\mu^{\eta_n}_{\theta_n}(f) - \mu_{\theta_n}(f)\right| + \lim_{n\to\infty} \left|\mu_{\theta_n}(f) - \mu_{\theta^*}(f)\right| = 0$$

for all $f \in \mathcal{C}(\mathbb{S} \setminus \mathbb{C})$. By Portmanteau theorem for standard $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$-convergence, we yield the contradiction $\limsup_{n\to\infty} \mu^{\eta_n}_{\theta_n}(F) \leq \mu_{\theta^*}(F) \leq \sup_{\theta\in\Theta} \mu_\theta(F)$. In summary, we have established the claim $\limsup_{\eta\downarrow 0} \sup_{\theta\in\Theta} \mu^\eta_\theta(F) \leq \sup_{\theta\in\Theta} \mu_\theta(F)$ for all closed F bounded away from $\mathbb{C}$. The same approach can also be applied to show $\liminf_{\eta\downarrow 0} \inf_{\theta\in\Theta} \mu^\eta_\theta(G) \geq \inf_{\theta\in\Theta} \mu_\theta(G)$ for all open G bounded away from $\mathbb{C}$. This concludes the proof. □

To facilitate the application of Theorem 2.2, we introduce the concept of asymptotic equivalence between two families of random objects. Specifically, we consider a generalized version of asymptotic equivalence over $\mathbb{S} \setminus \mathbb{C}$, which is equivalent to definition 2.9 in [19].

Definition 3.1. *Let X_n and Y_n be random elements taking values in a complete separable metric space $(\mathbb{S}, \boldsymbol{d})$. Let ϵ_n be a sequence of positive real numbers. Let $\mathbb{C} \subseteq \mathbb{S}$ be Borel measurable. X_n is said to be* ***asymptotically equivalent to Y_n in $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$ with respect to ϵ_n*** *if for any $\Delta > 0$ and any $B \in \mathscr{S}_\mathbb{S}$ bounded away from $\mathbb{C}$,*

$$\lim_{n\to\infty} \epsilon_n^{-1} \mathbf{P}\Big(\boldsymbol{d}\big(X_n, Y_n\big)\mathbb{I}\big(X_n \in B \text{ or } Y_n \in B\big) > \Delta\Big) = 0.$$

In case that $\mathbb{C} = \emptyset$, Definition 3.1 simply degenerates to the standard notion of asymptotic equivalence; see Definition 1 of [56]. The following lemma demonstrates the application of the asymptotic equivalence and is plays an important role in our analysis below.

Lemma 3.2 (Lemma 2.11 of [19])**.** *Let X_n and Y_n be random elements taking values in a complete separable metric space $(\mathbb{S}, \boldsymbol{d})$ and let $\mathbb{C} \subseteq \mathbb{S}$ be Borel measurable. Suppose that $\epsilon_n^{-1}\mathbf{P}(X_n \in \cdot) \to \mu(\cdot)$ in $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$ for some sequence of positive real numbers ϵ_n. If X_n is asymptotically equivalent to Y_n when bounded away from $\mathbb{C}$ with respect to ϵ_n, then $\epsilon_n^{-1}\mathbf{P}(Y_n \in \cdot) \to \mu(\cdot)$ in $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$.*

3.2 Technical Lemmas for Theorems 2.3 and 2.4

Our analysis hinges on the separation of *large noises* among $(\boldsymbol{Z}_j)_{j\geq 1}$ from the rest, and we pay special attention to $\boldsymbol{Z}_j$'s with norm large enough such that $\eta \left\|\boldsymbol{Z}_j\right\|$ exceed some prefixed threshold level $\delta > 0$. To be more concrete, for any $i \geq 1$ and $\eta, \delta > 0$, define the i^{th} arrival time of "large noises" and its size as

$$\tau^{>\delta}_i(\eta) \triangleq \min\{n > \tau^{>\delta}_{i-1}(\eta) : \ \eta \left\|\boldsymbol{Z}_j\right\| > \delta\}, \quad \tau^{>\delta}_0(\eta) = 0 \tag{3.2}$$

$$\boldsymbol{W}^{>\delta}_i(\eta) \triangleq \boldsymbol{Z}_{\tau^{>\delta}_i(\eta)}. \tag{3.3}$$

For any $\delta > 0$ and $k = 1, 2, \cdots$, note that

$$\begin{aligned}\mathbf{P}\Big(\tau^{>\delta}_k(\eta) \leq \lfloor 1/\eta \rfloor\Big) &\leq \mathbf{P}\Big(\tau^{>\delta}_j(\eta) - \tau^{>\delta}_{j-1}(\eta) \leq \lfloor 1/\eta \rfloor \ \ \forall j \in [k]\Big) \\ &= \Big[\sum_{i=1}^{\lfloor 1/\eta \rfloor} \big(1 - H(\delta/\eta)\big)^{i-1} H(\delta/\eta)\Big]^k \leq \Big[\sum_{i=1}^{\lfloor 1/\eta \rfloor} H(\delta/\eta)\Big]^k \\ &\leq \Big[1/\eta \cdot H(\delta/\eta)\Big]^k. \end{aligned} \tag{3.4}$$

Recall the definition of filtration $\mathbb{F} = (\mathcal{F}_j)_{j\geq 0}$ where $\mathcal{F}_j$ is the σ-algebra generated by $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_j$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$. In the next lemma, we establish a uniform asymptotic concentration bound for the weighted sum of Z_i's where the weights are adapted to the filtration $\mathbb{F}$. For any $M \in (0, \infty)$, let $\boldsymbol{\Gamma}_M$

denote the collection of families of random matrices $\mathbf{V}_j = (V_{j;p,q})_{p\in[m],q\ \in[d]}$ taking values in $\mathbb{R}^{m\times d}$, over which we will prove the uniform asymptotics:

$$\mathbf{\Gamma}_M \triangleq \Big\{(\mathbf{V}_j)_{j\geq 0} \text{ is adapted to } \mathbb{F}:\ \|\mathbf{V}_j\| \leq M\ \forall j\geq 0 \text{ almost surely}\Big\}. \tag{3.5}$$

Lemma 3.3. *Let Assumption 1 hold.*

(a) Given any $M>0$, $N>0$, $t>0$, and $\epsilon>0$, there exists $\delta_0=\delta_0(\epsilon,M,N,t)>0$ such that

$$\lim_{\eta\downarrow 0}\eta^{-N}\sup_{(\boldsymbol{V}_i)_{i\geq 0}\in\mathbf{\Gamma}_M}\mathbf{P}\bigg(\max_{j\leq\lfloor t/\eta\rfloor\wedge\big(\tau_1^{>\delta}(\eta)-1\big)}\eta\bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i\bigg\|>\epsilon\bigg)=0\qquad\forall\delta\in(0,\delta_0).$$

(b) Furthermore, let Assumption 3 hold. For each $i\geq 1$, let

$$A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\triangleq\left\{\max_{j\in I_i(\eta,\delta)}\eta\left\|\sum_{n=\tau_{i-1}^{>\delta}(\eta)+1}^{j}\boldsymbol{\sigma}\big(\boldsymbol{X}_{n-1}^{\eta|b}(\boldsymbol{x})\big)\boldsymbol{Z}_n\right\|\leq\epsilon\right\}; \tag{3.6}$$

$$I_i(\eta,\delta)\triangleq\Big\{j\in\mathbb{N}:\ \tau_{i-1}^{>\delta}(\eta)+1\leq j\leq\big(\tau_i^{>\delta}(\eta)-1\big)\wedge\lfloor 1/\eta\rfloor\Big\}. \tag{3.7}$$

Here we adopt the convention that (under $b=\infty$)

$$A_i(\eta,\infty,\epsilon,\delta,x)\triangleq\left\{\max_{j\in I_i(\eta,\delta)}\eta\left\|\sum_{n=\tau_{i-1}^{>\delta}(\eta)+1}^{j}\boldsymbol{\sigma}\big(\boldsymbol{X}_{n-1}^{\eta}(\boldsymbol{x})\big)\boldsymbol{Z}_n\right\|\leq\epsilon\right\}.$$

For any $k\geq 0$, $N>0$, $\epsilon>0$ and $b\in(0,\infty]$, there exists $\delta_0=\delta_0(\epsilon,N)>0$ such that

$$\lim_{\eta\downarrow 0}\eta^{-N}\sup_{\boldsymbol{x}\in\mathbb{R}^m}\mathbf{P}\bigg(\Big(\bigcap_{i=1}^{k}A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\Big)^c\bigg)=0\qquad\forall\delta\in(0,\delta_0).$$

Proof. (a) Choose some β such that $\frac{1}{2\wedge\alpha}<\beta<1$. Let

$$\boldsymbol{Z}_i^{(1)}\triangleq\boldsymbol{Z}_i\mathbb{I}\Big\{\|\boldsymbol{Z}_i\|\leq\frac{1}{\eta^\beta}\Big\},\qquad\hat{\boldsymbol{Z}}_i^{(1)}\triangleq\boldsymbol{Z}_i^{(1)}-\mathbf{E}\boldsymbol{Z}_i^{(1)},\qquad\boldsymbol{Z}_i^{(2)}\triangleq\boldsymbol{Z}_i\mathbb{I}\Big\{\|\boldsymbol{Z}_i\|\in(\frac{1}{\eta^\beta},\frac{\delta}{\eta}]\Big\}.$$

Note that $\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i=\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i^{(1)}+\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i^{(2)}$ on $j<\tau_1^{>\delta}(\eta)$, and hence,

$$\begin{aligned}
&\max_{j\leq\lfloor t/\eta\rfloor\wedge\big(\tau_1^{>\delta}(\eta)-1\big)}\eta\bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i\bigg\|\\
&\leq\max_{j\leq\lfloor t/\eta\rfloor}\eta\bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i^{(1)}\bigg\|+\max_{j\leq\lfloor t/\eta\rfloor}\eta\bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i^{(2)}\bigg\|\\
&\leq\max_{j\leq\lfloor t/\eta\rfloor}\eta\bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\mathbf{E}\boldsymbol{Z}_i^{(1)}\bigg\|+\max_{j\leq\lfloor t/\eta\rfloor}\eta\bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\widehat{\boldsymbol{Z}}_i^{(1)}\bigg\|+\max_{j\leq\lfloor t/\eta\rfloor}\eta\bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i^{(2)}\bigg\|.
\end{aligned}$$

Therefore, it suffices to show the existence of δ_0 such that for any $\delta \in (0, \delta_0)$,

$$\limsup_{\eta \downarrow 0} \sup_{(\boldsymbol{V}_i)_{i\geq 0} \in \boldsymbol{\Gamma}_M} \max_{j \leq \lfloor t/\eta \rfloor} \eta \left\| \sum_{i=1}^{j} \boldsymbol{V}_{i-1} \mathbf{E} \boldsymbol{Z}_i^{(1)} \right\| < \frac{\epsilon}{3}, \tag{3.8}$$

$$\lim_{\eta \downarrow 0} \eta^{-N} \sup_{(\boldsymbol{V}_i)_{i\geq 0} \in \boldsymbol{\Gamma}_M} \mathbf{P}\left(\max_{j \leq \lfloor t/\eta \rfloor} \eta \left\| \sum_{i=1}^{j} \boldsymbol{V}_{i-1} \widehat{\boldsymbol{Z}}_i^{(1)} \right\| > \frac{\epsilon}{3} \right) = 0, \tag{3.9}$$

$$\lim_{\eta \downarrow 0} \eta^{-N} \sup_{(\boldsymbol{V}_i)_{i\geq 0} \in \boldsymbol{\Gamma}_M} \mathbf{P}\left(\max_{j \leq \lfloor t/\eta \rfloor} \eta \left\| \sum_{i=1}^{j} \boldsymbol{V}_{i-1} \boldsymbol{Z}_i^{(2)} \right\| > \frac{\epsilon}{3} \right) = 0. \tag{3.10}$$

For (3.8), first observe that

$$\begin{aligned} \left\| \mathbf{E} \boldsymbol{Z}_i^{(1)} \right\| &= \left\| \mathbf{E} \boldsymbol{Z}_i \mathbb{I}\{ \| \boldsymbol{Z}_i \| > 1/\eta^\beta \} \right\| \qquad \text{due to } \mathbf{E}\boldsymbol{Z}_i = \boldsymbol{0} \\ &\leq \mathbf{E}\Big[\| \boldsymbol{Z}_i \| \, \mathbb{I}\{ \| \boldsymbol{Z}_i \| > 1/\eta^\beta \} \Big] \\ &= \mathbf{E}\Big[(\| \boldsymbol{Z}_i \| - 1/\eta^\beta) \mathbb{I}\{ \| \boldsymbol{Z}_i \| - 1/\eta^\beta > 0 \} \Big] + 1/\eta^\beta \cdot \mathbf{P}(\| \boldsymbol{Z}_i \| > 1/\eta^\beta). \end{aligned}$$

Since $(\|\boldsymbol{Z}_i\| - 1/\eta^\beta)\mathbb{I}\{\|\boldsymbol{Z}_i\| - 1/\eta^\beta > 0\}$ is non-negative,

$$\begin{aligned} \mathbf{E}(\| \boldsymbol{Z}_i \| - 1/\eta^\beta) \mathbb{I}\{ \| \boldsymbol{Z}_i \| - 1/\eta^\beta > 0 \} &= \int_0^\infty \mathbf{P}\big((\| \boldsymbol{Z}_i \| - 1/\eta^\beta) \mathbb{I}\{ \| \boldsymbol{Z}_i \| - 1/\eta^\beta \} > x \big) dx \\ &= \int_0^\infty \mathbf{P}(\| \boldsymbol{Z}_i \| - 1/\eta^\beta > x) dx = \int_{1/\eta^\beta}^\infty \mathbf{P}(\| \boldsymbol{Z} \| > x) dx. \end{aligned}$$

Recall that $H(x) = \mathbf{P}(\|\boldsymbol{Z}\| > x) \in \mathcal{RV}_{-\alpha}(x)$ as $x \to \infty$. Therefore, from Karamata's theorem,

$$\left\| \mathbf{E} \boldsymbol{Z}_i^{(1)} \right\| \leq \int_{1/\eta^\beta}^\infty \mathbf{P}(\| \boldsymbol{Z} \| > x) dx + 1/\eta^\beta \cdot \mathbf{P}(\| \boldsymbol{Z} \| > 1/\eta^\beta) \in \mathcal{RV}_{(\alpha-1)\beta}(\eta) \tag{3.11}$$

as $\eta \downarrow 0$. Therefore, there exists some $\eta_0 = \eta_0(t, M, \epsilon) > 0$ such that for any $\eta \in (0, \eta_0)$, we have $t \cdot M \cdot \left\| \mathbf{E}\boldsymbol{Z}_i^{(1)} \right\| < \epsilon/3$, and hence for any $(\boldsymbol{V}_i)_{i\geq 0} \in \boldsymbol{\Gamma}_M$ and $\eta \in (0, \eta_0)$,

$$\max_{j \leq \lfloor t/\eta \rfloor} \eta \left\| \sum_{i=1}^{j} \boldsymbol{V}_{i-1} \mathbf{E} \boldsymbol{Z}_i^{(1)} \right\| \leq \lfloor t/\eta \rfloor \cdot M \cdot \eta \left\| \mathbf{E} \boldsymbol{Z}_i^{(1)} \right\| < \epsilon/3,$$

from which we immediately get (3.8).

Next, for (3.9), recall our convention that vectors in Euclidean spaces are understood as row vectors (unless specified otherwise), and write $\boldsymbol{V}_t = (V_{t;l,k})_{l \in [m], k \in [d]}$, $\widehat{\boldsymbol{Z}}_t = (\widehat{Z}_{t;1}, \cdots, \widehat{Z}_{t;d})$. Since

$$\left\| \sum_{i=1}^{j} \boldsymbol{V}_{i-1} \widehat{\boldsymbol{Z}}_i \right\| = \sqrt{ \sum_{l=1}^{m} \left(\sum_{i=1}^{j} \sum_{k=1}^{d} V_{i-1;l,k} \hat{Z}_{i,k} \right)^2 },$$

to prove (3.9), it suffices to show that

$$\lim_{\eta \downarrow 0} \eta^{-N} \sup_{(\boldsymbol{V}_i)_{i\geq 0} \in \boldsymbol{\Gamma}_M} \mathbf{P}\left(\max_{j \leq \lfloor t/\eta \rfloor} \eta |Y_{l,k}(j; \boldsymbol{V})| > \frac{\epsilon}{3\sqrt{md^2}} \right) = 0 \qquad \forall l \in [m],\ k \in [d], \tag{3.12}$$

where

$$Y_{l,k}(j; \boldsymbol{V}) \triangleq \sum_{i=1}^{j} V_{i-1;l,k} \hat{Z}_{i,k}.$$

To proceed, we fix a sufficiently large p satisfying

$$p \geq 1, \quad p > \frac{2N}{\beta}, \quad p > \frac{2N}{1-\beta}, \quad p > \frac{2N}{(\alpha-1)\beta} > \frac{2N}{(2\alpha-1)\beta}, \tag{3.13}$$

and some $l \in [m]$, $k \in [d]$. Note that for $(\boldsymbol{V}_i)_{i\geq 0} \in \boldsymbol{\Gamma}_M$ and $\eta > 0$, $\{V_{i-1;l,k}\widehat{Z}^{(1)}_{i;k} : i \geq 1\}$ is a martingale difference sequence. Therefore, $\big(Y_{l,k}(j;\boldsymbol{V})\big)_{j\geq 0}$ is a martingale, and

$$\begin{aligned}
&\mathbf{E}\left[\left(\max_{j\leq \lfloor t/\eta\rfloor} \eta\Big|Y_{l,k}(j;\boldsymbol{V})\Big|\right)^p\right] \\
&\leq c_1\mathbf{E}\left[\left(\sum_{i=1}^{\lfloor t/\eta\rfloor}\Big(\eta V_{i-1;l,k}\widehat{Z}^{(1)}_{i;k}\Big)^2\right)^{p/2}\right] \\
&\leq c_1 M^p\mathbf{E}\left[\left(\sum_{i=1}^{\lfloor t/\eta\rfloor}\Big(\eta\widehat{Z}^{(1)}_{i;k}\Big)^2\right)^{p/2}\right] \qquad \text{due to } \|\boldsymbol{V}_s\| \leq M \text{ for all } s \geq 0 \\
&\leq c_1c_2M^p\mathbf{E}\left[\left(\max_{j\leq\lfloor t/\eta\rfloor}\Big|\sum_{i=1}^{j}\eta\widehat{Z}^{(1)}_{i;k}\Big|\right)^p\right] \leq \underbrace{c_1c_2\left(\frac{p}{p-1}\right)^p}_{\triangleq c'} M^p\mathbf{E}\left[\left|\sum_{i=1}^{\lfloor t/\eta\rfloor}\eta\widehat{Z}^{(1)}_{i;k}\right|^p\right]
\end{aligned} \tag{3.14}$$

for some $c_1, c_2 > 0$ that only depend on p and won't vary with $(\boldsymbol{V}_i)_{i\geq 0}$ and η. The first and third inequalities are from the uppper and lower bounds of Burkholder-Davis-Gundy inequality (Theorem 48, Chapter IV of [53]), respectively, and the fourth inequality is from Doob's maximal inequality. It then follows from Bernstein's inequality that for any $\eta > 0$ and any $s \in [0,t], y \geq 1$

$$\begin{aligned}
\mathbf{P}\left(\left|\sum_{j=1}^{\lfloor s/\eta\rfloor}\eta\widehat{Z}^{(1)}_{j;l,k}\right|^p > \eta^{2N}y\right) &= \mathbf{P}\left(\left|\sum_{j=1}^{\lfloor s/\eta\rfloor}\eta\widehat{Z}^{(1)}_{j;l,k}\right| > \eta^{\frac{2N}{p}}y^{1/p}\right) \\
&\leq 2\exp\left(-\frac{\frac{1}{2}\eta^{\frac{4N}{p}}\sqrt[p]{y^2}}{\frac{1}{3}\eta^{1-\beta+\frac{2N}{p}}\sqrt[p]{y} + \frac{t}{\eta}\cdot\eta^2\cdot\mathbf{E}\big[(\widehat{Z}^{(1)}_{1;k})^2\big]}\right).
\end{aligned} \tag{3.15}$$

Our next goal is to show that $\frac{t}{\eta}\cdot\eta^2\cdot\mathbf{E}\big[(\widehat{Z}^{(1)}_{1;k})^2\big] < \frac{1}{3}\eta^{1-\beta+\frac{2N}{p}}$ for any $\eta > 0$ small enough. First, due to $(a+b)^2 \leq 2a^2 + 2b^2$,

$$\mathbf{E}\big[(\widehat{Z}^{(1)}_{1;k})^2\big] = \mathbf{E}\big[\big(Z^{(1)}_{1;k} - \mathbf{E}Z^{(1)}_{1;k}\big)^2\big] \leq 2\mathbf{E}\big[\big(Z^{(1)}_{1;k}\big)^2\big] + 2\big[\mathbf{E}Z^{(1)}_{1;k}\big]^2 \leq 2\mathbf{E}\left[\left\|\boldsymbol{Z}^{(1)}_1\right\|^2\right] + 2\left[\mathbf{E}\left\|\boldsymbol{Z}^{(1)}_1\right\|\right]^2.$$

Also, it has been shown earlier that $\mathbf{E}\left\|\boldsymbol{Z}^{(1)}_1\right\| \in \mathcal{RV}_{(\alpha-1)\beta}(\eta)$, and hence $\left[\mathbf{E}\left\|\boldsymbol{Z}^{(1)}_1\right\|\right]^2 \in \mathcal{RV}_{2(\alpha-1)\beta}(\eta)$. From our choice of $p > \frac{2N}{(2\alpha-1)\beta}$ in (3.13), we have $1 + 2(\alpha-1)\beta > 1 - \beta + \frac{2N}{p}$, thus implying

$$\frac{t}{\eta}\cdot\eta^2\cdot 2\left[\mathbf{E}\left\|\boldsymbol{Z}^{(1)}_1\right\|\right]^2 < \frac{1}{6}\eta^{1-\beta+\frac{2N}{p}}$$

for any $\eta > 0$ sufficiently small. Next, $\mathbf{E}\big[\left\|\boldsymbol{Z}^{(1)}_1\right\|^2\big] = \int_0^\infty 2x\mathbf{P}(\left\|\boldsymbol{Z}^{(1)}_1\right\| > x)dx = \int_0^{1/\eta^\beta} 2x\mathbf{P}(\|\boldsymbol{Z}_1\| > x)dx$. If $\alpha \in (1,2]$, then Karamata's theorem implies $\int_0^{1/\eta^\beta} 2x\mathbf{P}(\|\boldsymbol{Z}_1\| > x)dx \in \mathcal{RV}_{-(2-\alpha)\beta}(\eta)$ as

$\eta \downarrow 0$. Given our choice of p in (3.13), one can see that $1-(2-\alpha)\beta > 1-\beta+\frac{2N}{p}$. As a result, for any $\eta > 0$ small enough we have $\frac{t}{\eta}\cdot\eta^2\cdot 2\mathbf{E}\Big[\Big\|\boldsymbol{Z}_1^{(1)}\Big\|^2\Big] < \frac{1}{6}\eta^{1-\beta+\frac{2N}{p}}$. If $\alpha > 2$, then $\lim_{\eta\downarrow 0}\int_0^{1/\eta^\beta} 2x\mathbf{P}(\|\boldsymbol{Z}_1\| > x)dx = \int_0^\infty 2x\mathbf{P}(\|\boldsymbol{Z}_1\| > x)dx < \infty$. Also, (3.13) implies that $1-\beta+\frac{2N}{p} < 1$. As a result, for any $\eta > 0$ small enough we have $\frac{t}{\eta}\cdot\eta^2\cdot 2\mathbf{E}\Big[\Big\|\boldsymbol{Z}_1^{(1)}\Big\|^2\Big] < \frac{1}{6}\eta^{1-\beta+\frac{2N}{p}}$. In summary,

$$\frac{t}{\eta}\cdot\eta^2\cdot\mathbf{E}\big[(\widehat{Z}_{1;k}^{(1)})^2\big] < \frac{1}{3}\eta^{1-\beta+\frac{2N}{p}} \tag{3.16}$$

holds for any $\eta > 0$ small enough. Along with (3.15), we yield that for any $\eta > 0$ small enough,

$$\mathbf{P}\left(\Big|\sum_{j=1}^{\lfloor s/\eta\rfloor}\eta\widehat{Z}_{j;l,k}^{(1)}\Big|^p > \eta^{2N}y\right) \leq 2\exp\left(\frac{-\frac{1}{2}y^{1/p}}{\frac{2}{3}\eta^{1-\beta-\frac{2N}{p}}}\right) \leq 2\exp\Big(-\frac{3}{4}y^{1/p}\Big) \qquad \forall y \geq 1,$$

where the last inequality is due to our choice of p in (3.13) that $1-\beta-\frac{2N}{p} > 0$. Moreover, since $\int_0^\infty \exp\big(-\frac{3}{4}y^{1/p}\big)dy < \infty$, one can see the existence of some $C_p^{(1)} < \infty$ such that $\mathbf{E}\Big|\sum_{j=1}^{\lfloor t/\eta\rfloor}\eta\widehat{Z}_{j;l,k}^{(1)}\Big|^p\Big/\eta^{2N} < C_p^{(1)}$ for all $\eta > 0$ small enough. Combining this bound, (3.14), and Markov inequality, for all $\eta > 0$ small enough,

$$\begin{aligned}\mathbf{P}\left(\max_{j\leq\lfloor t/\eta\rfloor}\eta|Y_{l,k}(j;\boldsymbol{V})| > \frac{\epsilon}{3\sqrt{md^2}}\right) &\leq \frac{\mathbf{E}\Big[\max_{j\leq\lfloor t/\eta\rfloor}\Big|\sum_{i=1}^{j}\eta Y_{l,k}(j,\boldsymbol{V})\Big|^p\Big]}{\epsilon^p/(3\sqrt{md^2})^p} \\ &\leq \frac{c'M^p\mathbf{E}\Big|\sum_{j=1}^{\lfloor t/\eta\rfloor}\eta\widehat{Z}_{i;k}^{(1)}\Big|^p}{\epsilon^p/(3\sqrt{md^2})^p} \leq \frac{c'M^p\cdot C_p^{(1)}}{\epsilon^p/(3\sqrt{md^2})^p}\cdot\eta^{2N}\end{aligned}$$

holds uniformly for all $(\boldsymbol{V}_i)_{i\geq 0}\in\boldsymbol{\Gamma}_M$. This proves (3.12) and hence (3.9).

Finally, for (3.10), recall that we have chosen β in such a way that $\alpha\beta-1 > 0$. Fix a constant $J = \lceil\frac{N}{\alpha\beta-1}\rceil+1$, and define $I(\eta) \triangleq \#\{i \leq \lfloor t/\eta\rfloor:\ \boldsymbol{Z}_i^{(2)} \neq 0\}$. Besides, fix $\delta_0 = \frac{\epsilon}{3MJ}$. For any $\delta\in(0,\delta_0)$ and $(\boldsymbol{V}_i)_{i\geq 0}\in\boldsymbol{\Gamma}_M$, note that on event $\{I(\eta) < J\}$, we must have $\max_{j\leq\lfloor t/\eta\rfloor}\eta\Big\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i^{(2)}\Big\| < \eta\cdot M\cdot J\cdot\delta_0/\eta < MJ\delta_0 < \epsilon/3$. On the other hand,

$$\mathbf{P}\big(I(\eta)\geq J\big) \leq \binom{\lfloor t/\eta\rfloor}{J}\cdot\Big(H\big(1/\eta^\beta\big)\Big)^J \leq (t/\eta)^J\cdot\Big(H\big(1/\eta^\beta\big)\Big)^J \in \mathcal{RV}_{J(\alpha\beta-1)}(\eta) \text{ as } \eta\downarrow 0.$$

Lastly, the choice of $J = \lceil\frac{N}{\alpha\beta-1}\rceil+1$ guarantees that $J(\alpha\beta-1) > N$, and hence,

$$\lim_{\eta\downarrow 0}\sup_{(V_i)_{i\geq 0}\in\boldsymbol{\Gamma}_M}\mathbf{P}\left(\max_{j\leq\lfloor t/\eta\rfloor}\eta\left\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i^{(2)}\right\| > \frac{\epsilon}{3}\right)\Big/\eta^N \leq \lim_{\eta\downarrow 0}\sup_{(V_i)_{i\geq 0}\in\boldsymbol{\Gamma}_M}\mathbf{P}\big(I(\eta)\geq J\big)\Big/\eta^N = 0.$$

This concludes the proof of part (a).

(b) To ease notations, in this proof we write $\boldsymbol{X}^\eta = \boldsymbol{X}^{\eta|\infty}$ for the cases where $b=\infty$. Due to Assumption 3, it holds for any $\boldsymbol{x}\in\mathbb{R}^m$ and any $\eta > 0, n \geq 0$ that $\Big\|\boldsymbol{\sigma}\big(\boldsymbol{X}_n^{\eta|b}(\boldsymbol{x})\big)\Big\| \leq C$, so $\{\boldsymbol{\sigma}(\boldsymbol{X}_i^{\eta|b}(\boldsymbol{x}))\}_{i\geq 0}\in\boldsymbol{\Gamma}_C$. By strong Markov property at stopping times $\big(\tau_j^{>\delta}(\eta)\big)_{j\geq 1}$,

$$\sup_{\boldsymbol{x}\in\mathbb{R}^m}\mathbf{P}\left(\Big(\bigcap_{i=1}^{k}A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\Big)^c\right) \leq \sum_{i=1}^{k}\sup_{\boldsymbol{x}\in\mathbb{R}^m}\mathbf{P}\left(\Big(A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\Big)^c\right)$$

$$\leq k \cdot \sup_{(\boldsymbol{V}_i)_{i\geq 0}\in\boldsymbol{\Gamma}_C} \mathbf{P}\Bigg(\max_{j\leq \lfloor 1/\eta\rfloor \wedge \left(\tau_1^{>\delta}(\eta)-1\right)} \eta \Bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i\Bigg\| > \epsilon\Bigg)$$

where $C<\infty$ is the constant in Assumption 3 and the set $\boldsymbol{\Gamma}_C$ is defined in (3.5). Thanks to part (a), one can find some $\delta_0=\delta_0(\epsilon,C,N)\in(0,\bar{\delta})$ such that

$$\sup_{(\boldsymbol{V}_i)_{i\geq 0}\in\boldsymbol{\Gamma}_C} \mathbf{P}\Bigg(\max_{j\leq \lfloor 1/\eta\rfloor \wedge \left(\tau_1^{>\delta}(\eta)-1\right)} \eta \Bigg\|\sum_{i=1}^{j}\boldsymbol{V}_{i-1}\boldsymbol{Z}_i\Bigg\| > \epsilon\Bigg) = \boldsymbol{o}(\eta^N)$$

(as $\eta\downarrow 0$) for any $\delta\in(0,\delta_0)$, which concludes the proof of part (b). □

Next, for any $c>\delta>0$, we study the law of $\big(\tau_j^{>\delta}(\eta)\big)_{j\geq 1}$ and $\big(\boldsymbol{W}_j^{>\delta}(\eta)\big)_{j\geq 1}$ conditioned on event

$$E^{\delta}_{c,k}(\eta) \triangleq \Big\{\tau_k^{>\delta}(\eta)\leq \lfloor 1/\eta\rfloor < \tau_{k+1}^{>\delta}(\eta);\ \eta\left\|\boldsymbol{W}_j^{>\delta}(\eta)\right\| > c\ \ \forall j\in[k]\Big\}. \tag{3.17}$$

The intuition is that, on event $E^{\delta}_{c,k}(\eta)$, among the first $\lfloor 1/\eta\rfloor$ steps there are exactly k "large" jumps, all of which has size larger than c/η. Next, for each $c>0$, we consider a random vector $\boldsymbol{W}^*(c)$ in $\mathbb{R}^d$ with $\|\boldsymbol{W}^*(c)\|>c$ almost surely, whose polar coordinates $\big(R^*(c),\boldsymbol{\Theta}^*(c)\big)\triangleq\Big(\|\boldsymbol{W}^*(c)\|, \frac{\boldsymbol{W}^*(c)}{\|\boldsymbol{W}^*(c)\|}\Big)$ admit the law

$$\mathbf{P}\Big(\big(R^*(c),\boldsymbol{\Theta}^*(c)\big)\in\ \cdot\ \Big) = \big(\bar{\nu}_\alpha|_{(c,\infty)}\times\mathbf{S}\big)(\cdot). \tag{3.18}$$

Here, recall the definition of the measure ν_α in (2.5) and the measure $\mathbf{S}$ in Assumption 1, and note that $\alpha>1$ is the heavy-tail index in Assumption 1. For any $c>0$, we set

$$\bar{\nu}_\alpha|_{(c,\infty)}(\ \cdot\ ) \triangleq c^\alpha\cdot\nu_\alpha\Big(\ \cdot\ \cap(c,\infty)\Big).$$

to be the restricted and normalized (as a probability measure) version of ν_α over (c,∞). Let $\big(\boldsymbol{W}_j^*(c)\big)_{j\geq 1}$ be a sequence of iid copies of $\boldsymbol{W}^*(c)$. Also, for $(U_j)_{j\geq 1}$, a sequence of iid copies of Unif$(0,1)$ that is also independent of $\big(\boldsymbol{W}_j^*(c)\big)_{j\geq 1}$, let $U_{(1;k)}\leq U_{(2;k)}\leq\cdots\leq U_{(k;k)}$ be the order statistics of $(U_j)_{j=1}^k$. For any random element X and any Borel measureable set A, let $\mathscr{L}(X)$ be the law of X, and $\mathscr{L}(X|A)$ be the conditional law of X given event A.

Lemma 3.4. *Let Assumption 1 hold. For any $\delta>0, c\geq\delta$ and $k\in\mathbb{Z}^+$,*

$$\lim_{\eta\downarrow 0}\frac{\mathbf{P}\big(E^{\delta}_{c,k}(\eta)\big)}{\lambda^k(\eta)} = \frac{1/c^{\alpha k}}{k!}, \tag{3.19}$$

and

$$\begin{aligned}&\mathscr{L}\Big(\eta\boldsymbol{W}_1^{>\delta}(\eta),\eta\boldsymbol{W}_2^{>\delta}(\eta),\cdots,\eta\boldsymbol{W}_k^{>\delta}(\eta),\eta\tau_1^{>\delta}(\eta),\eta\tau_2^{>\delta}(\eta),\cdots,\eta\tau_k^{>\delta}(\eta)\Big| E^{\delta}_{c,k}(\eta)\Big)\\ \to&\mathscr{L}\Big(\boldsymbol{W}_1^*(c),\boldsymbol{W}_2^*(c),\cdots,\boldsymbol{W}_k^*(c),U_{(1;k)},U_{(2;k)},\cdots,U_{(k;k)}\Big)\ \text{as}\ \eta\downarrow 0.\end{aligned}$$

Proof. Note that $\big(\tau_i^{>\delta}(\eta)\big)_{i\geq 1}$ is independent of $\big(\boldsymbol{W}_i^{>\delta}(\eta)\big)_{i\geq 1}$. Therefore, $\mathbf{P}\big(E^{\delta}_{c,k}(\eta)\big)=\mathbf{P}\big(\tau_k^{>\delta}(\eta)\leq \lfloor 1/\eta\rfloor<\tau_{k+1}^{>\delta}(\eta)\big)\cdot\Big(\mathbf{P}\big(\eta\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\|>c\big)\Big)^k$. Recall that $H(x)=\mathbf{P}(\|\boldsymbol{Z}\|>x)$. Observe that

$$\begin{aligned}\mathbf{P}\Big(\tau_k^{>\delta}(\eta)\leq\lfloor 1/\eta\rfloor<\tau_{k+1}^{>\delta}(\eta)\Big) &= \mathbf{P}\Big(\#\{j\leq\lfloor 1/\eta\rfloor:\ \eta|Z_j|>\delta\}=k\Big)\\ &= \underbrace{\binom{\lfloor 1/\eta\rfloor}{k}}_{\triangleq q_1(\eta)}\underbrace{\Big(1-H(\delta/\eta)\Big)^{\lfloor 1/\eta\rfloor-k}}_{\triangleq q_2(\eta)}\underbrace{\big(H(\delta/\eta)\big)^k}_{\triangleq q_3(\eta)}.\end{aligned} \tag{3.20}$$

For $q_1(\eta)$, note that

$$\lim_{\eta\downarrow 0}\frac{q_1(\eta)}{1/\eta^k} = \frac{\big(\lfloor 1/\eta\rfloor\big)\big(\lfloor 1/\eta\rfloor - 1\big)\cdots\big(\lfloor 1/\eta\rfloor - k + 1\big)/k!}{1/\eta^k} = \frac{1}{k!}. \tag{3.21}$$

Also, since $(\lfloor 1/\eta\rfloor - k)\cdot H(\delta/\eta) = \boldsymbol{o}(1)$ as $\eta\downarrow 0$, we have that $\lim_{\eta\downarrow 0} q_2(\eta) = 1$. Lastly, note that

$$\mathbf{P}\big(\eta\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\| > c\big) = H(c/\eta)\Big/H(\delta/\eta),$$

and hence,

$$\lim_{\eta\downarrow 0}\frac{q_3(\eta)\cdot\Big(\mathbf{P}\big(\eta\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\| > c\big)\Big)^k}{\big(H(1/\eta)\big)^k} = \lim_{\eta\downarrow 0}\frac{\big(H(\delta/\eta)\big)^k\cdot\Big(H(c/\eta)\Big/H(\delta/\eta)\Big)^k}{\big(H(1/\eta)\big)^k} = \lim_{\eta\downarrow 0}\frac{\big(H(c/\eta)\big)^k}{\big(H(1/\eta)\big)^k} = 1/c^{\alpha k} \tag{3.22}$$

Plugging (3.21) and (3.22) into (3.20), we obtain (3.19).

Next, we move onto the proof of the weak convergence. We use $\big(R_1^{>\delta}(\eta),\boldsymbol{\Theta}_1^{>\delta}(\eta)\big) \triangleq \Big(\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\|,\frac{\boldsymbol{W}_1^{>\delta}(\eta)}{\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\|}\Big)$ to denote the polar coordinates of $\boldsymbol{W}_1^{>\delta}(\eta)$. Observe the following weak convergence:

$$\begin{aligned}
&\mathbf{P}\Big(\big(\eta R_1^{>\delta}(\eta),\boldsymbol{\Theta}_1^{>\delta}(\eta)\big)\in\ \cdot\ \Big|\ \eta R_1^{>\delta}(\eta) > c\Big)\\
&= \frac{\mathbf{P}\Big(\big(\eta R_1^{>\delta}(\eta),\boldsymbol{\Theta}_1^{>\delta}(\eta)\big)\in\ \cdot\ \cap\big((c,\infty)\cap\mathfrak{N}_d\big)\Big)}{\mathbf{P}\big(\eta\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\| > c\big)}\\
&= \frac{\mathbf{P}\Big(\big(\eta R,\boldsymbol{\Theta}\big)\in\ \cdot\ \cap\big((c,\infty)\cap\mathfrak{N}_d\big)\Big)\Big/\mathbf{P}(\eta\left\|\boldsymbol{Z}\right\| > \delta)}{\mathbf{P}(\eta\left\|\boldsymbol{Z}\right\| > c)\big/\mathbf{P}(\eta\left\|\boldsymbol{Z}\right\| > \delta)} \qquad \text{with } (R,\boldsymbol{\Theta}) = (\left\|\boldsymbol{Z}\right\|,\frac{\boldsymbol{Z}}{\left\|\boldsymbol{Z}\right\|})\\
&= \frac{\mathbf{P}\Big(\big(\eta R,\boldsymbol{\Theta}\big)\in\ \cdot\ \cap\big((c,\infty)\cap\mathfrak{N}_d\big)\Big)}{\mathbf{P}(\eta\left\|\boldsymbol{Z}\right\| > 1)}\cdot\frac{\mathbf{P}(\eta\left\|\boldsymbol{Z}\right\| > 1)}{\mathbf{P}(\eta\left\|\boldsymbol{Z}\right\| > c)} = \frac{\mathbf{P}\Big(\big(\eta R,\boldsymbol{\Theta}\big)\in\ \cdot\ \cap\big((c,\infty)\cap\mathfrak{N}_d\big)\Big)}{H(\eta^{-1})}\cdot\frac{H(\eta^{-1})}{H(c\cdot\eta^{-1})}\\
&\Rightarrow (\bar{\nu}_\alpha|_{(c,\infty)}\times\mathbf{S})(\cdot) \qquad \text{as } \eta\downarrow 0 \text{ by Assumption 1.}
\end{aligned}$$

As a result, we must have $\mathscr{L}\Big(\eta\boldsymbol{W}_1^{>\delta}(\eta),\eta\boldsymbol{W}_2^{>\delta}(\eta),\cdots,\eta\boldsymbol{W}_k^{>\delta}(\eta)\Big|E^\delta_{c,k}(\eta)\Big)\to\mathscr{L}\Big(\boldsymbol{W}_1^*(c),\cdots,\boldsymbol{W}_k^*(c)\Big)$. Moreover, one can easily see that, conditioned on the event $E^\delta_{c,k}(\eta)$, the sequences $\eta\boldsymbol{W}_1^{>\delta}(\eta),\cdots,\eta\boldsymbol{W}_k^{>\delta}(\eta)$ and $\eta\tau_1^{>\delta}(\eta),\cdots,\eta\tau_k^{>\delta}(\eta)$ are conditionally independent. Therefore, as $\eta\downarrow 0$, the limit of the conditional law $\mathscr{L}\Big(\eta\boldsymbol{W}_1^{>\delta}(\eta),\cdots,\eta\boldsymbol{W}_k^{>\delta}(\eta)\Big|E^\delta_{c,k}(\eta)\Big)$ is also independent from that of $\mathscr{L}\Big(\eta\tau_1^{>\delta}(\eta),\cdots,\eta\tau_k^{>\delta}(\eta)\Big|E^\delta_{c,k}(\eta)\Big)$, and it only remains to show that

$$\mathscr{L}\Big(\eta\tau_1^{>\delta}(\eta),\eta\tau_2^{>\delta}(\eta),\cdots,\eta\tau_k^{>\delta}(\eta)\Big|E^\delta_{c,k}(\eta)\Big)\to\mathscr{L}\Big(U_{(1;k)},\cdots,U_{(k;k)}\Big).$$

Note that since both $\{\eta\tau_i^{>\delta}(\eta): i = 1,\ldots,k\}$ and $\{U_{(i):k}: i = 1,\ldots,k\}$ are sorted in an ascending order, the joint CDFs are completely characterized by $\{t_i: i = 1,\ldots,k\}$'s such that $0\le t_1\le t_2\le\cdots\le t_k\le 1$. For any such $(t_1,\cdots,t_k)\in[0,t]^k$, note that

$$\begin{aligned}
&\mathbf{P}\Big(\eta\tau_1^{>\delta}(\eta) > t_1,\ \eta\tau_2^{>\delta}(\eta) > t_2,\ \cdots,\eta\tau_k^{>\delta}(\eta) > t_k\ \Big|\ E^\delta_{c,k}(\eta)\Big)\\
&= \mathbf{P}\Big(\eta\tau_1^{>\delta}(\eta) > t_1,\ \eta\tau_2^{>\delta}(\eta) > t_2,\ \cdots,\eta\tau_k^{>\delta}(\eta) > t_k\ \Big|\ \tau_k^{>\delta}(\eta) < \lfloor 1/\eta\rfloor < \tau_{k+1}^{>\delta}(\eta)\Big)\\
&= \frac{\mathbf{P}\Big(\eta\tau_1^{>\delta}(\eta) > t_1,\ \eta\tau_2^{>\delta}(\eta) > t_2,\ \cdots,\eta\tau_k^{>\delta}(\eta) > t_k;\ \tau_k^{>\delta}(\eta) < \lfloor 1/\eta\rfloor < \tau_{k+1}^{>\delta}(\eta)\Big)}{\mathbf{P}\Big(\tau_k^{>\delta}(\eta) < \lfloor 1/\eta\rfloor < \tau_{k+1}^{>\delta}(\eta)\Big)}
\end{aligned}$$

and observe that

$$\begin{aligned}
&\frac{\mathbf{P}\Big(\eta\tau_1^{>\delta}(\eta) > t_1,\ \eta\tau_2^{>\delta}(\eta) > t_2,\ \cdots,\eta\tau_k^{>\delta}(\eta) > t_k;\ \tau_k^{>\delta}(\eta) < \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta)\Big)}{\mathbf{P}\Big(\tau_k^{>\delta}(\eta) < \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta)\Big)} \\
&= \frac{\big|\boldsymbol{E}^\eta\big| \cdot q_2(\eta)q_3(\eta)}{q_1(\eta)q_2(\eta)q_3(\eta)} = \big|\boldsymbol{E}^\eta\big|\Big/q_1(\eta)
\end{aligned}$$

where $\boldsymbol{E}^\eta \triangleq \Big\{(s_1,\cdots,s_k) \in \{1,2,\cdots,\lfloor 1/\eta \rfloor - 1\}^k:\ \eta s_j > t_j\ \forall j \in [k];\ s_1 < s_2 < \cdots < s_k\Big\}$. Note that

$$|\boldsymbol{E}^\eta| = \sum_{s_k=\lfloor\frac{t_k}{\eta}\rfloor+1}^{\lfloor 1/\eta\rfloor - 1}\ \sum_{s_{k-1}=\lfloor\frac{t_{k-1}}{\eta}\rfloor+1}^{s_k-1}\ \sum_{s_{k-2}=\lfloor\frac{t_{k-2}}{\eta}\rfloor+1}^{s_{k-1}-1} \cdots \sum_{s_2=\lfloor\frac{t_2}{\eta}\rfloor+1}^{s_3-1}\ \sum_{s_1=\lfloor\frac{t_1}{\eta}\rfloor+1}^{s_2-1} 1.$$

Together with (3.21), we obtain

$$\begin{aligned}
\lim_{\eta\downarrow 0}\big|\boldsymbol{E}^\eta\big|\Big/q_1(\eta) &= (k!)\cdot\lim_{\eta\downarrow 0}\frac{\big|\boldsymbol{E}^\eta\big|}{(1/\eta)^k} = (k!)\int_{t_k}^1\int_{t_{k-1}}^{s_k}\int_{t_{k-2}}^{s_{k-1}}\cdots\int_{t_2}^{s_3}\int_{t_1}^{s_2} ds_1 ds_2\cdots ds_k \\
&= \mathbf{P}\big(U_{(i;k)} > t_i\ \forall i \in [j]\big)
\end{aligned}$$

and conclude the proof. □

Next, we present useful results about mappings $h^{(k)}_{[0,T]}$ defined in (2.10)–(2.13) and $h^{(k)|b}_{[0,T]}$ defined in (2.21)–(2.24). These results will serve as crucial tools when establishing Theorems 2.3 and 2.4. First, recall the definitions of the sets $\mathbb{D}^{(k)}_A(r)$ and $\mathbb{D}^{(k)|b}_A(r)$ in (2.17) and (2.25), respectively. The first two results reveal useful properties of $\mathbb{D}^{(k)}_A(r)$ and $\mathbb{D}^{(k)|b}_A(r)$ when Assumptions 2 and 3 hold. As their proofs mostly rely on arguments and calculations independent of those in the other sections of our analyses, we collect the proofs of Lemmas 3.5 and 3.6 in Section C.

Lemma 3.5. *Let Assumptions 2 and 3 hold. Given some compact $A \subseteq \mathbb{R}^m$, some $B \in \mathscr{S}_{\mathbb{D}}$, and some $k \in \mathbb{N}$, $r > 0$, if B is bounded away from $\mathbb{D}^{(k-1)}_A(r)$, then there exist $\bar\epsilon > 0$ and $\bar\delta > 0$ such that the following claims hold:*

(a) For any $\boldsymbol{x} \in A$,

$$h^{(k)}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),\boldsymbol{t}\big) \in B^{\bar\epsilon} \qquad\Longrightarrow\qquad \|\boldsymbol{w}_j\| > \bar\delta\ \forall j \in [k];$$

(b) $\boldsymbol{d}_{J_1}\big(B^{\bar\epsilon},\mathbb{D}^{(k-1)}_A(r)\big) > 0$.

Lemma 3.6. *Let Assumptions 2 and 3 hold. Given some compact $A \subseteq \mathbb{R}^m$, some $B \in \mathscr{S}_{\mathbb{D}}$, and some $k \in \mathbb{N}$, $b, r > 0$, if B is bounded away from $\mathbb{D}^{(k-1)|b}_A(r)$, then there exist $\bar\epsilon > 0$ and $\bar\delta > 0$ such that the following claims hold:*

(a) for any $\boldsymbol{x} \in A$, $b > 0$, and any $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$ with $\max_{j\in[k]}\|\boldsymbol{v}_j\| \leq \bar\epsilon$,

$$\bar h^{(k)|b}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k),\boldsymbol{t}\big) \in B^{\bar\epsilon} \qquad\Longrightarrow\qquad \|\boldsymbol{w}_j\| > \bar\delta\ \forall j \in [k];$$

(b) $\boldsymbol{d}_{J_1}\big(B^{\bar\epsilon},\mathbb{D}^{(k-1)|b}_A(r)\big) > 0$.

The next lemma establishes a convergence result from the measure $\mathbf{C}^{(k)|b}$ defined in (2.26) to the measure $\mathbf{C}^{(k)}$ defined in (2.14). Again, we collect the proof in Section C.

Lemma 3.7. *Let Assumptions 2 and 3 hold. Let $k \in \mathbb{N}$, $r > 0$, and $A \subseteq \mathbb{R}^m$ be compact. For any $g \in \mathcal{C}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)}(r)\big)$ and $\boldsymbol{x} \in A$,*

$$\lim_{b \to \infty} \mathbf{C}^{(k)|b}(g;\boldsymbol{x}) = \mathbf{C}^{(k)}(g;\boldsymbol{x}).$$

In Lemma 3.8, we show that the image of $h^{(1)}$ (resp. $h^{(1)|b}$) provides good approximations of the sample path of $\boldsymbol{X}_j^{\eta}(\boldsymbol{x})$ (resp. $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$) up until $\tau_1^{>\delta}(\eta)$, i.e. the arrival time of the first "large noise"; see (3.2),(3.3) for the definition of $\tau_i^{>\delta}(\eta), \boldsymbol{W}_i^{>\delta}(\eta)$.

Lemma 3.8. *Let Assumptions 2 and 3 hold. Let $D, C \in [1,\infty)$ be the constants in Assumptions 2 and 3, respectively, and let $\rho \triangleq \exp(D)$.*

(a) For any $\epsilon, \delta, \eta > 0$ and any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, it holds on the event

$$\left\{ \max_{i \le \lfloor 1/\eta \rfloor \wedge \left(\tau_1^{>\delta}(\eta) - 1\right)} \eta \left\| \sum_{j=1}^{i} \boldsymbol{\sigma}\big(\boldsymbol{X}_{j-1}^{\eta}(\boldsymbol{x})\big)\boldsymbol{Z}_j \right\| \le \epsilon \right\}$$

that

$$\sup_{t \in [0,1]:\ t < \eta\tau_1^{>\delta}(\eta)} \left\| \xi_t - \boldsymbol{X}_{\lfloor t/\eta \rfloor}^{\eta}(\boldsymbol{x}) \right\| \le \rho \cdot \big(\epsilon + \|\boldsymbol{x} - \boldsymbol{y}\| + \eta C\big), \tag{3.23}$$

where

$$\xi = \begin{cases} h^{(1)}\big(\boldsymbol{y}, \eta\boldsymbol{W}_1^{>\delta}(\eta), \eta\tau_1^{>\delta}(\eta)\big) & \text{if } \eta\tau_1^{>\delta}(\eta) \le 1, \\ h^{(0)}(\boldsymbol{y}) & \text{if } \eta\tau_1^{>\delta}(\eta) > 1. \end{cases}$$

(b) For any $\gamma, b > 0$, $\epsilon \in (0,1)$, $\delta \in (0, \frac{b}{2C})$, $\eta \in (0, \frac{b \wedge 1}{2C})$, and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, it holds on the event

$$\left\{ \max_{i \le \lfloor 1/\eta \rfloor \wedge \left(\tau_1^{>\delta}(\eta) - 1\right)} \eta \left\| \sum_{j=1}^{i} \boldsymbol{\sigma}\big(\boldsymbol{X}_{j-1}^{\eta|b}(\boldsymbol{x})\big)\boldsymbol{Z}_j \right\| \le \epsilon \right\} \cap \left\{ \eta \left\| \boldsymbol{W}_1^{>\delta}(\eta) \right\| \le 1/\epsilon^{\gamma} \right\}$$

that

$$\sup_{t \in [0,1]:\ t < \eta\tau_1^{>\delta}(\eta)} \left\| \xi_t - \boldsymbol{X}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x}) \right\| \le \rho \cdot \big(\epsilon + \|\boldsymbol{x} - \boldsymbol{y}\| + \eta C\big), \tag{3.24}$$

$$\sup_{t \in [0,1]:\ t \le \eta\tau_1^{>\delta}(\eta)} \left\| \xi_t - \boldsymbol{X}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x}) \right\| \le \rho D \cdot \big(\epsilon + \|\boldsymbol{x} - \boldsymbol{y}\| + 2\eta C\big) \cdot \epsilon^{-\gamma} \tag{3.25}$$

where

$$\xi = \begin{cases} h^{(1)|b}\big(\boldsymbol{y}, \eta\boldsymbol{W}_1^{>\delta}(\eta), \eta\tau_1^{>\delta}(\eta)\big) & \text{if } \eta\tau_1^{>\delta}(\eta) \le 1, \\ h^{(0)|b}(\boldsymbol{y}) & \text{if } \eta\tau_1^{>\delta}(\eta) > 1. \end{cases}$$

Proof. (a) Recall that $\boldsymbol{y}_t(\boldsymbol{x})$ defined in (2.28) is the solution to ODE $d\boldsymbol{y}_t(\boldsymbol{x})/dt = \boldsymbol{a}\big(\boldsymbol{y}_t(\boldsymbol{x})\big)$ under initial condition $\boldsymbol{y}_0(\boldsymbol{x}) = \boldsymbol{x}$. By definition of ξ, we have $\xi_t = \boldsymbol{y}_t(\boldsymbol{y})$ for any $t \in [0,1]$ with $t < \eta\tau_1^{>\delta}(\eta)$. Also, since $\tau_1^{>\delta}(\eta)$ only takes integer values, we know that $\eta\tau_1^{>\delta}(\eta) \le 1 \iff \tau_1^{>\delta}(\eta) \le \lfloor 1/\eta \rfloor$ and $\eta\tau_1^{>\delta}(\eta) > 1 \iff \tau_1^{>\delta}(\eta) > \lfloor 1/\eta \rfloor$.

Let $A \triangleq \Big\{ \max_{i \le \lfloor 1/\eta \rfloor \wedge \left(\tau_1^{>\delta}(\eta) - 1\right)} \eta \left\| \sum_{j=1}^{i} \boldsymbol{\sigma}\big(\boldsymbol{X}_{j-1}^{\eta}(\boldsymbol{x})\big)\boldsymbol{Z}_j \right\| \le \epsilon \Big\}$. Let $\boldsymbol{x}_{\cdot}^{\eta}(\cdot)$ be the deterministic process defined in (C.13). Applying discrete version of Gronwall's inequality (see, for example, Lemma A.3 of [41]) we know that on event A,

$$\left\| \boldsymbol{x}_j^{\eta}(\boldsymbol{x}) - \boldsymbol{X}_j^{\eta}(\boldsymbol{x}) \right\| \le \epsilon \cdot \exp(\eta D \cdot \lfloor 1/\eta \rfloor) \le \rho\epsilon \qquad \forall j \le \lfloor 1/\eta \rfloor \wedge \big(\tau_1^{>\delta}(\eta) - 1\big). \tag{3.26}$$

On the other hand, since $\xi_t = \boldsymbol{y}_t(\boldsymbol{y})$ for all $t < \eta\tau_1^{>\delta}(\eta)$, by applying Lemma C.5 we get

$$\sup_{t\in[0,1]:\ t<\eta\tau_1^{>\delta}(\eta)} \left\|\xi_t - \boldsymbol{x}^\eta_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\right\| \leq \big(\eta C + \|\boldsymbol{x}-\boldsymbol{y}\|\big)\cdot\rho. \tag{3.27}$$

Combining (3.26) and (3.27), we get

$$\sup_{t\in[0,1]:t<\eta\tau_1^{>\delta}(\eta)} \left\|\xi_t - \boldsymbol{X}^\eta_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\right\| \leq \rho\cdot\Big(\epsilon + \|\boldsymbol{x}-\boldsymbol{y}\| + \eta C\Big). \tag{3.28}$$

(b) Note that for any $\boldsymbol{x}\in\mathbb{R}^m$ and any $t\in[0,1]$ with $t<\eta\tau_1^{>\delta}(\eta)$,

$$h^{(0)|b}(\boldsymbol{x})(t) = h^{(0)}(\boldsymbol{x})(t) = h^{(1)|b}\big(\boldsymbol{x},\eta\boldsymbol{W}_1^{>\delta}(\eta),\eta\tau_1^{>\delta}(\eta)\big)(t) = h^{(1)}\big(\boldsymbol{x},\eta\boldsymbol{W}_1^{>\delta}(\eta),\eta\tau_1^{>\delta}(\eta)\big)(t) = \boldsymbol{y}_t(\boldsymbol{x}).$$

Also, for any $\boldsymbol{w}\in\mathbb{R}^d$ with $\|\boldsymbol{w}\|\leq\delta<\frac{b}{2C}$ and any $\boldsymbol{x}\in\mathbb{R}^m$ note that $\varphi_b\Big(\eta\boldsymbol{a}(\boldsymbol{x})+\boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{w}\Big) = \eta\boldsymbol{a}(\boldsymbol{x})+\boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{w}$ due to $\eta\|\boldsymbol{a}(\boldsymbol{x})\|\leq\eta C<\frac{b}{2}$ and $\|\boldsymbol{\sigma}(\boldsymbol{x})\|\,\|\boldsymbol{w}\|\leq C\delta<b/2$ (recall our choice of $\eta C<\frac{b}{2}\wedge 1$ and $\delta<\frac{b}{2C}$). As a result, $\boldsymbol{X}^\eta_j(\boldsymbol{x}) = \boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ for all $\boldsymbol{x}\in\mathbb{R}^m$ and $j<\tau_1^{>\delta}(\eta)$. It then follows directly from (3.28) that on event $\Big\{\max_{i\leq\lfloor 1/\eta\rfloor\wedge\big(\tau_1^{>\delta}(\eta)-1\big)}\eta\left\|\sum_{j=1}^i\boldsymbol{\sigma}\big(\boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x})\big)\boldsymbol{Z}_j\right\|\leq\epsilon\Big\}$, we have

$$\sup_{t\in[0,1]:t<\eta\tau_1^{>\delta}(\eta)} \left\|\xi_t - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\right\| \leq \rho\cdot\big(\epsilon + \|\boldsymbol{x}-\boldsymbol{y}\| + \eta C\big).$$

A direct consequence is (we write $\boldsymbol{y}(u;\boldsymbol{x}) = \boldsymbol{y}_u(\boldsymbol{x})$, $\boldsymbol{y}(s-;\boldsymbol{x}) = \lim_{u\uparrow s}\boldsymbol{y}_u(\boldsymbol{x})$, and $\xi(t)=\xi_t$ in this proof)

$$\left\|\boldsymbol{y}(\eta\tau_1^{>\delta}(\eta)-;\boldsymbol{y}) - \boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\right\| \leq \rho\cdot\big(\epsilon+\|\boldsymbol{x}-\boldsymbol{y}\|+\eta C\big). \tag{3.29}$$

Therefore,

$$\begin{aligned}
&\left\|\xi\big(\eta\tau_1^{>\delta}(\eta)\big) - \boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)}(\boldsymbol{x})\right\| \\
&= \Bigg\|\boldsymbol{y}(\eta\tau_1^{>\delta}(\eta)-;\boldsymbol{y}) + \varphi_b\Big(\eta\boldsymbol{\sigma}\Big(\boldsymbol{y}(\eta\tau_1^{>\delta}(\eta)-;\boldsymbol{y})\Big)\boldsymbol{W}_1^{>\delta}(\eta)\Big) \\
&\qquad - \Big[\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x}) + \varphi_b\Big(\eta\boldsymbol{a}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big) + \eta\boldsymbol{\sigma}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big)\boldsymbol{W}_1^{>\delta}(\eta)\Big)\Big]\Bigg\| \\
&\leq \left\|\boldsymbol{y}(\eta\tau_1^{>\delta}(\eta)-;\boldsymbol{y}) - \boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\right\| \\
&\qquad + \underbrace{\left\|\varphi_b\Big(\eta\boldsymbol{\sigma}\Big(\boldsymbol{y}(\eta\tau_1^{>\delta}(\eta)-;\boldsymbol{y})\Big)\boldsymbol{W}_1^{>\delta}(\eta)\Big) - \varphi_b\Big(\eta\boldsymbol{\sigma}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big)\boldsymbol{W}_1^{>\delta}(\eta)\Big)\right\|}_{\triangleq I_1} \\
&\qquad + \underbrace{\left\|\varphi_b\Big(\eta\boldsymbol{\sigma}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big)\boldsymbol{W}_1^{>\delta}(\eta)\Big) - \varphi_b\Big(\eta\boldsymbol{a}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big) + \eta\boldsymbol{\sigma}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big)\boldsymbol{W}_1^{>\delta}(\eta)\Big)\right\|}_{\triangleq I_2}.
\end{aligned}$$

First, due to $\|\varphi_b(\boldsymbol{x})-\varphi_b(\boldsymbol{y})\|\leq\|\boldsymbol{x}-\boldsymbol{y}\|$,

$$\begin{aligned}
I_1 &\leq \eta\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\|\cdot\left\|\boldsymbol{\sigma}\Big(\boldsymbol{y}(\eta\tau_1^{>\delta}(\eta)-;\boldsymbol{y})\Big) - \boldsymbol{\sigma}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big)\right\| \\
&\leq \eta\left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\|\cdot D\cdot\left\|\boldsymbol{y}(\eta\tau_1^{>\delta}(\eta)-;\boldsymbol{y}) - \boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\right\| \qquad \text{by Assumption 2}
\end{aligned}$$

$$\leq \rho D\big(\epsilon + \|\boldsymbol{x} - \boldsymbol{y}\| + \eta C\big) \cdot \eta \left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\| \qquad \text{by (3.29)}$$
$$\leq \rho D\big(\epsilon + \|\boldsymbol{x} - \boldsymbol{y}\| + \eta C\big) \cdot \epsilon^{-\gamma} \qquad \text{on event } \Big\{\eta \left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\| \leq 1/\epsilon^{\gamma}\Big\}.$$

Similarly, we can get $I_2 \leq \left\|\eta \boldsymbol{a}\Big(\boldsymbol{X}^{\eta|b}_{\tau_1^{>\delta}(\eta)-1}(\boldsymbol{x})\Big)\right\| \leq \eta C$. In summary, on event $\Big\{\eta \left\|\boldsymbol{W}_1^{>\delta}(\eta)\right\| \leq 1/\epsilon^{\gamma}\Big\}$,

$$\sup_{t \in [0,1]:\ t \leq \eta\tau_1^{>\delta}(\eta)} \left\|\xi_t - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x})\right\| \leq \rho D\big(\epsilon + \|\boldsymbol{x} - \boldsymbol{y}\| + \eta C\big) \cdot \epsilon^{-\gamma} + \eta C$$
$$\leq \rho D\big(\epsilon + \|\boldsymbol{x} - \boldsymbol{y}\| + 2\eta C\big) \cdot \epsilon^{-\gamma}.$$

This concludes the proof of part (b). □

By applying Lemma 3.8 inductively, the next result establishes the conditions under which the image of the mapping $h^{(k)|b}$ approximates the path of $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$.

Lemma 3.9. *Let Assumptions 2 and 3 hold. Let $A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})$ be defined as in (3.6). For any $k \in \mathbb{N}$, $\boldsymbol{x} \in \mathbb{R}^m$, $b > 0$, $\epsilon \in (0,1)$, $\delta \in (0, \frac{b}{2C})$, and $\eta \in (0, \frac{b \wedge \epsilon}{2C})$, it holds on event*

$$\Big(\bigcap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\Big) \cap \Big\{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta)\Big\} \cap \Big\{\eta \left\|\boldsymbol{W}_i^{>\delta}(\eta)\right\| \leq 1/\epsilon^{\frac{1}{2k}}\ \forall i \in [k]\Big\}$$

that
$$\sup_{t \in [0,1]} \left\|\xi_t - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x})\right\| < (2\rho D)^{k+1}\sqrt{\epsilon},$$

where
$$\xi \triangleq h^{(k)|b}\Big(\boldsymbol{x}, \big(\eta \boldsymbol{W}_1^{>\delta}(\eta), \cdots, \eta \boldsymbol{W}_k^{>\delta}(\eta)\big), \big(\eta\tau_1^{>\delta}(\eta), \cdots, \eta\tau_k^{>\delta}(\eta)\big)\Big),$$

$\rho = \exp(D) \geq 1$, $D \in [1, \infty)$ is the Lipschitz coefficient in Assumption 2, and $C \geq 1$ is the constant in Assumption 3.

Proof. It is straightforward to see the claim is an immeidate corollary of (3.25) in Lemma 3.8 when applied inductively (in particular, with $\gamma = \frac{1}{2k}$, and note that due to our choice of η, we have $2\eta C < \epsilon$). To avoid repetition, we omit the details. e □

To conclude, Lemma 3.10 provides tools for verifying the sequential compactness condition (2.1) for measures $\mathbf{C}^{(k)}(\,\cdot\,; \boldsymbol{x})$ and $\mathbf{C}^{(k)|b}(\,\cdot\,; \boldsymbol{x})$ when we restrict $\boldsymbol{x}$ over a compact set A.

Lemma 3.10. *Let Assumptions 2 and 3 hold. Let $T, r > 0$ and $k \geq 1$. Let $A \subseteq \mathbb{R}^m$ be compact.*

(a) For any $\boldsymbol{x}_n \in A$ and $\boldsymbol{x}^* \in A$ such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$,

$$\lim_{n\to\infty} \mathbf{C}^{(k)}(f; \boldsymbol{x}_n) = \mathbf{C}^{(k)}(f; \boldsymbol{x}^*) \qquad \forall f \in \mathcal{C}\big(\mathbb{D}[0,T] \setminus \mathbb{D}_A^{(k-1)}[0,T](r)\big).$$

(b) Let $b > 0$. For any $\boldsymbol{x}_n \in A$ and $\boldsymbol{x}^* \in A$ such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$,

$$\lim_{n\to\infty} \mathbf{C}^{(k)|b}(f; \boldsymbol{x}_n) = \mathbf{C}^{(k)|b}(f; \boldsymbol{x}^*) \qquad \forall f \in \mathcal{C}\big(\mathbb{D}[0,T] \setminus \mathbb{D}_A^{(k-1)|b}[0,T](r)\big).$$

Proof. For convenience we consider the case $T = 1$, but the proof can easily extend for arbitrary $T > 0$.
(a) Pick some $f \in \mathcal{C}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)}(r)\big)$, and let $\phi(\boldsymbol{x}) \triangleq \mathbf{C}^{(k)}(f; \boldsymbol{x})$. We argue that $\phi(\cdot)$ is a continuous function using Dominated Convergence theorem. First, from the continuity of f and $h^{(k)}$ (see Lemma C.4), for any sequence $\boldsymbol{y}_n \in \mathbb{R}^m$ with $\lim_{n\to\infty} \boldsymbol{y}_n = \boldsymbol{y} \in \mathbb{R}^m$, we have

$$\lim_{m\to\infty} f\Big(h^{(k)}(\boldsymbol{y}_m, \mathbf{W}, \boldsymbol{t})\Big) = f\Big(h^{(k)}(\boldsymbol{y}, \mathbf{W}, \boldsymbol{t})\Big) \qquad \forall \mathbf{W} \in \mathbb{R}^{d \times k},\ \boldsymbol{t} \in (0,1)^{k\uparrow}.$$

Next, by applying Lemma 3.5 onto $B = \text{supp}(f)$, which is bounded away from $\mathbb{D}_A^{(k-1)}(r)$, we find $\bar{\delta} > 0$ such that $h^{(k)}\big(\boldsymbol{x}, (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k), \boldsymbol{t}\big) \in B \Longrightarrow \|\boldsymbol{w}_j\| > \bar{\delta}\ \forall j \in [k]$. As a result,

$$\Big| f\Big(h^{(k)}\big(\boldsymbol{x}, (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k), \boldsymbol{t}\big)\Big)\Big| \le \|f\| \cdot \mathbb{I}\big\{ \|\boldsymbol{w}_j\| > \bar{\delta}\ \forall j \in [k]\big\}.$$

Also, note that $\int \mathbb{I}\big\{ \|\boldsymbol{w}_j\| > \bar{\delta}\ \forall j \in [k]\big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^k(d\mathbf{W}) \times \mathcal{L}_1^{k\uparrow}(d\boldsymbol{t}) \le 1/\bar{\delta}^{k\alpha} < \infty$. This allows us to apply Dominated Convergence theorem and obtain

$$\lim_{n\to\infty} \phi(\boldsymbol{x}_n) = \lim_{n\to\infty} \mathbf{C}^{(k)}(f; \boldsymbol{x}_n) = \mathbf{C}^{(k)}(f; \boldsymbol{x}^*) = \phi(\boldsymbol{x}^*),$$

which the proof of part (a).

(b) The proof is almost identical. The only differences are that we apply Lemma C.3 (resp. Lemma 3.6) instead of Lemma C.4 (resp. Lemma 3.5) so we omit the details. $\square$

3.3 Proofs of Theorems 2.3 and 2.4

In the proofs of Theorems 2.3 and 2.4 below, without loss of generality we focus on the case where $T = 1$. But we note that the proof for the cases with arbitrary $T > 0$ is nearly identical. Recall that, to simplify notations, we write $\boldsymbol{X}^\eta(\boldsymbol{x}) = \boldsymbol{X}^\eta_{[0,1]}(\boldsymbol{x}) = \{\boldsymbol{X}^\eta_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) :\ t \in [0,1]\}$, and $\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) = \boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}) = \{\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) :\ t \in [0,1]\}$.

3.3.1 Proof of Theorem 2.3

Recall the notion of uniform $\mathbb{M}$-convergence introduced in Definition 2.1. At first glance, the uniform version of $\mathbb{M}$-convergence stated in Theorem 2.3 and 2.4 is stronger than the standard $\mathbb{M}$-convergence introduced in [42]. Nevertheless, under the conditions stated in Theorem 2.3 or 2.4 regarding the initial values of $\boldsymbol{X}^\eta$ or $\boldsymbol{X}^{\eta|b}$, we can show that it suffices to prove the standard notion of $\mathbb{M}$-convergence. In particular, the proofs to Theorem 2.3 and 2.4 hinge on the following key proposition for $\boldsymbol{X}^{\eta|b}$.

Proposition 3.11. *Let η_n be a sequence of strictly positive real numbers with $\lim_{n\to\infty} \eta_n = 0$. Let compact set $A \subseteq \mathbb{R}^m$ and $\boldsymbol{x}_n, \boldsymbol{x}^* \in A$ be such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$. Under Assumptions 1 and 2, it holds for all $k \in \mathbb{N}$ and $b, r > 0$ that*

$$\mathbf{P}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \in \ \cdot\ \big)/\lambda^k(\eta_n) \to \mathbf{C}^{(k)|b}\big(\ \cdot\ ; \boldsymbol{x}^*\big) \text{ in } \mathbb{M}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)|b}(r)\big) \text{ as } n \to \infty.$$

As the first application of Proposition 3.11, in Section 3.3.1 we prepare a similar result for the unclipped dynamics $\boldsymbol{X}^\eta$ defined in (2.3) and (2.16), which will be the key tool in our proof of Theorem 2.3.

Proposition 3.12. *Let η_n be a sequence of strictly positive real numbers with $\lim_{n\to\infty} \eta_n = 0$. Let compact set $A \subseteq \mathbb{R}^m$ and $\boldsymbol{x}_n, \boldsymbol{x}^* \in A$ be such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$. Under Assumptions 1, 2, and 3, it holds for all $k \in \mathbb{N}$ and $r > 0$ that*

$$\mathbf{P}\big(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n) \in \ \cdot\ \big)/\lambda^k(\eta_n) \to \mathbf{C}^{(k)}\big(\ \cdot\ ; \boldsymbol{x}^*\big) \text{ in } \mathbb{M}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)}(r)\big) \text{ as } n \to \infty.$$

Proof. Fix some $k = 0, 1, 2, \cdots$, $r > 0$, and some $g \in \mathcal{C}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)}(r)\big)$. By virtue of the Portmanteau theorem for $\mathbb{M}$-convergence (see theorem 2.1 of [42]), it suffices to show that

$$\lim_{n\to\infty} \mathbf{E}\big[g\big(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\big)\big]/\lambda^k(\eta_n) = \mathbf{C}^{(k)}(g; \boldsymbol{x}^*).$$

To this end, we let $B \triangleq \text{supp}(g)$ and observe that for any $n \ge 1$ and any $\delta, b > 0$,

$$\mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\big)\Big]$$

$$
\begin{aligned}
&= \mathbf{E}\Big[g(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n))\mathbb{I}\Big\{\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\in B\Big\}\Big] \\
&= \mathbf{E}\Big[g(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n))\mathbb{I}\{\tau^{>\delta}_{k+1}(\eta_n) < \lfloor 1/\eta_n\rfloor;\ \boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\in B\}\Big] \\
&\quad + \mathbf{E}\Big[g(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n))\mathbb{I}\Big\{\tau^{>\delta}_{k}(\eta_n) > \lfloor 1/\eta_n\rfloor;\ \boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\in B\Big\}\Big] \\
&\quad + \mathbf{E}\Bigg[g(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n))\mathbb{I}\Bigg\{\tau^{>\delta}_{k}(\eta_n) \le \lfloor 1/\eta_n\rfloor < \tau^{>\delta}_{k+1}(\eta_n);\ \eta_n\left\|\boldsymbol{W}^{>\delta}_j(\eta_n)\right\| > \frac{b}{2C}\ \text{ for some } j\in[k];\ \boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\in B\Bigg\}\Bigg] \\
&\quad + \mathbf{E}\Bigg[\underbrace{g(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n))\mathbb{I}\Bigg\{\tau^{>\delta}_{k}(\eta_n) \le \lfloor 1/\eta_n\rfloor < \tau^{>\delta}_{k+1}(\eta_n);\ \eta_n\left\|\boldsymbol{W}^{>\delta}_j(\eta_n)\right\| \le \frac{b}{2C}\ \forall j\in[k];\ \boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\in B\Bigg\}}_{\triangleq I_*(n,b,\delta)}\Bigg],
\end{aligned}
$$

where $C \ge 1$ is the constant in Assumption 3 such that $\|\boldsymbol{a}(\boldsymbol{x})\| \vee \|\boldsymbol{\sigma}(\boldsymbol{x})\| \le C$ for any $\boldsymbol{x}$, and $\tau^{>\delta}_j(\eta)$'s, $\boldsymbol{W}^{>\delta}_j(\eta)$'s are defined in (3.2) and (3.3). Now, we focus on the term $I_*(n,b,\delta)$. For any n large enough, we have $\eta_n \cdot \sup_{\boldsymbol{x}\in\mathbb{R}^m}\|\boldsymbol{a}(\boldsymbol{x})\| \le \eta_n C \le b/2$. As a result,for such n and any $\delta \in (0, \frac{b}{2C})$, on the event

$$
\widetilde{A}(n,b,\delta) \triangleq \Bigg\{\tau^{>\delta}_{k}(\eta_n) \le \lfloor 1/\eta_n\rfloor < \tau^{>\delta}_{k+1}(\eta_n);\ \eta_n\left\|\boldsymbol{W}^{>\delta}_j(\eta_n)\right\| \le \frac{b}{2C}\ \forall j\in[k];\ \boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\in B\Bigg\},
$$

the norm of the step-size (before truncation) $\eta\boldsymbol{a}\big(\boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x})\big)\boldsymbol{Z}_j$ of $\boldsymbol{X}^{\eta|b}_j$ is less than b for each $j \le \lfloor 1/\eta_n\rfloor$, and hence $\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n) = \boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)$. This observation leads to the following upper bound: Given any $b > 0$ and $\delta \in (0, \frac{b}{2C})$, it holds for any n large enough that

$$
\begin{aligned}
\mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\big)\Big] &\le \|g\|\underbrace{\mathbf{P}\big(\tau^{>\delta}_{k+1}(\eta_n) \le \lfloor 1/\eta_n\rfloor\big)}_{\triangleq p_1(n,\delta)} \\
&\quad + \|g\|\underbrace{\mathbf{P}\big(\tau^{>\delta}_{k}(\eta_n) > \lfloor 1/\eta_n\rfloor;\ \boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\in B\big)}_{\triangleq p_2(n,\delta)} \\
&\quad + \|g\|\underbrace{\mathbf{P}\bigg(\tau^{>\delta}_{k}(\eta_n) \le \lfloor 1/\eta_n\rfloor < \tau^{>\delta}_{k+1}(\eta_n);\ \eta_n\left\|\boldsymbol{W}^{>\delta}_j(\eta_n)\right\| > \frac{b}{2C}\ \text{for some } j\in[k]\bigg)}_{\triangleq p_3(n,b,\delta)} \\
&\quad + \mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big].
\end{aligned}
$$

Similarly, given any n large enough, any $b > 0$ and any $\delta \in (0, \frac{b}{2C})$, we have the lower bound

$$
\begin{aligned}
\mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n)\big)\Big] &\ge \mathbf{E}\big[I_*(n,b,\delta)\big] \\
&= \mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\mathbb{I}\Big(\widetilde{A}(n,b,\delta)\Big)\Big] \qquad \text{due to } \boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n) = \boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \text{ on } \widetilde{A}(n,b,\delta) \\
&\ge \mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big] - \|g\|\,\mathbf{P}\Big(\big(\widetilde{A}(n,b,\delta)\big)^c\Big) \\
&\ge \mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big] - \|g\|\cdot\big[p_1(n,\delta) + p_2(n,\delta) + p_3(n,b,\delta)\big].
\end{aligned}
$$

We claim that there exists some $\delta > 0$ such that

$$
\lim_{n\to\infty} p_1(n,\delta)/\lambda^k(\eta_n) = 0, \tag{3.30}
$$

$$
\lim_{n\to\infty} p_2(n,\delta)/\lambda^k(\eta_n) = 0. \tag{3.31}
$$

Furthermore, we claim that for any $b > 0$,

$$\limsup_{n\to\infty} p_3(n,b,\delta)/\lambda^k(\eta_n) \leq \psi_\delta(b) \triangleq \frac{k}{\delta^{\alpha k}} \cdot \big(\frac{\delta}{2C}\big)^\alpha \cdot \frac{1}{b^\alpha}. \tag{3.32}$$

Note that $\lim_{b\to\infty}\psi_\delta(b) = 0$. Lastly, by Lemma 3.7,

$$\lim_{b\to\infty} \mathbf{C}^{(k)|b}(g;\boldsymbol{x}^*) = \mathbf{C}^{(k)}(g;\boldsymbol{x}^*). \tag{3.33}$$

Then by combining (3.30)–(3.32) with the upper and lower bounds above for $\mathbf{E}\big[g(\boldsymbol{X}^{\eta_n}(x_n))\big]$, we see that for any b large enough (such that $\frac{b}{2C} > \delta$),

$$\lim_{n\to\infty}\frac{\mathbf{E}\big[g(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n))\big]}{\lambda^k(\eta_n)} - \|g\|\,\psi_\delta(b) \leq \lim_{n\to\infty}\frac{\mathbf{E}\big[g(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n))\big]}{\lambda^k(\eta_n)} \leq \lim_{n\to\infty}\frac{\mathbf{E}\big[g(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n))\big]}{\lambda^k(\eta_n)} + \|g\|\,\psi_\delta(b),$$
$$\Longrightarrow -\|g\|\,\psi_\delta(b) + \mathbf{C}^{(k)|b}(g;\boldsymbol{x}^*) \leq \lim_{n\to\infty}\frac{\mathbf{E}\big[g(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n))\big]}{\lambda^k(\eta_n)} \leq \|g\|\,\psi_\delta(b) + \mathbf{C}^{(k)|b}(g;\boldsymbol{x}^*).$$

In the last line of the display, we applied Proposition 3.11. Letting b tend to ∞ and applying the limit (3.33), we conclude the proof. Now, it only remains to prove (3.30) (3.31) (3.32).

Proof of Claim (3.30):

We show that this claim holds for any $\delta > 0$. Applying (3.4), we see that $p_1(n,\delta) \leq \big(H(\frac{\delta}{\eta_n})/\eta_n\big)^{k+1}$ holds for any $\delta > 0$ and any $n \geq 1$. Due to the regularly varying nature of $H(\cdot)$, we then yield $\limsup_{n\to\infty}\frac{p_1(n,\delta)}{\lambda^{k+1}(\eta_n)} \leq 1/\delta^{\alpha(k+1)} < \infty$. Therefore,

$$\limsup_{n\to\infty}\frac{p_1(n,\delta)}{\lambda^k(\eta_n)} \leq \limsup_{n\to\infty}\frac{p_1(n,\delta)}{\lambda^{k+1}(\eta_n)}\cdot\lim_{n\to\infty}\lambda(\eta_n) \leq \frac{1}{\delta^{\alpha(k+1)}}\cdot\lim_{n\to\infty}\frac{H(1/\eta_n)}{\eta_n} = 0$$

due to $\frac{H(1/\eta)}{\eta} = \lambda(\eta) \in \mathcal{RV}_{\alpha-1}(\eta)$ as $\eta \downarrow 0$ and $\alpha > 1$.

Proof of Claim (3.31):

We claim the existence of some $\epsilon > 0$ such that

$$\Big\{\tau_k^{>\delta}(\eta) > \lfloor 1/\eta\rfloor;\ \boldsymbol{X}^\eta(\boldsymbol{x}) \in B\Big\} \cap \Bigg(\bigcap_{i=1}^{k+1} A_i(\eta,\infty,\epsilon,\delta,\boldsymbol{x})\Bigg) = \emptyset \quad \forall \boldsymbol{x}\in A,\ \delta > 0,\ \eta \in (0,\frac{\epsilon}{C\rho}) \tag{3.34}$$

where $D, C \in [1,\infty)$ are the constants in Assumptions 2 and 3 respectively, $\rho \triangleq \exp(D)$, and event $A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})$ is defined in (3.6). Then for any $\delta > 0$,

$$\limsup_{n\to\infty} p_2(n,\delta)/\lambda^k(\eta_n) \leq \limsup_{n\to\infty}\sup_{x\in A}\mathbf{P}\Bigg(\Bigg(\bigcap_{i=1}^{k+1}A_i(\eta_n,\infty,\epsilon,\delta,\boldsymbol{x})\Bigg)^c\Bigg)\Big/\lambda^k(\eta_n).$$

Applying Lemma 3.3 (b) with $N > k(\alpha-1)$, we conclude that claim (3.31) holds for all $\delta > 0$ small enough. Now, it only remains to find $\epsilon > 0$ that satisfies condition (3.34). To this end, we first recall that the set $B = \text{supp}(g)$ is bounded away from $\mathbb{D}_A^{(k-1)}(r)$. By Lemma 3.5, there is $\bar\epsilon > 0$ such that $\boldsymbol{d}_{J_1}\big(B^{\bar\epsilon},\mathbb{D}_A^{(k-1)}(r)\big) > \bar\epsilon$. W.l.o.g. we pick $\bar\epsilon$ small enough such that $\bar\epsilon \in (0,r)$. Next, we show that (3.34) holds for any $\epsilon > 0$ small enough with $(\rho+1)\epsilon < \bar\epsilon$. To see why, we fix some $\boldsymbol{x}\in A$, $\delta > 0$ and $\eta \in (0,\frac{\epsilon}{C\rho})$. Define a process $\breve{\boldsymbol{X}}^{\eta,\delta}(\boldsymbol{x}) \triangleq \big\{\breve{\boldsymbol{X}}_t^{\eta,\delta}(\boldsymbol{x}):\ t\in[0,1]\big\}$ as the solution to (under initial condition $\breve{\boldsymbol{X}}_0^{\eta,\delta}(\boldsymbol{x}) = \boldsymbol{x}$)

$$\frac{d\breve{\boldsymbol{X}}_t^{\eta,\delta}(\boldsymbol{x})}{dt} = \boldsymbol{a}\big(\breve{\boldsymbol{X}}_t^{\eta,\delta}(\boldsymbol{x})\big) \qquad \forall t \geq 0,\ t \notin \{\eta\tau_j^{>\delta}(\eta):\ j\geq 1\},$$

$$\breve{\boldsymbol{X}}^{\eta,\delta}_{\eta\tau_i^{>\delta}(\eta)}(\boldsymbol{x}) = \boldsymbol{X}^{\eta}_{\tau_i^{>\delta}(\eta)}(\boldsymbol{x}) \qquad \forall j \geq 1.$$

On event $\big(\cap_{i=1}^{k+1} A_i(\eta,\infty,\epsilon,\delta,.x)\big) \cap \{\tau_k^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\}$, observe that

$$\begin{aligned}
&\boldsymbol{d}_{J_1}\Big(\breve{\boldsymbol{X}}^{\eta,\delta}(\boldsymbol{x}), \boldsymbol{X}^{\eta}(\boldsymbol{x})\Big) \\
&\leq \sup_{t \in \big[0,\eta\tau_1^{>\delta}(\eta)\big) \cup \big[\eta\tau_1^{>\delta}(\eta),\eta\tau_2^{>\delta}(\eta)\big) \cup \cdots \cup \big[\eta\tau_k^{>\delta}(\eta),\eta\tau_{k+1}^{>\delta}(\eta)\big)} \Big\| \breve{\boldsymbol{X}}^{\eta,\delta}_t(\boldsymbol{x}) - \boldsymbol{X}^{\eta}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) \Big\| \\
&\leq \rho \cdot \big(\epsilon + \eta C\big) \leq \rho\epsilon + \epsilon < \bar{\epsilon} \qquad \text{because of (3.23) of Lemma 3.8.}
\end{aligned}$$

In the last line of the display above, we applied $\eta < \frac{\epsilon}{C\rho}$ and our choice of $(\rho+1)\epsilon < \bar{\epsilon}$. However, from the display above, we also learned that on $\{\tau_k^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\}$, we have $\breve{\boldsymbol{X}}^{\eta,\delta}(x) \in \mathbb{D}_A^{(k-1)}(\bar{\epsilon}) \subseteq \mathbb{D}_A^{(k-1)}(r)$; recall that we picked $\bar{\epsilon} \in (0,r)$. As a result, on event $\big(\cap_{i=1}^{k+1} A_i(\eta,\infty,\epsilon,\delta,\boldsymbol{x})\big) \cap \{\tau_k^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\}$ we must have $\boldsymbol{d}_{J_1}\big(\mathbb{D}_A^{(k-1)}(r), \boldsymbol{X}^{\eta}(\boldsymbol{x})\big) < \bar{\epsilon}$, and hence $\boldsymbol{X}^{\eta}(\boldsymbol{x}) \notin B$ due to the fact that $\boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)}(r)\big) > \bar{\epsilon}$. This verifies (3.34).

Proof of Claim (3.32):

Due to the independence between $\big(\tau_i^{>\delta}(\eta) - \tau_{j-1}^{\eta}(\delta)\big)_{j\geq 1}$ and $\big(\boldsymbol{W}_i^{>\delta}(\eta)\big)_{j\geq 1}$,

$$\begin{aligned}
p_3(n,b,\delta) &= \mathbf{P}\Big(\tau_k^{>\delta}(\eta_n) < \lfloor 1/\eta_n \rfloor < \tau_{k+1}^{>\delta}(\eta_n)\Big) \cdot \mathbf{P}\bigg(\eta_n \left\|\boldsymbol{W}_j^{>\delta}(\eta_n)\right\| > \frac{b}{2C} \text{ for some } j \in [k]\bigg) \\
&\leq \mathbf{P}\Big(\tau_k^{>\delta}(\eta_n) \leq \lfloor 1/\eta_n \rfloor\Big) \cdot \sum_{j=1}^{k} \mathbf{P}\bigg(\eta_n \left\|\boldsymbol{W}_j^{>\delta}(\eta_n)\right\| > \frac{b}{2C}\bigg) \\
&\leq \Big(\frac{H(\delta/\eta_n)}{\eta_n}\Big)^k \cdot k \cdot \frac{H\big(\frac{b}{2C}\cdot\frac{1}{\eta_n}\big)}{H\big(\delta\cdot\frac{1}{\eta_n}\big)}.
\end{aligned}$$

Due to $H(x) \in \mathcal{RV}_{-\alpha}(x)$ as $x \to \infty$ (see Assumption 1), we conclude that $\limsup_{n\to\infty} \frac{p_4(n,b,\delta)}{\lambda^k(\eta_n)} \leq \frac{k}{\delta^{\alpha k}} \cdot \big(\frac{\delta}{2C}\big)^{\alpha} \cdot \frac{1}{b^{\alpha}} = \psi_\delta(b)$.

□

With Proposition 3.12 in our arsenal, we prove Theorem 2.3.

Proof of Theorem 2.3. For simplicity of notations we focus on the case where $T = 1$, but the proof below can be easily generalized for arbitrary $T > 0$.

We first prove the uniform $\mathbb{M}$-convergence. Specifically, we proceed with a proof by contradiction. Fix some $r > 0$ and $k \in \mathbb{N}$, and suppose that there is some $f \in \mathcal{C}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)}(r)\big)$, some sequence $\eta_n > 0$ with limit $\lim_{n\to\infty} \eta_n = 0$, some sequence $\boldsymbol{x}_n \in A$, and $\epsilon > 0$ such that

$$\left|\mu_n^{(k)}(f) - \mathbf{C}^{(k)}(f;\boldsymbol{x}_n)\right| > \epsilon \ \forall n \geq 1 \qquad \text{with } \mu_n^{(k)}(\cdot) \triangleq \mathbf{P}\big(\boldsymbol{X}^{\eta_n}(\boldsymbol{x}_n) \in \ \cdot\ \big)/\lambda^k(\eta_n).$$

Since $A \subseteq \mathbb{R}^m$ is compact, by picking a proper subsequence we can assume w.l.o.g. that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$ for some $\boldsymbol{x}^* \in A$. This allows us to apply Proposition 3.12 and yield $\lim_{n\to\infty}\left|\mu_n^{(k)}(f) - \mathbf{C}^{(k)}(f;\boldsymbol{x}^*)\right| = 0$. On the other hand, using part (a) of Lemma 3.10, we get $\lim_{n\to\infty}\left|\mathbf{C}^{(k)}(f;\boldsymbol{x}_n) - \mathbf{C}^{(k)}(f;\boldsymbol{x}^*)\right| = 0$. Therefore, we arrive at the contradiction

$$\lim_{n\to\infty}\left|\mu_n^{(k)}(f) - \mathbf{C}^{(k)}(f;\boldsymbol{x}_n)\right| \leq \lim_{n\to\infty}\left|\mu_n^{(k)}(f) - \mathbf{C}^{(k)}(f;\boldsymbol{x}^*)\right| + \lim_{n\to\infty}\left|\mathbf{C}^{(k)}(f;\boldsymbol{x}^*) - \mathbf{C}^{(k)}(f;\boldsymbol{x}_n)\right| = 0$$

and conclude the proof of the uniform $\mathbb{M}$-convergence claim.

Next, we prove the uniform sample-path large deviations stated in (2.18). Part (a) of Lemma 3.10 verifies the compactness condition (2.1) for the family of measures $\{\mathbf{C}^{(k)}(\ \cdot\ ;\boldsymbol{x}) : \ \boldsymbol{x} \in A\}$.

In light of the Portmanteau theorem for uniform $\mathbb{M}$-convergence (i.e., Theorem 2.2), most claims follow directly from the uniform $\mathbb{M}$-convergence established above, and it only remains to verify that $\sup_{\boldsymbol{x}\in A}\mathbf{C}^{(k)}\big(B^-;\boldsymbol{x}\big) < \infty$. To do so, note that B^- is bounded away from $\mathbb{D}_A^{(k-1)}(r)$. This allows us to apply Lemma 3.5 and find $\bar{\epsilon} > 0$, $\bar{\delta} > 0$ such that, for any $\boldsymbol{x} \in A$ and $\boldsymbol{t} \in (0,1]^{k\uparrow}$,

$$h^{(k)}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),\boldsymbol{t}\big) \in B^{\bar{\epsilon}} \Longrightarrow \|\boldsymbol{w}_j\| > \bar{\delta}\ \forall j \in [k].$$

Then by the definition of $\mathbf{C}^{(k)}$ in (2.14),

$$\begin{aligned}\sup_{\boldsymbol{x}\in A}\mathbf{C}^{(k)}(B^-;\boldsymbol{x}) &= \sup_{\boldsymbol{x}\in A}\int \mathbb{I}\Big\{h^{(k)}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),\boldsymbol{t}\big) \in B^-\Big\}\big((\nu_\alpha\times\mathbf{S})\circ\Phi\big)^k(d\mathbf{W})\times\mathcal{L}_1^{k\uparrow}(d\boldsymbol{t}q \\ &\le \int \mathbb{I}\Big\{\ \|\boldsymbol{w}_j\| > \bar{\delta}\ \forall j\in[k]\Big\}\big((\nu_\alpha\times\mathbf{S})\circ\Phi\big)^k(d\mathbf{W})\times\mathcal{L}_1^{k\uparrow}(d\boldsymbol{t}) \le 1/\bar{\delta}^{k\alpha} < \infty.\end{aligned}$$

This concludes the proof. □

3.3.2 Proof of Theorem 2.4

Aside from Proposition 3.11, another key tool in our proof of Theorem 2.4 is the following "truncated" version of the drift and diffusion coefficients $\boldsymbol{a}(\cdot),\boldsymbol{\sigma}(\cdot)$. Given any $M \ge 1$, let

$$\boldsymbol{a}_M(\boldsymbol{x}) \triangleq \begin{cases}\boldsymbol{a}\Big(M\cdot\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\Big) & \text{if } \|\boldsymbol{x}\| > M,\\ \boldsymbol{a}(\boldsymbol{x}) & \text{otherwise.}\end{cases}\qquad \boldsymbol{\sigma}_M(\boldsymbol{x}) \triangleq \begin{cases}\boldsymbol{\sigma}\Big(M\cdot\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\Big) & \text{if } \|\boldsymbol{x}\| > M,\\ \boldsymbol{\sigma}(\boldsymbol{x}) & \text{otherwise.}\end{cases} \tag{3.35}$$

That is, we project $\boldsymbol{x}$ onto the closed ball $\{\boldsymbol{x}\in\mathbb{R}^m:\ \|\boldsymbol{x}\|\le M\}$. For any $\boldsymbol{a}(\cdot),\boldsymbol{\sigma}(\cdot)$ satisfying Assumption 2, one can see that $\boldsymbol{a}_M(\cdot),\boldsymbol{\sigma}_M(\cdot)$ will satisfy Assumptions 2 and 3. Similarly, recall the definition of the mapping $\bar{h}^{(k)|b}$ in (2.21)-(2.23). We also consider its "truncated" counterpart by defining the mapping $\bar{h}^{(k)|b}_{M\downarrow}:\mathbb{R}^m\times\mathbb{R}^{d\times k}\times\mathbb{R}^{m\times k}\times(0,1]^{k\uparrow}\to\mathbb{D}$ as follows. Given any $\boldsymbol{x}\in\mathbb{R}^m$, $\mathbf{W} = (\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k)\in\mathbb{R}^{d\times k}$, $\mathbf{V} = (\boldsymbol{v}_1,\cdots,\boldsymbol{v}_j)\in\mathbb{R}^{m\times k}$, $\boldsymbol{t} = (t_1,\cdots,t_k)\in(0,1]^{k\uparrow}$, let $\xi = \bar{h}^{(k)|b}_{M\downarrow}(\boldsymbol{x},\mathbf{W},\mathbf{V},\boldsymbol{t})$ be the solution to

$$\xi_0 = \boldsymbol{x}; \tag{3.36}$$

$$\frac{d\xi_t}{dt} = \boldsymbol{a}_M(\xi_t) \qquad \forall t\in[0,1],\ t\ne t_1,t_2,\cdots,t_k; \tag{3.37}$$

$$\xi_t = \xi_{t-} + \boldsymbol{v}_j + \varphi_b\big(\boldsymbol{\sigma}_M(\xi_{t-}+\boldsymbol{v}_j)\boldsymbol{w}_j\big) \qquad \text{if } t = t_j \text{ for some } j\in[k]. \tag{3.38}$$

Define mapping $h^{(k)|b}_{M\downarrow}:\mathbb{R}^m\times\mathbb{R}^{d\times k}\times(0,1]^{k\uparrow}\to\mathbb{D}$ by

$$h^{(k)|b}_{M\downarrow}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),\boldsymbol{t}\big) \triangleq \bar{h}^{(k)|b}_{M\downarrow}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),(\mathbf{0},\cdots,\mathbf{0}),\boldsymbol{t}\big). \tag{3.39}$$

Also, recall that $\bar{B}_r(\boldsymbol{x})$ is the closed ball with radius r centered at $\boldsymbol{x}$, and set

$$\mathbb{D}^{(k)|b}_{A;M\downarrow}(r) \triangleq \bar{h}^{(k)|b}_{M\downarrow}\Big(A\times\mathbb{R}^{m\times k}\times\big(\bar{B}_r(\mathbf{0})\big)^k\times(0,1]^{k\uparrow}\Big). \tag{3.40}$$

In short, the difference between $\bar{h}^{(k)|b}_{M\downarrow}$ and $\bar{h}^{(k)|b}$ is that, when constructing $\bar{h}^{(k)|b}_{M\downarrow}$, we use the truncated drift and diffusion coefficients $\boldsymbol{a}_M(\cdot)$ and $\boldsymbol{\sigma}_M(\cdot)$.

The main idea for our proof of Theorem 2.4 is as follows. For large enough $M > 0$, one can show that it is very unlikely for the truncated dynamics $\boldsymbol{X}^{\eta|b}(\boldsymbol{x})$ to exit from the the ball $\bar{B}_r(\mathbf{0}) = \{\boldsymbol{y}:\|\boldsymbol{y}\|\le M\}$. Therefore, it suffices to study the $\mathbb{M}$-convergence and large deviation limits of a modified version of $\boldsymbol{X}^{\eta|b}(\boldsymbol{x})$, where we use $\boldsymbol{a}_M$ and $\boldsymbol{\sigma}_M$ for the drift and diffusion coefficients, instead of $\boldsymbol{a}$ and $\boldsymbol{\sigma}$. Since $\boldsymbol{a}_M$ and $\boldsymbol{\sigma}_M$ automatically satisfy the boundedness condition in Assumption 3, we essentially reduce the problem to a simpler one, whose proof is almost identical to that of Theorem 2.3 and builds upon the technical tools developed in Section 3.2 again.

Proof of Theorem 2.4. First, we argue that the proof is almost identical to that of Theorem 2.3 if Assumption 3 also holds. In particular, the proof-by-contradiction approach in Theorem 2.3 can be applied here to establish the uniform $\mathbb{M}$-convergence. The only difference is that we apply Proposition 3.11 (resp., part (*b*) of Lemma 3.10) instead of Proposition 3.12 (resp., part (*a*) of Lemma 3.10). Similarly, the proof to the uniform sample-path large deviations stated in (2.27) is almost identical to that of (2.18) in Theorem 2.3. The only difference is that we apply part (*b*) of Lemma 3.10 (resp., Lemma 3.6) instead of part (*a*) of Lemma 3.10 (resp., Lemma 3.5). To avoid repetition we omit the details.

In the remainder of this proof, we discuss how to extend the proof and cover the case where Assumption 3 is dropped. To prove the uniform $\mathbb{M}$-convergence claim, we proceed again with a proof by contradiction. Fix some $b, r > 0, k \in \mathbb{N}$, and suppose that there are some $g \in \mathcal{C}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)}(r)\big)$, some sequence $\eta_n > 0$ with limit $\lim_{n\to\infty} \eta_n = 0$, some sequence $\boldsymbol{x}_n \in A$, and $\epsilon > 0$ such that

$$\big|\mu_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g; \boldsymbol{x}_n)\big| > \epsilon \ \forall n \geq 1 \qquad \text{with } \mu_n^{(k)}(\cdot) \triangleq \mathbf{P}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \in \ \cdot \ \big)\big/\lambda^k(\eta_n). \tag{3.41}$$

By the compactness of A, we can pick a sub-sequence if needed and w.l.o.g. assume that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$ for some $\boldsymbol{x}^* \in A$. Next, let $B \triangleq \text{supp}(g)$ and note that B is bounded away from $\mathbb{D}_A^{(k-1)|b}(r)$. Applying Corollary C.2, we can fix some M_0 such that the following claim holds for any $M \geq M_0$: for any $\xi = \bar{h}_{M\downarrow}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t})$ with $\boldsymbol{t} = (t_1, \cdots, t_k) \in (0,1]^{k\uparrow}$, $\mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\mathbf{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$ with $\max_{j\in[d]} \|\boldsymbol{v}_j\| \leq r$, and $\boldsymbol{x} \in A$,

$$\xi = \bar{h}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t}) = \bar{h}_{M\downarrow}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t}) \qquad \text{and} \qquad \sup_{t\in[0,1]} \|\xi_t\| \leq M_0. \tag{3.42}$$

Here, recall that the mappings $\bar{h}_{M\downarrow}^{(k)|b}$ and $h_{M\downarrow}^{(k)|b}$ are defined in (3.36)–(3.39). Now, we fix some $M \geq M_0 + 1$ and recall the definitions of $\boldsymbol{a}_M$, $\boldsymbol{\sigma}_M$ in (3.35). Define the stochastic processes $\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x}) \triangleq \big\{\widetilde{\boldsymbol{X}}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x}) : \ t \in [0,1]\big\}$ by

$$\widetilde{\boldsymbol{X}}_j^{\eta|b}(\boldsymbol{x}) = \widetilde{\boldsymbol{X}}_{j-1}^{\eta|b}(\boldsymbol{x}) + \varphi_b\Big(\eta \boldsymbol{a}_M\big(\widetilde{\boldsymbol{X}}_{j-1}^{\eta|b}(\boldsymbol{x})\big) + \eta \boldsymbol{\sigma}_M\big(\widetilde{\boldsymbol{X}}_{j-1}^{\eta|b}(\boldsymbol{x})\big)\boldsymbol{Z}_j\Big) \ \ \forall j \geq 1 \tag{3.43}$$

under initial condition $\widetilde{\boldsymbol{X}}_0^{\eta|b}(\boldsymbol{x}) = \boldsymbol{x}$. In particular, by comparing the definition of $\widetilde{\boldsymbol{X}}_j^{\eta|b}(\boldsymbol{x})$ with that of $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ in (2.19), one can see that (for any $\boldsymbol{x} \in \mathbb{R}^m, \eta > 0$)

$$\sup_{t\in[0,1]} \Big\|\widetilde{\boldsymbol{X}}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x})\Big\| > M \iff \sup_{t\in[0,1]} \Big\|\boldsymbol{X}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x})\Big\| > M, \tag{3.44}$$

$$\sup_{t\in[0,1]} \Big\|\boldsymbol{X}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x})\Big\| \leq M \implies \boldsymbol{X}^{\eta|b}(\boldsymbol{x}) = \widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x}). \tag{3.45}$$

Now, we observe a few facts. First, define measure

$$\widetilde{\mathbf{C}}^{(k)|b}(\ \cdot\ ; \boldsymbol{x}) \triangleq \int \mathbb{I}\Big\{h_{M\downarrow}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t}) \in \ \cdot\ \Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^k(d\mathbf{W}) \times \mathcal{L}_T^{k\uparrow}(d\boldsymbol{t}).$$

Due to (3.42), we must have

$$\widetilde{\mathbf{C}}^{(k)|b}(\ \cdot\ ; \boldsymbol{x}) = \mathbf{C}^{(k)|b}(\ \cdot\ ; \boldsymbol{x}) \qquad \forall \boldsymbol{x} \in A. \tag{3.46}$$

Next, recall that $\boldsymbol{a}_M$ and $\boldsymbol{\sigma}_M$ satisfy Assumption 3. Then as has been established at the beginning of the proof, we have the following uniform $\mathbb{M}$-convergence for $\widetilde{\boldsymbol{X}}^{\eta|b}(x)$:

$$\lambda^{-k}(\eta)\mathbf{P}\big(\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x}) \in \ \cdot\ \big) \to \widetilde{\mathbf{C}}^{(k)|b}(\ \cdot\ ; \boldsymbol{x}) = \mathbf{C}^{(k)|b}(\ \cdot\ ; \boldsymbol{x}) \quad \text{in } \mathbb{M}\Big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)|b}(r)\Big) \text{ uniformly in } \boldsymbol{x} \text{ on } A \tag{3.47}$$

as $\eta \downarrow 0$. By Definition 2.1, for the function $g \in \mathcal{C}\big(\mathbb{D} \setminus \mathbb{D}_A^{(k-1)}(r)\big)$ fixed above, we now have

$$\lim_{n\to\infty} \big|\tilde{\mu}_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\big| = 0 \qquad \text{with } \tilde{\mu}_n^{(k)}(\cdot) \triangleq \mathbf{P}\big(\widetilde{\boldsymbol{X}}^{\eta_n|b}(\boldsymbol{x}_n) \in \ \cdot\ \big)\big/\lambda^k(\eta_n). \tag{3.48}$$

On the other hand, for any $n \geq 1$ (recall that $B = \text{supp}(g)$)

$$\begin{aligned}\mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big] &= \mathbf{E}\Bigg[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\mathbb{I}\Big\{\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \in B;\ \sup_{t\in[0,1]}\Big\|\boldsymbol{X}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| \leq M\Big\}\Bigg] \\ &\quad + \mathbf{E}\Bigg[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\mathbb{I}\Big\{\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \in B;\ \sup_{t\in[0,1]}\Big\|\boldsymbol{X}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\Big\}\Bigg].\end{aligned} \tag{3.49}$$

The following bound then follows immediately from (3.44) and (3.45):

$$\bigg|\mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big] - \mathbf{E}\Big[g\big(\widetilde{\boldsymbol{X}}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big]\bigg| \leq \|g\|\,\mathbf{P}\Bigg(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\Bigg). \tag{3.50}$$

Furthermore, we claim that

$$\lim_{n\to\infty} \lambda^{-k}(\eta_n)\mathbf{P}\Bigg(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\Bigg) = 0. \tag{3.51}$$

Then observe that

$$\begin{aligned}&\limsup_{n\to\infty}\Big|\mu_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\ &\leq \limsup_{n\to\infty}\Big|\mu_n^{(k)}(g) - \tilde{\mu}_n^{(k)}(g)\Big| + \limsup_{n\to\infty}\Big|\tilde{\mu}_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\ &\leq \limsup_{n\to\infty}\lambda^{-k}(\eta_n)\mathbf{P}\Bigg(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\Bigg) + 0 \qquad \text{due to (3.50) and (3.48)} \\ &= 0 \qquad \text{due to (3.51)}.\end{aligned}$$

In summary, we end up with a clear contradiction to (3.41), thus allowing us to conclude the proof. Now, it only remains to prove claim (3.51).

Proof of Claim (3.51):

Let $E \triangleq \{\xi \in \mathbb{D} :\ \sup_{t\in[0,1]}\|\xi_t\| > M\}$. Suppose we can show that E is bounded away from $\mathbb{D}_A^{(k)|b}(r)$, then by applying the uniform $\mathbb{M}$-convergence established above in (3.47) for $\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x})$, we get $\limsup_{n\to\infty}\mathbf{P}\Big(\widetilde{\boldsymbol{X}}^{\eta_n|b}(\boldsymbol{x}_n) \in E\Big)\Big/\lambda^{k+1}(\eta_n) < \infty$, which then implies (3.51). To see why E is bounded away from $\mathbb{D}_A^{(k)|b}(r)$, note that by (3.42),

$$\xi \in \mathbb{D}_A^{(k)|b}(r) \Longrightarrow \sup_{t\in[0,1]}\|\xi_t\| \leq M_0 \leq M - 1$$

due to our choice of $M \geq M_0 + 1$ at the beginning. Therefore, we yield $\boldsymbol{d}_{J_1}\big(\mathbb{D}_A^{(k)|b}(r), E\big) \geq 1$ and conclude the proof. □

3.3.3 Proof of Proposition 3.11

As has been demonstrated earlier, Proposition 3.11 lays the foundation for the sample path large deviations of heavy-tailed stochastic difference equations. In Section 3.3.3, we provide the proof of Proposition 3.11. Analogous to the proof of Theorem 2.4 above, we show that it suffices to prove the seemingly more restrictive results stated below in Proposition 3.13, where we impose the the boundedness condition in Assumption 3.

Proposition 3.13. *Let η_n be a sequence of strictly positive real numbers with $\lim_{n\to\infty}\eta_n = 0$. Let compact set $A \subseteq \mathbb{R}^m$ and $\boldsymbol{x}_n, \boldsymbol{x}^* \in A$ be such that $\lim_{n\to\infty}\boldsymbol{x}_n = \boldsymbol{x}^*$. Under Assumptions 1, 2, and 3, it holds for all $k \in \mathbb{N}$ and $b, r > 0$ that*

$$\mathbf{P}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \in \ \cdot\ \big)/\lambda^k(\eta_n) \to \mathbf{C}^{(k)|b}\big(\ \cdot\ ;\boldsymbol{x}^*\big) \ \textit{in}\ \mathbb{M}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)\ \textit{as}\ n\to\infty.$$

Proof of Proposition 3.11. The proof is almost identical to the second half of the proof for Theorem 2.4. Specifically, we fix some $M \geq M_0 + 1$ with M_0 specified in (3.42), and we arbitrarily pick some $g \in \mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)}(r)\big)$. Besides, define the stochastic processes $\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x}) \triangleq \big\{\widetilde{\boldsymbol{X}}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}):\ t\in[0,1]\big\}$ by (3.43). By repeating the arguments in the proof for Theorem 2.4, we yield (3.46) and (3.50) again. Next, by applying Proposition 3.13 onto $\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x})$, we again obtain (3.48) and (3.51) (in particular, for the claim (3.51), note that at the end of the proof for Theorem 2.4 we have already shown that $\{\xi\in\mathbb{D}:\ \sup_{t\in[0,1]}\|\xi_t\| > M\}$ is bounded away from $\mathbb{D}_A^{(k)|b}(r)$). Now, for $\mu_n^{(k)}(\cdot) \triangleq \mathbf{P}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\in\ \cdot\ \big)/\lambda^k(\eta_n)$, observe that

$$\begin{aligned}
&\lim_{n\to\infty}\Big|\mu_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\
&\leq \limsup_{n\to\infty}\Big|\mu_n^{(k)}(g) - \tilde{\mu}_n^{(k)}(g)\Big| + \limsup_{n\to\infty}\Big|\tilde{\mu}_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\
&\leq \limsup_{n\to\infty}\lambda^{-k}(\eta_n)\mathbf{P}\left(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\right) + 0 \qquad \text{due to (3.50) and (3.48)} \\
&= 0 \qquad \text{due to (3.51).}
\end{aligned}$$

By the Portmanteau theorem for $\mathbb{M}$-convergence (see theorem 2.1 of [42]) and the arbitrariness of the function $g\in\mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)}(r)\big)$, we conclude the proof. □

The rest of Section 3.3.3 is devoted to establishing Proposition 3.13. In light of Lemma 3.2, one approach to Proposition 3.13 is to construct some process $\hat{\boldsymbol{X}}^{\eta|b;(k)}$ that is not only asymptotically equivalent to $\boldsymbol{X}^{\eta|b}$ (as $\eta\downarrow 0$) but also (under the right scaling) converges to $\mathbf{C}^{(k)|b}$. To properly introduce the process $\hat{\boldsymbol{X}}^{\eta|b;(k)}$, we set a few notations. For any $j\geq 1$ and $n\geq j$ let

$$\mathcal{J}_{\boldsymbol{Z}}(c,n) \triangleq \#\Big\{i\in[n]:\ \|\boldsymbol{Z}_i\|\geq c\Big\};\qquad \boldsymbol{Z}^{(j)}(\eta) \triangleq \max\Big\{c\geq 0:\ \mathcal{J}_{\boldsymbol{Z}}(c,\lfloor 1/\eta\rfloor)\geq j\Big\}. \tag{3.52}$$

In other words, $\mathcal{J}_{\boldsymbol{Z}}(c,n)$ counts the number of elements in $\{\boldsymbol{Z}_i: i\in[n]\}$ with a norm larger than c, and $\boldsymbol{Z}^{(j)}(\eta)$ identifies the value of the j^{th} largest element in $\{\|\boldsymbol{Z}_i\|: i\leq\lfloor 1/\eta\rfloor\}$. Moreover, let

$$\tau_i^{(j)}(\eta) \triangleq \min\Big\{k > \tau_{i-1}^{(j)}(\eta):\ \|\boldsymbol{Z}_k\|\geq \boldsymbol{Z}^{(j)}(\eta)\Big\},\quad \boldsymbol{W}_i^{(j)}(\eta)\triangleq \boldsymbol{Z}_{\tau_i^{(j)}(\eta)}\quad \forall i=1,2,\cdots,j \tag{3.53}$$

with the convention that $\tau_0^{(j)}(\eta) = 0$. Note that $\big(\tau_i^{(j)}(\eta), \boldsymbol{W}_i^{(j)}(\eta)\big)_{i\in[j]}$ record the arrival time and size of the top j elements (in terms of L_2 norm) of $\{\boldsymbol{Z}_i: i\in[n]\}$. In case that there are ties between the values of $\{\|\boldsymbol{Z}_i\|: i\leq\lfloor 1/\eta\rfloor\}$, under our definition we always pick the first j elements. Now, for any $j\geq 1$ and any $\eta, b > 0, \boldsymbol{x}\in\mathbb{R}^m$, we define $\hat{\boldsymbol{X}}^{\eta|b;(j)}(\boldsymbol{x}) \triangleq \big\{\hat{\boldsymbol{X}}_t^{\eta|b;(j)}(\boldsymbol{x}):\ t\in[0,1]\big\}$ as the solution to

$$\frac{d\hat{\boldsymbol{X}}_t^{\eta|b;(j)}(\boldsymbol{x})}{dt} = \boldsymbol{a}\big(\hat{\boldsymbol{X}}_t^{\eta|b;(j)}(\boldsymbol{x})\big)\qquad \forall t\in[0,1],\ t\notin\big\{\eta\tau_i^{(j)}(\eta):\ i\in[j]\big\}, \tag{3.54}$$

$$\hat{\boldsymbol{X}}_t^{\eta|b;(j)}(\boldsymbol{x}) = \hat{\boldsymbol{X}}_{t-}^{\eta|b;(j)}(\boldsymbol{x}) + \varphi_b\Big(\eta\boldsymbol{\sigma}\big(\hat{\boldsymbol{X}}_{t-}^{\eta|b;(j)}(\boldsymbol{x})\big)\boldsymbol{W}_i^{(j)}(\eta)\Big)\qquad \text{if } t = \eta\tau_i^{(j)}(\eta) \text{ for some } i\in[j] \tag{3.55}$$

with initial condition $\hat{\boldsymbol{X}}_0^{\eta|b;(j)}(\boldsymbol{x}) = \boldsymbol{x}$. For the case $j=0$, we adopt the convention that

$$d\hat{\boldsymbol{X}}_t^{\eta|b;(0)}(\boldsymbol{x})/dt = \boldsymbol{a}\big(\hat{\boldsymbol{X}}_t^{\eta|b;(0)}(\boldsymbol{x})\big)\ \forall t\in[0,1]$$

with $\hat{\boldsymbol{X}}^{\eta|b;(0)}_0(\boldsymbol{x}) = \boldsymbol{x}$. First, by definition of the mapping $h^{(k)|b}$ in (2.21)–(2.24), we have

$$\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}) = h^{(k)|b}\Big(\boldsymbol{x}, \big(\eta \boldsymbol{W}^{(k)}_1(\eta), \cdots, \eta \boldsymbol{W}^{(k)}_k(\eta)\big), \big(\eta\tau^{(k)}_1(\eta), \cdots, \eta\tau^{(k)}_k(\eta)\big)\Big). \tag{3.56}$$

Furthermore, the following property is central to our proof: for any $\eta, b > 0$, $k \in \mathbb{N}$, and $\boldsymbol{x} \in \mathbb{R}^m$,

$$\text{on event } \Big\{\tau^{>\delta}_k(\eta) \leq \lfloor 1/\eta \rfloor < \tau^{>\delta}_{k+1}(\eta)\Big\}, \text{ we have } \hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}) = h^{(k)|b}\big(\boldsymbol{x}, \eta \mathbf{W}^{>\delta}(\eta), \eta\boldsymbol{\tau}^{>\delta}(\eta)\big) \tag{3.57}$$

with $\mathbf{W}^{>\delta}(\eta) = (\boldsymbol{W}^{>\delta}_1(\eta), \cdots, \boldsymbol{W}^{>\delta}_k(\eta))$ and $\boldsymbol{\tau}^{>\delta}(\eta) = (\tau^{>\delta}_1(\eta), \cdots, \tau^{>\delta}_k(\eta))$.

We first state two results that allow us to apply Lemma 3.2.

Proposition 3.14. *Let η_n be a sequence of strictly positive real numbers with $\lim_{n\to\infty} \eta_n = 0$. Let compact set $A \subseteq \mathbb{R}^m$ and $\boldsymbol{x}_n, \boldsymbol{x}^* \in A$ be such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$. Under Assumptions 1, 2, 3, it holds for all $k \in \mathbb{N}$ and $b, r > 0$ that $\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)$ is asymptotically equivalent to $\hat{\boldsymbol{X}}^{\eta_n|b;(k)}(\boldsymbol{x}_n)$ (as $n \to \infty$) in $\mathbb{M}\big(\mathbb{D} \setminus \mathbb{D}^{(k)|b}_A(r)\big)$ w.r.t. $\lambda^k(\eta_n)$.*

Proposition 3.15. *Let η_n be a sequence of strictly positive real numbers with $\lim_{n\to\infty} \eta_n = 0$. Let compact set $A \subseteq \mathbb{R}^m$ and $\boldsymbol{x}_n, \boldsymbol{x}^* \in A$ be such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$. Under Assumptions 1, 2, 3, it holds for all $k \in \mathbb{N}$ and $b, r > 0$ that*

$$\mathbf{P}\Big(\hat{\boldsymbol{X}}^{\eta_n|b;(k)}(\boldsymbol{x}_n) \in \ \cdot\ \Big)\Big/\lambda^k(\eta_n) \to \mathbf{C}^{(k)|b}\big(\ \cdot\ ; \boldsymbol{x}^*\big) \text{ in } \mathbb{M}\big(\mathbb{D} \setminus \mathbb{D}^{(k-1)|b}_A(r)\big) \text{ as } n \to \infty,$$

where $\mathbf{C}^{(k)|b}$ is the measure defined in (2.26).

Proof of Proposition 3.13. In light of Lemma 3.2, it is a direct corollary of Propositions 3.14 and 3.15. ☐

Now, it only remains to prove Propositions 3.14 and 3.15.

Proof of Proposition 3.14. Fix some $b, r > 0, k \in \mathbb{N}$, and some sequence of strictly positive real numbers η_n with $\lim_{n\to\infty} \eta_n = 0$. Also, fix a compact set $A \subseteq \mathbb{R}^m$ and $\boldsymbol{x}_n, \boldsymbol{x}^* \in A$ such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$. Besides, we arbitrarily pick some $\Delta > 0$ and some $B \in \mathscr{S}_{\mathbb{D}}$ that is bounded away from $\mathbb{D}^{(k-1)|b}_A(r)$. By Definition 3.1, it suffices to show that

$$\lim_{n\to\infty} \mathbf{P}\Big(\boldsymbol{d}_{J_1}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n), \hat{\boldsymbol{X}}^{\eta_n|b;(k)}(\boldsymbol{x}_n)\big)\mathbb{I}\big\{\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \text{ or } \hat{\boldsymbol{X}}^{\eta_n|b;(k)}(\boldsymbol{x}_n) \in B\big\} > \Delta\Big)\Big/\lambda^k(\eta_n) = 0. \tag{3.58}$$

By Lemma 3.6, there are some $\bar{\epsilon} \in (0, r)$ and $\bar{\delta} > 0$ such that

- for any $\boldsymbol{x} \in A$ and $b > 0$, and any $(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$ with $\max_{j\in[k]} \|\boldsymbol{v}_j\| \leq \bar{\epsilon}$,

$$\bar{h}^{(k)|b}\big(\boldsymbol{x}, (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k), \boldsymbol{t}\big) \in B^{\bar{\epsilon}} \qquad \Longrightarrow \qquad \|\boldsymbol{w}_i\| > \bar{\delta}\ \forall i \in [k]; \tag{3.59}$$

- furthermore,

$$\boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}^{(k-1)|b}_A(r)\big) > \bar{\epsilon}. \tag{3.60}$$

Meanwhile, for any $\eta, \delta, \epsilon > 0$ and $\boldsymbol{x} \in A$, let

$$B_0 \triangleq \Big\{\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \in B \text{ or } \hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}) \in B;\ \boldsymbol{d}_{J_1}\big(\boldsymbol{X}^{\eta|b}(\boldsymbol{x}), \hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})\big) > \Delta\Big\},$$
$$B_1 \triangleq \Big\{\tau^{>\delta}_{k+1}(\eta) > \lfloor 1/\eta \rfloor\Big\},$$

$$B_2 \triangleq \Big\{ \tau^{>\delta}_k(\eta) \leq \lfloor 1/\eta \rfloor \Big\},$$
$$B_3 \triangleq \Big\{ \eta \left\| \boldsymbol{W}^{>\delta}_i(\eta) \right\| > \bar{\delta} \text{ for all } i \in [k] \Big\},$$
$$B_4 \triangleq \Big\{ \eta \left\| \boldsymbol{W}^{>\delta}_i(\eta) \right\| \leq 1/\epsilon^{\frac{1}{2k}} \text{ for all } i \in [k] \Big\}.$$

We have the following decomposition of events:

$$\begin{aligned} B_0 = {} & (B_0 \cap B_1^c) \cup (B_0 \cap B_1 \cap B_2^c) \cup (B_0 \cap B_1 \cap B_2 \cap B_3^c) \\ & \cup (B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4^c) \cup (B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4). \end{aligned} \tag{3.61}$$

To proceed, let $\rho = \exp(D)$ and $D \in [1, \infty)$ is the Lipschitz coefficient in Assumption 2. For any $\epsilon > 0$ small enough such that

$$(2\rho D)^{k+1}\sqrt{\epsilon} < \Delta, \qquad 2\rho\epsilon < \bar{\epsilon}, \qquad \epsilon \in (0,1), \tag{3.62}$$

we claim that

$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1^c \Big) \Big/ \lambda^k(\eta) = 0, \tag{3.63}$$
$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2^c \Big) \Big/ \lambda^k(\eta) = 0, \tag{3.64}$$
$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2 \cap B_3^c \Big) \Big/ \lambda^k(\eta) = 0, \tag{3.65}$$
$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4^c \Big) \Big/ \lambda^k(\eta) \leq \bar{\delta}^{-k\alpha} \cdot \epsilon^{\frac{\alpha}{2k}}, \tag{3.66}$$
$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4 \Big) \Big/ \lambda^k(\eta) = 0, \tag{3.67}$$

if we pick $\delta > 0$ sufficiently small. Under such δ, by the decomposition of event B_0 in (3.61), we yield

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\big(B_0\big) / \lambda^k(\eta) \leq \bar{\delta}^{-k\alpha} \cdot \epsilon^{\frac{\alpha}{2k}}$$

for all $\epsilon > 0$ small enough. Note that $\bar{\delta} > 0$ is the constant fixed in (3.59). Driving $\epsilon \downarrow 0$, we conclude the proof of (3.58). The remainder of this proof is devoted to claims (3.63)–(3.67).

Proof of (3.63):

For any $\delta > 0$, (3.4) implies that $\sup_{x \in A} \mathbf{P}(B_0 \cap B_1^c) \leq \mathbf{P}(B_1^c) \leq \big(\eta^{-1} H(\delta\eta^{-1})\big)^{k+1} = \boldsymbol{O}\big(\lambda^{k+1}(\eta)\big) = \boldsymbol{o}\big(\lambda^k(\eta)\big)$.

Proof of (3.64):

It suffices to show that (for all $\delta > 0$ small enough)

$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(\underbrace{B_0 \cap \{\tau^{>\delta}_k(\eta) > \lfloor 1/\eta \rfloor\}}_{\triangleq \widetilde{B}} \Big) \Big/ \lambda^k(\eta) = 0$$

In particular, we only consider $\delta \in (0, \bar{\delta} \wedge \frac{b}{2})$ with $\bar{\delta}$ characterized in (3.59) and $C \geq 1$ being the constant in Assumption 3. On event $\{\tau^{>\delta}_k(\eta) > \lfloor 1/\eta \rfloor\}$ we have $\#\big\{i \in \big[\lfloor 1/\eta \rfloor\big] : \ \eta \|\boldsymbol{Z}_i\| > \delta\big\} < k$. By the definition of $\boldsymbol{Z}^{(k)}(\eta)$ in (3.52) and the definition of $\boldsymbol{W}^{(k)}_i(\eta)$ in (3.53), we then get $\min_{i \in [k]} \eta \left\| \boldsymbol{W}^{(k)}_i(\eta) \right\| \leq \delta < \bar{\delta}$. In light of (3.56) and (3.59), we yield $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}) \notin B^{\bar{\epsilon}}$ on event $\{\tau^{>\delta}_k(\eta) > \lfloor 1/\eta \rfloor\}$. As a result,

$$\widetilde{B} \subseteq \{\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \in B\} \cap \{\tau^{>\delta}_k(\eta) > \lfloor 1/\eta \rfloor\} \qquad \forall \boldsymbol{x} \in A.$$

Furthermore, let event $A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})$ be defined as in (3.6). We claim that

$$\{\boldsymbol{X}^{\eta|b}(x) \in B\} \cap \{\tau_k^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\} \cap \big(\cap_{i=1}^k A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\big) = \emptyset \tag{3.68}$$

holds for all $\eta > 0$ small enough with $\eta < \min\{\frac{b\wedge 1}{2C}, \frac{\epsilon}{C}\}$, all $\delta \in (0, \frac{b}{2C})$, and all $\boldsymbol{x} \in A$. Then

$$\limsup_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in A} \mathbf{P}\big(\widetilde{B}\big)\Big/\lambda^k(\eta) \leq \limsup_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in A} \mathbf{P}\left(\Big(\bigcap_{i=1}^k A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\Big)^c\right)\Big/\lambda^k(\eta).$$

To conclude the proof, one only need to apply Lemma 3.3 (b) with some $N > k(\alpha - 1)$ (recall that $\lambda(\eta) \in \mathcal{RV}_{k(\alpha-1)}(\eta)$ as $\eta \downarrow 0$).

Now, we prove claim (3.68) for any $\eta \in \big(0, \min\{\frac{b\wedge 1}{2C}, \frac{\epsilon}{C}\}\big)$, $\delta \in (0, \frac{b}{2C})$, and $\boldsymbol{x} \in A$. Define the stochastic process $\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}) \triangleq \big\{\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x}) :\ t \in [0, 1]\big\}$ as the solution to

$$\frac{d\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x})}{dt} = \boldsymbol{a}\big(\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x})\big) \qquad \forall t \in [0, \infty) \setminus \{\eta\tau_j^{>\delta}(\eta) :\ j \geq 1\}, \tag{3.69}$$

$$\breve{\boldsymbol{X}}^{\eta|b;\delta}_{\eta\tau_j^{>\delta}(\eta)}(\boldsymbol{x}) = \boldsymbol{X}^{\eta|b}_{\tau_j^{>\delta}(\eta)}(\boldsymbol{x}) \qquad \forall j \geq 1, \tag{3.70}$$

under the initial condition $\breve{\boldsymbol{X}}_0^{\eta|b;\delta}(\boldsymbol{x}) = \boldsymbol{x}$. For any $j \geq 1$, observe that on event $\big(\cap_{i=1}^j A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\big) \cap \{\tau_j^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\}$,

$$\begin{aligned}
&\boldsymbol{d}_{J_1}\Big(\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}), \boldsymbol{X}^{\eta|b}(\boldsymbol{x})\Big)\\
&\leq \sup_{t\in[0,1]} \Big\|\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x}) - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x})\Big\|\\
&\leq \sup_{t\in\big[0,\eta\tau_1^{>\delta}(\eta)\big)\cup\big[\eta\tau_1^{>\delta}(\eta),\eta\tau_2^{>\delta}(\eta)\big)\cup\cdots\cup\big[\eta\tau_{j-1}^{>\delta}(\eta),\eta\tau_j^{>\delta}(\eta)\big)} \Big\|\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x}) - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x})\Big\|\\
&\leq \rho\cdot\big(\epsilon + \eta C\big) \leq 2\rho\epsilon < \bar{\epsilon} \qquad \text{by (3.24) of Lemma 3.8.}
\end{aligned} \tag{3.71}$$

In the last line of the display above, note that (i) our choices of $\eta < \frac{b\wedge 1}{2C}$ and $\delta < \frac{b}{2C}$ allow us to apply part (b) of Lemma 3.8, and (ii) the inequalities then follow from the choice of $\eta < \frac{\epsilon}{C}$ above and the choice of $2\rho\epsilon < \bar{\epsilon}$ in (3.62). Moreover, recall that we have fixed $\bar{\epsilon} < r$ at the beginning of the proof, and note that (3.71) confirms (under the choice of $j = k$) that on event

$$\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}) \in \mathbb{D}_A^{(k-1)|b}(\bar{\epsilon}) \subseteq \mathbb{D}_A^{(k-1)|b}(r) \qquad \text{and} \qquad \boldsymbol{d}_{J_1}\Big(\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}), \boldsymbol{X}^{\eta|b}(\boldsymbol{x})\Big) < \bar{\epsilon}.$$

In light of (3.60), this implies that on event $\big(\cap_{i=1}^k A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\big) \cap \{\tau_k^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\}$, we must have $\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \notin B^{\bar{\epsilon}}$, thus concluding the proof of claim (3.68).

Proof of (3.65):

On event $B_1 \cap B_2 = \{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta)\}$, recall that (3.57) holds. Furthermore, on B_3^c, there is some $i \in [k]$ such that $\eta\,\|\boldsymbol{W}_i^{>\delta}(\eta)\| \leq \bar{\delta}$. Combining (3.57) with the choice of $\bar{\delta}$ in (3.59), we get that for all $\boldsymbol{x} \in A$, it holds on event $B_1 \cap B_2 \cap B_3^c$ that $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}) \notin B$, and hence

$$\begin{aligned}
&B_0 \cap B_1 \cap B_2 \cap B_3^c\\
&\quad\subseteq \{\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \in B\} \cap \Big\{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta);\ \eta\,\|\boldsymbol{W}_i^{>\delta}(\eta)\| \leq \bar{\delta} \text{ for some } i \in [k]\Big\}.
\end{aligned}$$

Furthermore, we claim that for all $\boldsymbol{x} \in A$, $\delta \in (0, \bar{\delta} \wedge \frac{b}{2C})$ and $\eta \in \big(0, \min\{\frac{b\wedge 1}{2C}, \bar{\delta}\}\big)$,

$$\begin{aligned}
&\{\boldsymbol{X}^{\eta|b}(x) \in B\} \cap \Big\{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta);\ \eta\,\|\boldsymbol{W}_i^{>\delta}(\eta)\| \leq \bar{\delta} \text{ for some } i \in [k]\Big\}\\
&\qquad\qquad \cap \left(\bigcap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\right) = \emptyset.
\end{aligned} \tag{3.72}$$

Then for any $\delta \in (0, \bar{\delta} \wedge \frac{b}{2})$,

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2 \cap B_3^c\Big) \Big/ \lambda^k(\eta) \leq \limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Bigg(\Big(\bigcap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\Big)^c\Bigg) \Bigg/ \lambda^k(\eta).$$

Applying Lemma 3.3 (b) with some $N > k(\alpha - 1)$, we conclude the proof of (3.65).

Now, it remains to prove the claim (3.72) for any $\boldsymbol{x} \in A$, $\delta \in (0, \bar{\delta} \wedge \frac{b}{2C})$ and $\eta \in \big(0, \min\{\frac{b \wedge 1}{2C}, \bar{\delta}\}\big)$. First, on this event, there exists some $J \in [k]$ such that $\eta \left\| \boldsymbol{W}_J^{>\delta}(\eta) \right\| \leq \bar{\delta}$. Next, recall the definition of the process $\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x})$ in (3.69)–(3.70). Applying (3.71) with $j = k+1$, we get that $\big(\cap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\big) \cap \{\tau_{k+1}^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\}$,

$$\boldsymbol{d}_{J_1}\Big(\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}), \boldsymbol{X}^{\eta|b}(\boldsymbol{x})\Big) \leq \sup_{t \in [0,1]} \left\| \breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x}) - \boldsymbol{X}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x}) \right\| < 2\rho\epsilon < \bar{\epsilon}. \tag{3.73}$$

This further confirms that, on the said event, there exists some $\mathbf{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \mathbb{R}^{m \times k}$ with $\|\boldsymbol{v}_j\| \leq \bar{\epsilon} < r$ (recall that we have fixed $\bar{\epsilon} < r$ at the beginning of the proof) such that

$$\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}) = \bar{h}^{(k)|b}\Big(\boldsymbol{x}, \big(\eta \boldsymbol{W}_1^{>\delta}(\eta), \cdots, \eta \boldsymbol{W}_k^{>\delta}(\eta)\big), \mathbf{V}, \big(\eta \tau_1^{>\delta}(\eta), \cdots, \eta \tau_k^{>\delta}(\eta)\big)\Big),$$

where the mapping $\bar{h}^{(k)|b}$ is defined in (2.10)–(2.12). Due to $\eta \left\| \boldsymbol{W}_J^{>\delta}(\eta) \right\| \leq \bar{\delta}$, it follows from (3.59) that $\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}) \notin B^{\bar{\epsilon}}$. Then by (3.60) and (3.73), we must have $\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \notin B$ on the event $\big\{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta);\ \eta \left\| \boldsymbol{W}_i^{>\delta}(\eta) \right\| \leq \bar{\delta}$ for some $i \in [k]\big\} \cap \big(\bigcap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})\big)$, thus verifying claim (3.72).

Proof of (3.66):

Recall that $H(x) = \mathbf{P}(\|\boldsymbol{Z}\| > x)$. Due to

$$\begin{aligned}
& B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4^c \\
& \subseteq \Big\{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta)\Big\} \cap \Big\{\eta \left\| \boldsymbol{W}_i^{>\delta}(\eta) \right\| > \bar{\delta}\ \forall i \in [k];\ \eta \left\| \boldsymbol{W}_i^{>\delta}(\eta) \right\| > 1/\epsilon^{\frac{1}{2k}} \text{ for some } i \in [k]\Big\}.
\end{aligned}$$

and the independence between $\big(\tau_i^{>\delta}(\eta)\big)_{i \in [k]}$ and $\big(\boldsymbol{W}_i^{>\delta}(\eta)\big)_{i \in [k]}$, we get

$$\begin{aligned}
& \limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \frac{\mathbf{P}\big(B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4^c\big)}{\lambda^k(\eta)} \\
& \leq \lim_{\eta \downarrow 0} \frac{1}{\lambda^k(\eta)} \cdot \Big(\eta^{-1} H(\delta \eta^{-1})\Big)^k \cdot k \cdot \left(\frac{H(\bar{\delta}\eta^{-1})}{H(\delta \eta^{-1})}\right)^{k-1} \cdot \frac{H(\epsilon^{-\frac{1}{2k}} \eta^{-1})}{H(\delta \eta^{-1})} \qquad \text{by (3.4)} \\
& = \lim_{\eta \downarrow 0} \frac{1}{\lambda^k(\eta)} \cdot \Big(\eta^{-1} H(\eta^{-1})\Big)^k \cdot k \cdot \left(\frac{H(\bar{\delta}\eta^{-1})}{H(\eta^{-1})}\right)^{k-1} \cdot \frac{H(\epsilon^{-\frac{1}{2k}} \eta^{-1})}{H(\eta^{-1})} \\
& = k \cdot \lim_{\eta \downarrow 0} \left(\frac{H(\bar{\delta}\eta^{-1})}{H(\eta^{-1})}\right)^{k-1} \cdot \frac{H(\epsilon^{-\frac{1}{2k}} \eta^{-1})}{H(\eta^{-1})} \qquad \text{recall that } \lambda(\eta) = \eta^{-1} H(\eta^{-1}) \\
& = \bar{\delta}^{-k\alpha} \cdot \epsilon^{\frac{\alpha}{2k}} \qquad \text{due to } H(x) \in \mathcal{RV}_{-\alpha}(x) \text{ as } x \to \infty; \text{ see Assumption 1.}
\end{aligned}$$

Proof of (3.67):

We only consider $\delta \in (0, \frac{b}{2C})$. On event $B_1 \cap B_2 = \{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta)\}$, $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})$ admits the expression in (3.57). Then by applying Lemma 3.9 we yield that for any $\boldsymbol{x} \in A$ and any $\eta \in (0, \frac{\epsilon \wedge b}{2C})$, the inequality

$$\boldsymbol{d}_{J_1}\Big(\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}), \boldsymbol{X}^{\eta|b}(\boldsymbol{x})\Big) \leq \sup_{t \in [0,1]} \left\| \hat{\boldsymbol{X}}_t^{\eta|b;(k)}(\boldsymbol{x}) - \boldsymbol{X}_{\lfloor t/\eta \rfloor}^{\eta|b}(\boldsymbol{x}) \right\| < (2\rho D)^{k+1}\sqrt{\epsilon},$$

holds on event $\Big(\bigcap_{i=1}^{k+1} A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\Big)$. Due to our choice of $(2\rho D)^{k+1}\sqrt{\epsilon} < \Delta$ in (3.62), we get $\Big(\bigcap_{i=1}^{k+1} A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\Big) \cap B_1 \cap B_2 \cap B_3 \cap B_4 \cap B_0 = \emptyset$. Therefore,

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in A}\mathbf{P}\Big(B_1\cap B_2\cap B_3\cap B_0\Big)\Big/\lambda^k(\eta) \leq \limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in A}\mathbf{P}\Bigg(\Big(\bigcap_{i=1}^{k+1} A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\Big)^c\Bigg)\Bigg/\lambda^k(\eta).$$

Again, by applying Lemma 3.3 (b) with some $N > k(\alpha-1)$, we conclude the proof. ☐

Recall that $\big(\boldsymbol{W}_j^*(c)\big)_{j\geq 1}$ is a sequence of iid copies of $\boldsymbol{W}^*(c)$ defined in (3.18), and $\big(U_{(j:k)}\big)_{j\in[k]}$ are the order statistics of k samples of Unif$(0,1)$. In order to prove Proposition 3.15, we prepare a lemma regarding a weak convergence on events $E^\delta_{c,k}(\eta) = \big\{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta\rfloor < \tau_{k+1}^{>\delta}(\eta);\ \eta\big\|\boldsymbol{W}_j^{>\delta}(\eta)\big\| > c\ \ \forall j\in[k]\big\}$ defined in (3.17).

Lemma 3.16. *Let Assumption 1 hold. Let $A \subseteq \mathbb{R}^m$ be a compact set. Let bounded function $\Psi : \mathbb{R}^m \times \mathbb{R}^{d\times k}\times(0,1]^{k\uparrow}\to\mathbb{R}$ be continuous on $\mathbb{R}^m\times\mathbb{R}^{d\times k}\times(0,1)^{k\uparrow}$. For any $\delta>0$, $c>\delta$ and $k\in\mathbb{N}$,*

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in A}\left|\frac{\mathbf{E}\Big[\Psi\Big(\boldsymbol{x},\big(\eta\boldsymbol{W}_1^{>\delta}(\eta),\cdots,\eta\boldsymbol{W}_k^{>\delta}(\eta)\big),\big(\eta\tau_1^{>\delta}(\eta),\cdots,\eta\tau_k^{>\delta}(\eta)\big)\Big)\mathbb{I}_{E^\delta_{c,k}(\eta)}\Big]}{\lambda^k(\eta)} - \frac{(1/c^{\alpha k})\psi_{c,k}(\boldsymbol{x})}{k!}\right| = 0$$

where $\psi_{c,k}(\boldsymbol{x}) \triangleq \mathbf{E}\Big[\Psi\Big(\boldsymbol{x},\big(\boldsymbol{W}_1^*(c),\cdots,\boldsymbol{W}_k^*(c)\big),\big(U_{(1;k)},\cdots,U_{(k;k)}\big)\Big)\Big]$.

Proof. Fix some $\delta>0, c>\delta$ and $k\in\mathbb{N}$. We proceed with a proof by contradiction. Suppose there exist some $\epsilon>0$, some sequence $\boldsymbol{x}_n\in A$, and some sequence $\eta_n\downarrow 0$ such that

$$\Big|\lambda^{-k}(\eta_n)\mathbf{E}\Big[\Psi\big(\boldsymbol{x}_n,\eta_n\mathbf{W}^{\eta_n},\eta_n\boldsymbol{\tau}^{\eta_n}\big)\mathbb{I}_{E^\delta_{c,k}(\eta_n)}\Big] - (1/k!)\cdot c^{-\alpha k}\cdot\psi_{c,k}(\boldsymbol{x}_n)\Big| > \epsilon\quad\forall n\geq 1 \tag{3.74}$$

where $\mathbf{W}^\eta \triangleq (\boldsymbol{W}_1^{>\delta}(\eta),\cdots,\boldsymbol{W}_k^{>\delta}(\eta))$, $\boldsymbol{\tau}^\eta\triangleq(\tau_1^{>\delta}(\eta),\cdots,\tau_k^{>\delta}(\eta))$. Since A is compact, by picking a sub-sequence if needed we can assume w.l.o.g. that $\boldsymbol{x}_n\to\boldsymbol{x}^*$ for some $\boldsymbol{x}^*\in A$. Now, observe that

$$\begin{aligned}
&\lim_{n\to\infty}\lambda^{-k}(\eta_n)\mathbf{E}\Big[\Psi\big(\boldsymbol{x}_n,\eta_n\mathbf{W}^{\eta_n},\eta_n\boldsymbol{\tau}^{\eta_n}\big)\mathbb{I}_{E^\delta_{c,k}(\eta_n)}\Big]\\
&=\Big[\lim_{n\to\infty}\lambda^{-k}(\eta_n)\mathbf{P}\Big(E^\delta_{c,k}(\eta_n)\Big)\Big]\cdot\lim_{n\to\infty}\mathbf{E}\Big[\Psi\big(\boldsymbol{x}_n,\eta_n\mathbf{W}^{\eta_n},\eta_n\boldsymbol{\tau}^{\eta_n}\big)\Big|E^\delta_{c,k}(\eta_n)\Big]\\
&=(1/k!)\cdot c^{-\alpha k}\cdot\psi_{c,k}(\boldsymbol{x}^*)\qquad\text{by Lemma 3.4, }\boldsymbol{x}_n\to\boldsymbol{x}^*\text{, and continuous mapping theorem.}
\end{aligned}$$

However, by Bounded Convergence theorem, we see that $\psi_{c,k}$ is also continuous, and hence $\psi_{c,k}(\boldsymbol{x}_n)\to\phi_{c,k}(\boldsymbol{x}^*)$. This leads to a contradiction with (3.74) and allows us to conclude the proof. ☐

We are now ready to prove Proposition 3.15.

Proof of Proposition 3.15. Fix some $b,r>0$, $k\in\mathbb{N}$, and $g\in\mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)$; i.e. $g:\mathbb{D}\to[0,\infty)$ is continuous and bounded with support $B\triangleq\text{supp}(g)$ bounded away from $\mathbb{D}_A^{(k-1)|b}(r)$. By Lemma 3.6, we can fix some $\bar\epsilon\in(0,r)$ and $\bar\delta>0$ such that

- for any $\boldsymbol{x}\in A$ and $b>0$,
$$h^{(k)|b}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),\boldsymbol{t}\big)\in B^{\bar\epsilon}\Longrightarrow\|\boldsymbol{w}_j\|>\bar\delta\ \ \forall j\in[k]; \tag{3.75}$$
- $\boldsymbol{d}_{J_1}\big(B^{\bar\epsilon},\mathbb{D}_A^{(k-1)|b}(r)\big)>\bar\epsilon$.

Fix some $\delta \in (0, \bar{\delta} \wedge \frac{b}{2})$, and observe that for any $\eta > 0$ and $\boldsymbol{x} \in A$,

$$\begin{aligned}
& g\Big(\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})\Big) \\
& = \underbrace{g\Big(\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})\Big)\mathbb{I}\Big\{\tau^{>\delta}_{k+1}(\eta) \leq \lfloor 1/\eta \rfloor\Big\}}_{\triangleq I_1(\eta,\boldsymbol{x})} + \underbrace{g\Big(\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})\Big)\mathbb{I}\Big\{\tau^{>\delta}_{k}(\eta) > \lfloor 1/\eta \rfloor\Big\}}_{\triangleq I_2(\eta,\boldsymbol{x})} \\
& \quad + \underbrace{g\Big(\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})\Big)\mathbb{I}\Big\{\tau^{>\delta}_{k}(\eta) \leq \lfloor 1/\eta \rfloor < \tau^{>\delta}_{k+1}(\eta);\ \eta\left\|\boldsymbol{W}^{>\delta}_{j}(\eta)\right\| \leq \bar{\delta} \text{ for some } j \in [k]\Big\}}_{\triangleq I_3(\eta,\boldsymbol{x})} \\
& \quad + \underbrace{g\Big(\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})\Big)\mathbb{I}\Big(E^{\delta}_{\bar{\delta},k}(\eta)\Big)}_{\triangleq I_4(\eta,\boldsymbol{x})}.
\end{aligned}$$

For $I_1(\eta, \boldsymbol{x})$, it follows from (3.4) that $\sup_{\boldsymbol{x}\in\mathbb{R}^m} \mathbf{E}[I_1(\eta,\boldsymbol{x})] \leq \|g\| \cdot \Big[\frac{1}{\eta} \cdot H(\delta/\eta)\Big]^{k+1}$. Therefore, $\lim_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in A} \mathbf{E}[I_1(\eta,\boldsymbol{x})]\Big/\big(\eta^{-1}H(\eta^{-1})\big)^k \leq \frac{\|g\|}{\delta^{\alpha(k+1)}} \cdot \lim_{n\to\infty} \frac{H(1/\eta)}{\eta} = 0$ due to $H(x) \in \mathcal{RV}_{-\alpha}(x)$ and $\alpha > 1$.

Next, for term $I_2(\eta, \boldsymbol{x})$, it has been shown in the proof of (3.64) for Proposition 3.14 that, for all $\delta \in (0, \bar{\delta} \wedge \frac{b}{2})$ and $\boldsymbol{x} \in A$, it holds on event $\{\tau^{>\delta}_k(\eta) > \lfloor 1/\eta \rfloor\}$ that $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}) \notin B^{\bar{\epsilon}}$, and hence $I_2(\eta, \boldsymbol{x}) = 0$.

For the term $I_3(\eta, \boldsymbol{x})$, on event $\{\tau^{>\delta}_k(\eta) \leq \lfloor 1/\eta \rfloor < \tau^{>\delta}_{k+1}(\eta)\}$ the process $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})$ admits the expression in (3.57). In particular, since there is some $i \in [k]$ such that $\eta\left\|\boldsymbol{W}^{>\delta}_i(\eta)\right\| \leq \bar{\delta}$, by (3.75) we must have $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x}) \notin B$, and hence $I_3(\eta, \boldsymbol{x}) = 0$.

Lastly, for the term $I_4(\eta, \boldsymbol{x})$, on event $E^{\delta}_{\bar{\delta},k}(\eta)$ the process $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})$ would again admit the form in (3.57). As a result, for any $\eta > 0$ and $\boldsymbol{x} \in A$, we have

$$\mathbf{E}[I_4(\eta,\boldsymbol{x})] = \mathbf{E}\Big[\Psi\big(\boldsymbol{x}, \eta\mathbf{W}^{\eta}, \eta\boldsymbol{\tau}^{\eta}\big)\mathbb{I}_{E^{\delta}_{\bar{\delta},k}(\eta)}\Big],$$

where $\mathbf{W}^{\eta} \triangleq (\boldsymbol{W}^{>\delta}_1(\eta), \cdots, \boldsymbol{W}^{>\delta}_k(\eta))$, $\boldsymbol{\tau}^{\eta} \triangleq (\tau^{>\delta}_1(\eta), \cdots, \tau^{>\delta}_k(\eta))$, and $\Psi(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t}) \triangleq g\big(h^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t})\big)$. Besides, let $\psi(\boldsymbol{x}) \triangleq \mathbf{E}\Big[\Psi\Big(\boldsymbol{x}, \big(\boldsymbol{W}^*_1(c), \cdots, \boldsymbol{W}^*_k(c)\big), \big(U_{(1;k)}, \cdots, U_{(k;k)}\big)\Big)\Big]$. First, the continuity of mapping Ψ on $\mathbb{R}^m \times \mathbb{R}^{d\times k} \times (0,1)^{k\uparrow}$ follows directly from the continuity of g and $h^{(k)|b}$ (see Lemma C.3). Besides, $\|\Psi\| \leq \|g\| < \infty$, so $\Psi(\cdot)$ is also bounded. By Bounded Convergence Theorem, one can see that $\psi(\cdot)$ is also continuous. Also, $\|\psi\| \leq \|\Psi\| \leq \|g\| < \infty$. By Lemma 3.16,

$$\lim_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in A} \left|\lambda^{-k}(\eta)\mathbf{E}\Big[\Psi\big(\boldsymbol{x}, \eta\mathbf{W}^{\eta}, \eta\boldsymbol{\tau}^{\eta}\big)\mathbb{I}_{E^{\delta}_{\bar{\delta},k}(\eta)}\Big] - (1/k!) \cdot \bar{\delta}^{-\alpha k} \cdot \psi(\boldsymbol{x})\right| = 0.$$

Meanwhile, due to the continuity of $\psi(\cdot)$, for any $\boldsymbol{x}_n,\ \boldsymbol{x}^* \in A$ with $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$, we have $\lim_{n\to\infty} \psi(\boldsymbol{x}_n) = \psi(\boldsymbol{x}^*)$. To conclude the proof, we only need to show that

$$\frac{(1/\bar{\delta}^{\alpha k})\psi(\boldsymbol{x}^*)}{k!} = \mathbf{C}^{(k)|b}(g; \boldsymbol{x}^*). \tag{3.76}$$

To do so, recall the law of $\boldsymbol{W}^*(c)$ in (3.18). By definition of $\psi(\cdot)$,

$$\begin{aligned}
\psi(\boldsymbol{x}^*) = \int g\Big(h^{(k)|b}\big(\boldsymbol{x}^*, (w_1\boldsymbol{\theta}_1, \cdots, w_k\boldsymbol{\theta}_k), (t_1, \cdots, t_k)\big)\Big)\mathbb{I}\Big\{w_j > \bar{\delta}\ \forall j \in [k]\Big\} \\
\mathbf{P}\Big(U_{(1;k)} \in dt_1, \cdots, U_{(k;k)} \in dt_k\Big) \times \left(\bigtimes_{j=1}^{k}\Big(\bar{\delta}^{\alpha} \cdot \nu_{\alpha}(dw_j) \times \mathbf{S}(d\boldsymbol{\theta}_j)\Big)\right).
\end{aligned}$$

By (3.75), we have

$$g\Big(h^{(k)|b}\big(\boldsymbol{x}^*, (w_1\boldsymbol{\theta}_1, \cdots, w_k\boldsymbol{\theta}_k), (t_1, \cdots, t_k)\big)\Big)$$
$$= g\Big(h^{(k)|b}\big(\boldsymbol{x}^*, (w_1\boldsymbol{\theta}_1, \cdots, w_k\boldsymbol{\theta}_k), (t_1, \cdots, t_k)\big)\Big)\mathbb{I}\Big\{w_j > \bar{\delta}\ \forall j \in [k]\Big\}.$$

Besides, $\mathbf{P}\Big(U_{(1;k)} \in dt_1, \cdots, U_{(k;k)} \in dt_k\Big) = k! \cdot \mathbb{I}\{0 < t_1 < t_2 < \cdots < t_k < 1\}\mathcal{L}_1^{k\uparrow}(dt_1, \cdots, dt_k)$ where $\mathcal{L}_1^{k\uparrow}$ is the Lebesgue measure restricted on $(0,1)^{k\uparrow}$. As a result,

$$\psi(\boldsymbol{x}^*)$$
$$= k! \cdot \bar{\delta}^{\alpha k} \int g\Big(h^{(k)|b}\big(\boldsymbol{x}^*, (w_1\boldsymbol{\theta}_1, \cdots, w_k\boldsymbol{\theta}_k), \boldsymbol{t}\big)\Big)\left(\bigtimes_{j=1}^{k}\Big(\nu_\alpha(dw_j) \times \mathbf{S}(d\boldsymbol{\theta}_j)\Big)\right) \times \mathcal{L}_1^{k\uparrow}(d\boldsymbol{t})$$
$$= k! \cdot \bar{\delta}^{\alpha k} \cdot \mathbf{C}^{(k)|b}\big(g; \boldsymbol{x}^*\big)$$

by the definition of $\mathbf{C}^{(k)|b}$ in (2.26), thus verifying (3.76). □

4 Metastability Analysis

In this section, we collect the proofs for Section 2.3. Specifically, Section 4.1 develops the general framework for first exit analysis of Markov processes by establishing Theorem 2.9. Section 4.2 then applies the framework in the context of heavy-tailed stochastic difference equations and proves Theorem 2.6.

4.1 Proof of Theorem 2.9

Our proof of Theorem 2.9 hinges on the following proposition.

Proposition 4.1. *Suppose that Condition 1 holds.*

(i) If $C(\cdot)$ is a probability measure supported on I^c (i.e., $C(I^c) = 1$), then for each measurable set $B \subseteq \mathbb{S}$ and $t \geq 0$, there exists $\delta_{t,B}(\epsilon)$ such that

$$C(B^\circ) \cdot e^{-t} - \delta_{t,B}(\epsilon) \leq \liminf_{\eta \downarrow 0} \inf_{x \in A(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^\eta_{I(\epsilon)^c}(x) > t;\ V^\eta_{\tau_\epsilon}(x) \in B\big)$$
$$\leq \limsup_{\eta \downarrow 0} \sup_{x \in A(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^\eta_{I(\epsilon)^c}(x) > t;\ V^\eta_{\tau_\epsilon}(x) \in B\big) \leq C(B^-) \cdot e^{-t} + \delta_{t,B}(\epsilon)$$

for all sufficiently small $\epsilon > 0$, where $\delta_{t,B}(\epsilon) \to 0$ as $\epsilon \to 0$.

(ii) If $C(I^c) = 0$ (i.e., $C(\cdot)$ is trivially zero), then for each $t > 0$, there exists $\delta_t(\epsilon)$ such that

$$\limsup_{\eta \downarrow 0} \sup_{x \in A(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^\eta_{I(\epsilon)^c}(x) > t\big) \leq \delta_t(\epsilon)$$

for all $\epsilon > 0$ sufficiently small, where $\delta_t(\epsilon) \to 0$ as $\epsilon \to 0$.

Proof. Fix some measurable $B \subseteq \mathbb{S}$ and $t \geq 0$. Henceforth in the proof, given any choice of $0 < r < R$, we only consider $\epsilon \in (0, \epsilon_B)$ and T sufficiently large such that Condition 1 holds with T replaced with $\frac{1-r}{2}T$, $\frac{2-r}{2}T$, rT, and RT. Let

$$\rho^\eta_i(x) \triangleq \inf\Big\{j \geq \rho^\eta_{i-1}(x) + \lfloor rT/\eta \rfloor : V^\eta_j(x) \in A(\epsilon)\Big\}$$

where $\rho^\eta_0(x) = 0$. One can interpret these as the i^{th} asymptotic regeneration times after cooling period rT/η. We start with the following two observations: For any $0 < r < R$,

$$\begin{aligned}
\mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \in \big(RT/\eta,\, \rho^\eta_1(y)\big]\Big) &\leq \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \wedge \rho^\eta_1(y) > RT/\eta\Big)\\
&\leq \mathbf{P}\Big(V^\eta_j(y) \in I(\epsilon)\setminus A(\epsilon) \quad \forall j \in \big[\lfloor rT/\eta\rfloor,\, RT/\eta\big]\Big)\\
&\leq \sup_{z\in I(\epsilon)\setminus A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{(I(\epsilon)\setminus A(\epsilon))^c}(z) > \frac{R-r}{2}T/\eta\Big)\\
&= \gamma(\eta)T/\eta \cdot o(1), \qquad (4.1)
\end{aligned}$$

where the last equality is from (2.38) of Condition 1, and

$$\begin{aligned}
&\sup_{y\in A(\epsilon)} \mathbf{P}\Big(V^\eta_{\tau_\epsilon}(y) \in B;\ \tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)\\
&\leq \sup_{y\in A(\epsilon)} \mathbf{P}\Big(V^\eta_{\tau_\epsilon}(y) \in B;\ \tau^\eta_{I(\epsilon)^c}(y) \leq RT/\eta\Big) + \sup_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \in \big(RT/\eta,\, \rho^\eta_1(y)\big]\Big)\\
&\leq \sup_{y\in A(\epsilon)} \mathbf{P}\Big(V^\eta_{\tau_\epsilon}(y) \in B;\ \tau^\eta_{I(\epsilon)^c}(y) \leq RT/\eta\Big) + \gamma(\eta)T/\eta\cdot o(1)\\
&\leq \big(C(B^-) + \delta_B(\epsilon, RT) + o(1)\big)\cdot \gamma(\eta)RT/\eta, \qquad (4.2)
\end{aligned}$$

where the second inequaility is from (4.1) and the last equality is from (2.37) of Condition 1.

Proof of Case (i)**.**

We work with different choices of R and r for the lower and upper bounds. For the lower bound, we work with $R > r > 1$ and set $K = \left\lceil \frac{t/\gamma(\eta)}{T/\eta} \right\rceil$. Note that for $\eta \in \big(0, (r-1)T\big)$, we have $\lfloor rT/\eta\rfloor \geq T/\eta$ and hence $\rho^\eta_K(x) \geq K\lfloor rT/\eta\rfloor \geq t/\gamma(\eta)$. Note also that from the Markov property conditioning on $\mathcal{F}_{\rho^\eta_j(x)}$,

$$\begin{aligned}
&\inf_{x\in A(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^\eta_{I(\epsilon)^c}(x) > t;\ V^\eta_{\tau_\epsilon}(x) \in B\big)\\
&\geq \inf_{x\in A(\epsilon)} \mathbf{P}(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_K(x);\ V^\eta_{\tau_\epsilon}(x) \in B) = \inf_{x\in A(\epsilon)} \sum_{j=K}^\infty \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) \in \big(\rho^\eta_j(x),\, \rho^\eta_{j+1}(x)\big];\ V^\eta_{\tau_\epsilon}(x) \in B\Big)\\
&\geq \inf_{x\in A(\epsilon)} \sum_{j=K}^\infty \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) \in \big(\rho^\eta_j(x),\, \rho^\eta_j(x) + T/\eta\big];\ V^\eta_{\tau_\epsilon}(x) \in B\Big)\\
&\geq \inf_{x\in A(\epsilon)} \sum_{j=K}^\infty \inf_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq T/\eta;\ V^\eta_{\tau_\epsilon}(y) \in B\Big)\cdot \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\Big).\\
&\geq \inf_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq T/\eta;\ V^\eta_{\tau_\epsilon}(y) \in B\Big)\cdot \sum_{j=K}^\infty \inf_{x\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\Big). \qquad (4.3)
\end{aligned}$$

From the Markov property conditioning on $\mathcal{F}_{\rho^\eta_j(x)}$, the second term can be bounded as follows:

$$\begin{aligned}
&\sum_{j=K}^\infty \inf_{x\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\Big)\\
&\geq \sum_{j=0}^\infty \bigg(\inf_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) > \rho^\eta_1(y)\Big)\bigg)^{K+j} = \sum_{j=0}^\infty \bigg(1 - \sup_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)\bigg)^{K+j}\\
&= \frac{1}{\sup_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)} \cdot \bigg(1 - \sup_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)\bigg)^{\lceil \frac{t/\gamma(\eta)}{T/\eta}\rceil}
\end{aligned}$$

$$\geq \frac{1}{\big(1+\delta_{\mathbb{S}}(\epsilon,RT)+o(1)\big)\cdot\gamma(\eta)RT/\eta}\cdot\bigg(1-\big(1+\delta_{\mathbb{S}}(\epsilon,RT)+o(1)\big)\cdot\gamma(\eta)RT/\eta\bigg)^{\frac{t/\gamma(\eta)}{T/\eta}+1}. \tag{4.4}$$

where the last inequality is from (4.2) with $B=\mathbb{S}$. From (4.3), (4.4), and (2.36) of Condition 1, we have

$$\begin{aligned}
&\liminf_{\eta\downarrow 0}\inf_{x\in A(\epsilon)}\mathbf{P}\big(\gamma(\eta)\tau^{\eta}_{I(\epsilon)^c}(x)>t;\ V^{\eta}_{\tau_\epsilon}(x)\in B\big)\\
&\geq \liminf_{\eta\downarrow 0}\frac{C(B^\circ)-\delta_B(\epsilon,T)+o(1)}{\big(1+\delta_{\mathbb{S}}(\epsilon,RT)+o(1)\big)\cdot R}\cdot\bigg(1-\big(1+\delta_{\mathbb{S}}(\epsilon,RT)+o(1)\big)\cdot\gamma(\eta)RT/\eta\bigg)^{\frac{R\cdot t}{\gamma(\eta)RT/\eta}+1}.\\
&\geq \frac{C(B^\circ)-\delta_B(\epsilon,T)}{1+\delta_{\mathbb{S}}(\epsilon,RT)}\cdot\exp\Big(-\big(1+\delta_{\mathbb{S}}(\epsilon,RT)\big)\cdot R\cdot t\Big).
\end{aligned}$$

By taking limit $T\to\infty$ and then considering an R arbitrarily close to 1, it is straightforward to check that the desired lower bound holds.

Moving on to the upper bound, we set $R=1$ and fix an arbitrary $r\in(0,1)$. Set $k=\Big\lfloor\frac{t/\gamma(\eta)}{T/\eta}\Big\rfloor$ and note that

$$\begin{aligned}
\sup_{x\in A(\epsilon)}\mathbf{P}\big(\gamma(\eta)\tau^{\eta}_{I(\epsilon)^c}(x)>t;\ V^{\eta}_{\tau_\epsilon}(x)\in B\big) &= \sup_{x\in A(\epsilon)}\mathbf{P}\big(\tau^{\eta}_{I(\epsilon)^c}(x)>t/\gamma(\eta);\ V^{\eta}_{\tau_\epsilon}(x)\in B\big)\\
&= \underbrace{\sup_{x\in A(\epsilon)}\mathbf{P}\big(\tau^{\eta}_{I(\epsilon)^c}(x)>t/\gamma(\eta)\geq\rho^{\eta}_k(x);\ V^{\eta}_{\tau_\epsilon}(x)\in B\big)}_{\text{(I)}}\\
&\quad+\underbrace{\sup_{x\in A(\epsilon)}\mathbf{P}\big(\tau^{\eta}_{I(\epsilon)^c}(x)>t/\gamma(\eta);\ \rho^{\eta}_k(x)>t/\gamma(\eta);\ V^{\eta}_{\tau_\epsilon}(x)\in B\big)}_{\text{(II)}}
\end{aligned}$$

We first show that (II) vanishes as $\eta\to 0$. Our proof hinges on the following claim:

$$\big\{\tau^{\eta}_{I(\epsilon)^c}(x)>t/\gamma(\eta);\ \rho^{\eta}_k(x)>t/\gamma(\eta)\big\}\ \subseteq\ \bigcup_{j=1}^{k}\big\{\tau^{\eta}_{I(\epsilon)^c}(x)\wedge\rho^{\eta}_j(x)-\rho^{\eta}_{j-1}(x)\geq T/\eta\big\}$$

Proof of the claim: Suppose that $\tau^{\eta}_{I(\epsilon)^c}(x)>t/\gamma(\eta)$ and $\rho^{\eta}_k(x)>t/\gamma(\eta)$. Let $k^*\triangleq\max\{j\geq 1:\rho^{\eta}_j(x)\leq t/\gamma(\eta)\}$. Note that $k^*<k$. We consider two cases separately: (i) $\rho^{\eta}_{k^*}(x)/k^*>(t/\gamma(\eta)-T/\eta)/k^*$ and (ii) $\rho^{\eta}_{k^*}(x)\leq t/\gamma(\eta)-T/\eta$. In case of (i), since $\rho^{\eta}_{k^*}(x)/k^*$ is the average of $\{\rho^{\eta}_j(x)-\rho^{\eta}_{j-1}(x):j=1,\ldots,k^*\}$, there exists $j^*\leq k^*$ such that

$$\rho^{\eta}_{j^*}(x)-\rho^{\eta}_{j^*-1}(x)>\frac{t/\gamma(\eta)-T/\eta}{k^*}\geq\frac{kT/\eta-T/\eta}{k-1}=T/\eta$$

Note that since $\rho^{\eta}_{j^*}(x)\leq\rho^{\eta}_{k^*}(x)\leq t/\gamma(\eta)\leq\tau^{\eta}_{I(\epsilon)^c}(x)$, this proves the claim for case (i). For case (ii), note that

$$\rho^{\eta}_{k^*+1}(x)\wedge\tau^{\eta}_{I(\epsilon)^c}(x)-\rho^{\eta}_{k^*}(x)\geq t/\gamma(\eta)-(t/\gamma(\eta)-T/\eta)=T/\eta,$$

which proves the claim.

Now, with the claim in hand, we have that

$$\begin{aligned}
\text{(II)} &\leq \sum_{j=1}^{k}\sup_{x\in A(\epsilon)}\mathbf{P}\big(\tau^{\eta}_{I(\epsilon)^c}(x)\wedge\rho^{\eta}_j(x)-\rho^{\eta}_{j-1}(x)\geq T/\eta\big)\\
&= \sum_{j=1}^{k}\sup_{x\in A(\epsilon)}\mathbf{E}\Big[\mathbf{P}\big(\tau^{\eta}_{I(\epsilon)^c}(x)\wedge\rho^{\eta}_j(x)-\rho^{\eta}_{j-1}(x)\geq T/\eta\big|\mathcal{F}_{\rho^{\eta}_{j-1}(x)}\big)\Big]
\end{aligned}$$

$$\leq \sum_{j=1}^{k} \sup_{y\in A(\epsilon)} \mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(y) \wedge \rho^\eta_1(y) \geq T/\eta\big)$$
$$\leq \frac{t}{\gamma(\eta)T/\eta}\cdot \gamma(\eta)T/\eta \cdot o(1) = o(1)$$

for sufficiently large T's, where the last inequality is from the definition of k and (4.1). We are now left with bounding (I) from above.

$$\text{(I)} = \sup_{x\in A(\epsilon)} \mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(x) > t/\gamma(\eta) \geq \rho^\eta_K(x);\ V^\eta_{\tau_\epsilon}(x)\in B\big) \leq \sup_{x\in A(\epsilon)} \mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_K(x);\ V^\eta_{\tau_\epsilon}(x)\in B\big)$$
$$= \sum_{j=k}^{\infty} \sup_{x\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) \in \big(\rho^\eta_j(x),\, \rho^\eta_{j+1}(x)\big];\ V^\eta_{\tau_\epsilon}(x)\in B\Big)$$
$$= \sum_{j=k}^{\infty} \sup_{x\in A(\epsilon)} \mathbf{E}\Big[\mathbf{E}\Big[\mathbb{I}\{V^\eta_{\tau_\epsilon}(x)\in B\}\cdot \mathbb{I}\{\tau^\eta_{I(\epsilon)^c}(x) \leq \rho^\eta_{j+1}(x)\}\Big|\mathcal{F}_{\rho^\eta_j(x)}\Big]\cdot \mathbb{I}\{\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\}\Big]$$
$$\leq \sum_{j=k}^{\infty} \sup_{x\in A(\epsilon)} \mathbf{E}\Big[\sup_{y\in A(\epsilon)} \mathbf{P}\Big(V^\eta_{\tau_\epsilon}(y)\in B;\ \tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)\cdot \mathbb{I}\{\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\}\Big]$$
$$= \sup_{y\in A(\epsilon)} \mathbf{P}\Big(V^\eta_{\tau_\epsilon}(y)\in B;\ \tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)\cdot \sum_{j=k}^{\infty} \sup_{x\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\Big)$$

The first term can be bounded via (4.2) with $R = 1$:

$$\sup_{y\in A(\epsilon)} \mathbf{P}\Big(V^\eta_{\tau_\epsilon}(y)\in B;\ \tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)$$
$$\leq \big(C(B^-) + \delta_B(\epsilon, T) + o(1)\big)\cdot \gamma(\eta)T/\eta + \frac{1-r}{2}\cdot \gamma(\eta)T/\eta\cdot o(1)$$

whereas the second term is bounded via (2.36) of Condition 1 as follows:

$$\sum_{j=k}^{\infty} \sup_{x\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\Big)$$
$$\leq \sum_{j=0}^{\infty}\bigg(\sup_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) > \lfloor rT/\eta\rfloor\Big)\bigg)^{k+j} = \sum_{j=0}^{\infty}\bigg(1 - \inf_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq rT/\eta\Big)\bigg)^{k+j}$$
$$\leq \frac{1}{\inf_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y)\leq rT/\eta\Big)}\cdot \bigg(1 - \inf_{y\in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq rT/\eta\Big)\bigg)^{\frac{t/\gamma(\eta)}{T/\eta}-1}$$
$$= \frac{1}{r\cdot\big(1-\delta_B(\epsilon, rT) + o(1)\big)\cdot\gamma(\eta)T/\eta}\cdot\Big(1 - r\cdot\big(1-\delta_B(\epsilon, rT) + o(1)\big)\cdot \gamma(\eta)T/\eta\Big)^{\frac{t}{\gamma(\eta)T/\eta}-1}$$

Therefore,

$$\limsup_{\eta\downarrow 0}\sup_{x\in A(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^\eta_{I(\epsilon)^c}(x) > t;\ V^\eta_{\tau_\epsilon}(x)\in B\big) \leq \frac{C(B^-)+\delta_B(\epsilon, T)}{r\cdot(1-\delta_B(\epsilon, rT))}\cdot \exp\Big(-r\cdot\big(1-\delta_B(\epsilon, rT)\big)\cdot t\Big).$$

Again, taking $T\to\infty$ and considering r arbitrarily close to 1, we can check that the desired upper bound holds.

Proof of Case (ii)**.**

We work with $R = 1$ and set $K = \left\lceil \frac{t/\gamma(\eta)}{T/\eta} \right\rceil$. Again, for $\eta \in \big(0, (r-1)T\big)$, we have $\lfloor rT/\eta \rfloor \geq T/\eta$ and hence $\rho^\eta_K(x) \geq K\lfloor rT/\eta \rfloor \geq t/\gamma(\eta)$. By the Markov property conditioning on $\mathcal{F}_{\rho^\eta_j(x)}$,

$$
\begin{aligned}
&\sup_{x \in A(\epsilon)} \mathbf{P}\big(\gamma(\eta)\tau^\eta_{I(\epsilon)^c}(x) \leq t\big) \\
&\leq \sup_{x \in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) \leq \rho^\eta_K(x)\Big) = \sup_{x \in A(\epsilon)} \sum_{j=1}^K \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) \in \big(\rho^\eta_{j-1}(x), \rho^\eta_j(x)\big]\Big) \\
&\leq \sum_{j=1}^K \sup_{y \in A(\epsilon)} \bigg(1 - \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big)\bigg)^{j-1} \cdot \sup_{y \in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big) \\
&\leq K \cdot \sup_{y \in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq \rho^\eta_1(y)\Big) \leq K \cdot \big(\delta_{I^c}(\epsilon, T) + o(1)\big) \cdot \gamma(\eta)T/\eta \\
&\qquad \text{by (4.2) (with } B = I^c\text{) and the running assumption of Case } (ii) \text{ that } C(\cdot) \equiv 0 \\
&\leq \frac{2t/\gamma(\eta)}{T/\eta} \cdot \big(\delta_{I^c}(\epsilon, T) + o(1)\big) \cdot \gamma(\eta)T/\eta \qquad \text{for all } \eta \text{ small enough under } K = \lceil \frac{t/\gamma(\eta)}{T/\eta} \rceil \\
&= 2t \cdot \big(\delta_{I^c}(\epsilon, T) + o(1)\big).
\end{aligned}
$$

Lastly, by Condition 1 (specifically, $\lim_{\epsilon \downarrow 0} \lim_{T \uparrow \infty} \delta_{I^c}(\epsilon, T) = 0$ in Definition 2.8), we verify the upper bounds in Case (ii) and conclude the proof. □

Now, we are ready to prove Theorem 2.9.

Proof of Theorem 2.9. We focus on the proof of Case (i) since the proof of Case (ii) is almost identical, with the only key difference being that we apply part (ii) of Proposition 4.1 instead of part (i).

We first claim that for any $\epsilon, \epsilon' > 0$, $t \geq 0$, and measurable $B \subseteq \mathbb{S}$,

$$
\begin{aligned}
C(B^\circ) \cdot e^{-t} - \delta_{t,B}(\epsilon) &\leq \liminf_{\eta \downarrow 0} \inf_{x \in I(\epsilon')} \mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_{I(\epsilon)^c}(x) > t,\, V^\eta_{\tau_\epsilon}(x) \in B\Big) \\
&\leq \limsup_{\eta \downarrow 0} \sup_{x \in I(\epsilon')} \mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_{I(\epsilon)^c}(x) > t,\, V^\eta_{\tau_\epsilon}(x) \in B\Big) \leq C(B^-) \cdot e^{-t} + \delta_{t,B}(\epsilon)
\end{aligned} \tag{4.5}
$$

where $\delta_{t,B}(\epsilon)$ is characterized in part (i) of Proposition 4.1 such that $\delta_{t,B}(\epsilon) \to 0$ as $\epsilon \to 0$. Now, note that for any measurable $B \subseteq I^c$,

$$
\begin{aligned}
&\mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_{I^c}(x) > t,\, V^\eta_\tau(x) \in B\Big) \\
&= \underbrace{\mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_{I^c}(x) > t,\, V^\eta_\tau(x) \in B,\, V^\eta_{\tau_\epsilon}(x) \in I\Big)}_{\text{(I)}} + \underbrace{\mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_{I^c}(x) > t,\, V^\eta_\tau(x) \in B,\, V^\eta_{\tau_\epsilon}(x) \notin I\Big)}_{\text{(II)}}
\end{aligned}
$$

and since

$$
\text{(I)} \leq \mathbf{P}\Big(V^\eta_{\tau_\epsilon}(x) \in I\Big) \qquad \text{and} \qquad \text{(II)} = \mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_\epsilon(x) > t,\, V^\eta_{\tau_\epsilon}(x) \in B \setminus I\Big),
$$

we have that

$$
\begin{aligned}
\liminf_{\eta \downarrow 0} \inf_{x \in I(\epsilon')} \mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_{I^c}(x) > t,\, V^\eta_\tau(x) \in B\Big) &\geq \liminf_{\eta \downarrow 0} \inf_{x \in I(\epsilon')} \mathbf{P}\Big(\gamma(\eta) \cdot \tau^\eta_\epsilon(x) > t,\, V^\eta_{\tau_\epsilon}(x) \in B \setminus I\Big) \\
&\geq C\big((B \setminus I)^\circ\big) \cdot e^{-t} - \delta_{t,B \setminus I}(\epsilon) \\
&= C(B^\circ) \cdot e^{-t} - \delta_{t,B \setminus I}(\epsilon)
\end{aligned}
$$

due to $B \subseteq I^c$, and

$$\begin{aligned}
&\limsup_{\eta\downarrow 0}\sup_{x\in I(\epsilon')}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_{I^c}(x)>t,\,V^\eta_\tau(x)\in B\Big)\\
&\leq \limsup_{\eta\downarrow 0}\sup_{x\in I(\epsilon')}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_{\epsilon}(x)>t,\,V^\eta_{\tau_\epsilon}(x)\in B\setminus I\Big)+\limsup_{\eta\downarrow 0}\sup_{x\in I(\epsilon')}\mathbf{P}\Big(V^\eta_{\tau_\epsilon}(x)\in I\Big)\\
&\leq C\big((B\setminus I)^-\big)\cdot e^{-t}+\delta_{t,B\setminus I}(\epsilon)+C(I^-)+\delta_{0,I}(\epsilon)\\
&= C(B^-)\cdot e^{-t}+\delta_{t,B\setminus I}(\epsilon)+\delta_{0,I}(\epsilon).
\end{aligned}$$

Taking $\epsilon\to 0$, we arrive at the desired lower and upper bounds of the theorem. Now we are left with the proof of the claim (4.5) is true. Note that for any $x \in I$,

$$\begin{aligned}
&\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_\epsilon(x)>t,\,V^\eta_{\tau_\epsilon}(x)\in B\Big)\\
&=\mathbf{E}\bigg[\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_\epsilon(x)>t,\,V^\eta_{\tau_\epsilon}(x)\in B\Big|\mathcal{F}_{\tau^\eta_{A(\epsilon)}(x)}\Big)\cdot\Big(\mathbb{I}\{\tau^\eta_{A(\epsilon)}(x)\leq T/\eta\}+\mathbb{I}\{\tau^\eta_{A(\epsilon)}(x)>T/\eta\}\Big)\bigg]
\end{aligned}\tag{4.6}$$

Fix an arbitrary $s > 0$, and note that from the Markov property,

$$\begin{aligned}
&\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_\epsilon(x)>t,\,V^\eta_{\tau_\epsilon}(x)\in B\Big)\\
&\leq\mathbf{E}\bigg[\sup_{y\in A(\epsilon)}\mathbf{P}\Big(\tau^\eta_\epsilon(y)>t/\gamma(\eta)-T/\eta,\,V^\eta_{\tau_\epsilon}(y)\in B\Big)\cdot\mathbb{I}\{\tau^\eta_{A(\epsilon)}(x)\leq T/\eta\}\bigg]+\mathbf{P}\Big(\tau^\eta_{A(\epsilon)}(x)>T/\eta\Big)\\
&\leq\sup_{y\in A(\epsilon)}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_\epsilon(y)>t-s,\,V^\eta_{\tau_\epsilon}(y)\in B\Big)+\mathbf{P}\Big(\tau^\eta_{A(\epsilon)}(x)>T/\eta\Big)
\end{aligned}$$

for sufficiently small η's; here, we applied $\gamma(\eta)/\eta \to 0$ as $\eta \downarrow 0$ in the last inequality. In light of part (i) of Proposition 4.1, by taking $T\to\infty$ we yield

$$\limsup_{\eta\downarrow 0}\sup_{x\in I(\epsilon')}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_\epsilon(x)>t,\,V^\eta_{\tau_\epsilon}(x)\in B\Big)\leq C(B^-)\cdot e^{-(t-s)}+\delta_{t,B}(\epsilon)$$

Considering an arbitrarily small $s > 0$, we get the upper bound of the claim (4.5). For the lower bound, again from (4.6) and the Markov property,

$$\begin{aligned}
&\liminf_{\eta\downarrow 0}\inf_{x\in I(\epsilon')}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_\epsilon(x)>t,\,V^\eta_{\tau_\epsilon}(x)\in B\Big)\\
&\geq\liminf_{\eta\downarrow 0}\inf_{x\in I(\epsilon')}\mathbf{E}\bigg[\inf_{y\in A(\epsilon)}\mathbf{P}\Big(\tau^\eta_\epsilon(y)>t/\gamma(\eta),\,V^\eta_{\tau_\epsilon}(y)\in B\Big)\cdot\mathbb{I}\{\tau^\eta_{A(\epsilon)}(x)\leq T/\eta\}\bigg]\\
&\geq\liminf_{\eta\downarrow 0}\inf_{y\in A(\epsilon)}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_\epsilon(y)>t,\,V^\eta_{\tau_\epsilon}(y)\in B\Big)\cdot\inf_{x\in I(\epsilon')}\mathbf{P}\big(\tau^\eta_{A(\epsilon)}(x)\leq T/\eta\big)\\
&\geq C(B^\circ)-\delta_{t,B}(\epsilon),
\end{aligned}$$

which is the desired lower bound of the claim (4.5). This concludes the proof. □

4.2 Proof of Theorem 2.6

In this section, we apply the framework developed in Section 2.3.2 and prove Theorem 2.6. Throughout this section, we impose Assumptions 1, 2, and 4. Besides, we fix a few useful constants. Recall the definition of the discretized width metric $\mathcal{J}^I_b$ defined in (2.33). To prove Theorem 2.6, in this section we fix some $b > 0$ such that the conditions in Theorem 2.6 hold. This allows us to fix some $\breve{\epsilon} > 0$ small enough such that

$$\bar{B}_{\breve{\epsilon}}(\mathbf{0})\subseteq I,\qquad \boldsymbol{a}(\boldsymbol{x})\boldsymbol{x}<0\ \forall\boldsymbol{x}\in\bar{B}_{\breve{\epsilon}}(\mathbf{0})\setminus\{\mathbf{0}\},\quad \inf\Big\{\|\boldsymbol{x}-\boldsymbol{y}\|:\ \boldsymbol{x}\in I^c,\ \boldsymbol{y}\in\mathcal{G}^{(\mathcal{J}^I_b-1)|b}(\breve{\epsilon})\Big\}>0.\tag{4.7}$$

Here, $\bar{B}_r(\boldsymbol{x}) = \{\boldsymbol{x} :\ \|\boldsymbol{x}\| \leq r\}$ is the closed ball with radius r centered at $\boldsymbol{x}$. An direct implication of the first condition in (4.7) is the following positive invariance property under the gradient field $\boldsymbol{a}(\cdot)$: for any $r \in (0, \breve{\epsilon}]$,

$$\boldsymbol{y}_t(\boldsymbol{x}) \in \bar{B}_r(\boldsymbol{0}) \qquad \forall \boldsymbol{x} \in \bar{B}_r(\boldsymbol{0}). \tag{4.8}$$

Next, for any $\epsilon \in (0, \breve{\epsilon})$, let

$$\breve{I}(\epsilon) \triangleq \Big\{ \boldsymbol{x} \in I :\ \big\|\boldsymbol{y}_{1/\epsilon}(\boldsymbol{x})\big\| < \breve{\epsilon} \Big\} \tag{4.9}$$

with the ODE $\boldsymbol{y}_t(x)$ defined in (2.28). By Gronwall's inequality, it is easy to see that $\breve{I}(\epsilon)$ is an open set. Meanwhile, by Assumption 4, given any $\boldsymbol{x} \in I$ we must have $\boldsymbol{x} \in \breve{I}(\epsilon)$ for all $\epsilon > 0$ small enough. As a result, the collection of open sets $\{\breve{I}(\epsilon) :\ \epsilon \in (0, \breve{\epsilon})\}$ provides a covering for I:

$$\bigcup_{\epsilon \in (0, \breve{\epsilon})} \breve{I}(\epsilon) = I.$$

Next, recall that we use $I_\epsilon = \{\boldsymbol{y} \in \mathbb{R}^n :\ \|\boldsymbol{x} - \boldsymbol{y}\| < \epsilon \implies \boldsymbol{x} \in I\}$ to denote the ϵ-shrinkage of the set I. Given any $\epsilon > 0$, note that I_ϵ is an open set and, by definition, its closure I_ϵ^- is still bounded away from I^c, i.e., $\|\boldsymbol{x} - \boldsymbol{y}\| \geq \epsilon$ for all $\boldsymbol{x} \in I_\epsilon^-$, $\boldsymbol{y} \in I^c$. Then from the continuity of $\boldsymbol{a}(\cdot)$ (see Assumption 2), the boundedness of set I and hence $I_\epsilon^- \subseteq I$, as well as property (4.8), we know that given any $\epsilon > 0$, the claim

$$\|\boldsymbol{y}_T(\boldsymbol{x})\| < \breve{\epsilon} \qquad \forall \boldsymbol{x} \in I_\epsilon^-$$

holds for all $T > 0$ large enough. This confirms that given $\epsilon > 0$, it holds for all $\epsilon' > 0$ small enough that

$$I_\epsilon^- \subseteq \breve{I}(\epsilon'). \tag{4.10}$$

As a direct consequence of the discussion above, we highlight another important property of the sets $\mathcal{G}^{(k)|b}(\epsilon)$ defined in (2.32). For any $k \in \mathbb{N}$, $b > 0$, and $\epsilon \geq 0$, let

$$\bar{\mathcal{G}}^{(k)|b}(\epsilon) \triangleq \Big\{ \boldsymbol{y}_t(\boldsymbol{x}) :\ \boldsymbol{x} \in \mathcal{G}^{(k)|b}(\epsilon),\ t \geq 0 \Big\}, \tag{4.11}$$

where $\boldsymbol{y}_\cdot(\boldsymbol{x})$ is the ODE defined in (2.28). First, due to (4.10) and the fact that $\mathcal{G}^{(\mathcal{J}_b^I - 1)|b}(\breve{\epsilon})$ is bounded away from I^c (see (4.7)), given any $\epsilon \in (0, \breve{\epsilon}]$, it holds for all $\epsilon' > 0$ small enough that $\mathcal{G}^{(\mathcal{J}_b^I - 1)|b}(\epsilon) \subseteq \breve{I}(\epsilon')$. Furthermore, we claim that $\bar{\mathcal{G}}^{(\mathcal{J}_b^I - 1)|b}(\breve{\epsilon})$ is also bounded away from I^c, i.e.,

$$\inf \Big\{ \|\boldsymbol{x} - \boldsymbol{z}\| :\ \boldsymbol{x} \in \bar{\mathcal{G}}^{(\mathcal{J}_b^I - 1)|b}(\breve{\epsilon}),\ \boldsymbol{z} \in I^c \Big\} > 0. \tag{4.12}$$

Again, this can be argued with a proof by contradiction. Suppose there exist sequences $\boldsymbol{x}'_n \in \bar{\mathcal{G}}^{(\mathcal{J}_b^I - 1)|b}(\breve{\epsilon})$ and $\boldsymbol{z}_n \notin I$ such that $\|\boldsymbol{x}'_n - \boldsymbol{z}_n\| \leq 1/n$. By definition of $\bar{\mathcal{G}}^{(\mathcal{J}_b^I - 1)|b}(\breve{\epsilon})$, there exist sequences $\boldsymbol{x}_n \in \mathcal{G}^{(\mathcal{J}_b^I - 1)|b}(\breve{\epsilon})$ and $t_n \geq 0$ such that $\boldsymbol{x}'_n = \boldsymbol{y}_{t_n}(\boldsymbol{x}_n)$ for all $n \geq 1$. Furthermore, recall that we have $\mathcal{G}^{(\mathcal{J}_b^I - 1)|b}(\breve{\epsilon}) \subseteq \breve{I}(\epsilon)$ for $\epsilon > 0$ small enough. On the other hand, by the definition of $\breve{I}(\epsilon)$ in (4.9) and the property (4.8), it holds for all $n \geq 1$ that $\boldsymbol{y}_t(\boldsymbol{x}_n) \in \bar{B}_{\breve{\epsilon}}(\boldsymbol{0})\ \forall t \geq 1/\epsilon$. Since $\boldsymbol{z}_n \notin I$ and $\bar{B}_{\breve{\epsilon}}(\boldsymbol{0}) \subseteq I$ (see (4.7)), we must have $t_n < 1/\breve{\epsilon}$ for all n. Together with the boundedness of I, by picking a sub-sequence if necessary, we can w.l.o.g. assume that $\boldsymbol{x}_n \to \boldsymbol{x}^*$ for some $\boldsymbol{x}^* \in \big(\mathcal{G}^{(\mathcal{J}_b^I - 1)|b}\big)^- \subset I$ and $t_n \to t^*$ for some $t^* \in [0, 1/\breve{\epsilon}]$. Since $\boldsymbol{x}^* \in I$, by Assumption 4 we must have $\boldsymbol{y}_{t^*}(\boldsymbol{x}^*) \in I$. By the continuity of the flow (specifically, using Gronwall's inequality) and the fact that I is an open set, we have $\boldsymbol{z}_n = \boldsymbol{y}_{t_n}(\boldsymbol{x}_n) \in I$ for all n large enough. This contradicts our choice that $\boldsymbol{z}_n \notin I$ for all n, thus establishing (4.12). Now, by (4.7), (4.8), and (4.12), we can fix some $\bar{\epsilon} > 0$ small enough such that the following claims hold:

$$\bar{B}_{\bar{\epsilon}}(\mathbf{0}) \subseteq I_{\bar{\epsilon}}, \tag{4.13}$$

$$r \in (0, \bar{\epsilon}],\ \boldsymbol{x} \in \bar{B}_r(\mathbf{0}) \quad \Longrightarrow \quad \boldsymbol{y}_t(\boldsymbol{x}) \in \bar{B}_r(\mathbf{0})\ \forall t \geq 0, \tag{4.14}$$

$$\inf\Big\{ \|\boldsymbol{x} - \boldsymbol{z}\| :\ \boldsymbol{x} \in \bar{\mathcal{G}}^{(\mathcal{J}_b^I - 1)|b}(2\bar{\epsilon}),\ \boldsymbol{z} \notin I_{\bar{\epsilon}} \Big\} > \bar{\epsilon}. \tag{4.15}$$

Moving on, let

$$\boldsymbol{t}_{\boldsymbol{x}}(\epsilon) \triangleq \inf\Big\{ t \geq 0 :\ \boldsymbol{y}_t(\boldsymbol{x}) \in \bar{B}_\epsilon(\mathbf{0}) \Big\}$$

be the hitting time of the closed ball $\bar{B}_\epsilon(\mathbf{0})$ for the ODE $\boldsymbol{y}_t(\boldsymbol{x})$, and let

$$\boldsymbol{t}(\epsilon) \triangleq \sup\Big\{ \boldsymbol{t}_{\boldsymbol{x}}(\epsilon) :\ \boldsymbol{x} \in I_\epsilon^- \Big\} \tag{4.16}$$

be the upper bound for the hitting times $\boldsymbol{t}_{\boldsymbol{x}}(\epsilon)$ over $\boldsymbol{x} \in I_\epsilon^-$. Again, from the continuity of $\boldsymbol{a}(\cdot)$, the contraction of $\boldsymbol{y}_t(\boldsymbol{x})$ around the origin (see Assumption 4 and its implication (4.14)), and the boundedness of I and hence I_ϵ^-, we have $\boldsymbol{t}(\epsilon) < \infty$ for any $\epsilon > 0$. Besides, by definition of $\boldsymbol{t}(\cdot)$, we have

$$\boldsymbol{y}_t(\boldsymbol{x}) \in \bar{B}_\epsilon(\mathbf{0}) \qquad \forall \boldsymbol{x} \in I_\epsilon^-,\ t \geq \boldsymbol{t}(\epsilon). \tag{4.17}$$

Furthermore, by repeating the arguments for (4.12), one can show that (for all $\epsilon > 0$)

$$\inf\Big\{ \|\boldsymbol{y}_t(\boldsymbol{x}) - \boldsymbol{z}\| :\ \boldsymbol{x} \in I_\epsilon^-,\ t \geq 0,\ \boldsymbol{z} \notin I \Big\} > 0. \tag{4.18}$$

Specifically, for the constant $\bar{\epsilon} > 0$ fixed in (4.13)–(4.15), by (4.18) we can find some $\bar{c} \in (0,1)$ such that

$$\Big\{ \boldsymbol{y}_t(\boldsymbol{x}) :\ \boldsymbol{x} \in I_{\bar{\epsilon}}^-,\ t \geq 0 \Big\} \subseteq I_{\bar{c}\bar{\epsilon}}. \tag{4.19}$$

Recall that we use E^- and E° to denote the closure and interior of any Borel set E. In our analysis below, we make use of the following inequality in Lemma 4.2. We collect its proof in Section D, together with the proofs of other useful properties regarding measures $\breve{\mathbf{C}}^{(k)|b}$.

Lemma 4.2. *Let $\bar{t}, \bar{\delta} \in (0, \infty)$ be the constants characterized in part (b) of Lemma D.2. Given $\Delta \in (0, \bar{\epsilon})$, there exists $\epsilon_0 = \epsilon_0(\Delta) > 0$ such that for any $\epsilon \in (0, \epsilon_0]$, $T \geq \bar{t}$, and Borel measurable $B \subseteq (I_\epsilon)^c$,*

$$(T - \bar{t}) \cdot \Big(\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(B_\Delta) - \breve{\boldsymbol{c}}(\epsilon_0) \Big) \leq \inf_{\boldsymbol{x}:\ \|\boldsymbol{x}\| \leq \epsilon} \mathbf{C}^{(\mathcal{J}_b^I)|b}_{[0,T]}\bigg(\Big(\breve{E}(\epsilon, B, T) \Big)^\circ;\ \boldsymbol{x} \bigg)$$
$$\leq \sup_{\boldsymbol{x}:\ \|\boldsymbol{x}\| \leq \epsilon} \mathbf{C}^{(\mathcal{J}_b^I)|b}_{[0,T]}\bigg(\Big(\breve{E}(\epsilon, B, T) \Big)^-;\ \boldsymbol{x} \bigg) \leq T \cdot \Big(\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(B^\Delta) + \breve{\boldsymbol{c}}(\epsilon_0) \Big)$$

where

$$\breve{E}(\epsilon, B, T) \triangleq \Big\{ \xi \in \mathbb{D}[0,T] :\ \exists t \leq T \text{ s.t. } \xi_t \in B \text{ and } \xi_s \in I(\epsilon)\ \forall s \in [0,t) \Big\}, \tag{4.20}$$

$$\breve{\boldsymbol{c}}(\epsilon) \triangleq \mathcal{J}_b^I \cdot (\bar{t})^{\mathcal{J}_b^I - 1} \cdot (\bar{\delta})^{-\alpha \cdot (\mathcal{J}_b^I - 1)} \cdot \epsilon^{\frac{\alpha}{2\mathcal{J}_b^I}}. \tag{4.21}$$

To see how we apply the framework developed in Section 2.3.2, let us specialize Condition 1 to a setting where $\mathbb{S} = \mathbb{R}$, $A(\epsilon) = \{\boldsymbol{x} \in \mathbb{R}^m :\ \|\boldsymbol{x}\| < \epsilon\}$, and the covering $I(\epsilon) = I_\epsilon$. Let $V_j^\eta(x) =$

$\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$. Meanwhile, for $C_b^I = \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\big(I^c\big)$, it is shown in Lemma D.3 that $C_b^I < \infty$. Now, recall that $H(\cdot) = \mathbf{P}(\|\boldsymbol{Z}_1\| > \cdot)$ and $\lambda(\eta) = \eta^{-1}H(\eta^{-1})$. Recall that in Theorem 2.6, we consider two cases: (i) $C_b^I \in (0,\infty)$, and (ii) $C_b^I = 0$. We first discuss our choices in Case (i). When $C_b^I > 0$, we set

$$C(\ \cdot\ ) \triangleq \frac{\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\ \cdot\ \setminus I)}{C_b^I}, \qquad \gamma(\eta) \triangleq C_b^I \cdot \eta \cdot \big(\lambda(\eta)\big)^{\mathcal{J}_b^I}. \tag{4.22}$$

The regularity conditions in Theorem 2.6 dictate that $\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\partial I) = 0$, and hence $C(\partial I) = 0$. Besides, note that $C(\cdot)$ is a probability measure and $\gamma(\eta)T/\eta = C_b^I T \cdot \big(\lambda(\eta)\big)^{\mathcal{J}_b^I}$. Besides, this corresponds to Case (i) for the location measure in the definition of asymptotic atoms; see the discussion before Definition 2.8.

The application of the framework developed in Section 2.3.2 (specifically, Theorem 2.9) hinges on the verification of (2.36)–(2.39). We start by verifying (2.36) and (2.37). First, given any Borel measurable $B \subseteq \mathbb{R}$, we specify the choice of function $\delta_B(\epsilon,T)$ in Condition 1. From the continuity of measures, we get $\lim_{\Delta\downarrow 0} \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^{\Delta}\cap I^c)\setminus(B^{-}\cap I^c)\Big) = 0$ and $\lim_{\Delta\downarrow 0} \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^{\circ}\cap I^c)\setminus(B_{\Delta}\cap I^c)\Big) = 0$. This allows us to fix a sequence $(\Delta^{(n)})_{n\geq 1}$ such that $\Delta^{(n+1)} \in (0, \Delta^{(n)}/2)$ and

$$\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^{\Delta^{(n)}} \cap I^c) \setminus (B^{-} \cap I^c)\Big) \vee \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^{\circ} \cap I^c) \setminus (B_{\Delta^{(n)}} \cap I^c)\Big) \leq 1/2^n \tag{4.23}$$

for each $n \geq 1$. Next, recall the definition of set $\breve{E}(\epsilon, B, T)$ in Lemma 4.2, and let $\widetilde{B}(\epsilon) \triangleq B \setminus I_\epsilon$. Using Lemma 4.2, we are able to fix another sequence of strictly decreasing positive real numbers $(\epsilon^{(n)})_{n\geq 1}$ such that $\epsilon^{(n)} \in (0, \bar{\epsilon}]$ $\forall n \geq 1$ and for any $n \geq 1$, $\epsilon \in (0, \epsilon^{(n)}]$, we have

$$\sup_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon} \mathbf{C}_{[0,T]}^{(\mathcal{J}_b^I)|b}\bigg(\Big(\breve{E}\big(\epsilon, \widetilde{B}(\epsilon), T\big)\Big)^{-};\ \boldsymbol{x}\bigg) \leq T \cdot \bigg(\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(\big(B \setminus I_\epsilon\big)^{\Delta^{(n)}}\Big) + \breve{\boldsymbol{c}}(\epsilon^{(n)})\bigg), \tag{4.24}$$

$$\inf_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon} \mathbf{C}_{[0,T]}^{(\mathcal{J}_b^I)|b}\bigg(\Big(\breve{E}\big(\epsilon, \widetilde{B}(\epsilon), T\big)\Big)^{\circ};\ \boldsymbol{x}\bigg) \geq (T - \bar{t}) \cdot \bigg(\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(\big(B \setminus I_\epsilon\big)_{\Delta^{(n)}}\Big) - \breve{\boldsymbol{c}}(\epsilon^{(n)})\bigg). \tag{4.25}$$

Besides, note that given any $\epsilon \in (0, \epsilon^{(1)}]$, there uniquely exists some $n = n_\epsilon \geq 1$ such that $\epsilon \in (\epsilon^{(n+1)}, \epsilon^{(n)}]$. This allows us to set

$$\begin{aligned} &\breve{\delta}_B(\epsilon,T) \\ &= T \cdot \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^{\Delta^{(n)}} \cap I^c)\setminus(B^{-} \cap I^c)\Big) \vee \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^{\circ} \cap I^c)\setminus(B_{\Delta^{(n)}} \cap I^c)\Big) \vee \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^{\epsilon+\Delta^{(n)}}\Big) \\ &\qquad + T \cdot \breve{\boldsymbol{c}}(\epsilon^{(n)}) + \bar{t} \cdot \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\big(B^{\circ} \setminus I\big), \end{aligned} \tag{4.26}$$

where $\breve{\boldsymbol{c}}(\cdot)$ is defined in (4.21). Also, let $\delta_B(\epsilon,T) \triangleq \breve{\delta}_B(\epsilon,T)/(C_b^I \cdot T)$. By (4.23) and $\breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(B \setminus I) \leq \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(I^c) < \infty$, we get

$$\lim_{T\to\infty} \delta_B(\epsilon,T) \leq \frac{1}{C_b^I} \cdot \bigg[\breve{\boldsymbol{c}}(\epsilon^{(n)}) + \frac{1}{2^n} \vee \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^{\epsilon+\Delta^{(n)}}\Big)\bigg],$$

where n is the unique positive integer satisfying $\epsilon \in (\epsilon^{(n+1)}, \epsilon^{(n)}]$. Moreover, as $\epsilon \downarrow 0$ we get $n_\epsilon \to \infty$. Since ∂I is closed, we get $\cap_{r>0}(\partial I)^r = \partial I$, which then implies $\lim_{r\downarrow 0} \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^r\Big) = \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\partial I) = 0$ due to continuity of measures. Also, by definition of $\breve{\boldsymbol{c}}$ in (4.21), we have $\lim_{\epsilon\downarrow 0} \breve{\boldsymbol{c}}(\epsilon) = 0$. In summary, we have verified that $\lim_{\epsilon\downarrow 0} \lim_{T\to\infty} \delta_B(\epsilon,T) = 0$.

Next, in case that $C_b^I = 0$, we set

$$C(\cdot) \equiv 0, \qquad \gamma(\eta) \triangleq \eta\big(\lambda(\eta)\big)^{\mathcal{J}_b^I}, \qquad \delta_B(\epsilon,T) \triangleq \breve{\delta}_B(\epsilon,T)/T.$$

The calculations above again verify that $\lim_{\epsilon\downarrow 0}\lim_{T\to\infty}\delta_B(\epsilon,T)=0$.

Now, we are ready to verify conditions (2.36) and (2.37). Specifically, we introduce stopping time

$$\tau^{\eta|b}_{\epsilon}(\boldsymbol{x}) \triangleq \min\big\{j\geq 0:\ \boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})\notin I_\epsilon\big\}. \tag{4.27}$$

Lemma 4.3 (Verifying conditions (2.36) and (2.37))**.** *Let $\bar t$ be characterized as in Lemma 4.2. Given any measurable $B\subseteq\mathbb{R}$, any $\epsilon\in(0,\bar\epsilon]$ small enough, and any $T>\bar t$,*

$$\begin{aligned} C(B^\circ)-\delta_B(\epsilon,T) &\leq \liminf_{\eta\downarrow 0}\inf_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\frac{\mathbf{P}\Big(\tau^{\eta|b}_\epsilon(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}_\epsilon(\boldsymbol{x})}(\boldsymbol{x})\in B\Big)}{\gamma(\eta)T/\eta} \\ &\leq \limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\frac{\mathbf{P}\Big(\tau^{\eta|b}_\epsilon(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}_\epsilon(\boldsymbol{x})}(\boldsymbol{x})\in B\Big)}{\gamma(\eta)T/\eta}\leq C(B^-)+\delta_B(\epsilon,T). \end{aligned}$$

Proof. Recall that

(*i*) in case that $C^I_b\in(0,\infty)$, we have $\gamma(\eta)T/\eta = C^I_bT\cdot\big(\lambda(\eta)\big)^{\mathcal{J}^I_b}$, $C(\cdot)=\check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(\ \cdot\ \setminus I)/C^I_b$, and $\delta_B(\epsilon,T)=\check\delta_B(\epsilon,T)/(C^I_b\cdot T)$;

(*ii*) in case that $C^I_b=0$, we have $\gamma(\eta)T/\eta = T\cdot\big(\lambda(\eta)\big)^{\mathcal{J}^I_b}$, $C(\cdot)\equiv 0$, and $\delta_B(\epsilon,T)=\check\delta_B(\epsilon,T)/T$.

In both cases, by rearranging the terms, it suffices to show that

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\frac{\mathbf{P}\Big(\tau^{\eta|b}_\epsilon(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}_\epsilon(\boldsymbol{x})}(\boldsymbol{x})\in B\Big)}{\big(\lambda(\eta)\big)^{\mathcal{J}^I_b}}\leq T\cdot\check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(B^-\setminus I)+\check\delta_B(\epsilon,T), \tag{4.28}$$

$$\liminf_{\eta\downarrow 0}\inf_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\frac{\mathbf{P}\Big(\tau^{\eta|b}_\epsilon(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}_\epsilon(\boldsymbol{x})}(\boldsymbol{x})\in B\Big)}{\big(\lambda(\eta)\big)^{\mathcal{J}^I_b}}\geq T\cdot\check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(B^\circ\setminus I)-\check\delta_B(\epsilon,T). \tag{4.29}$$

Recall the definition of set $\check E(\epsilon,\cdot,T)$ in (4.20). Let $\widetilde B(\epsilon)\triangleq B\setminus I_\epsilon$. Note that

$$\Big\{\tau^{\eta|b}_\epsilon(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}_\epsilon(\boldsymbol{x})}(\boldsymbol{x})\in B\Big\}=\Big\{\tau^{\eta|b}_\epsilon(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}_\epsilon(\boldsymbol{x})}(\boldsymbol{x})\in \widetilde B(\epsilon)\Big\}=\Big\{\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x})\in\check E\big(\epsilon,\widetilde B(\epsilon),T\big)\Big\}.$$

For any $\epsilon\in(0,\bar\epsilon)$ and $\xi\in\check E(\epsilon,\widetilde B(\epsilon),T)$, there exists $t\in[0,T]$ such that $\xi_t\notin I(\epsilon)$. On the other hand, recall that we use $\bar B_\epsilon(\mathbf{0})$ to denote the closed ball with radius ϵ centered at the origin. By part (*a*) of Lemma D.2, given $\epsilon\in(0,\bar\epsilon]$, it holds for all $\xi\in\mathbb{D}^{(\mathcal{J}^I_b-1)|b}_{\bar B_\epsilon(\mathbf{0})}[0,T](\epsilon)$ that $\xi_t\in I^-_{2\bar\epsilon}\ \forall t\in[0,T]$. Therefore, the claim

$$\boldsymbol{d}^{[0,T]}_{J_1}\bigg(\check E\big(\epsilon,\widetilde B(\epsilon),T\big),\ \mathbb{D}^{(\mathcal{J}^I_b-1)|b}_{\bar B_\epsilon(\mathbf{0})}[0,T](\epsilon)\bigg)\geq\bar\epsilon$$

for all $\epsilon\in(0,\bar\epsilon]$. Next, recall the strictly decreasing positive real number sequence $(\epsilon^{(n)})_{n\geq 1}$ specified in (4.24)–(4.25). For all $\epsilon>0$ small enough we have $\epsilon\in(0,\epsilon^{(1)}]$, so for such ϵ we can set $n=n_\epsilon$ as the unique positive integer such that $\epsilon\in(\epsilon^{(n+1)},\epsilon^{(n)}]$. It then follows from Theorem 2.4 that

$$\begin{aligned}\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\frac{\mathbf{P}\Big(\tau^{\eta|b}_\epsilon(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}_\epsilon(\boldsymbol{x})}(\boldsymbol{x})\in B\Big)}{\big(\lambda(\eta)\big)^{\mathcal{J}^I_b}} &\leq \sup_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\mathbf{C}^{(\mathcal{J}^I_b)|b}_{[0,T]}\bigg(\Big(\check E\big(\epsilon,\widetilde B(\epsilon),T\big)\Big)^-;\boldsymbol{x}\bigg) \\ &\leq T\cdot\bigg(\check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}\Big((B\setminus I_\epsilon)^{\Delta^{(n)}}\Big)+\check{\boldsymbol{c}}(\epsilon^{(n)})\bigg),\end{aligned} \tag{4.30}$$

where we applied property (4.24) in the last inequality. Furthermore,

$$\begin{aligned}
\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B\setminus I_\epsilon)^{\Delta^{(n)}}\Big) &\leq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B^{\Delta^{(n)}}\cup (I_\epsilon^c)^{\Delta^{(n)}}\Big) \qquad \text{due to } (E\cup F)^\Delta\subseteq E^\Delta\cup F^\Delta \\
&= \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B^{\Delta^{(n)}}\cup (I_\epsilon^c)^{\Delta^{(n)}}\cap I^c\Big) + \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B^{\Delta^{(n)}}\cup (I_\epsilon^c)^{\Delta^{(n)}}\cap I\Big) \\
&\leq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B^{\Delta^{(n)}}\setminus I\Big) + \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((I_\epsilon^c)^{\Delta^{(n)}}\cap I\Big) \\
&\leq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B^{\Delta^{(n)}}\setminus I\Big) + \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^{\epsilon+\Delta^{(n)}}\Big) \\
&\leq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B^{-}\setminus I\Big) + \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^{\Delta^{(n)}}\cap I^c)\setminus(B^-\cap I^c)\Big) + \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^{\epsilon+\Delta^{(n)}}\Big)
\end{aligned}$$

By definition of $\check{\delta}_B$ in (4.26) and the choice of $C(\cdot)$ in (4.22), we can plug this bound back into (4.30) and yield the upper bound (4.28). Similarly, by Theorem 2.4 and the property (4.25), we obtain (for all ϵ small enough)

$$\begin{aligned}
\liminf_{\eta\downarrow 0}\inf_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\frac{\mathbf{P}\Big(\tau_\epsilon^{\eta|b}(\boldsymbol{x})\leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau_\epsilon^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x})\in B\Big)}{\big(\lambda(\eta)\big)^{\mathcal{J}_b^I}} &\geq \inf_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon}\mathbf{C}^{(\mathcal{J}_b^I)|b}_{[0,T]}\bigg(\Big(\check{E}\big(\epsilon,\widetilde{B}(\epsilon),T\big)\Big)^\circ;\boldsymbol{x}\bigg) \\
&\geq (T-\bar{t})\cdot\bigg(\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B\setminus I_\epsilon)_{\Delta^{(n)}}\Big)-\check{\boldsymbol{c}}(\epsilon^{(n)})\bigg).
\end{aligned} \tag{4.31}$$

Furthermore, from the preliminary bound $(E\cap F)_\Delta\supseteq E_\Delta\cap F_\Delta$ we get

$$\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B\setminus I_\epsilon)_{\Delta^{(n)}}\Big)\geq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B\setminus I)_{\Delta^{(n)}}\Big)\geq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B_{\Delta^{(n)}}\cap I^c_{\Delta^{(n)}}\Big).$$

Together with the fact that $B_\Delta\setminus I = B_\Delta\cap I^c\subseteq \Big(B_\Delta\cap (I^c)_\Delta\Big)\cup\Big(I^c\setminus(I^c)_\Delta\Big)$, we yield

$$\begin{aligned}
\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B\setminus I_\epsilon)_{\Delta^{(n)}}\Big) &\geq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B_{\Delta^{(n)}}\setminus I\Big) - \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(I^c\setminus I^c_{\Delta^{(n)}}\Big) \\
&\geq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B_{\Delta^{(n)}}\setminus I\Big) - \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^{\Delta^{(n)}}\Big) \\
&\geq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big(B^\circ\setminus I\Big) - \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((B^\circ\cap I^c)\setminus(B_{\Delta^{(n)}}\cap I^c)\Big) - \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^{\Delta^{(n)}}\Big).
\end{aligned}$$

Plugging this bound back into (4.31), we establish the lower bound (4.29) and conclude the proof. □

The next two results verify conditions (2.38) and (2.39). Let

$$R_\epsilon^{\eta|b}(\boldsymbol{x})\triangleq\min\Big\{j\geq 0:\ \Big\|\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})\Big\|<\epsilon\Big\} \tag{4.32}$$

be the first time $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ returns to the ϵ-neighborhood of the origin. Under our choice of $A(\epsilon)=\{\boldsymbol{x}\in\mathbb{R}^m:\ \|\boldsymbol{x}\|<\epsilon\}$ and $I(\epsilon)=I_\epsilon$, the event $\{\tau^\eta_{(I(\epsilon)\setminus A(\epsilon))^c}(x)>T/\eta\}$ in condition (2.38) means that $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})\in I_\epsilon\backslash\{\boldsymbol{x}:\|\boldsymbol{x}\|<\epsilon\}$ for all $j\leq T/\eta$. Also, recall the definition of $\boldsymbol{t}(\cdot)$ in (4.16) and that $\gamma(\eta)T/\eta = C_b^I T\cdot\big(\lambda(\eta)\big)^{\mathcal{J}_b^I}$. Therefore, to verify condition (2.38), it suffices to prove the following result.

Lemma 4.4 (Verifying condition (2.38))**.** *Given $k\geq 1$ and $\epsilon\in(0,\bar{\epsilon})$, it holds for all $T\geq k\cdot\boldsymbol{t}(\epsilon/2)$ that*

$$\lim_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in I_\epsilon}\frac{1}{\lambda^{k-1}(\eta)}\mathbf{P}\Big(\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})\in I_\epsilon\setminus\{\boldsymbol{x}:\|\boldsymbol{x}\|<\epsilon\}\quad\forall j\leq T/\eta\Big)=0.$$

Proof. In this proof, we write $\xi(t) = \xi_t$ for any $\xi \in \mathbb{D}[0,T]$, and set $B_\epsilon(\mathbf{0}) = \{\boldsymbol{x} \in \mathbb{R}^m : \ \|\boldsymbol{x}\| < \epsilon\}$. Note that $\{\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) \in I_\epsilon \setminus B_\epsilon(\mathbf{0}) \ \ \forall j \leq T/\eta\} = \{\boldsymbol{X}^{\eta|b}_{[0,T]}(x) \in E(\epsilon)\}$ where

$$E(\epsilon) \triangleq \Big\{\xi \in \mathbb{D}[0,T] : \ \xi(t) \in I_\epsilon \setminus B_\epsilon(\mathbf{0}) \ \ \forall t \in [0,T]\Big\}.$$

Recall the definition of $\mathbb{D}^{(k)|b}_A[0,T](\epsilon)$ in (2.25). We claim that $E(\epsilon)$ is bounded away from $\mathbb{D}^{(k-1)|b}_{I^-_\epsilon}[0,T](\epsilon)$. This allows us to apply Theorem 2.4 and conclude that

$$\sup_{x \in I_\epsilon} \mathbf{P}\Big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in E(\epsilon)\Big) = \boldsymbol{O}\big(\lambda^k(\eta)\big) = \boldsymbol{o}\big(\lambda^{k-1}(\eta)\big) \quad \text{as } \eta \downarrow 0.$$

Now, it only remains to verify that $E(\epsilon)$ is bounded away from $\mathbb{D}^{(k-1)|b}_{I^-_\epsilon}[0,T](\epsilon)$, which can be established if we show the existence of some $\delta > 0$ such that

$$\boldsymbol{d}^{[0,T]}_{J_1}(\xi, \xi') \geq \delta > 0 \qquad \forall \xi \in \mathbb{D}^{(k-1)|b}_{I^-_\epsilon}[0,T](\epsilon), \ \xi' \in E(\epsilon). \tag{4.33}$$

First, by definition of $E(\epsilon)$, we have $\xi'_t \in I_\epsilon \ \forall t \in [0,T]$ for any $\xi' \in E(\epsilon)$. Note that $\inf\{\|\boldsymbol{x} - \boldsymbol{y}\| : \ \boldsymbol{x} \in I_\epsilon, \boldsymbol{y} \notin I_{\epsilon/2}\} \geq \epsilon/2$. Therefore, if $\xi_t \notin I_{\epsilon/2}$ for some $t \leq T$, we must have $\boldsymbol{d}^{[0,T]}_{J_1}(\xi,\xi') \geq \epsilon/2 > 0$. Now suppose that $\xi_t \in I_{\epsilon/2}$ for all $t \leq T$. Due to $\xi \in \mathbb{D}^{(k-1)|b}_{I^-_\epsilon}[0,T](\epsilon)$, there is some $\boldsymbol{x} \in I^-_\epsilon$, $\mathbf{W} \in \mathbb{R}^{d \times (k-1)}$, $\mathbf{V} \in \big(\bar{B}_\epsilon(\mathbf{0})\big)^{k-1}$, and $(t_1, \cdots, t_{k-1}) \in (0,T]^{k-1\uparrow}$ such that $\xi = \bar{h}^{(k-1)|b}_{[0,T]}\big(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, (t_1, \cdots, t_{k-1})\big)$. With the convention that $t_0 = 0$ and $t_k = T$, we have

$$\xi(t) = \boldsymbol{y}_{t - t_{j-1}}\big(\xi(t_{j-1})\big) \qquad \forall t \in [t_{j-1}, t_j). \tag{4.34}$$

for each $j \in [k]$. Here, $\boldsymbol{y}_\cdot(x)$ is the ODE defined in (2.28). Due to the assumption $T \geq k \cdot \boldsymbol{t}(\epsilon/2)$, there must be some $j \in [k]$ such that $t_j - t_{j-1} \geq \boldsymbol{t}(\epsilon/2)$. However, due to the running assumption that $\xi(t) \in I_{\epsilon/2} \ \forall t \in [0,T]$, we have $\xi(t_{j-1}) \in I_{\epsilon/2}$. Combining (4.34) along with property (4.17), we get $\lim_{t \uparrow t_j} \xi(t) \in \bar{B}_{\epsilon/2}(\mathbf{0}) \subset B_\epsilon(\mathbf{0})$. On the other hand, by definition of $E(\epsilon)$, we have $\xi'(t) \notin B_\epsilon(\mathbf{0})$ for all $t \in [0,T]$, which implies $\boldsymbol{d}^{[0,T]}_{J_1}(\xi,\xi') \geq \frac{\epsilon}{2}$. This concludes the proof. □

Lastly, we establish condition (2.39). Note that the first visit time $\tau^\eta_{A(\epsilon)}(x)$ therein coincides with $R^{\eta|b}_\epsilon(x)$ defined in (4.32) under our choice of $A(\epsilon) = \{\boldsymbol{x} \in \mathbb{R}^m : \ \|\boldsymbol{x}\| < \epsilon\}$.

Lemma 4.5 (Verifying condition (2.39))**.** *Let $\boldsymbol{t}(\cdot)$ be defined as in (4.16) and*

$$E(\eta, \epsilon, \boldsymbol{x}) \triangleq \left\{R^{\eta|b}_\epsilon(\boldsymbol{x}) \leq \frac{\boldsymbol{t}(\epsilon/2)}{\eta}; \ \boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) \in I \ \forall j \leq R^{\eta|b}_\epsilon(\boldsymbol{x})\right\}.$$

It holds for all $\epsilon \in (0, \bar{\epsilon})$ that $\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I^-_\epsilon} \mathbf{P}\Big(\big(E(\eta,\epsilon,\boldsymbol{x})\big)^c\Big) = 0$.

Proof. In this proof, we write $B_\epsilon(\mathbf{0}) = \{\boldsymbol{x} \in \mathbb{R}^m : \ \|\boldsymbol{x}\| < \epsilon\}$ and $I(\epsilon) = I_\epsilon$. Note that $\big(E(\eta,\epsilon,\boldsymbol{x})\big)^c \subseteq \{\boldsymbol{X}^{\eta|b}_{[0,\boldsymbol{t}(\epsilon/2)]}(\boldsymbol{x}) \in E^*_1(\epsilon) \cup E^*_2(\epsilon)\}$, where

$$E^*_1(\epsilon) \triangleq \Big\{\xi \in \mathbb{D}[0,\boldsymbol{t}(\epsilon/2)] : \ \xi(t) \notin B_\epsilon(\mathbf{0}) \ \forall t \in [0, \boldsymbol{t}(\epsilon/2)]\Big\},$$

$$E^*_2(\epsilon) \triangleq \Big\{\xi \in \mathbb{D}[0,\boldsymbol{t}(\epsilon/2)] : \ \exists 0 \leq s \leq t \leq \boldsymbol{t}(\epsilon/2) \ s.t. \ \xi(t) \in B_\epsilon(\mathbf{0}), \ \xi(s) \notin I\Big\}.$$

Recall the definition of $\mathbb{D}^{(k)|b}_A[0,T](\epsilon)$ in (2.25). We claim that both $E^*_1(\epsilon)$ and $E^*_2(\epsilon)$ are bounded away from

$$\mathbb{D}^{(0)|b}_{(I(\epsilon))^-}[0,\boldsymbol{t}(\epsilon/2)] = \Big\{\{\boldsymbol{y}_t(\boldsymbol{x}) : \ t \in [0,\boldsymbol{t}(\epsilon/2)]\} : \ \boldsymbol{x} \in \big(I(\epsilon)\big)^-\Big\}.$$

To see why, note that $\inf\{\|\boldsymbol{x}-\boldsymbol{y}\| : \boldsymbol{x} \in I(\epsilon), \boldsymbol{y} \notin I(\epsilon/2)\} \geq \epsilon/2$. Meanwhile, properties (4.14) and (4.17) imply that $\boldsymbol{y}_{\boldsymbol{t}(\epsilon/2)}(\boldsymbol{x}) \in \bar{B}_{\epsilon/2}(\mathbf{0})$ for all $\boldsymbol{x} \in \big(I(\epsilon)\big)^-$. Therefore,

$$\boldsymbol{d}_{J_1}^{[0,\boldsymbol{t}(\epsilon/2)]}\bigg(\mathbb{D}^{(0)|b}_{(I(\epsilon))^-}[0,\boldsymbol{t}(\epsilon/2)],\ E_1^*(\epsilon)\bigg) \geq \frac{\epsilon}{2} > 0, \tag{4.35}$$

Meanwhile, by property (4.18), we immediately get

$$\boldsymbol{d}_{J_1}^{[0,\boldsymbol{t}(\epsilon/2)]}\bigg(\mathbb{D}^{(0)|b}_{(I(\epsilon))^-}[0,\boldsymbol{t}(\epsilon/2)],\ E_2^*(\epsilon)\bigg) \geq \delta > 0. \tag{4.36}$$

This allows us to apply Theorem 2.4 and obtain

$$\sup_{\boldsymbol{x}\in(I(\epsilon))^-}\mathbf{P}\bigg(\big(E(\eta,\epsilon,\boldsymbol{x})\big)^c\bigg) \leq \sup_{\boldsymbol{x}\in(I(\epsilon))^-}\mathbf{P}\bigg(\boldsymbol{X}^{\eta|b}_{[0,\boldsymbol{t}(\epsilon/2)]}(\boldsymbol{x}) \in E_1^*(\epsilon)\cup E_2^*(\epsilon)\bigg) = \boldsymbol{O}\big(\lambda(\eta)\big)$$

as $\eta \downarrow 0$. To conclude the proof, one only needs to note that $\lambda(\eta) \in \mathcal{RV}_{\alpha-1}(\eta)$ (with $\alpha > 1$) and hence $\lim_{\eta\downarrow 0}\lambda(\eta) = 0$. □

We conclude this section with the proof of Theorem 2.6.

Proof of Theorem 2.6. First, it is established in Lemma D.3 that $C_b^I < \infty$. Next, since Lemmas 4.3–4.5 verify Condition 1, Theorem 2.6 follows immediately from Theorem 2.9. □

References

[1] P. Baldi and L. Caramellino. General freidlin–wentzell large deviations and positive diffusions. *Statistics & Probability Letters*, 81(8):1218–1229, 2011.

[2] M. Bazhba, J. Blanchet, C.-H. Rhee, and B. Zwart. Sample path large deviations for Lévy processes and random walks with Weibull increments. *The Annals of Applied Probability*, 30(6):2695 – 2739, 2020.

[3] M. Bazhba, C.-H. Rhee, and B. Zwart. Large deviations for stochastic fluid networks with weibullian tails. *Queueing Systems*, 102(1):25–52, 2022.

[4] A. S. Bedi, A. Parayil, J. Zhang, M. Wang, and A. Koppel. On the sample complexity and metastability of heavy-tailed policy search in continuous control. *Journal of Machine Learning Research*, 25(39):1–58, 2024.

[5] H. Bernhard, B. Das, et al. Heavy-tailed random walks, buffered queues and hidden large deviations. *Bernoulli*, 26(1):61–92, 2020.

[6] A. A. Borovkov. Large deviation principles for random walks with regularly varying distributions of jumps. *Siberian mathematical journal*, 52(3):402–410, 2011.

[7] A. A. Borovkov and K. A. Borovkov. *Asymptotic Analysis of Random Walks: Heavy-Tailed Distributions*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2008.

[8] A. A. Borovkov and A. A. Mogul'skii. On large deviation principles in metric spaces. *Siberian Mathematical Journal*, 51:989–1003, 2010.

[9] A. Bovier and F. Den Hollander. *Metastability: a potential-theoretic approach*, volume 351. Springer, 2016.

[10] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability and low lying spectra in reversible markov chains. *Communications in mathematical physics*, 228:219–255, 2002.

[11] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes. i. sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc.(JEMS)*, 6(4):399–424, 2004.

[12] A. Budhiraja and P. Dupuis. A variational representation for positive functionals of infinite dimensional brownian motion. *Probability and mathematical statistics-Wroclaw University*, 20(1):39–61, 2000.

[13] A. Budhiraja, P. Dupuis, and V. Maroulas. Large deviations for infinite dimensional stochastic dynamical systems. *The Annals of Probability*, 36(4):1390 – 1420, 2008.

[14] O. Catoni and R. Cerf. The exit path of a markov chain with rare transitions. *ESAIM: Probability and Statistics*, 1:95–144, 1997.

[15] S. Cerrai and M. Röckner. Large deviations for stochastic reaction-diffusion systems with multiplicative noise and non-Lipshitz reaction term. *The Annals of Probability*, 32(1B):1100 – 1139, 2004.

[16] A. V. Chechkin, V. Y. Gonchar, J. Klafter, and R. Metzler. Barrier crossing of a lévy flight. *Europhysics Letters*, 72(3):348, 2005.

[17] A. V. Chechkin, O. Y. Sliusarenko, R. Metzler, and J. Klafter. Barrier crossing driven by lévy noise: Universality and the role of noise intensity. *Physical Review E*, 75(4):041101, 2007.

[18] B. Chen, J. Blanchet, C.-H. Rhee, and B. Zwart. Efficient rare-event simulation for multiple jump events in regularly varying random walks and compound poisson processes. *Mathematics of Operations Research*, 44(3):919–942, 2019.

[19] B. Chen, C.-H. Rhee, and B. Zwart. Sample-path large deviations for a class of heavy-tailed markov additive processes. *arXiv preprint arXiv:2010.10751*, 2023.

[20] P. Chigansky and R. Liptser. *The Freidlin-Wentzell LDP with Rapidly Growing Coeffcients*, pages 197–219. World Scientific, 2007.

[21] A. Debussche, M. Högele, and P. Imkeller. *The dynamics of nonlinear reaction-diffusion equations with small Lévy noise*, volume 2085. Springer, 2013.

[22] D. Denisov, A. B. Dieker, and V. Shneer. Large deviations for random walks under subexponentiality: The big-jump domain. *The Annals of Probability*, 36(5):1946 – 1991, 2008.

[23] C. Donati-Martin. Large deviations for wishart processes. *Probability and Mathematical Statistics*, 28, 2008.

[24] C. Donati-Martin, A. Rouault, M. Yor, and M. Zani. Large deviations for squares of bessel and ornstein-uhlenbeck processes. *Probability Theory and Related Fields*, 129:261–289, 2004.

[25] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 2011.

[26] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013.

[27] H. Eyring. The activated complex in chemical reactions. *The Journal of Chemical Physics*, 3(2):107–115, 1935.

[28] S. Foss, D. Korshunov, S. Zachary, et al. *An introduction to heavy-tailed and subexponential distributions*, volume 6. Springer, 2011.

[29] M. Högele and I. Pavlyukevich. Metastability in a class of hyperbolic dynamical systems perturbed by heavy-tailed lévy type noise. *Stochastics and Dynamics*, 15(03):1550019, 2015.

[30] H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publications de l'Institut Mathématique*, 80(94):121–140, 2006.

[31] H. Hult, F. Lindskog, T. Mikosch, and G. Samorodnitsky. Functional large deviations for multivariate regularly varying random walks. *The Annals of Applied Probability*, 15(4):2651 – 2680, 2005.

[32] P. Imkeller and I. Pavlyukevich. First exit times of sdes driven by stable lévy processes. *Stochastic Processes and their Applications*, 116(4):611–642, 2006.

[33] P. Imkeller and I. Pavlyukevich. Metastable behaviour of small noise lévy-driven diffusions. *ESAIM: PS*, 12:412–437, 2008.

[34] P. Imkeller, I. Pavlyukevich, and M. Stauch. First exit times of non-linear dynamical systems in $\mathbb{R}^d$ perturbed by multifractal Lévy noise. *Journal of Statistical Physics*, 141(1):94–119, 2010.

[35] P. Imkeller, I. Pavlyukevich, and T. Wetzel. First exit times for Lévy-driven diffusions with exponentially light jumps. *The Annals of Probability*, 37(2):530 – 564, 2009.

[36] P. Imkeller, I. Pavlyukevich, and T. Wetzel. The hierarchy of exit times of Lévy-driven Langevin equations. *The European Physical Journal Special Topics*, 191(1):211–222, 2010.

[37] G. Kallianpur and J. Xiong. Stochastic differential equations in infinite dimensional spaces. *Lecture Notes-Monograph Series*, 26:iii–342, 1995.

[38] Y. Kifer. A Discrete-Time Version of the Wentzell-Friedlin Theory. *The Annals of Probability*, 18(4):1676 – 1692, 1990.

[39] F. Klebaner and O. Zeitouni. The exit problem for a class of period doubling systems. *Ann. Appl. Probab*, 4:1188–1205, 1994.

[40] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

[41] R. Kruse. *Strong and weak approximation of semilinear stochastic evolution equations*. Springer, 2014.

[42] F. Lindskog, S. I. Resnick, J. Roy, et al. Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Probability Surveys*, 11:270–314, 2014.

[43] A. D. W. M. I. Freidlin. *Random Perturbations of Dynamical Systems*. Springer New York, NY, 1998.

[44] M. T. Mohan. Wentzell–freidlin large deviation principle for stochastic convective brinkman–forchheimer equations. *Journal of Mathematical Fluid Mechanics*, 23(3):62, 2021.

[45] G. J. Morrow and S. Sawyer. Large deviation results for a class of markov chains arising from population genetics. *The Annals of Probability*, 17(3):1124–1146, 1989.

[46] A. Nagaev. Limit theorems for large deviations where cramér's conditions are violated. *Fiz-Mat. Nauk*, 7:17–22, 1969.

[47] A. V. Nagaev. On a property of sums of independent random variables. *Theory of Probability & Its Applications*, 22(2):326–338, 1978.

[48] T. H. Nguyen, U. Simsekli, M. Gurbuzbalaban, and G. RICHARD. First exit time analysis of stochastic gradient descent under heavy-tailed gradient noise. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[49] E. Olivieri and M. E. Vares. *Large deviations and metastability.* Number 100. Cambridge University Press, 2005.

[50] I. Pavlyukevich. First exit times of solutions of stochastic differential equations driven by multiplicative lévy noise with heavy tails. *Stochastics and Dynamics*, 11(2–3):495–519, 2011.

[51] I. PAVLYUKEVICH. First exit times of solutions of stochastic differential equations driven by multiplicative lÉvy noise with heavy tails. *Stochastics and Dynamics*, 11(02n03):495–519, 2011.

[52] C. Penland and B. D. Ewald. On modelling physical systems with stochastic models: diffusion versus lévy processes. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1875):2455–2474, 2008.

[53] P. E. Protter. *Stochastic integration and differential equations.* Springer, 2005.

[54] S. Resnick. On the foundations of multivariate heavy-tail analysis. *Journal of Applied Probability*, 41(A):191–212, 2004.

[55] S. I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling.* Springer Science & Business Media, 2007.

[56] C.-H. Rhee, J. Blanchet, B. Zwart, et al. Sample path large deviations for lévy processes and random walks with regularly varying increments. *The Annals of Probability*, 47(6):3551–3605, 2019.

[57] M. Röckner, T. Zhang, and X. Zhang. Large deviations for stochastic tamed 3d navier-stokes equations. *Applied Mathematics and Optimization*, 61:267–285, 2010.

[58] K.-i. Sato, S. Ken-Iti, and A. Katok. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge university press, 1999.

[59] E. Scalas, R. Gorenflo, and F. Mainardi. Fractional calculus and continuous-time finance. *Physica A: Statistical Mechanics and its Applications*, 284(1-4):376–384, 2000.

[60] A. Shwartz and A. Weiss. *Large deviations for performance analysis: queues, communication and computing*, volume 5. CRC Press, 1995.

[61] R. B. Sowers. Large deviations for a reaction-diffusion equation with non-gaussian perturbations. *The Annals of Probability*, 20(1):504–537, 1992.

[62] A. D. Ventsel' and M. I. Freidlin. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, 25(1):1, feb 1970.

[63] X. Wang, S. Oh, and C.-H. Rhee. Eliminating sharp minima from SGD with truncated heavy-tailed noise. In *International Conference on Learning Representations*, 2022.

[64] X. Wang and C.-H. Rhee. Global dynamics of heavy-tailed SGD in nonconvex loss landscape: characterization and control. https://chrhee.github.io/papers/WangRhee24b.pdf. *Working Paper.*

A Results under General Scaling

Below, we present results analogous to those in Section 2 under a general scaling. Specifically, throughout this section we define $(\boldsymbol{X}^{\eta}_j(\boldsymbol{x}))_{j\geq 0}$ and $(\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}))_{j\geq 0}$ with the recursions in (1.3). Here, we consider $\gamma \in (\frac{1}{2\wedge\alpha}, \infty)$ where $\alpha > 1$ is the tail index in Assumption 1. Let

$$\lambda(\eta;\gamma) = \eta^{-1}H(\eta^{-\gamma}).$$

We adopt the notations $\mathbf{C}^{(k)|b}$, $\mathbb{D}^{(k)|b}_A(\epsilon)$, $\boldsymbol{X}^{\eta|b}(\boldsymbol{x})$, etc., as described in Section 2. First, we present the sample-path large deviations.

Theorem A.1. *Let Assumptions 1 and 2 hold. Let* $\gamma \in (\frac{1}{2\wedge\alpha}, \infty)$.

(a) For any $k \in \mathbb{N}$*, any* $b, T, \epsilon > 0$*, and any compact* $A \subseteq \mathbb{R}^m$*,*

$$\lambda^{-k}(\eta;\gamma)\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in \ \cdot\ \big) \to \mathbf{C}^{(k)|b}_{[0,T]}(\ \cdot\ ;\boldsymbol{x}) \quad \text{in } \mathbb{M}\Big(\mathbb{D}[0,T] \setminus \mathbb{D}^{(k-1)|b}_A[0,T](\epsilon)\Big) \text{ uniformly in } \boldsymbol{x} \text{ on } A$$

as $\eta \downarrow 0$*. Furthermore, for any* $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ *that is bounded away from* $\mathbb{D}^{(k-1)|b}_A[0,T](\epsilon)$*,*

$$\begin{aligned}\inf_{\boldsymbol{x}\in A} \mathbf{C}^{(k)|b}_{[0,T]}\big(B^\circ;\boldsymbol{x}\big) &\leq \liminf_{\eta\downarrow 0} \frac{\inf_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta;\gamma)} \\ &\leq \limsup_{\eta\downarrow 0} \frac{\sup_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta;\gamma)} \leq \sup_{\boldsymbol{x}\in A} \mathbf{C}^{(k)|b}_{[0,T]}\big(B^-;\boldsymbol{x}\big) < \infty.\end{aligned}$$

(b) Furthermore, suppose that Assumption 3 holds. For any $k \in \mathbb{N}$*,* $T, \epsilon > 0$*, and any compact* $A \subseteq \mathbb{R}^m$ *that*

$$\lambda^{-k}(\eta;\gamma)\mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in \ \cdot\ \big) \to \mathbf{C}^{(k)}_{[0,T]}(\ \cdot\ ;\boldsymbol{x}) \quad \text{in } \mathbb{M}\Big(\mathbb{D}[0,T] \setminus \mathbb{D}^{(k-1)}_A[0,T](\epsilon)\Big) \text{ uniformly in } \boldsymbol{x} \text{ on } A$$

as $\eta \downarrow 0$*. Furthermore, for any* $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ *that is bounded away from* $\mathbb{D}^{(k-1)}_A[0,T](\epsilon)$*,*

$$\begin{aligned}\inf_{\boldsymbol{x}\in A} \mathbf{C}^{(k)}_{[0,T]}\big(B^\circ;\boldsymbol{x}\big) &\leq \liminf_{\eta\downarrow 0} \frac{\inf_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta;\gamma)} \\ &\leq \limsup_{\eta\downarrow 0} \frac{\sup_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta;\gamma)} \leq \sup_{\boldsymbol{x}\in A} \mathbf{C}^{(k)}_{[0,T]}\big(B^-;\boldsymbol{x}\big) < \infty.\end{aligned}$$

The corresponding conditional limit theorem is identical to Corollary 2.5, under the condition that $\gamma \in (\frac{1}{2\wedge\alpha}, \infty)$, so we skip the details. Lastly, we present the metastability analysis. Let $I \subseteq \mathbb{R}^m$ be an open set such that $\boldsymbol{0} \in \boldsymbol{I}$ and Assumption 4 holds. Let the first exit times $\tau^\eta(\boldsymbol{x})$ and $\tau^{\eta|b}(\boldsymbol{x})$ be defined as in (2.29). We adopt the notations $\mathcal{J}^I_b$, $\mathcal{G}^{(k)|b}(\epsilon)$, $\check{\mathbf{C}}^{k|b}$, etc., as described in Section 2.3.

Theorem A.2. *Let Assumptions 1, 2, and 4 hold. Let* $\gamma \in (\frac{1}{2\wedge\alpha}, \infty)$.

(a) Let $b > 0$ *such that* $\mathcal{J}^I_b < \infty$*. Suppose that* I^c *is bounded away from* $\mathcal{G}^{(\mathcal{J}^I_b - 1)|b}(\epsilon)$ *for some (and hence all)* $\epsilon > 0$ *small enough, and* $\check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(\partial I) = 0$*. Then* $C^I_b \triangleq \check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(I^c) < \infty$*. Furthermore, if* $C^I_b \in (0,\infty)$*, then for any* $\epsilon > 0$*,* $t \geq 0$*, and measurable set* $B \subseteq I^c$*,*

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in I_\epsilon} \mathbf{P}\bigg(C^I_b\eta\cdot\lambda^{\mathcal{J}^I_b}(\eta;\gamma)\tau^{\eta|b}(\boldsymbol{x}) > t;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg) \leq \frac{\check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(B^-)}{C^I_b}\cdot\exp(-t),$$

$$\liminf_{\eta\downarrow 0}\inf_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(C^I_b\eta\cdot\lambda^{\mathcal{J}^I_b}(\eta;\gamma)\tau^{\eta|b}(\boldsymbol{x})>t;\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x})\in B\bigg)\geq\frac{\check{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(B^\circ)}{C^I_b}\cdot\exp(-t).$$

Otherwise, we must have $C^I_b=0$, *and*

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(\eta\cdot\lambda^{\mathcal{J}^I_b}(\eta;\gamma)\tau^{\eta|b}(\boldsymbol{x})\leq t\bigg)=0\qquad\forall\epsilon>0,\ t\geq 0.$$

(b) *Suppose that* $\check{\mathbf{C}}(\partial I)=0$. *Then* $C^I_\infty\triangleq\check{\mathbf{C}}(I^c)<\infty$. *Furthermore, if* $C^I_\infty>0$, *then for any* $t\geq 0$ *and measurable set* $B\subseteq I^c$,

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(C^I_\infty\eta\cdot\lambda(\eta;\gamma)\tau^{\eta}(\boldsymbol{x})>t;\ \boldsymbol{X}^{\eta}_{\tau^{\eta}(\boldsymbol{x})}(\boldsymbol{x})\in B\bigg)\leq\frac{\check{\mathbf{C}}(B^-)}{C^I_\infty}\cdot\exp(-t),$$
$$\liminf_{\eta\downarrow 0}\inf_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(C^I_\infty\eta\cdot\lambda(\eta;\gamma)\tau^{\eta}(\boldsymbol{x})>t;\ \boldsymbol{X}^{\eta}_{\tau^{\eta}(\boldsymbol{x})}(\boldsymbol{x})\in B\bigg)\geq\frac{\check{\mathbf{C}}(B^\circ)}{C^I_\infty}\cdot\exp(-t).$$

Otherwise, we must have $C^I_\infty=0$, *and*

$$\limsup_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in I_\epsilon}\mathbf{P}\bigg(\eta\cdot\lambda(\eta;\gamma)\tau^{\eta}(\boldsymbol{x})\leq t\bigg)=0\qquad\forall\epsilon>0,\ t\geq 0.$$

The proofs for results in this section will be almost identical to those presented in the main paper. We omit the details to avoid repetition.

B Results for Lévy-Driven Stochastic Differential Equations

In this section, we collect the results for stochastic differential equations driven by Lévy processes with regularly varying increments. Specifically, any one-dimensional Lévy process $\boldsymbol{L}=\{\boldsymbol{L}_t:\ t\geq 0\}$ can be characterized by its generating triplet (c_L,σ_L,ν) where $c_{\boldsymbol{L}}\in\mathbb{R}^m$ is the drift parameter, the positive semi-definite matrix $\boldsymbol{\Sigma}_{\boldsymbol{L}}\in\mathbb{R}^{m\times m}$ is the magnitude of the Brownian motion term in $\boldsymbol{L}_t$, and ν is the Lévy measure of the Lévy process $\boldsymbol{L}_t$ characterizing the intensity of jumps in $\boldsymbol{L}_t$. More precisely, we have the following Lévy–Itô decomposition

$$\boldsymbol{L}_t\overset{d}{=}\boldsymbol{c_L}t+\boldsymbol{\Sigma}^{1/2}_{\boldsymbol{L}}\boldsymbol{B}_t+\int_{\|\boldsymbol{x}\|\leq 1}\boldsymbol{x}\big[N([0,t]\times d\boldsymbol{x})-t\nu(dx)\big]+\int_{\|\boldsymbol{x}\|>1}\boldsymbol{x}N([0,t]\times d\boldsymbol{x})\qquad\text{(B.1)}$$

where $\boldsymbol{B}$ is a standard Brownian motion in $\mathbb{R}^m$, the measure ν satisfies $\int(\|\boldsymbol{x}\|^2\wedge 1)\nu(d\boldsymbol{x})<\infty$, and N is a Poisson random measure independent of $\boldsymbol{B}$ with intensity measure $\mathcal{L}_\infty\times\nu$. See Chapter 4 of [58] for details. We impose the following assumption that characterizes the heavy-tailedness in the increments of $\boldsymbol{L}_t$.

Assumption 5. $\mathbf{E}\boldsymbol{L}_1=\mathbf{0}$. *Besides, there exist* $\alpha>1$ *and a probability measure* $\mathbf{S}(\cdot)$ *on the unit sphere of* $\mathbb{R}^d$ *such that*

- $H_L(x)\triangleq\nu\Big(\big\{\boldsymbol{y}\in\mathbb{R}^d:\ \|\boldsymbol{y}\|>x\big\}\Big)\in\mathcal{RV}_{-\alpha}(x)$ *as* $x\to\infty$,
- *As* $r\to\infty$,

$$\frac{\big(\nu\circ\Phi_r^{-1}\big)(\cdot)}{H_L(r)}\to\nu_\alpha\times\mathbf{S}\qquad\text{in }\mathbb{M}\Big(\big([0,\infty)\times\mathfrak{N}_d\big)\setminus\big(\{0\}\times\mathfrak{N}_d\big)\Big),$$

where $\mathfrak{N}_d$ *is the unit sphere of* $\mathbb{R}^d$, *the measure* $\big(\nu\circ\Phi_r^{-1}\big)$ *is defined by*

$$\big(\nu\circ\Phi_r^{-1}\big)(\cdot)\triangleq\nu\Big(\Phi^{-1}(r^{-1}\cdot,\cdot)\Big),$$

i.e. $\big(\nu\circ\Phi_r^{-1}\big)(A\times B)=\nu\big(\Phi^{-1}(r^{-1}A,B)\big)$ *for all Borel sets* $A\subseteq(0,\infty)$ *and* $B\subseteq\mathfrak{N}_d$.

Consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$ satisfying the usual hypotheses stated in Chapter I, [53] and supporting the Lévy process $\boldsymbol{L}$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_t$ is the σ-algebra generated by $\{\boldsymbol{L}_s : s \in [0,t]\}$. For $\eta \in (0,1]$ and $\beta \geq 0$, define the scaled process

$$\bar{\boldsymbol{L}}^\eta \triangleq \big\{ \bar{\boldsymbol{L}}^\eta_t = \eta \boldsymbol{L}_{t/\eta^\beta} : \ t \geq 0 \big\}, \tag{B.2}$$

and let $\boldsymbol{Y}^\eta_t(\boldsymbol{x})$ be the solution to SDE

$$\boldsymbol{Y}^\eta_0(\boldsymbol{x}) = \boldsymbol{x}, \qquad d\boldsymbol{Y}^\eta_t(\boldsymbol{x}) = \boldsymbol{a}\big(\boldsymbol{Y}^\eta_{t-}(\boldsymbol{x})\big)dt + \boldsymbol{\sigma}\big(\boldsymbol{Y}^\eta_{t-}(\boldsymbol{x})\big)d\bar{\boldsymbol{L}}^\eta_t. \tag{B.3}$$

Henceforth in Section B, we consider $\beta \in [0, 2 \wedge \alpha)$ where $\alpha > 1$ is the tail index in Assumption 5. Below, we state the results regarding the sample-path large deviations and metastability of $\boldsymbol{Y}^\eta_t(\boldsymbol{x})$.

B.1 Sample Path Large Deviations

Recall the definitions of the mapping $h^{(k)}_{[0,T]}$ in (2.10)–(2.12) as well as the measure $\mathbf{C}^{(k)}_{[0,T]}(\ \cdot\ ; x)$ in (2.14). Also, recall the notion of uniform $\mathbb{M}$-convergence introduced in Definition 2.1. Define $\boldsymbol{Y}^\eta_{[0,T]}(x) = \{Y^\eta_t(x) : \ t \in [0,T]\}$ as a random element in $\mathbb{D}[0,T]$. In case that $T = 1$, we suppress $[0,1]$ and write $\boldsymbol{Y}^\eta(x)$. The next result characterizes the sample-path large deviations for $\boldsymbol{Y}^\eta_{[0,T]}(x)$ by establishing $\mathbb{M}$-convergence that is uniform in the initial condition x. The proofs are almost identical to those of $X^\eta_j(x)$ and hence omitted to avoid repetition. Recall that $H_L(x) = \nu((\infty, -x) \cup (x, \infty))$. Let

$$\lambda_L(\eta; \beta) \triangleq \eta^{-\beta} H_L(\eta^{-1})$$

where $\beta \in [0, 2 \wedge \alpha)$ determines the time scaling in (B.2).

Theorem B.1. *Under Assumptions 2, 3, and 5, it holds for any $\beta \in [0, 2 \wedge \alpha)$, $T, \epsilon > 0$, $k \in \mathbb{N}$, and any compact set $A \subseteq \mathbb{R}^m$ that*

$$\lambda_L^{-k}(\eta; \beta)\mathbf{P}\big(\boldsymbol{Y}^\eta_{[0,T]}(\boldsymbol{x}) \in \ \cdot\ \big) \to \mathbf{C}^{(k)}_{[0,T]}(\ \cdot\ ; \boldsymbol{x}) \quad \text{in } \mathbb{M}\Big(\mathbb{D}[0,T] \setminus \mathbb{D}^{(k-1)}_A[0,T](\epsilon)\Big) \text{ uniformly in } \boldsymbol{x} \text{ on } A$$

as $\eta \to 0$. Furthermore, for any $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ that is bounded away from $\mathbb{D}^{(k-1)}_A[0,T](\epsilon)$,

$$\begin{aligned}\inf_{\boldsymbol{x} \in A} \mathbf{C}^{(k)}_{[0,T]}\big(B^\circ; \boldsymbol{x}\big) &\leq \liminf_{\eta \downarrow 0} \frac{\inf_{\boldsymbol{x} \in A} \mathbf{P}\big(\boldsymbol{Y}^\eta_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k_L(\eta; \beta)} \\ &\leq \limsup_{\eta \downarrow 0} \frac{\sup_{\boldsymbol{x} \in A} \mathbf{P}\big(\boldsymbol{Y}^\eta_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k_L(\eta; \beta)} \leq \sup_{\boldsymbol{x} \in A} \mathbf{C}^{(k)}_{[0,T]}\big(B^-; \boldsymbol{x}\big) < \infty.\end{aligned}$$

Analogous to the truncated dynamics $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$, we introduce processes $\boldsymbol{Y}^{\eta|b}_t(\boldsymbol{x})$ that can be seen as a modulated version of $\boldsymbol{Y}^\eta_t(\boldsymbol{x})$ where all jumps are truncated under the threshold value b. More generally, we consider a sequence of stochastic processes $\big(\boldsymbol{Y}^{\eta|b;(k)}_t(\boldsymbol{x})\big)_{k \geq 0}$. First, for any $\boldsymbol{x} \in \mathbb{R}^m$ and $t \geq 0$, let

$$d\boldsymbol{Y}^{\eta|b;(0)}_t(\boldsymbol{x}) \triangleq \boldsymbol{a}\big(\boldsymbol{Y}^{\eta|b;(0)}_{t-}(\boldsymbol{x})\big)dt + \boldsymbol{\sigma}\big(\boldsymbol{Y}^{\eta|b;(0)}_{t-}(\boldsymbol{x})\big)d\bar{\boldsymbol{L}}_t. \tag{B.4}$$

Next, building upon the process $Y^{\eta|b;(0)}_t(x)$, we define

$$\tau^{\eta|b;(1)}_Y(\boldsymbol{x}) \triangleq \min\Big\{ t > 0 : \Big\|\boldsymbol{\sigma}\big(\boldsymbol{Y}^{\eta|b;(0)}_{t-}(\boldsymbol{x})\big)\Delta\bar{\boldsymbol{L}}^\eta_t\Big\| = \Big\|\Delta\boldsymbol{Y}^{\eta|b;(0)}_t(\boldsymbol{x})\Big\| > b \Big\}, \tag{B.5}$$

$$\boldsymbol{W}^{\eta|b;(1)}_Y(\boldsymbol{x}) \triangleq \Delta\boldsymbol{Y}^{\eta|b;(0)}_{\tau^{\eta|b;(1)}_Y(\boldsymbol{x})}(\boldsymbol{x}) \tag{B.6}$$

as the arrival time and size of the first jump in $\boldsymbol{Y}_t^{\eta|b;(0)}(\boldsymbol{x})$ that is larger than b. Furthermore, we define (for any $k \geq 1$)

$$\boldsymbol{Y}^{\eta|b;(k)}_{\tau_Y^{\eta|b;(k)}(\boldsymbol{x})}(\boldsymbol{x}) \triangleq \boldsymbol{Y}^{\eta|b;(k)}_{\tau_Y^{\eta|b;(k)}(\boldsymbol{x})-}(\boldsymbol{x}) + \varphi_b\Big(\boldsymbol{W}_Y^{\eta|b;(k)}(\boldsymbol{x})\Big), \tag{B.7}$$

$$d\boldsymbol{Y}_t^{\eta|b;(k)}(\boldsymbol{x}) \triangleq \boldsymbol{a}\big(\boldsymbol{Y}_{t-}^{\eta|b;(k)}(\boldsymbol{x})\big)dt + \boldsymbol{\sigma}\big(\boldsymbol{Y}_{t-}^{\eta|b;(k)}(\boldsymbol{x})\big)d\bar{\boldsymbol{L}}_t^{\eta} \qquad \forall t > \tau_Y^{\eta|b;(k)}(\boldsymbol{x}), \tag{B.8}$$

$$\tau_Y^{\eta|b;(k+1)}(\boldsymbol{x}) \triangleq \min\Big\{t > \tau_Y^{\eta|b;(k)}(\boldsymbol{x}) : \Big\|\sigma\big(\boldsymbol{Y}_{t-}^{\eta|b;(k)}(\boldsymbol{x})\big)\Delta\bar{\boldsymbol{L}}_t^{\eta}\Big\| > b\Big\}, \tag{B.9}$$

$$\boldsymbol{W}_Y^{\eta|b;(k+1)}(\boldsymbol{x}) \triangleq \Delta\boldsymbol{Y}^{\eta|b;(k)}_{\tau_Y^{\eta|b;(k+1)}(\boldsymbol{x})}(\boldsymbol{x}) \tag{B.10}$$

Lastly, for any $t \geq 0, b > 0$ and $x \in \mathbb{R}$, we define (under convention $\tau_Y^{\eta|b;(0)}(\boldsymbol{x}) = 0$)

$$\boldsymbol{Y}_t^{\eta|b}(\boldsymbol{x}) \triangleq \sum_{k\geq 0} \boldsymbol{Y}_t^{\eta|b;(k)}(\boldsymbol{x}) \cdot \mathbb{I}\Big\{t \in \Big[\tau_Y^{\eta|b;(k)}(\boldsymbol{x}), \tau_Y^{\eta|b;(k+1)}(\boldsymbol{x})\Big)\Big\} \tag{B.11}$$

and let $\boldsymbol{Y}_{[0,T]}^{\eta|b}(x) \triangleq \big\{\boldsymbol{Y}_t^{\eta|b}(\boldsymbol{x}) :\ t \in [0,T]\big\}$. By definition, for any $t \geq 0, b > 0, k \geq 0$ and $x \in \mathbb{R}$,

$$\boldsymbol{Y}_t^{\eta|b}(\boldsymbol{x}) = \boldsymbol{Y}_t^{\eta|b;(k)}(\boldsymbol{x}) \qquad \Longleftrightarrow \qquad t \in \Big[\tau_Y^{\eta|b;(k)}(\boldsymbol{x}), \tau_Y^{\eta|b;(k+1)}(\boldsymbol{x})\Big). \tag{B.12}$$

In case that $T = 1$, we suppress $[0,1]$ and write $\boldsymbol{Y}^{\eta|b}(\boldsymbol{x})$. The next theorem presents the sample path large deviations for $\boldsymbol{Y}_t^{\eta|b}(\boldsymbol{x})$. Once again, the proof is omitted as it closely resembles that of $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$.

Theorem B.2. *Under Assumptions 2 and 5, it holds for any $\beta \in [0, 2\wedge\alpha)$, any $b, T, \epsilon > 0,\ k \in \mathbb{N}$, and any compact set $A \subseteq \mathbb{R}^m$ that*

$$\lambda_L^{-k}(\eta;\beta)\mathbf{P}\big(\boldsymbol{Y}_{[0,T]}^{\eta|b}(\boldsymbol{x}) \in\ \cdot\ \big) \to \mathbf{C}_{[0,T]}^{(k)|b}(\ \cdot\ ;\boldsymbol{x}) \text{ in } \mathbb{M}\Big(\mathbb{D}[0,T]\setminus\mathbb{D}_A^{(k-1)|b}[0,T](\epsilon)\Big) \text{ uniformly in } \boldsymbol{x} \text{ on } A$$

as $\eta \to 0$. Furthermore, for any $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ that is bounded away from $\mathbb{D}_A^{(k-1)|b}[0,T](\epsilon)$,

$$\inf_{\boldsymbol{x}\in A} \mathbf{C}_{[0,T]}^{(k)|b}\big(B^\circ;\boldsymbol{x}\big) \leq \liminf_{\eta\downarrow 0} \frac{\inf_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{Y}_{[0,T]}^{\eta|b}(\boldsymbol{x}) \in B\big)}{\lambda_L^k(\eta;\beta)}$$
$$\leq \limsup_{\eta\downarrow 0} \frac{\sup_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{Y}_{[0,T]}^{\eta|b}(\boldsymbol{x}) \in B\big)}{\lambda_L^k(\eta;\beta)} \leq \sup_{\boldsymbol{x}\in A} \mathbf{C}_{[0,T]}^{(k)|b}\big(B^-;\boldsymbol{x}\big) < \infty.$$

To conclude this subsection, we present the conditional limit results for $\boldsymbol{Y}^{\eta}$ and $\boldsymbol{Y}^{\eta|b}$.

Corollary B.3. *Let Assumptions 2 and 5 hold. Let $\beta \in [0, 2\wedge\alpha)$.*

(i) For some $b, \epsilon > 0,\ k \in \mathbb{N},\ \boldsymbol{x} \in \mathbb{R}^m$, and measurable $B \subseteq \mathbb{D}$, suppose that B is bounded away from $\mathbb{D}_{\{\boldsymbol{x}\}}^{(k-1)|b}(\epsilon)$ and $\mathbf{C}^{(k)|b}(B^\circ;\boldsymbol{x}) = \mathbf{C}^{(k)|b}(B^-;\boldsymbol{x}) > 0$. Then

$$\mathbf{P}\big(\boldsymbol{Y}_{[0,1]}^{\eta|b}(\boldsymbol{x}) \in \cdot \mid \boldsymbol{Y}_{[0,1]}^{\eta|b}(\boldsymbol{x}) \in B\big) \Rightarrow \frac{\mathbf{C}^{(k)|b}(\ \cdot \cap B; x)}{\mathbf{C}^{(k)|b}(B;x)} \qquad \text{as } \eta \downarrow 0.$$

(ii) Furthermore, suppose that Assumption 3 holds. For some $k \in \mathbb{N},\ \boldsymbol{x} \in \mathbb{R}^m$, and measurable $B \subseteq \mathbb{D}$, suppose that B is bounded away from $\mathbb{D}_{\{\boldsymbol{x}\}}^{(k-1)}(\epsilon)$ and $\mathbf{C}^{(k)}(B^\circ;\boldsymbol{x}) = \mathbf{C}^{(k)}(B^-;\boldsymbol{x}) > 0$. Then

$$\mathbf{P}\big(\boldsymbol{Y}_{[0,1]}^{\eta}(x) \in \cdot \mid \boldsymbol{Y}_{[0,1]}^{\eta}(x) \in B\big) \Rightarrow \frac{\mathbf{C}^{(k)}(\ \cdot \cap B; x)}{\mathbf{C}^{(k)}(B;x)} \qquad \text{as } \eta \downarrow 0.$$

B.2 Metastability Analysis

Consider some open set $I \subseteq \mathbb{R}^m$ such that $\mathbf{0} \in \boldsymbol{I}$ and Assumption 4 holds. Define stopping times

$$\tau_Y^\eta(\boldsymbol{x}) \triangleq \inf\{t \geq 0 : \boldsymbol{Y}_t^\eta(\boldsymbol{x}) \notin I\}, \qquad \tau_Y^{\eta|b}(\boldsymbol{x}) \triangleq \inf\{t \geq 0 : \boldsymbol{Y}_t^{\eta|b}(\boldsymbol{x}) \notin I\}.$$

as the first exit times of $\boldsymbol{Y}_t^\eta(\boldsymbol{x})$ and $\boldsymbol{Y}_t^{\eta|b}(\boldsymbol{x})$ from I, respectively. The following result characterizes the asymptotic law of the first exit times $\tau_Y^\eta(\boldsymbol{x})$ and $\tau_Y^{\eta|b}(\boldsymbol{x})$ using the measures $\check{\mathbf{C}}^{(k)|b}(\cdot)$ defined in (2.34) and $\check{\mathbf{C}}(\cdot)$ defined in (2.35). We omit the proof due to its similarity to that of Theorem 2.6.

Theorem B.4. *Let Assumptions 2, 4, and 5 hold. Let $\beta \in [0, 2 \wedge \alpha)$.*

(a) Let $b > 0$ such that $\mathcal{J}_b^I < \infty$. Suppose that I^c is bounded away from $\mathcal{G}^{(\mathcal{J}_b^I - 1)|b}(\epsilon)$ for some (and hence all) $\epsilon > 0$ small enough, and $\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\partial I) = 0$. Then $C_b^I \triangleq \check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(I^c) < \infty$. Furthermore, if $C_b^I \in (0, \infty)$, then for any $\epsilon > 0$, $t \geq 0$, and measurable set $B \subseteq I^c$,

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\left(C_b^I \lambda_L^{\mathcal{J}_b^I}(\eta; \beta) \tau_Y^{\eta|b}(\boldsymbol{x}) > t;\ \boldsymbol{Y}^{\eta|b}_{\tau_Y^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B \right) \leq \frac{\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(B^-)}{C_b^I} \cdot \exp(-t),$$

$$\liminf_{\eta \downarrow 0} \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\left(C_b^I \lambda_L^{\mathcal{J}_b^I}(\eta; \beta) \tau_Y^{\eta|b}(\boldsymbol{x}) > t;\ \boldsymbol{Y}^{\eta|b}_{\tau_Y^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B \right) \geq \frac{\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(B^\circ)}{C_b^I} \cdot \exp(-t).$$

Otherwise, we must have $C_b^I = 0$, and

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\left(\lambda_L^{\mathcal{J}_b^I}(\eta; \gamma) \tau_Y^{\eta|b}(\boldsymbol{x}) \leq t \right) = 0 \qquad \forall \epsilon > 0,\ t \geq 0.$$

(b) Suppose that $\check{\mathbf{C}}(\partial I) = 0$. Then $C_\infty^I \triangleq \check{\mathbf{C}}(I^c) < \infty$. Furthermore, if $C_\infty^I > 0$, then for any $t \geq 0$ and measurable set $B \subseteq I^c$,

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\left(C_\infty^I \lambda_L(\eta; \beta) \tau_Y^\eta(\boldsymbol{x}) > t;\ \boldsymbol{Y}^\eta_{\tau_Y^\eta(\boldsymbol{x})}(\boldsymbol{x}) \in B \right) \leq \frac{\check{\mathbf{C}}(B^-)}{C_\infty^I} \cdot \exp(-t),$$

$$\liminf_{\eta \downarrow 0} \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\left(C_\infty^I \lambda_L(\eta; \beta) \tau_Y^\eta(\boldsymbol{x}) > t;\ \boldsymbol{Y}^\eta_{\tau_Y^\eta(\boldsymbol{x})}(\boldsymbol{x}) \in B \right) \geq \frac{\check{\mathbf{C}}(B^\circ)}{C_\infty^I} \cdot \exp(-t).$$

Otherwise, we must have $C_\infty^I = 0$, and

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\left(\lambda(_L \eta; \gamma) \tau_Y^\eta(\boldsymbol{x}) \leq t \right) = 0 \qquad \forall \epsilon > 0,\ t \geq 0.$$

C Properties of Mappings $\bar{h}_{[0,T]}^{(k)}$ and $\bar{h}_{[0,T]}^{(k)|b}$

In this section, we collect a few useful results about the mapping $\bar{h}_{[0,T]}^{(k)}$ defined in (2.10)–(2.12) and $\bar{h}_{[0,T]}^{(k)|b}$ defined in (2.21)–(2.23), and provide the proof of Lemmas 3.5, 3.6, and 3.7.

For any $\xi \in \mathbb{D}$, let $\|\xi\| \triangleq \sup_{t \in [0,1]} \|\xi(t)\|$. Also, recall the definition of $\mathbb{D}_A^{(k)|b}(r)$ in (2.25). Lemma C.1 shows that $\|\xi\|$ is uniformly bounded for all $\xi \in \mathbb{D}_A^{(k)|b}(r)$.

Lemma C.1. *Let Assumption 2 hold. Given $k \in \mathbb{N}$, $b, r > 0$, and a compact set $A \subseteq \mathbb{R}^m$, there exists $M = M(k, b, r, A) < \infty$ such that $\|\xi\| \leq M\ \forall \xi \in \mathbb{D}_A^{(k)|b}(r)$.*

Proof. Fix some $\boldsymbol{x}_0 \in A$, and let $\xi^*(t) = \boldsymbol{y}_t(\boldsymbol{x}_0)$; see (2.28). Let $N = r + \sup_{\boldsymbol{x},\boldsymbol{y}\in A} \|\boldsymbol{x} - \boldsymbol{y}\| \vee b$ and $\rho = \exp(D) \geq 1$ where $D \in [1, \infty)$ is the Lipschitz coefficient in Assumption 2.

By arbitrarily picking an element from $\mathbb{D}_A^{(k)|b}(r)$, we get some $\xi = \bar{h}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t})$ with $\boldsymbol{x} \in A$, $\mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\mathbf{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$, $\boldsymbol{t} = (t_1, \cdots, t_k) \in (0, 1]^{k\uparrow}$. By Assumption 2 and Gronwall's inequality, we get $\sup_{t\in[0,t_1)} \|\xi^*(t) - \xi(t)\| \leq \|\boldsymbol{x} - \boldsymbol{x}_0\| \cdot \exp(Dt_1) \leq \rho N$. Since $\xi^*(t)$ is continuous, and $\|\xi(t_1) - \xi(t_1-)\| \leq b + r$ (see the definition of φ_b in (2.23)), we get $\sup_{t\in[0,t_1]} \|\xi^*(t) - \xi(t)\| \leq \rho N + b + r \leq 2\rho N$.

Next, we proceed by induction. Adopt the convention that $t_{k+1} = 1$, and suppose that for some $j = 1, 2, \cdots, k$,

$$\sup_{t\in[0,t_j]} \|\xi^*(t) - \xi(t)\| \leq \underbrace{(2\rho)^j N}_{\triangleq M_j}.$$

Then from Gronwall's inequality again, we get $\|\xi^*(t) - \xi(t)\| \leq \rho A_j$ for any $t \in [t_j, t_{j+1})$. Due to the continuity of ξ^* and the truncation threshold b and the upper bound $\|\boldsymbol{v}_j\| \leq r$ at step (2.23), we have

$$\|\xi(t_{j+1}) - \xi^*(t_{j+1})\| \leq \rho M_j + b + r \leq 2\rho M_j \leq M_{j+1}.$$

Therefore, $\sup_{t\in[0,t_{j+1}]} \|\xi^*(t) - \xi(t)\| \leq M_{j+1}$. By induction, we can conclude the proof with $M = M_{k+1} + \|\xi^*\| = (2\rho)^{k+1}N + \|\xi^*\|$. □

Recall the definitions of $\boldsymbol{a}_M, \boldsymbol{\sigma}_M$ in (3.35), the mapping $\bar{h}_{M\downarrow}^{(k)|b}$ in (3.36)–(3.38), and sets $\mathbb{D}_{A;M\downarrow}^{(k)|b}(r)$ in (3.40). Next, we present a corollary of the boundedness of $\mathbb{D}_A^{(k)|b}(r)$ established in Lemma C.1.

Corollary C.2. *Let Assumption 2 hold. Let $b, r > 0$, $k \in \mathbb{N}$. Let $A \subseteq \mathbb{R}^m$ be compact. There exists $M_0 \in (0, \infty)$ such that for any $M \geq M_0$,*

- $\sup_{t\leq 1} \|\xi_t\| \leq M_0$ $\forall \xi \in \mathbb{D}_A^{(k)|b}(r) \cup \mathbb{D}_{A;M\downarrow}^{(k)|b}(r)$;
- *For any $\boldsymbol{t} = (t_1, \cdots, t_k) \in (0, 1]^{k\uparrow}$, $\boldsymbol{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\boldsymbol{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$ with $\max_{j\in[d]} \|\boldsymbol{v}_j\| \leq r$, and $\boldsymbol{x} \in A$,*

$$\bar{h}^{(k)|b}(\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{V}, \boldsymbol{t}) = \bar{h}_{M\downarrow}^{(k)|b}(\boldsymbol{x}, \boldsymbol{W}, \boldsymbol{V}, \boldsymbol{t}).$$

Proof. The claims follow immediately from Lemma C.1, as well as the fact that $\bar{h}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t}) \in \mathbb{D}_A^{(k)|b}(r)$ and $\xi = \bar{h}_{M\downarrow}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t}) \in \mathbb{D}_{A;M\downarrow}^{(k)|b}(r)$. □

Next, we study the continuity of mappings $\bar{h}_{[0,T]}^{(k)}$ and $\bar{h}_{[0,T]}^{(k)|b}$.

Lemma C.3. *Let Assumption 2 hold. Given any $b, T > 0$ and any $k \in \mathbb{N}$, the mapping $\bar{h}_{[0,T]}^{(k)|b}$ is continuous on $\mathbb{R}^m \times \mathbb{R}^{d\times k} \times \mathbb{R}^{m\times k} \times (0, T)^{k\uparrow}$.*

Proof. To ease notations we focus on the case where $T = 1$, but the proof is identical for any $T > 0$. Arbitrarily pick some $b > 0$ and $k \in \mathbb{N}$, some $\boldsymbol{x}^* \in \mathbb{R}^m$, $\mathbf{W}^* = (\boldsymbol{w}_1^*, \cdots, \boldsymbol{w}_k^*) \in \mathbb{R}^{d\times k}$, $\mathbf{V}^* = (\boldsymbol{v}_1^*, \cdots, \boldsymbol{v}_k^*) \in \mathbb{R}^{m\times k}$, and $\boldsymbol{t}^* = (t_1^*, \cdots, t_k^*) \in (0, 1)^{k\uparrow}$. Let $\xi^* = \bar{h}^{(k)|b}(\boldsymbol{x}^*, \mathbf{W}^*, \mathbf{V}^*, \boldsymbol{t}^*)$. Also, fix some $\epsilon \in (0, 1)$. It suffices to show the existence of some $\delta \in (0, 1)$ such that $\boldsymbol{d}_{J_1}(\xi^*, \xi') < \epsilon$ for all $\xi' = \bar{h}^{(k)|b}(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}')$ with $\boldsymbol{x}' \in \mathbb{R}^m$, $\mathbf{W}' = (\boldsymbol{w}_1', \cdots, \boldsymbol{w}_k') \in \mathbb{R}^{d\times k}$, $\mathbf{V}' = (\boldsymbol{v}_1', \cdots, \boldsymbol{v}_k') \in \mathbb{R}^{m\times k}$, $\boldsymbol{t}' = (t_1', \cdots, t_k') \in (0, 1)^{k\uparrow}$ satisfying

$$\|\boldsymbol{x}^* - \boldsymbol{x}'\| < \delta, \qquad \left\|\boldsymbol{w}_j' - \boldsymbol{w}_j^*\right\| \vee \left\|\boldsymbol{v}_j^* - \boldsymbol{v}_j'\right\| \vee |t_j' - t_j^*| < \delta \ \forall j \in [k]. \tag{C.1}$$

We start by setting some constants and notations. First, by Corollary C.2, it follows for any $M \in (0,\infty)$ large enough that

$$\|\xi^*\| + 1 < M \qquad \text{and} \qquad \|\xi'\| + 1 < M \quad \forall \xi' = \bar{h}^{(k)|b}(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}') \text{ satisfying (C.1).} \tag{C.2}$$

By picking an even larger M if necessary, we can ensure that $M \geq 1 + \max_{j\in[k]} \left\|\boldsymbol{w}_j^*\right\|$. In this proof, we write $\boldsymbol{a}^* = \boldsymbol{a}_M$, $\boldsymbol{\sigma}^* = \boldsymbol{\sigma}_M$ (see (3.35) for definitions). Fix the constant

$$C^* \triangleq \sup_{\boldsymbol{y}: \|\boldsymbol{y}\| \leq M} \|\boldsymbol{a}(y)\| \vee \|\boldsymbol{\sigma}(y)\| \vee 1 < \infty.$$

We also write $h^* = \bar{h}_{M\downarrow}^{(k)|b}$ in this proof; see (3.36)–(3.38) for definitions. The choice of M ensures that $\xi^* = h^*(\boldsymbol{x}^*, \mathbf{W}^*, \mathbf{V}^*, \boldsymbol{t}^*)$ and, under condition (C.1), $\xi' = h^*(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}')$.

Let $\rho \triangleq \exp(D) \geq 1$ where $D \in [1,\infty)$ is the Lipschitz coefficient in Assumption 2. Let $R_0 = 1$,

$$R_j \triangleq (7C^* + \rho R_{j-1} + 1)(DM + 1) + C^* \qquad \forall j \geq 1. \tag{C.3}$$

We pick some $\tilde{\delta} > 0$ small enough such that

$$2\tilde{\delta} < 1 \wedge \epsilon; \qquad R_{k+1}\tilde{\delta} < \epsilon. \tag{C.4}$$

Also, by picking $\delta > 0$ small enough, it is guaranteed that (under convention $t_0^* = t_0' = 0$, $t_{k+1}^* = t_{k+1}' = 1$)

$$\delta < \tilde{\delta} \vee 1; \qquad \max_{j\in[k]} \left| \frac{t_{j+1}^* - t_j^*}{t_{j+1}' - t_j'} - 1 \right| < \tilde{\delta} \ \forall \boldsymbol{t}' = (t_1', \cdots, t_k') \in (0,1)^{k\uparrow},\ \max_{j\in[k]} |t_j' - t_j^*| < \delta. \tag{C.5}$$

Now it only remains to show that, under the current the choice of δ, the bound $\boldsymbol{d}_{J_1}(\xi, \xi') < \epsilon$ follows from condition (C.1). To do so, we fix some ξ' satisfying condition (C.1). Define $\lambda : [0,1] \to [0,1]$ as

$$\lambda(u) = \begin{cases} 0 & \text{if } u = 0 \\ t_j^* + \frac{t_{j+1}^* - t_j^*}{t_{j+1}' - t_j'} \cdot (u - t_j') & \text{if } u \in (t_j', t_{j+1}'] \text{ for some } j = 0, 1, \cdots, k. \end{cases}$$

For any $u \in (0,1)$, let $j \in \{0, 1, \cdots, k\}$ be such that $u \in (t_j', t_{j+1}']$. Observe that

$$\begin{aligned} |\lambda(u) - u| &= \left| t_j^* + \frac{t_{j+1}^* - t_j^*}{t_{j+1}' - t_j'} \cdot (u - t_j') - u \right| = \left| t_j^* + \frac{t_{j+1}^* - t_j^*}{t_{j+1}' - t_j'} \cdot v - (v + t_j') \right| \quad \text{with } v \triangleq u - t_j' \\ &\leq |t_j^* - t_j'| + \left| \frac{t_{j+1}^* - t_j^*}{t_{j+1}' - t_j'} - 1 \right| \cdot v \\ &\leq \tilde{\delta} + \tilde{\delta} \cdot 1 < \epsilon. \end{aligned} \tag{C.6}$$

In summary, we have shown that $\sup_{u\in[0,1]} |\lambda(u) - u| < \epsilon$. Moving on, we prove that

$$\sup_{u\in[0,1]} \left\| \xi^*\big(\lambda(u)\big) - \xi'(u) \right\| < \epsilon$$

using an inductive argument. First, Assumption 2 allows us to apply Gronwall's inequality and get $\sup_{u\in(0, t_1^* \wedge t_1')} \|\xi^*(u) - \xi'(u)\| \leq \exp\big(D \cdot (t_1^* \wedge t_1')\big) \|\boldsymbol{x}^* - \boldsymbol{x}'\| \leq \rho\delta$. As a result, for any $u \in (0, t_1^* \wedge t_1')$,

$$\begin{aligned} \left\| \xi^*\big(\lambda(u)\big) - \xi'(u) \right\| &= \left\| \xi^*\left(\frac{t_1^*}{t_1'} \cdot u\right) - \xi'(u) \right\| \leq \left\| \xi^*\left(\frac{t_1^*}{t_1'} \cdot u\right) - \xi^*(u) \right\| + \|\xi'(u) - \xi^*(u)\| \\ &\leq \left\| \xi^*\left(\frac{t_1^*}{t_1'} \cdot u\right) - \xi^*(u) \right\| + \rho\delta \end{aligned}$$

$$
\begin{aligned}
&\le \sup_{\boldsymbol{y}\in\mathbb{R}^m} \|\boldsymbol{a}^*(\boldsymbol{y})\| \cdot \left|\frac{t_1^*}{t_1'} - 1\right| \cdot u + \rho\delta \qquad \text{by (C.2)} \\
&\le C^*\widetilde{\delta} + \rho\widetilde{\delta} = (C^* + \rho)\widetilde{\delta} \qquad \text{due to (C.5)}.
\end{aligned}
$$

In case that $t_1' \le t_1^*$, we get $\sup_{u\in(0,t_1')} \left\|\xi^*\big(\lambda(u)\big) - \xi'(u)\right\| < (C^* + \rho)\widetilde{\delta}$ directly. In case that $t_1^* < t_1'$, due to $\xi' = h^*(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}')$ as well as the bounds in (C.5)(C.6), for any $u \in [t_1^*, t_1')$ we have

$$
\begin{aligned}
\|\xi'(u) - \xi'(t_1^*)\| &\le \sup_{\boldsymbol{y}\in\mathbb{R}^m} \|\boldsymbol{a}^*(\boldsymbol{y})\| \cdot |u - t_1^*| < C^*\widetilde{\delta}; \\
\left\|\xi^*\big(\lambda(u)\big) - \xi^*\big(\lambda(t_1^*)\big)\right\| &\le \sup_{\boldsymbol{y}\in\mathbb{R}^m} \|\boldsymbol{a}^*(\boldsymbol{y})\| \cdot \left|\lambda(u) - \lambda(t_1^*)\right| < 5C^*\widetilde{\delta}.
\end{aligned}
$$

As a result, $\sup_{u\in(0,t_1')} \left\|\xi^*\big(\lambda(u)\big) - \xi'(u)\right\| < (7C^* + \rho)\widetilde{\delta}$. In addition, due to $\|\varphi_b(\boldsymbol{x}) - \varphi_b(\boldsymbol{y})\| \le \|\boldsymbol{x} - \boldsymbol{y}\|$,

$$
\begin{aligned}
&\left\|\xi^*\big(\lambda(t_1')\big) - \xi'(t_1')\right\| \\
&= \left\|\xi^*\big(\lambda(t_1'-)\big) + \boldsymbol{v}_1^* + \varphi_b\Big(\boldsymbol{\sigma}^*\Big(\xi^*\big(\lambda(t_1'-)\big) + \boldsymbol{v}_1^*\Big)\boldsymbol{w}_1^*\Big) - \xi'(t_1'-) - \boldsymbol{v}_1' - \varphi_b\Big(\boldsymbol{\sigma}^*\Big(\xi'(t_1'-) + \boldsymbol{v}_1'\Big)\boldsymbol{w}_1'\Big)\right\| \\
&\le \left\|\xi^*\big(\lambda(t_1'-)\big) - \xi'(t_1'-)\right\| + \|\boldsymbol{v}_1^* - \boldsymbol{v}_1'\| + \left\|\boldsymbol{\sigma}^*\Big(\xi^*\big(\lambda(t_1'-)\big) + \boldsymbol{v}_1^*\Big)\boldsymbol{w}_1^* - \boldsymbol{\sigma}^*\Big(\xi'(t_1'-) + \boldsymbol{v}_1'\Big)\boldsymbol{w}_1'\right\| \\
&\le \left\|\xi^*\big(\lambda(t_1'-)\big) - \xi'(t_1'-)\right\| + \|\boldsymbol{v}_1^* - \boldsymbol{v}_1'\| + \left\|\boldsymbol{\sigma}^*\Big(\xi^*\big(\lambda(t_1'-)\big) + \boldsymbol{v}_1^*\Big) - \boldsymbol{\sigma}^*\Big(\xi'(t_1'-) + \boldsymbol{v}_1'\Big)\right\| \cdot \|\boldsymbol{w}_1^*\| \\
&\qquad\qquad + \left\|\boldsymbol{\sigma}^*\Big(\xi'(t_1'-) + \boldsymbol{v}_1'\Big)\right\| \cdot \|\boldsymbol{w}_1' - \boldsymbol{w}_1^*\| \\
&\le \left\|\xi^*\big(\lambda(t_1'-)\big) - \xi'(t_1'-)\right\| + \delta + \left\|\boldsymbol{\sigma}^*\Big(\xi^*\big(\lambda(t_1'-)\big) + \boldsymbol{v}_1^*\Big) - \boldsymbol{\sigma}^*\Big(\xi'(t_1'-) + \boldsymbol{v}_1'\Big)\right\| \cdot M + C^*\delta \\
&\le (7C^* + \rho)\widetilde{\delta} + \delta + M \cdot D \cdot \Big(\left\|\xi^*\big(\lambda(t_1'-)\big) - \xi'(t_1'-)\right\| + \|\boldsymbol{v}_1^* - \boldsymbol{v}_1'\| \Big) + C^*\delta \qquad \text{due to Assumption 2} \\
&= (7C^* + \rho)\widetilde{\delta} + \delta + DM\big((7C^* + \rho)\widetilde{\delta} + \delta\big) + C^*\delta \\
&\le \big[(7C^* + \rho + 1)(DM + 1) + C^*\big]\widetilde{\delta} \qquad \text{by our choice of } \delta < \widetilde{\delta} \text{ in (C.4)(C.5)}.
\end{aligned}
$$

In summary, we yield $\sup_{u\in[0,t_1']} \left\|\xi^*\big(\lambda(u)\big) - \xi'(u)\right\| \le \big[(7C^* + \rho + 1)(DM + 1) + C^*\big]\widetilde{\delta} = R_1\widetilde{\delta}$; see definitions in (C.3). Now, suppose that for some $j = 1, 2, \cdots, k$, we have $\sup_{u\in[0,t_j']} \left\|\xi^*\big(\lambda(u)\big) - \xi'(u)\right\| \le R_j\widetilde{\delta}$. By repeating the calculations above, one can obtain that $\sup_{u\in[0,t_{j+1}']} \left\|\xi^*\big(\lambda(u)\big) - \xi'(u)\right\| \le R_{j+1}\widetilde{\delta}$. To conclude, note that $R_{k+1}\widetilde{\delta} < \epsilon$ by our choice of parameters in (C.4). □

Lemma C.4. *Let Assumption 2 and 3 hold. Given any $k \in \mathbb{N}$ and $T > 0$, the mapping $\bar{h}^{(k)}_{[0,T]}$ is continuous on $\mathbb{R}^m \times \mathbb{R}^{d\times k} \times \mathbb{R}^{m\times k} \times (0, T)^{k\uparrow}$.*

Proof. To ease notations we focus on the case where $T = 1$, but the proof is identical for any $T > 0$. Arbitrarily pick some $k \in \mathbb{N}$, $\boldsymbol{x}^* \in \mathbb{R}^m$, $\mathbf{W}^* = (\boldsymbol{w}_1^*, \cdots, \boldsymbol{w}_k^*) \in \mathbb{R}^{d\times k}$, $\mathbf{V}^* = (\boldsymbol{v}_1^*, \cdots, \boldsymbol{v}_k^*) \in \mathbb{R}^{m\times k}$, and $\boldsymbol{t}^* = (t_1^*, \cdots, t_k^*) \in (0, 1)^{k\uparrow}$. We claim the existence of some $b = b(\boldsymbol{x}^*, \mathbf{W}^*, \mathbf{V}^*, \boldsymbol{t}^*) > 0$ such that for any $\delta \in (0, 1)$, $\boldsymbol{x}' \in \mathbb{R}^m$, $\mathbf{W}' = (\boldsymbol{w}_1', \cdots, \boldsymbol{w}_k') \in \mathbb{R}^{d\times k}$, $\mathbf{V}' = (\boldsymbol{v}_1', \cdots, \boldsymbol{v}_k') \in \mathbb{R}^{m\times k}$, and $\boldsymbol{t}' \in (0, 1)^{k\uparrow}$ satisfying

$$
\|\boldsymbol{x}^* - \boldsymbol{x}'\| < \delta, \qquad \left\|\boldsymbol{w}_j' - \boldsymbol{w}_j^*\right\| \vee \left\|\boldsymbol{v}_j^* - \boldsymbol{v}_j'\right\| \vee |t_j' - t_j^*| < \delta \ \forall j \in [k]. \tag{C.7}
$$

we have $\bar{h}^{(k)}(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}') = \bar{h}^{(k)|b}(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}')$. Then the continuity of $\bar{h}^{(k)}$ follows immediately from the continuity of $\bar{h}^{(k)|b}$ established in Lemma C.3.

Now, it only remains to find such $b > 0$. In particular, we can simply set $b = C \cdot \big(\max\{\left\|\boldsymbol{w}_j^*\right\| : j \in [k]\} + 1\big)$ where $C \ge 1$ is the constant in Assumption 3 satisfying $\sup_{\boldsymbol{y}\in\mathbb{R}^m} \|\boldsymbol{\sigma}(\boldsymbol{y})\| \le C$. Indeed, given any $\delta \in (0,1)$ and $\boldsymbol{x}' \in \mathbb{R}^m$, $\mathbf{W}' \in \mathbb{R}^{d\times k}$, $\mathbf{V}' \in \mathbb{R}^{m\times k}$, and $\boldsymbol{t}' \in (0,1)^{k\uparrow}$ satisfying (C.7), for $\xi' = \bar{h}^{(k)|b}(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}')$ we have

$$\left\|\boldsymbol{\sigma}\big(\xi'(t'_j-) + \boldsymbol{v}_j\big)\boldsymbol{w}'_j\right\| \leq C\cdot\big(\max\{\|\boldsymbol{w}^*_i\|:\ i\in[k]\} + \delta\big) < b \qquad \forall j \in [d].$$

As a result, the truncation operator φ_b at step (2.23) is not in effect, and hence $\xi' = \bar{h}^{(k)}(\boldsymbol{x}', \mathbf{W}', \mathbf{V}', \boldsymbol{t}')$. This concludes the proof. □

Next, we move onto the proofs of Lemmas 3.5, 3.6, and 3.7.

Proof of Lemma 3.5. The claims are trivial if A or B is an empty set. Also, the claims are trivially true if $k = 0$ (note that in (b) we would have $\mathbb{D}_A^{(-1)}(r) = \emptyset$). Therefore, in this proof we focus on the case where $A \neq \emptyset$, $B \neq \emptyset$, and $k \geq 1$.

Since B is bounded away from $\mathbb{D}_A^{(k-1)}(r)$ under $\boldsymbol{d}_{J_1}$, there exists $\bar{\epsilon} > 0$ such that $\boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)}(r)\big) > 0$ so that part (b) is satisfied. Next, we show that there exists $\bar{\delta} > 0$, which together with $\bar{\epsilon}$ satisfies (a). Let $D \in [1,\infty)$ be the Lipschitz coefficient in Assumption 2. Besides, recall the constant $C \in (1,\infty)$ in Assumption 3 that satisfies $\sup_{\boldsymbol{x}\in\mathbb{R}}\|\boldsymbol{\sigma}(\boldsymbol{x})\| \leq C$. Let $\rho \triangleq \exp(D)$ and

$$\bar{\delta} \triangleq \frac{\bar{\epsilon}}{\rho C + 1}. \tag{C.8}$$

Note that $\bar{\delta} < \bar{\epsilon}$. To show that the claim (a) holds for such $\bar{\epsilon}$ and $\bar{\delta}$, we proceed with proof by contradiction. Suppose that there is some $\boldsymbol{t} = (t_1,\cdots,t_k) \in (0,1]^{k\uparrow}$, $\mathbf{W} = (\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, and $\boldsymbol{x}_0 \in A$ such that $\xi \triangleq h^{(k)}(\boldsymbol{x}_0, \mathbf{W}, \boldsymbol{t}) \in B^{\bar{\epsilon}}$ yet $\|\boldsymbol{w}_J\| \leq \bar{\delta}$ for some $J = 1,2,\cdots,k$. We construct $\xi' \in \mathbb{D}_A^{(k-1)}(r)$ such that $\boldsymbol{d}_{J_1}(\xi',\xi) < \bar{\epsilon}$. Specifically, we focus on the case where $J < k$, since the proof when $J = k$ is almost identical but only slightly simpler. Now, recall the definition of $h^{(0)}_{\boldsymbol{\sigma}}(\cdot)$ given below (2.12), and construct ξ' as

$$\xi'(s) \triangleq \begin{cases} \xi(s) & s \in [0, t_J) \\ h^{(0)}(\xi'(t_J-))(s - t_J) & s \in [t_J, t_{J+1}) \\ \xi(s) & s \in [t_{J+1}, t]. \end{cases}$$

That is, ξ' is driven by the same ODE as ξ on $[t_J, t_{J+1})$, except that at the beginning of the intervals, ξ' starts from $\xi(t_J-)$ instead of $\xi(t_J)$. On the other hand, ξ' coincides with ξ outside of $[t_J, t_{J+1})$. To bound the distance between ξ and ξ', note that from Assumption 3, we have $\|\xi(t_J) - \xi(t_J-)\| = \|\boldsymbol{\sigma}(\xi(t_J-))\boldsymbol{w}_J\| \leq C\bar{\delta}$. Then using Gronwall's inequality, we get

$$\begin{aligned} \|\xi(s) - \xi'(s)\| &\leq \exp\big((t_{J+1} - t_J)D\big)\,\|\xi(t_J) - \xi'(t_J-)\| \\ &\leq \rho\,\|\xi(t_J) - \xi'(t_J-)\| \\ &\leq \rho C\bar{\delta} < \bar{\epsilon} \end{aligned} \tag{C.9}$$

for all $s \in [t_J, t_{J+1})$. This shows that $\boldsymbol{d}_{J_1}(\xi,\xi') < \bar{\epsilon}$. However, this cannot be the case since $\xi \in B^{\bar{\epsilon}}$, $\xi' \in \mathbb{D}_A^{(k-1)}(r)$, and we chose $\bar{\epsilon}$ such that $\boldsymbol{d}_{J_1}(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)}) > 0$. This concludes the proof for the case with $J < k$. The proof for the case where $J = k$ is almost identical. The only difference is that ξ' is set to be $\xi'(s) = \xi(s)$ for all $s < t_k$, and $\xi'(s) = h^{(0)}\big(\xi'(t_k-)\big)(s - t_k)$ for all $s \in [t_k, 1]$, □

Proof of Lemma 3.6. Similar to Lemma 3.5, all claims hold trivially if A or B is empty, or if $k = 0$. In this proof, we focus on the case where $A \neq \emptyset$, $B \neq \emptyset$, and $k \geq 1$.

We start by fixing some constant. Since B is bounded away from $\mathbb{D}_A^{(k-1)|b}(r)$, we can fix some $\bar{\epsilon} > 0$ such that $\boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)|b}(r)\big) > 0$ to conclude the proof of part (b). Next, let $D \in [1,\infty)$ be the Lipschitz coefficient in Assumption 2. Besides, recall the constant $C \in (1,\infty)$ in Assumption 3 that satisfies $\sup_{\boldsymbol{x}\in\mathbb{R}} \|\boldsymbol{\sigma}(\boldsymbol{x})\| \leq C$. Let $\rho \triangleq \exp(D)$. By picking an even smaller $\bar{\epsilon} > 0$ if necessary, we can w.l.o.g. assume that

$$2\rho\bar{\epsilon} < r \qquad \text{and} \qquad \boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)|b}(r)\big) > 2\rho\bar{\epsilon}. \tag{C.10}$$

Let

$$\bar{\delta} \triangleq \bar{\epsilon}/C. \tag{C.11}$$

To prove that part (a) holds for such $\bar{\delta}$, we proceed with a proof by contradiction. Arbitrarily pick some $\boldsymbol{x} \in A$, $\mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\mathbf{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$ with $\max_{j\in[k]} \|\boldsymbol{v}_j\| \leq \bar{\epsilon}$, $\boldsymbol{t} = (t_1, \cdots, t_k) \in (0,1)^{k\uparrow}$, and $b > 0$. For $\xi_b = \bar{h}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, \boldsymbol{t})$, suppose that $\xi_b \in B^{\bar{\epsilon}}$ yet there is some $J \in [k]$ such that $\|\boldsymbol{w}_J\| \leq \bar{\delta}$. Next, construct $\xi \in \mathbb{D}$ as follows: (recall that $\boldsymbol{y}_\cdot(x)$ is the ODE defined in (2.28))

$$\xi(s) \triangleq \begin{cases} \xi_b(s) & s \in [0, t_J) \\ \boldsymbol{y}_{s-t_J}(\xi(t_J-)) & s \in [t_J, t_{J+1}) \\ \xi_b(s) & s \in [t_{J+1}, 1]. \end{cases}$$

That is, ξ is a modified version of ξ_b where the jump at time t_J is removed, but the two paths coincide on $[0,t_J) \cup [t_{J+1}, 1]$. Note that by Assumption 3,

$$\|\xi(t_J) - \xi_b(t_J)\| = \|\Delta\xi_b(t_J)\| \leq \|\boldsymbol{v}_J\| + \Big\|\varphi_b\Big(\boldsymbol{\sigma}\big(\xi_b(t_J-) + \boldsymbol{v}_J\big)\boldsymbol{w}_J\Big)\Big\| \leq \bar{\epsilon} + C\bar{\delta}.$$

Applying Gronwall's inequality, we then yield that for all $s \in [t_J, t_{J-1})$,

$$\begin{aligned} \|\xi_b(s) - \xi(s)\| &\leq \exp\big(D(s-t_J)\big)\cdot\|\xi(t_J) - \xi_b(t_J)\| \\ &\leq \rho\cdot\|\xi(t_J) - \xi_b(t_J)\| \qquad \text{where } \rho = \exp(D) \\ &\leq \rho(\bar{\epsilon} + C\bar{\delta}) = 2\rho\bar{\epsilon} \qquad \text{due to (C.11)}. \end{aligned}$$

This implies that $\boldsymbol{d}_{J_1}(\xi, \xi_b) \leq 2\rho\bar{\epsilon}$ and $\xi \in \mathbb{D}_A^{(k-1)|b}(2\rho\bar{\epsilon}) \subseteq \mathbb{D}_A^{(k-1)|b}(r)$; see (C.10). However, in light of the condition $\boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)|b}(r)\big) > 2\rho\bar{\epsilon}$ in (C.10), we arrive at the contraction that $\xi_b \notin B^{\bar{\epsilon}}$. This concludes the proof of part (a). □

Proof of Lemma 3.7. The proof relies on the following claim: for any $S \in \mathscr{S}_{\mathbb{D}}$ that is bounded away from $\mathbb{D}_A^{(k-1)}(r)$,

$$\lim_{b\to\infty} \mathbf{C}^{(k)|b}(S;\boldsymbol{x}) = \mathbf{C}^{(k)}(S;\boldsymbol{x}). \tag{C.12}$$

Then for any $g \in \mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)}(r)\big)$, we know that $B = \text{supp}(g)$ is bounded away from $\mathbb{D}_A^{(k-1)}(r)$. Also, given any $\Delta > 0$, an approximation to g using simple functions implies the existence of some $N \in \mathbb{N}$, some sequence of real numbers $\big(c_g^{(i)}\big)_{i=1}^N$, some sequence $\big(B_g^{(i)}\big)_{i=1}^N$ of Borel measurable sets on $\mathbb{D}$ that are bounded away from $\mathbb{D}_A^{(k-1)}(r)$ such that the following claims hold for $g^\Delta(\cdot) = \sum_{i=1}^N c_g^{(i)}\mathbb{I}\big(\,\cdot\, \in B_g^{(i)}\big)$:

$$B_g^{(i)} \subseteq B \;\; \forall i \in [N]; \qquad \big|g^\Delta(\xi) - g(\xi)\big| < \Delta \;\; \forall \xi \in \mathbb{D}.$$

Then

$$\begin{aligned} \limsup_{b\to\infty}\Big|\mathbf{C}^{(k)|b}(g;\boldsymbol{x}) - \mathbf{C}^{(k)}(g;\boldsymbol{x})\Big| &\leq \limsup_{b\to\infty}\Big|\mathbf{C}^{(k)|b}(g;\boldsymbol{x}) - \mathbf{C}^{(k)|b}(g^\Delta;\boldsymbol{x})\Big| \\ &\quad + \limsup_{b\to\infty}\Big|\mathbf{C}^{(k)|b}(g^\Delta;\boldsymbol{x}) - \mathbf{C}^{(k)}(g^\Delta;\boldsymbol{x})\Big| \\ &\quad + \limsup_{b\to\infty}\Big|\mathbf{C}^{(k)}(g^\Delta;\boldsymbol{x}) - \mathbf{C}^{(k)}(g;\boldsymbol{x})\Big| \end{aligned}$$

First, note that $\mathbf{C}^{(k)|b}(g^{\Delta};\boldsymbol{x}) = \sum_{i=1}^{N} c_g^{(i)}\mathbf{C}^{(k)|b}(B_g^{(i)};\boldsymbol{x})$ and $\mathbf{C}^{(k)}(g^{\Delta};\boldsymbol{x}) = \sum_{i=1}^{N} c_g^{(i)}\mathbf{C}^{(k)}(B_g^{(i)};\boldsymbol{x})$. Therefore, applying (C.12), we get $\limsup_{b\to\infty}\left|\mathbf{C}^{(k)|b}(g^{\Delta};\boldsymbol{x}) - \mathbf{C}^{(k)}(g^{\Delta};\boldsymbol{x})\right| = 0$. Next, note that $\left|\mathbf{C}^{(k)|b}(g^{\Delta};\boldsymbol{x})-\mathbf{C}^{(k)|b}(g;\boldsymbol{x})\right| \leq \Delta\cdot\mathbf{C}^{(k)|b}(B;\boldsymbol{x})$ and $\left|\mathbf{C}^{(k)}(g^{\Delta};\boldsymbol{x})-\mathbf{C}^{(k)}(g;\boldsymbol{x})\right| \leq \Delta\cdot\mathbf{C}^{(k)}(B;\boldsymbol{x})$. Thanks to (C.12) again, we get $\limsup_{b\to\infty}\left|\mathbf{C}^{(k)|b}(g;\boldsymbol{x}) - \mathbf{C}^{(k)}(g;\boldsymbol{x})\right| \leq 2\Delta\cdot\mathbf{C}^{(k)}(B;\boldsymbol{x})$. The arbitrariness of $\Delta > 0$ allows us to conclude the proof.

Now, we prove (C.12) using Dominated Convergence theorem. By the definition in (2.26),

$$\mathbf{C}^{(k)|b}(S;\boldsymbol{x}) \triangleq \int \mathbb{I}\Big\{h^{(k)|b}\big(\boldsymbol{x},\mathbf{W},\boldsymbol{t}\big)\in S\Big\}\big((\nu_\alpha\times\mathbf{S})\circ\Phi\big)^k(d\mathbf{W})\times\mathcal{L}_1^{k\uparrow}(d\boldsymbol{t}).$$

where $S\in\mathscr{S}_{\mathbb{D}}$ is bounded away from $\mathbb{D}_A^{(k-1)}(r)$. First, we fix some $\mathbf{W}\in\mathbb{R}^{d\times k}$ and $\boldsymbol{t}\in(0,1)^{k\uparrow}$ and $x_0\in\mathbb{R}$, and let $M\triangleq\max_{j\in[k]}\|\boldsymbol{w}_j\|$. For any $b > MC$ where $C\geq 1$ is the constant satisfying such that $\sup_{\boldsymbol{x}\in\mathbb{R}^m}\|\boldsymbol{a}(\boldsymbol{x})\|\vee\|\boldsymbol{\sigma}(\boldsymbol{x})\|\leq C$ (see Assumption 3), by the definitions of $h^{(k)}$ and $h^{(k)|b}$ it is easy to see that $h^{(k)|b}(\boldsymbol{x},\mathbf{W},\boldsymbol{t}) = h^{(k)}(\boldsymbol{x},\mathbf{W},\boldsymbol{t})$. This implies

$$\lim_{b\to\infty}\mathbb{I}\big\{h^{(k)|b}\big(\boldsymbol{x},\mathbf{W},\boldsymbol{t}\big)\in S\big\} = \mathbb{I}\big\{h^{(k)}\big(\boldsymbol{x},\mathbf{W},\boldsymbol{t}\big)\in S\big\}\qquad\forall\mathbf{W}\in\mathbb{R}^{d\times k},\ \boldsymbol{t}\in(0,1]^{k\uparrow}.$$

In order to apply Dominated Convergence theorem and conclude the proof of (C.12), it suffices to find an integrable function that dominates $\mathbb{I}\big\{h^{(k)|b}\big(\boldsymbol{x},\mathbf{W},\boldsymbol{t}\big)\in S\big\}$. Specifically, since S is bounded away from $\mathbb{D}_A^{(k-1)}(r)$, we can find some $\bar\epsilon > 0$ such that $\boldsymbol{d}_{J_1}\big(S,\mathbb{D}_A^{(k-1)}(r)\big) > \bar\epsilon$. Also, let $\rho = \exp(D)$ where $D\in[1,\infty)$ is the Lipschitz coefficient in Assumption 2. Fix some $\bar\delta < \frac{\bar\epsilon}{\rho C}$. By part (a) of Lemma 3.6, we get

$$\mathbb{I}\Big\{h^{(k)|b}\big(\boldsymbol{x},\mathbf{W},\boldsymbol{t}\big)\in S\Big\}\leq\mathbb{I}\Big\{\ \|\boldsymbol{w}_j\| > \bar\delta\ \forall j\in[k]\Big\}\qquad\forall b>0,\ \mathbf{W}\in\mathbb{R}^{d\times k},\ \boldsymbol{t}\in(0,1)^{k\uparrow}.$$

From $\int\mathbb{I}\big\{\ \|\boldsymbol{w}_j\| > \bar\delta\ \forall j\in[k]\big\}\big((\nu_\alpha\times\mathbf{S})\circ\Phi\big)^k(d\mathbf{W})\times\mathcal{L}_1^{k\uparrow}(d\boldsymbol{t})\leq 1/\bar\delta^{k\alpha}<\infty$, we conclude the proof. ☐

The following result will be applied in the proof of Lemma 3.8. Let $\boldsymbol{x}_j^\eta(x)$ be the solution to

$$\boldsymbol{x}_0^\eta(\boldsymbol{x}) = \boldsymbol{x},\qquad \boldsymbol{x}_j^\eta(\boldsymbol{x}) = \boldsymbol{x}_{j-1}^\eta(\boldsymbol{x}) + \eta\boldsymbol{a}\big(\boldsymbol{x}_{j-1}^\eta(\boldsymbol{x})\big)\quad\forall j\geq 1. \tag{C.13}$$

After proper scaling of the time parameter, $\boldsymbol{x}_j^\eta$ approximates $\boldsymbol{y}_t$ with small η. The next lemma is a direct result from Gronwall's inequality and bounds the distance between $\boldsymbol{x}^\eta_{\lfloor t/\eta\rfloor}(x)$ and $\boldsymbol{y}_t(y)$. For the sake of completeness we provide the proof.

Lemma C.5. *Let Assumptions 2 and 3 hold. For any $\eta>0, t>0$ and $x,y\in\mathbb{R}^m$,*

$$\sup_{s\in[0,t]}\left\|\boldsymbol{y}_s(y)-\boldsymbol{x}^\eta_{\lfloor s/\eta\rfloor}(x)\right\|\leq(\eta C+\|x-y\|)\exp(Dt)$$

where $D,C\in[1,\infty)$ are the constants in Assumptions 2 and 3 respectively.

Proof. For any $s\geq 0$ that is not an integer, let $\boldsymbol{x}_s^\eta(x)\triangleq\boldsymbol{x}^\eta_{\lfloor s\rfloor}(x)$ and $\boldsymbol{y}_s^\eta(y)\triangleq\boldsymbol{y}_{s\eta}(y)$. Now observe that (for any $s\geq 0$)

$$\boldsymbol{y}_s^\eta(y) = \boldsymbol{y}^\eta_{\lfloor s\rfloor}(y) + \eta\int_{\lfloor s\rfloor}^{s}\boldsymbol{a}(\boldsymbol{y}_u^\eta(y))du$$

$$\boldsymbol{y}^\eta_{\lfloor s\rfloor}(y) = y + \eta\int_0^{\lfloor s\rfloor}\boldsymbol{a}(\boldsymbol{y}_u^\eta(y))du$$

$$\boldsymbol{x}^\eta_{\lfloor s\rfloor}(y) = x + \eta\int_0^{\lfloor s\rfloor}\boldsymbol{a}(\boldsymbol{x}_u^\eta(y))du.$$

Let $\boldsymbol{b}(u) \triangleq \boldsymbol{y}^{\eta}_u(y) - \boldsymbol{x}^{\eta}_u(x)$. It suffices to show that $\sup_{u\in[0,t/\eta]} \|\boldsymbol{b}(u)\| \leq (\eta C + \|x-y\|)\exp(Dt)$. To this end, we observe that (for any $s > 0$)

$$\begin{aligned}
\|\boldsymbol{b}(s)\| &\leq \|\boldsymbol{b}(\lfloor s\rfloor)\| + \left\|\eta\int_{\lfloor s\rfloor}^{s} \boldsymbol{a}\big(\boldsymbol{y}^{\eta}_u(y)\big)du\right\| \leq \|\boldsymbol{b}(\lfloor s\rfloor)\| + \eta C \\
&\leq \eta\int_0^{\lfloor s\rfloor} \left\|\boldsymbol{a}\big(\boldsymbol{y}^{\eta}_u(y)\big) - \boldsymbol{a}\big(\boldsymbol{x}^{\eta}_u(x)\big)\right\| du + \|x-y\| + \eta C \\
&\leq \eta D\int_0^{s} \|\boldsymbol{b}(u)\|\, du + \|x-y\| + \eta C \qquad \text{due to Assumption 3.}
\end{aligned}$$

Apply Gronwall's inequality (see Theorem V.68 of [53]) to $\|\boldsymbol{b}(u)\|$ on interval $[0, t/\eta]$ and we conclude the proof. □

D Technical Results for Metastability Analysis

We first give the proof for Corollary 2.7. To do so, we provide some straightforward bounds for the law of geometric random variables.

Lemma D.1. *Let $a : (0,\infty) \to (0,\infty)$, $b : (0,\infty) \to (0,\infty)$ be two functions such that $\lim_{\epsilon\downarrow 0} a(\epsilon) = 0, \lim_{\epsilon\downarrow 0} b(\epsilon) = 0$. Let $\{U(\epsilon) : \epsilon > 0\}$ be a family of geometric RVs with success rate $a(\epsilon)$, i.e. $\mathbf{P}(U(\epsilon) > k) = (1-a(\epsilon))^k$ for $k \in \mathbb{N}$. For any $c > 1$, there exists $\epsilon_0 > 0$ such that*

$$\exp\Big(-\frac{c\cdot a(\epsilon)}{b(\epsilon)}\Big) \leq \mathbf{P}\Big(U(\epsilon) > \frac{1}{b(\epsilon)}\Big) \leq \exp\Big(-\frac{a(\epsilon)}{c\cdot b(\epsilon)}\Big) \quad \forall \epsilon \in (0,\epsilon_0).$$

Proof. Note that $\mathbf{P}(U(\epsilon) > \frac{1}{b(\epsilon)}) = \big(1-a(\epsilon)\big)^{\lfloor 1/b(\epsilon)\rfloor}$. By taking logarithm on both sides, we have

$$\ln \mathbf{P}\Big(U(\epsilon) > \frac{1}{b(\epsilon)}\Big) = \lfloor 1/b(\epsilon)\rfloor \ln\Big(1-a(\epsilon)\Big) = \frac{\lfloor 1/b(\epsilon)\rfloor}{1/b(\epsilon)}\frac{\ln\Big(1-a(\epsilon)\Big)}{-a(\epsilon)}\frac{-a(\epsilon)}{b(\epsilon)}.$$

Since $\lim_{x\to 0}\frac{\ln(1+x)}{x} = 1$, we know that for ϵ sufficiently small, we will have $-c\frac{a(\epsilon)}{b(\epsilon)} \leq \ln \mathbf{P}\Big(U(\epsilon) > \frac{1}{b(\epsilon)}\Big) \leq -\frac{a(\epsilon)}{c\cdot b(\epsilon)}$. By taking exponential on both sides, we conclude the proof. □

Proof of Corollary 2.7. That the value of $\boldsymbol{\sigma}(\cdot)$ and $\boldsymbol{a}(\cdot)$ outside of the domain I has no impact on the first exit analysis. Therefore, by modifying the value of $\boldsymbol{\sigma}(\cdot)$ and $\boldsymbol{a}(\cdot)$ outside of I, we can assume w.l.o.g. that

$$\|\boldsymbol{a}(\boldsymbol{x})\| \vee \|\boldsymbol{\sigma}(\boldsymbol{x})\| \leq C \qquad \forall \boldsymbol{x} \in \mathbb{R}^m. \tag{D.1}$$

for some $C \in (0,\infty)$. That is, we impose the boundedness condition in Assumption 3.

We start with a few observations. First, under any $\eta \in (0, \frac{b}{2C})$, on the event $\{\eta\|\boldsymbol{Z}_j\| \leq \frac{b}{2C}\ \forall j \leq t\}$ the norm of the step-size (before truncation) $\eta\boldsymbol{a}\big(\boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^{\eta|b}_{j-1}(\boldsymbol{x})\big)\boldsymbol{Z}_j$ of $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ is less than b for each $j \leq t$. Therefore, $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ and $\boldsymbol{X}^{\eta}_j(\boldsymbol{x})$ coincide for such j's. In other words, for any $\eta \in (0, \frac{b}{2C})$, on event $\big\{\eta\|\boldsymbol{Z}_j\| \leq \frac{b}{2C}\ \forall j \leq t\big\}$ we have

$$\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) = \boldsymbol{X}^{\eta}_j(\boldsymbol{x})\ \forall j \leq t. \tag{D.2}$$

More importantly, given any measurable $A \subseteq \mathbb{R}$ such that $r_A = \inf\{\|\boldsymbol{x}\| :\ \boldsymbol{x} \in A\} > 0$, we claim that

$$\lim_{b\to\infty} \check{\mathbf{C}}^{(1)|b}(A) = \check{\mathbf{C}}(A). \tag{D.3}$$

This claim follows from a simple application of the dominated convergence theorem. Indeed, by definition of $\check{\mathbf{C}}^{(1)|b}$, we get $\check{\mathbf{C}}^{(1)|b}(A) = \int \mathbb{I}\{\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}) \in A\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)(d\boldsymbol{w})$. For $f_b(\boldsymbol{w}) \triangleq \mathbb{I}\{\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}) \in A\}$, we first note that given $\boldsymbol{w} \in \mathbb{R}^m$, we have $f_b(\boldsymbol{w}) = f(\boldsymbol{w}) \triangleq \mathbb{I}\{\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w} \in A\}$ for all $b > \|\boldsymbol{w}\|\,\|\boldsymbol{\sigma}(\mathbf{0})\|$. Therefore, $\lim_{b\to\infty} f_b(\boldsymbol{w}) = f(\boldsymbol{w})$ holds for all $\boldsymbol{w} \in \mathbb{R}^m$. Next, due to $r_A > 0$ and (D.1), we have $f_b(w) \leq \mathbb{I}\{\|\boldsymbol{w}\| \geq r_A/C\}$ for all $b > 0$ and $\boldsymbol{w} \in \mathbb{R}^m$. Moreover $\int \mathbb{I}\{\|\boldsymbol{w}\| \geq r_A/C\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)(d\boldsymbol{w}) = (C/r_A)^\alpha < \infty$. This allows us to apply dominated convergence theorem and establish (D.3). Besides, for all b large enough, we have

$$C_b^I = \check{\mathbf{C}}^{(1)|b}(I^c) = \int \mathbb{I}\Big\{\varphi_b\big(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}\big) \in I^c\Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)(d\boldsymbol{w})$$
$$= \int \mathbb{I}\Big\{\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w} \in I^c\Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)(d\boldsymbol{w}) = \check{\mathbf{C}}(I^c) \triangleq C_\infty^I. \tag{D.4}$$

To see why, we only need to notice that, since I is a bounded set, it holds for all b large enough enough,

$$\varphi_b\big(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}\big) \notin I \qquad \Longleftrightarrow \qquad \boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w} \notin I.$$

Now, we fix $t \geq 0$ and $B \subseteq I^c$. Also, henceforth in the proof we only consider b large enough such that $C_\infty^I = C_b^I$. An immediate consequence of this choice of b is that $\mathcal{J}_b^I = 1$. We focus on the case where $C_\infty^I > 0$, but we stress that the proof for the case with $C_\infty^I = 0$ is almost identical. First, note that $\lambda(\eta) = \eta^{-1}H(\eta^{-1})$ and hence $\eta \cdot \lambda(\eta) = H(\eta^{-1})$. To analyze the probability of event $A(\eta, \boldsymbol{x}) = \{C_\infty^I H(\eta^{-1})\tau^\eta(\boldsymbol{x}) > t,\ \boldsymbol{X}^\eta_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x}) \in B\}$, we arbitrarily pick some $T > t$ and observe that

$$A(\eta, \boldsymbol{x}) = \underbrace{\Big\{C_\infty^I H(\eta^{-1})\tau^\eta(\boldsymbol{x}) \in (t, T],\ \boldsymbol{X}^\eta_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x}) \in B\Big\}}_{\triangleq A_1(\eta,\boldsymbol{x},T)} \cup \underbrace{\Big\{C_\infty^I H(\eta^{-1})\tau^\eta(\boldsymbol{x}) > T,\ \boldsymbol{X}^\eta_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x}) \in B\Big\}}_{\triangleq A_2(\eta,\boldsymbol{x},T)}. \tag{D.5}$$

Let $E_b(\eta, T) \triangleq \{\eta\,\|\boldsymbol{Z}_j\| \leq \frac{b}{2C}\ \forall j \leq \frac{T}{C_\infty^I H(\eta^{-1})}\}$. To analyze the probability of $A_1(\eta, \boldsymbol{x}, T)$, we further decompose the event as

$$A_1(\eta, \boldsymbol{x}, T) = \Big(A_1(\eta, \boldsymbol{x}, T) \cap E_b(\eta, T)\Big) \cup \Big(A_1(\eta, \boldsymbol{x}, T) \setminus E_b(\eta, T)\Big).$$

First, for all $\eta \in (0, \frac{b}{2C})$,

$$\begin{aligned}
&\mathbf{P}\Big(A_1(\eta, \boldsymbol{x}, T) \cap E_b(\eta, T)\Big) \\
&= \mathbf{P}\bigg(\Big\{C_b^I \eta \cdot \lambda(\eta)\tau^{\eta|b}(\boldsymbol{x}) \in (t, T],\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\Big\} \cap E_b(\eta, T)\bigg) \qquad \text{due to (D.2) and (D.4)} \\
&\leq \mathbf{P}\bigg(C_b^I \eta \cdot \lambda(\eta)\tau^{\eta|b}(\boldsymbol{x}) \in (t, T],\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg) \\
&= \mathbf{P}\bigg(C_b^I \eta \cdot \lambda(\eta)\tau^{\eta|b}(\boldsymbol{x}) > t,\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg) - \mathbf{P}\bigg(C_b^I \eta \cdot \lambda(\eta)\tau^{\eta|b}(\boldsymbol{x}) > T,\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg).
\end{aligned}$$

By Theorem 2.6 and observation (D.4), we get

$$\limsup_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in I_\epsilon} \mathbf{P}\Big(A_1(\eta, \boldsymbol{x}, T) \cap E_b(\eta, T)\Big) \leq \frac{\check{\mathbf{C}}^{(1)|b}(B^-)}{C_\infty^I} \cdot \exp(-t) - \frac{\check{\mathbf{C}}^{(1)|b}(B^\circ)}{C_\infty^I} \cdot \exp(-T). \tag{D.6}$$

Meanwhile, $\sup_{\boldsymbol{x}\in I_\epsilon} \mathbf{P}\big(A_1(\eta, \boldsymbol{x}, T) \setminus E_b(\eta, T)\big) \leq \mathbf{P}\big((E_b(\eta, T))^c\big) = \mathbf{P}\big(\eta\,\|\boldsymbol{Z}_j\| > \frac{b}{2C}$ for some $j \leq \frac{T}{C_\infty^I H(\eta^{-1})}\big)$. Applying Lemma D.1, we get

$$\limsup_{\eta\downarrow 0} \mathbf{P}\bigg(\eta\,\|\boldsymbol{Z}_j\| > \frac{b}{2C} \text{ for some } j \leq \frac{T}{C_\infty^I H(\eta^{-1})}\bigg) = 1 - \liminf_{\eta\downarrow 0} \mathbf{P}\bigg(\text{Geom}\Big(H\big(\frac{b}{\eta \cdot 2C}\big)\Big) > \frac{T}{C_\infty^I H(\eta^{-1})}\bigg)$$

$$\leq 1 - \lim_{\eta \downarrow 0} \exp\Big(- \frac{T \cdot H(\eta^{-1} \cdot \frac{b}{2C})}{C^I_\infty H(\eta^{-1})} \Big)$$
$$= 1 - \exp\Big(- \frac{T}{C^I_\infty} \cdot \Big(\frac{2C}{b}\Big)^\alpha \Big). \tag{D.7}$$

Similarly,

$$A_2(\eta, \boldsymbol{x}, T) \subseteq \Big\{ C^I_\infty H(\eta^{-1}) \tau^\eta(\boldsymbol{x}) > T \Big\}$$
$$= \Big(\Big\{ C^I_\infty H(\eta^{-1}) \tau^\eta(\boldsymbol{x}) > T \Big\} \cap E_b(\eta, T) \Big) \cup \Big(\Big\{ C^I_\infty H(\eta^{-1}) \tau^\eta(\boldsymbol{x}) > T \Big\} \setminus E_b(\eta, T) \Big).$$

On $\{C^I_\infty H(\eta^{-1})\tau^\eta(\boldsymbol{x}) > T\} \cap E_b(\eta, T)$, due to (D.2) we have $\tau^\eta(\boldsymbol{x}) = \tau^{\eta|b}(\boldsymbol{x})$. By Theorem 2.6 and (D.4)3 again, we get

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(\Big\{ C^I_\infty H(\eta^{-1}) \tau^\eta(\boldsymbol{x}) > T \Big\} \cap E_b(\eta, T) \Big)$$
$$\leq \limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(C^I_b \eta \cdot \lambda(\eta) \tau^{\eta|b}(\boldsymbol{x}) > T \Big) \leq \exp(-T). \tag{D.8}$$

Meanwhile, the limit of $\sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\big(C^I_\infty H(\eta^{-1})\tau^\eta(\boldsymbol{x}) > T\} \setminus E_b(\eta, T)\big)$ as $\eta \downarrow 0$ is again bounded by (D.7). Collecting (D.6), (D.7), and (D.8), we yield that for all $b > 0$ large enough and all $T > t$,

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(A(\eta, \boldsymbol{x}) \Big)$$
$$\leq \frac{\check{\mathbf{C}}^{(1)|b}(B^-)}{C^I_\infty} \cdot \exp(-t) - \frac{\check{\mathbf{C}}^{(1)|b}(B^\circ)}{C^I_\infty} \cdot \exp(-T) + \exp(-T) + 2 \cdot \Big[1 - \exp\Big(- \frac{T}{C^I_\infty} \cdot \Big(\frac{2C}{b}\Big)^\alpha \Big) \Big].$$

In light of claim (D.3), we send $b \to \infty$ and $T \to \infty$, and conclude the proof of the upper bound.

The lower bound can be established analogously. In particular, from the decomposition in (D.5),

$$\inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\big(A(\eta, \boldsymbol{x})\big)$$
$$\geq \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\big(A_1(\eta, \boldsymbol{x}, T)\big) \geq \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\big(A_1(\eta, \boldsymbol{x}, T) \cap E_b(\eta, T)\big)$$
$$= \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(\Big\{ C^I_b \eta \cdot \lambda(\eta) \tau^{\eta|b}(\boldsymbol{x}) \in (t, T],\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B \Big\} \cap E_b(\eta, T) \Big) \qquad \text{due to (D.2) and (D.4)}$$
$$\geq \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(C^I_b \eta \cdot \lambda(\eta) \tau^{\eta|b}(\boldsymbol{x}) \in (t, T],\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B \Big) - \mathbf{P}\Big(\big(E_b(\eta, T)\big)^c \Big)$$
$$\geq \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(C^I_b \eta \cdot \lambda(\eta) \tau^{\eta|b}(\boldsymbol{x}) > t,\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B \Big) - \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(C^I_b \eta \cdot \lambda(\eta) \tau^{\eta|b}(\boldsymbol{x}) > T,\ \boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B \Big)$$
$$- \mathbf{P}\Big(\big(E_b(\eta, T)\big)^c \Big).$$

By Theorem 2.6 and the limit in (D.7), we yield (for all $b > 0$ large enough and all $T > t$)

$$\liminf_{\eta \downarrow 0} \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\Big(A(\eta, \boldsymbol{x}) \Big) \leq \frac{\check{\mathbf{C}}^{(1)|b}(B^\circ)}{C^I_\infty} \cdot \exp(-t) - \frac{\check{\mathbf{C}}^{(1)|b}(B^-)}{C^I_\infty} \cdot \exp(-T) - \Big[1 - \exp\Big(- \frac{T}{C^I_\infty} \cdot \Big(\frac{2C}{b}\Big)^\alpha \Big) \Big].$$

Sending $b \to \infty$ and then $T \to \infty$, we conclude the proof of the lower bound. □

The remainder of this section collects important properties of the measure $\check{\mathbf{C}}^{(k)|b}(\cdot)$ defined in (2.34). In particular, the proof of Lemma 4.2 will be provided at the end of this section. Throughout the rest of this section, we impose Assumption 2 and 4, and fix some $b>0$ such that the conditions in Theorem 2.6 hold. We fix some $\bar{\epsilon}>0$ small enough such that the conditions in (4.13)–(4.15) hold.

Recall that $I_\epsilon = \{\boldsymbol{y}:\ \|\boldsymbol{x}-\boldsymbol{y}\|<\epsilon \implies \boldsymbol{x}\in I\}$ is the ϵ-shrinkage of the domain I, and that I^-_ϵ is the closure of I_ϵ. We first study the mapping $\breve{g}^{(k)|b}$ in (2.31), which is defined based on $\bar{h}^{(k)|b}_{[0,T]}$ and $h^{(k)|b}_{[0,T]}$ defined in (2.21)–(2.24).

Lemma D.2. *Let Assumptions 2 and 4 hold. Let $\bar{\epsilon}>0$ be the constant in (4.13)–(4.15). Let $C\in[1,\infty)$ be such that $\sup_{\boldsymbol{x}\in I^-}\|\boldsymbol{a}(\boldsymbol{x})\|\vee\|\boldsymbol{\sigma}(\boldsymbol{x})\|\leq C$. (Below, we adopt the convention that $t_0=0$.)*

(a) Given any $T>0$, the claim $\xi(t)\in I^-_{2\bar{\epsilon}}\ \forall t\in[0,T]$ holds for all $\xi\in\mathbb{D}^{(\mathcal{J}^I_b-1)|b}_{\bar{B}_{\bar{\epsilon}}}[0,T](\bar{\epsilon})$.

(b) Let $\bar{c}\in(0,1)$ be the constant fixed in (4.19). There exist $\bar{\delta}>0$ and $\bar{t}>0$ such that the following claim holds: Given any $T>0$ and $\boldsymbol{x}_0\in\mathbb{R}^m$ with $\|\boldsymbol{x}_0\|\leq\bar{\epsilon}$, if

$$\xi(t)\notin I_{\bar{c}\bar{\epsilon}}\qquad \text{for some } \xi=h^{(\mathcal{J}^I_b-1)|b}_{[0,T]}\Big(\boldsymbol{x}_0+\varphi_b\big(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_0\big),\,\boldsymbol{W},(t_1,\cdots,t_{\mathcal{J}^I_b-1})\Big),\ t\in[0,T], \tag{D.9}$$

where $\boldsymbol{W}=(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{\mathcal{J}^I_b-1})\in\mathbb{R}^{d\times\mathcal{J}^I_b-1}$, and $(t_1,\cdots,t_{\mathcal{J}^I_b-1})\in(0,T]^{\mathcal{J}^I_b-1\uparrow}$, then

(i) $\xi(t)\in I^-_{2\bar{\epsilon}}$ for all $t\in[0,t_{\mathcal{J}^I_b-1})$;

(ii) $\xi(t_{\mathcal{J}^I_b-1})\notin I_{\bar{\epsilon}}$;

(iii) $\|\xi(t)\|\geq\bar{\epsilon}$ for all $t\leq t_{\mathcal{J}^I_b-1}$;

(iv) $t_{\mathcal{J}^I_b-1}<\bar{t}$;

(v) $\|\boldsymbol{w}_j\|>\bar{\delta}$ for all $j=0,1,\cdots,\mathcal{J}^I_b-1$.

(c) Let $T>0$, $\boldsymbol{x}\in\mathbb{R}^m$, $\boldsymbol{W}=(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{\mathcal{J}^I_b})\in\mathbb{R}^{d\times\mathcal{J}^I_b}$, $(t_1,\cdots,t_{\mathcal{J}^I_b})\in(0,T]^{\mathcal{J}^I_b\uparrow}$, and $\epsilon\in(0,\bar{\epsilon})$. Let

$$\xi=h^{(\mathcal{J}^I_b)|b}_{[0,T]}\big(\boldsymbol{x},\boldsymbol{W},(t_1,\cdots,t_{\mathcal{J}^I_b})\big),$$
$$\breve{\xi}=h^{(\mathcal{J}^I_b-1)|b}_{[0,T]}\Big(\varphi_b\big(\boldsymbol{\sigma}(\boldsymbol{0})\boldsymbol{w}_1\big),(\boldsymbol{w}_2,\cdots,\boldsymbol{w}_{\mathcal{J}^I_b}),(t_2-t_1,t_3-t_1,\cdots,t_{\mathcal{J}^I_b}-t_1)\Big).$$

If $\|\xi(t_1-)\|<\epsilon$ and $\|\boldsymbol{w}_j\|\leq\epsilon^{-\frac{1}{2\mathcal{J}^I_b}}\ \forall j\in[\mathcal{J}^I_b]$, then

$$\sup_{t\in[t_1,t_{\mathcal{J}^I_b}]}\Big\|\xi(t)-\breve{\xi}(t-t_1)\Big\|\leq\Big(2\exp\big(D(t_{\mathcal{J}^I_b}-t_1)\big)\cdot D\Big)^{\mathcal{J}^I_b+1}\cdot\epsilon,$$

where $D\geq1$ is the constant in Assumption 2.

(d) Let $\bar{c}\in(0,1)$ be the constant fixed in (4.19). Given $\Delta>0$, there exists $\epsilon_0=\epsilon_0(\Delta)\in(0,\bar{\epsilon})$ such that the following claim holds: given $T>0$, $\boldsymbol{x}\in\mathbb{R}^m$, $\boldsymbol{W}=(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{\mathcal{J}^I_b})\in\mathbb{R}^{d\times\mathcal{J}^I_b}$, $(t_1,\cdots,t_{\mathcal{J}^I_b})\in(0,T]^{\mathcal{J}^I_b\uparrow}$, if $\|\boldsymbol{x}\|\leq\epsilon_0$ and $\max_{j\in[\mathcal{J}^I_b]}\|\boldsymbol{w}_j\|\leq\epsilon_0^{-\frac{1}{2\mathcal{J}^I_b}}$, then

$$\xi(t)\notin I_{\bar{c}\bar{\epsilon}}\text{ or }\breve{\xi}(t)\notin I_{\bar{c}\bar{\epsilon}}\text{ for some }t\in[t_1,T-t_1]\qquad\Longrightarrow\qquad\sup_{t\in[t_1,t_{\mathcal{J}^I_b}]}\Big\|\breve{\xi}(t-t_1)-\xi(t)\Big\|<\Delta,$$

where ξ and $\breve{\xi}$ are defined as in part (c).

Proof. Before the proof of the claims, we highlight two facts. First, Assumption 2 and I being a bounded set (so I^- is compact) imply the existence of $C \in (0,\infty)$ such that $\sup_{\boldsymbol{x}\in I^-} \|\boldsymbol{a}(\boldsymbol{x})\| \vee \|\boldsymbol{\sigma}(\boldsymbol{x})\| \leq C$. Without loss of generality, in the statement of Lemma D.2 we pick some $C \geq 1$. Next, one can see that the validity of all claims do not depend on the values of $\boldsymbol{\sigma}(\cdot)$ and $\boldsymbol{a}(\cdot)$ outside of I^-. Therefore, throughout this proof below we w.l.o.g. assume that

$$\|\boldsymbol{a}(\boldsymbol{x})\| \vee \|\boldsymbol{\sigma}(\boldsymbol{x})\| \leq C \qquad \forall \boldsymbol{x} \in \mathbb{R}^m. \tag{D.10}$$

for some $C \in [1,\infty)$. That is, we impose the boundedness condition in Assumption 3.

(a) Arbitrarily pick some $T > 0$ and $\xi \in \mathbb{D}^{(\mathcal{J}_b^I-1)|b}_{\bar{B}_{\bar{\epsilon}}}[0,T](\bar{\epsilon})$. To lighten notations, in the proof of part (a) we write $k = J_b^I$. By the definition of $\mathbb{D}^{(k-1)|b}_A(\epsilon)$ in (2.25), there are some $\boldsymbol{x}$ with $\|\boldsymbol{x}\| \leq \bar{\epsilon}$, some $(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{k-1}) \in \mathbb{R}^{d\times k-1}$, some $(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1}) \in \mathbb{R}^{m\times k-1}$ with $\max_{j\in[k-1]}\|\boldsymbol{v}_j\| \leq \bar{\epsilon}$, and $0 < t_1 < t_2 < \cdots < t_{k-1} < \infty$ such that

$$\xi = \bar{h}^{(k-1)|b}_{[0,T]}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{k-1}),(\boldsymbol{v}_1,\cdots,\boldsymbol{v}_{k-1}),(t_1,\cdots,t_{k-1})\big).$$

Given any $t \in [0,T]$, Let $j^* = j^*(t) = \max\{j = 0,1,\cdots,k-1 : \ t_j \leq t\}$. By definition of the mapping $\bar{h}^{(k-1)|b}_{[0,T]}$ in (2.21)–(2.23), we have $\xi(t) = \boldsymbol{y}_{t-t_{j^*}}\big(\xi(t_{j^*})\big)$ where $\boldsymbol{y}_\cdot(\boldsymbol{x})$ is the ODE under the vector field $\boldsymbol{a}(\cdot)$; see (2.28). By the definition of $\mathcal{G}^{(k)|b}(\epsilon)$ and $\bar{\mathcal{G}}^{(k)|b}$ in (2.32), (4.11), we then yield $\xi(t) \in \bar{\mathcal{G}}^{(k-1)|b}(2\bar{\epsilon})$. However, by property (4.15), we must have

$$\bar{\mathcal{G}}^{(k-1)|b}(2\bar{\epsilon}) \subseteq I^-_{2\bar{\epsilon}} \subseteq I_{\bar{\epsilon}}. \tag{D.11}$$

and hence $\xi(t) \in \bar{\mathcal{G}}^{(k-1)|b}(2\bar{\epsilon}) \subseteq I_{2\bar{\epsilon}}$. This concludes the proof.

(b) For simplicity, in the proof of part (b) we write $k = \mathcal{J}_b^I$. For claim (i), note that due to $\|\boldsymbol{x}_0\| \leq \bar{\epsilon}$, we have $\boldsymbol{x}_0 + \varphi_b(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_0) \in \mathcal{G}^{(1)|b}(2\bar{\epsilon})$. Moreover, for all $n = 0,1,\cdots,k-2$ (recall our convention of $t_0 = 0$), for the cadlag path ξ defined in (D.9) we have $\xi(t_n) \in \mathcal{G}^{(n+1)|b}(2\bar{\epsilon}) \subseteq \mathcal{G}^{(k-1)|b}(2\bar{\epsilon})$. As a result, for all $t \in [0,t_{k-1})$ we have $\xi(t) \in \bar{\mathcal{G}}^{(k-1)|b}(2\bar{\epsilon}) \subseteq I_{2\bar{\epsilon}}$ due to (D.11). This verifies claim (i).

For claim (ii), we proceed with a proof by contradiction and suppose that $\xi(t_{k-1}) \in I_{\bar{\epsilon}}$. By (4.19), we then get $\xi(t) = \boldsymbol{y}_{t-t_{k-1}}\big(\xi(t_{k-1})\big) \in I_{\bar{c}\bar{\epsilon}}$ for all $t \in [t_{k-1},T]$. Together with claim (i), we arrive at the contradiction that $\xi(t) \in I_{\bar{c}\bar{\epsilon}}$ for all $t \in [0,T]$.

For claim (iii), the fact $\|\xi(t_{k-1})\| \geq \bar{\epsilon}$ follows directly from claim (ii) and (4.13). For any $j = 1,\cdots,k-1$ and any $t \in [t_{j-1},t_j)$, we proceed with a proof by contradiction and suppose that $\|\xi(t)\| \leq \bar{\epsilon}$. Then we have $\|\xi(t_j-)\| \leq \bar{\epsilon}$ due to (4.14), and hence $\xi(t_j) \in \mathcal{G}^{(1)|b}(2\bar{\epsilon})$. As a result, we arrive at the contradiction that $\xi(t_{k-1}) \in \mathcal{G}^{(k-1)|b}(2\bar{\epsilon}) \subseteq I_{\bar{\epsilon}}$, due to $\mathcal{G}^{(k-1)|b}(2\bar{\epsilon}) \subseteq \bar{\mathcal{G}}^{(k-1)|b}(2\bar{\epsilon})$ and (D.11). This concludes the proof of claim (iii).

We prove claim (iv) for $\bar{t} \triangleq k \cdot \boldsymbol{t}(\bar{\epsilon}/2)$ where $\boldsymbol{t}(\epsilon)$ is defined in (4.16). Consider the following proof by contradiction. If $t_{k-1} \geq \bar{t} = (k-1)\cdot \boldsymbol{t}(\bar{\epsilon}/2)$, then there must be some $j = 1,2,\cdots,k-1$ such that $t_j - t_{j-1} \geq \bar{t}(\epsilon/2)$. By claim (i), we have $\xi(t_{j-1}) \in I^-_{2\bar{\epsilon}} \subseteq I_{\bar{\epsilon}/2}$. Using the property (4.17), we yield $\xi(t_j-) = \lim_{t\uparrow t_j}\xi(t) \in \bar{B}_{\bar{\epsilon}/2}(\boldsymbol{0})$, which implies $\|\xi(t)\| < \epsilon$ for all t less than but close enough to t_j and contradicts claim (iii). This concludes the proof of claim (iv).

Lastly, we prove claim (v) for $\bar{\delta} > 0$ small enough such that

$$\exp(D\bar{t}) \cdot C\bar{\delta} < \bar{\epsilon}, \qquad C\bar{\delta} < b,$$

where $D \geq 1$ is the Lipschitz coefficient in Assumption 2 and $C \geq 1$ is the constant in (D.10). Again, we consider a proof by contradiction. Suppose that for the cadlag path ξ in (D.9) there is some $j = 0,1,\cdots,k-1$ such that $\|\boldsymbol{w}_j\| < \bar{\delta}$. First, we consider the case where $j \leq k-2$. Then note that (for the proof of claim (v), we interpret $\xi(0-)$ as $\boldsymbol{x}_0$ while, by definition, $\xi(0) = \boldsymbol{x}_0 + \varphi_b\big(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_0\big)$), we have

$$\xi(t_j) - \xi(t_j-) = \varphi_b\big(\boldsymbol{\sigma}\big(\xi(t_j-)\boldsymbol{w}_j\big),$$

and hence $\|\xi(t_j) - \xi(t_j-)\| \le C\bar{\delta}$. By Gronwall's inequality, we then get

$$\left\|\boldsymbol{y}_{t-t_j}\big(\xi(t_j-)\big) - \xi(t)\right\| \le \exp\big(D(t-t_j)\big)\cdot C\bar{\delta} \qquad \forall t \in [t_j, t_{j+1}).$$

Recall that we currently focus on the case where $j \le k-2$. By claim (iv) and our choice of $\bar{\delta}$, we get $\exp\big(D(t-t_j)\big)\cdot C\bar{\delta} \le \exp\big(D\bar{t}\big)\cdot C\bar{\delta} < \bar{\epsilon}$ in the display above. This implies the existence of some $\xi' \in \mathbb{D}^{(k-1)|b}_{\bar{B}_{\bar{\epsilon}}(\mathbf{0})}(\bar{\epsilon})$ such that $\sup_{t\in[0,T]}\|\xi(t) - \xi'(t)\| < \bar{\epsilon}$. However, by results in part (a), we must have $\xi'(t) \in I^-_{2\bar{\epsilon}}$ $\forall t \in [0,T]$, which leads to $\xi(t) \in I^-_{\bar{\epsilon}}$ $\forall t \in [0,T]$. This contradicts the running assumption of part (b) that $\xi(t) \notin I_{\bar{c}\bar{\epsilon}}$ for some $t \in [0,T]$, and allows us to conclude the proof of claim (v) for the cases where $j \le k-2$. In case that $j = k-1$, by claim (i) we have $\xi(t_{k-1}-) = \lim_{t\uparrow t_{k-1}}\xi(t) \in I^-_{2\bar{\epsilon}}$. Meanwhile, by definition of the mapping $\bar{h}^{(k-1)|b}_{[0,T]}$, we have $\xi(t_{k-1}) = \xi(t_{k-1}-) + \varphi\Big(\boldsymbol{\sigma}\big(\xi(t_{k-1}-)\big)\boldsymbol{w}_{k-1}\Big)$. By $\|\boldsymbol{w}_{k-1}\| < \bar{\delta}$ and our choice of $\bar{\delta}$ above, we have $\Big\|\varphi\Big(\boldsymbol{\sigma}\big(\xi(t_{k-1}-)\big)\boldsymbol{w}_{k-1}\Big)\Big\| < \bar{\epsilon}$ and hence $\xi(t_{k-1}) \in I_{\bar{\epsilon}}$. Due to the contradiction with claim (ii), we conclude the proof.

(c) The proof is almost identical to that of Lemma 3.9 based on an inductive argument. We omit the details to avoid repetition.

(d) Let $\bar{t}$ be the constant specified in part (b). We claim that: if $\xi(t) \notin I_{\bar{c}\bar{\epsilon}}$ or $\breve{\xi}(t) \notin I_{\bar{c}\bar{\epsilon}}$ for some $t \in [0,T]$, then

$$\sup_{t\in[t_1, t_{\mathcal{J}^I_b}]}\left\|\breve{\xi}(t-t_1) - \xi(t)\right\| < \underbrace{\Big(2\exp\big(D\bar{t}\big)\cdot D\Big)^{\mathcal{J}^I_b+1}}_{\triangleq \rho^*}\cdot\epsilon_0 \qquad \forall \epsilon_0 \in (0,\bar{\epsilon}]. \tag{D.12}$$

As a result, claims of part (d) hold for any $\epsilon_0 \in (0,\bar{\epsilon})$ small enough such that $\rho^*\epsilon_0 < \Delta$. Now, it only remains to prove claim (D.12). Due to $\|\boldsymbol{x}\| = \|\xi(0)\| < \epsilon_0$ and (4.14), we have $\|\xi(t_1-)\| \le \epsilon_0$. This allows us to apply results in part (c) and get (recall our choice of $T = t_{\mathcal{J}^*_b} + 1$)

$$\sup_{t\in[t_1, t_{\mathcal{J}^I_b}]}\left\|\xi(t) - \breve{\xi}(t-t_1)\right\| \le \Big(2\exp\big(D(t_{\mathcal{J}^I_b} - t_1)\big)\cdot D\Big)^{\mathcal{J}^I_b+1}\cdot\epsilon_0,$$

Lastly, if $\xi(t) \notin I_{\bar{c}\bar{\epsilon}}$ for some $t \in [t_1,T]$, then $t_{\mathcal{J}^I_b} - t_1 < \bar{t}$ by claim (iv) of part (b). Likewise, if $\breve{\xi}(t) \notin I_{\bar{c}\bar{\epsilon}}$ for some $t \in [0,T]$, then we get $t_{\mathcal{J}^I_b} < \bar{t}$. In both cases, we get $t_{\mathcal{J}^*_b} - t_1 \le \bar{t}$. This concludes the proof. $\square$

The next lemma studies the mass the measure $\breve{\mathbf{C}}^{(k)|b}$ charges on the boundary of the domain I.

Lemma D.3. *Under Assumptions 2 and 4,* $\breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(I^c) < \infty$.

Proof. Let $\bar{\epsilon} > 0$ be such that the conditions in (4.13)–(4.15) hold. Let $\bar{t}$ and $\bar{\delta}$ be the constants characterized in Lemma D.2. Observe that (we write $\mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b})$)

$$\begin{aligned}
&\breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}\big(I^c\big)\\
&= \int \mathbb{I}\Big\{\breve{g}^{(\mathcal{J}^I_b-1)|b}\big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b}), (t_1, \cdots, t_{\mathcal{J}^I_b-1})\big) \notin I\Big\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad \big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b-1\uparrow}_\infty(dt_1, \cdots, dt_{\mathcal{J}^I_b-1})\\
&= \int \mathbb{I}\Big\{h^{(\mathcal{J}^I_b-1)|b}_{[0,1+t_{\mathcal{J}^I_b-1}]}\big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b}), (t_1, \cdots, t_{\mathcal{J}^I_b-1})\big)(t_{\mathcal{J}^I_b-1}) \notin I\Big\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad \big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b-1\uparrow}_\infty(dt_1, \cdots, dt_{\mathcal{J}^I_b-1})
\end{aligned}$$

$$\leq \int \mathbb{I}\Big\{ \ \|\boldsymbol{w}_j\| > \bar{\delta}\ \forall j \in [\mathcal{J}^I_b];\ t_{\mathcal{J}^I_b - 1} < \bar{t}\ \Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b - 1\uparrow}_\infty(dt_1, \cdots, dt_{\mathcal{J}^I_b - 1})$$

by part (b) of Lemma D.2

$$\leq \bar{t}^{\mathcal{J}^I_b - 1}/\bar{\delta}^{\alpha \mathcal{J}^I_b} < \infty.$$

This concludes the proof. □

To conclude, we provide the proof of Lemma 4.2.

Proof of Lemma 4.2. Let $\bar{c} \in (0,1)$ be the constant fixed in (4.19). By part (e) of Lemma D.2, for the fixed $\Delta \in (0, \bar{\epsilon})$, we are able to fix some $\epsilon_0 \in (0, \frac{\Delta}{2} \wedge \bar{c}\bar{\epsilon})$ such that the following claim holds: given $T > 0$, $\boldsymbol{x} \in \mathbb{R}^m$, $\mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b}) \in \mathbb{R}^{d \times \mathcal{J}^I_b}$, $(t_1, \cdots, t_{\mathcal{J}^I_b}) \in (0,T]^{\mathcal{J}^I_b \uparrow}$, if $\|\boldsymbol{x}\| \leq \epsilon_0$ and $\max_{j \in [\mathcal{J}^I_b]} \|\boldsymbol{w}_j\| \leq \epsilon_0^{-\frac{1}{2\mathcal{J}^I_b}}$, then

$$\xi(t) \notin I_{\bar{c}\bar{\epsilon}} \text{ or } \breve{\xi}(t) \notin I_{\bar{c}\bar{\epsilon}} \text{ for some } t \in [t_1, T - t_1] \quad \Longrightarrow \quad \sup_{t \in [t_1, t_{\mathcal{J}^I_b}]} \Big\| \breve{\xi}(t - t_1) - \xi(t) \Big\| < \Delta, \tag{D.13}$$

where

$$\xi = h^{(\mathcal{J}^I_b)|b}_{[0,T]}\big(\boldsymbol{x}, \mathbf{W}, (t_1, \cdots, t_{\mathcal{J}^I_b})\big),$$
$$\breve{\xi} = h^{(\mathcal{J}^I_b - 1)|b}_{[0,T]}\Big(\varphi_b\big(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1\big), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b}), (t_2 - t_1, t_3 - t_1, \cdots, t_{\mathcal{J}^I_b} - t_1)\Big).$$

Henceforth in the proof, we fix some $\epsilon \in (0, \epsilon_0]$ and $B \subseteq (I_\epsilon)^c$. Due to our choice of $\epsilon \leq \epsilon_0 < \bar{c}\bar{\epsilon}$, we have $B \subseteq (I_{\bar{c}\bar{\epsilon}})^c$. To prove the lower bound, let

$$\widetilde{E} = \Big\{ \xi \in \mathbb{D}[0,T] :\ \exists t \in [0,T] \text{ s.t. } \xi(t) \in B_{\Delta/2},\ \xi(s) \in I_{2\epsilon}\ \forall s \in [0,t) \Big\}.$$

For any $\xi \in \widetilde{E}$ and any ξ' with $\boldsymbol{d}^{[0,T]}_{J_1}(\xi, \xi') < \epsilon$, due to $\epsilon \leq \epsilon_0 < \Delta/2$, there must be some $t' \in [0,T]$ such that $\xi'(t') \in B$ and $\xi'(s) \in I_\epsilon\ \forall s \in [0, t')$. This implies that $\xi' \in \breve{E}(\epsilon, B, T)$, and hence

$$\widetilde{E} \subseteq \Big(\breve{E}(\epsilon, B, T)\Big)_\epsilon \subseteq \Big(\breve{E}(\epsilon, B, T)\Big)^\circ.$$

Therefore, for any $\boldsymbol{x} \in \mathbb{R}^m$ with $\|\boldsymbol{x}\| \leq \epsilon \leq \epsilon_0$,

$$\begin{aligned}
\mathbf{C}^{(\mathcal{J}^I_b)|b}_{[0,T]}\bigg(\Big(\breve{E}(\epsilon, B, T)\Big)^\circ;\ \boldsymbol{x}\bigg) &\geq \int \mathbb{I}\Big\{ h^{(\mathcal{J}^I_b)|b}_{[0,T]}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t}) \in \widetilde{E} \Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b \uparrow}_T(d\boldsymbol{t}) \\
&= \int \widetilde{\phi}_B(t_1, \boldsymbol{x}) \mathcal{L}_T(dt_1),
\end{aligned} \tag{D.14}$$

where $\mathcal{L}_T$ is the Lebesgue measure on $(0,T)$, $\mathcal{L}^{k\uparrow}_T$ is the k-fold ofq Lebesgue measure restricted on $\{(t_1, \cdots, t_k) \in (0,T)^k :\ t_1 < t_2 < \cdots < t_k\}$, and

$$\begin{aligned}
\widetilde{\phi}_B(t_1, \boldsymbol{x}) = \int \mathbb{I}\bigg\{ &\exists t \in [0,T] \text{ s.t. } h^{(\mathcal{J}^I_b)|b}_{[0,T]}\Big(\boldsymbol{x}, \mathbf{W}, (t_1, t_1 + u_2, t_1 + u_3, \cdots, t_1 + u_{\mathcal{J}^I_b})\Big)(t) \in B_{\Delta/2} \\
&\text{and } h^{(\mathcal{J}^I_b)|b}_{[0,T]}\Big(\boldsymbol{x}, \mathbf{W}, (t_1, t_1 + u_2, t_1 + u_3, \cdots, t_1 + u_{\mathcal{J}^I_b})\Big)(s) \in I_{2\epsilon}\ \forall s \in [0,t) \bigg\} \\
&\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b - 1\uparrow}_{T - t_1}(du_2, \cdots, du_{\mathcal{J}^I_b}).
\end{aligned}$$

Set $\boldsymbol{x}_0 = \lim_{t\uparrow t_1} \boldsymbol{y}_t(\boldsymbol{x})$, and note that

$$h^{(\mathcal{J}_b^I)|b}_{[0,T]}\Big(\boldsymbol{x}, (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (t_1, t_1+u_2, t_1+u_3, \cdots, t_1+u_{\mathcal{J}_b^I})\Big)(t_1+s)$$
$$= h^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\boldsymbol{x}_0 + \varphi_b(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, u_3, \cdots, u_{\mathcal{J}_b^I})\Big)(s) \qquad \forall s \in [0, T-t_1].$$

Therefore, for any $t_1 \in [0, T-\bar{t}]$ and $\boldsymbol{x}$ with $\|\boldsymbol{x}\| \leq \epsilon$, by property (4.14) we have $\|\boldsymbol{x}_0\| \leq \epsilon \leq \epsilon_0 \leq \Delta/2$, and

$$\begin{aligned}
&\widetilde{\phi}_B(t_1, x) \\
&\geq \inf_{\boldsymbol{x}_0:\ \|\boldsymbol{x}_0\| \leq \frac{\Delta}{2}} \int \mathbb{I}\Big\{\exists t \in [0, T-t_1] \text{ s.t. } h^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\boldsymbol{x}_0 + \varphi_b(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, \cdots, u_{\mathcal{J}_b^I})\Big)(t) \in B_{\Delta/2} \\
&\qquad \text{and } h^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\boldsymbol{x}_0 + \varphi_b(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, \cdots, u_{\mathcal{J}_b^I})\Big)(s) \in I_{2\epsilon}\ \forall s \in [0, t)\Big\} \\
&\qquad\qquad \big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}_b^I}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}_b^I-1\uparrow}_{T-t_1}(du_2, \cdots, du_{\mathcal{J}_b^I}) \\
&= \inf_{\boldsymbol{x}_0:\ \|\boldsymbol{x}_0\| \leq \frac{\Delta}{2}} \int \mathbb{I}\Big\{h^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\boldsymbol{x}_0 + \varphi_b(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, \cdots, u_{\mathcal{J}_b^I})\Big)(u_{\mathcal{J}_b^I}) \in B_{\Delta/2};\ \min_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| > \bar{\delta}\Big\} \\
&\qquad\qquad \big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}_b^I}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}_b^I-1\uparrow}_{T-t_1}(du_2, \cdots, du_{\mathcal{J}_b^I}) \\
&\qquad \text{by claims } (i), (ii), \text{ and } (v) \text{ in part } (b) \text{ of Lemma D.2} \\
&\geq \inf_{\boldsymbol{x}_0:\ \|\boldsymbol{x}_0\| \leq \frac{\Delta}{2}} \int \mathbb{I}\Big\{h^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\boldsymbol{x}_0 + \varphi_b(\boldsymbol{\sigma}(\boldsymbol{x}_0)\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, \cdots, u_{\mathcal{J}_b^I})\Big)(u_{\mathcal{J}_b^I}) \in B_{\Delta/2}; \\
&\qquad\qquad \min_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| > \bar{\delta},\ \max_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| \leq \epsilon_0^{-\frac{1}{2\mathcal{J}_b^I}}\Big\} \\
&\qquad\qquad \big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}_b^I}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}_b^I-1\uparrow}_{T-t_1}(du_2, \cdots, du_{\mathcal{J}_b^I}) \\
&\geq \int \mathbb{I}\Big\{h^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, \cdots, u_{\mathcal{J}_b^I})\Big)(u_{\mathcal{J}_b^I}) \in B_{\Delta}; \\
&\qquad\qquad \min_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| > \bar{\delta},\ \max_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| \leq \epsilon_0^{-\frac{1}{2\mathcal{J}_b^I}}\Big\} \\
&\qquad\qquad \big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}_b^I}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}_b^I-1\uparrow}_{T-t_1}(du_2, \cdots, du_{\mathcal{J}_b^I}) \\
&\qquad \text{by property (D.13)} \\
&= \int \mathbb{I}\Big\{\breve{g}^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, \cdots, u_{\mathcal{J}_b^I})\Big) \in B_{\Delta}; \\
&\qquad\qquad \min_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| > \bar{\delta},\ \max_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| \leq \epsilon_0^{-\frac{1}{2\mathcal{J}_b^I}}\Big\} \\
&\qquad\qquad \big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}_b^I}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}_b^I-1\uparrow}_{T-t_1}(du_2, \cdots, du_{\mathcal{J}_b^I}) \\
&\qquad \text{by the definition of } \breve{g}^{(k)|b} \text{ in (2.31)} \\
&= \int \mathbb{I}\Big\{\breve{g}^{(\mathcal{J}_b^I-1)|b}_{[0,T-t_1]}\Big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}_b^I}), (u_2, \cdots, u_{\mathcal{J}_b^I})\Big) \in B_{\Delta}; \\
&\qquad\qquad \min_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| > \bar{\delta},\ \max_{j \in [\mathcal{J}_b^I]} \|\boldsymbol{w}_j\| \leq \epsilon_0^{-\frac{1}{2\mathcal{J}_b^I}}\Big\} \\
&\qquad\qquad \big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^{\mathcal{J}_b^I}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}_b^I-1\uparrow}_{\bar{t}}(du_2, \cdots, du_{\mathcal{J}_b^I})
\end{aligned}$$

by claim (v) in part (b) of Lemma D.2

$$\geq \int \mathbb{I}\Big\{ \breve{g}^{(\mathcal{J}^I_b - 1)|b}_{[0,T-t_1]}\Big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b}), (u_2, \cdots, u_{\mathcal{J}^I_b})\Big) \in B_\Delta;\ \min_{j \in [\mathcal{J}^I_b]} \|\boldsymbol{w}_j\| > \bar{\delta} \Big\}$$

$$\Big((\nu_\alpha \times \mathbf{S}) \circ \Phi\Big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b - 1\uparrow}_{\bar{t}}(du_2, \cdots, du_{\mathcal{J}^I_b})$$

$$- \int \mathbb{I}\Big\{ \min_{j \in [\mathcal{J}^I_b]} \|\boldsymbol{w}_j\| > \bar{\delta},\ \max_{j \in [\mathcal{J}^I_b]} \|\boldsymbol{w}_j\| > \epsilon_0^{-\frac{1}{2\mathcal{J}^I_b}} \Big\} \Big((\nu_\alpha \times \mathbf{S}) \circ \Phi\Big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b - 1\uparrow}_{\bar{t}}(du_2, \cdots, du_{\mathcal{J}^I_b}).$$

We focus on the two integrals one the RHS of the last inequality in the display above. It is easy to see that the latter is upper bounded by

$$\breve{\boldsymbol{c}}(\epsilon_0) = \mathcal{J}^I_b \cdot (\bar{t})^{\mathcal{J}^I_b - 1} \cdot (\bar{\delta})^{-\alpha \cdot (\mathcal{J}^I_b - 1)} \cdot \epsilon_0^{\frac{\alpha}{2\mathcal{J}^I_b}}.$$

As for the former, using part (b) of Lemma D.2 and the fact that $B_\Delta \subseteq B \subseteq (I_{\bar{c}\bar{\epsilon}})^c$ again, we yield

$$\int \mathbb{I}\Big\{ \breve{g}^{(\mathcal{J}^I_b - 1)|b}_{[0,T-t_1]}\Big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b}), (u_2, \cdots, u_{\mathcal{J}^I_b})\Big) \in B_\Delta;\ \min_{j \in [\mathcal{J}^I_b]} \|\boldsymbol{w}_j\| > \bar{\delta} \Big\}$$

$$\Big((\nu_\alpha \times \mathbf{S}) \circ \Phi\Big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b - 1\uparrow}_{\bar{t}}(du_2, \cdots, du_{\mathcal{J}^I_b})$$

$$= \int \mathbb{I}\Big\{ \breve{g}^{(\mathcal{J}^I_b - 1)|b}_{[0,T-t_1]}\Big(\varphi_b(\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w}_1), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_{\mathcal{J}^I_b}), (u_2, \cdots, u_{\mathcal{J}^I_b})\Big) \in B_\Delta \}$$

$$\Big((\nu_\alpha \times \mathbf{S}) \circ \Phi\Big)^{\mathcal{J}^I_b}(d\mathbf{W}) \times \mathcal{L}^{\mathcal{J}^I_b - 1\uparrow}_{\infty}(du_2, \cdots, du_{\mathcal{J}^I_b})$$

$$= \breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(B_\Delta).$$

In summary, for any $\boldsymbol{x} \in \mathbb{R}^m$ with $\|\boldsymbol{x}\| \leq \epsilon$ and $t_1 \in [0, T - \bar{t}]$, we have shown that

$$\widetilde{\phi}_B(t_1, \boldsymbol{x}) \geq \breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(B_\Delta) - \breve{\boldsymbol{c}}(\epsilon_0).$$

Together with the trivial bound that $\widetilde{\phi}_B(t_1, \boldsymbol{x}) \geq 0$ for all $t_1 > T - \bar{t}$, we have in (D.14) that

$$\mathbf{C}^{(\mathcal{J}^I_b)|b}_{[0,T]}\bigg(\Big(\breve{E}(\epsilon, B, T)\Big)^\circ;\ \boldsymbol{x} \bigg) \geq (T - \bar{t}) \cdot \Big(\breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(B_\Delta) - \breve{\boldsymbol{c}}(\epsilon_0) \Big)$$

for all $\boldsymbol{x} \in \mathbb{R}^m$ with $\|\boldsymbol{x}\| \leq \epsilon$. This concludes the proof of the lower bound. The proof to the upper bound is almost identical, so we omit the details here. □