Submitted to the Annals of Applied Probability

LARGE DEVIATIONS AND METASTABILITY ANALYSIS FOR HEAVY-TAILED DYNAMICAL SYSTEMS

BY XINGYU WANG[1,a] AND CHANG-HAN RHEE[1,b]

[1]*Industrial Engineering and Management Sciences, Northwestern University*
[a]*xingyuwang2017@u.northwestern.edu*; [b]*chang-han.rhee@northwestern.edu*

This paper introduces novel frameworks for large deviations and metastability analysis in heavy-tailed stochastic dynamical systems. We develop and apply these frameworks within the context of stochastic difference equation $X^{\eta}_{j+1}(x) = X^{\eta}_{j}(x) + \eta a\big(X^{\eta}_{j}(x)\big) + \eta\sigma\big(X^{\eta}_{j}(x)\big)Z_{j+1}$ 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 $\varphi_b(x) = (x/\left\|x\right\|)\max\{\left\|x\right\|, b\}$. The truncation operator $\varphi_b(\cdot)$ is often introduced as a modulation mechanism in heavy-tailed systems, such as stochastic gradient descent algorithms in deep learning. We establish locally uniform sample-path large deviations for both processes and translate these asymptotics into precise characterizations of the joint distributions of the first exit times and exit locations. Our large deviations asymptotics are sharp enough to rigorously characterize *the catastrophe principle* by establishing the distributional limit of the sample paths conditional on the rare events of interest, thereby revealing the most likely paths through which rare events arise in heavy-tailed dynamical systems. Moreover the resulting limit theorem unveils a discrete hierarchy of phase transitions in the asymptotics as the truncation threshold b varies. Together, these developments lead to comprehensive heavy-tailed counterpart of the classical Freidlin-Wentzell theory. We present our results in the context of discrete time processes $X^{\eta}_{j+1}(x)$ and $X^{\eta|b}_{j+1}(x)$, as they more directly model the stochastic algorithms in deep learning that inspired this work. Nontheless, the same approach applies straightforwardly to continuous-time processes, and we include the corresponding results for the Lévy-driven SDEs in the appendix.

1. Introduction. Large deviations and metastability analysis in stochastic dynamical systems are deeply interconnected and have a rich history in probability theory and their applications. Since the pioneering works of Kramers and Eyring [32, 55, 41], which analyzed phase transitions in stochastic dynamical systems in the context of chemical reaction-rate theory, extensive theoretical advancements have been made. One of the most notable breakthroughs is the classical Freidlin-Wentzell theory [37, 38, 39], which introduced large deviations machinery to the analysis of exit times and global behaviors of small random perturbations of dynamical systems. Further extensions of this approach in the context of statistical physics were pioneered in [18] and described in detail in [68]. One of the key advantages of this approach—often called *the pathwise approach*—is its ability to describe in detail the scenarios that lead to phase transitions. In particular, the large deviations formalism at the sample-path level enables precise identification of the most likely paths out of the metastability sets. This ensures that, asymptotically, whenever the dynamical system escapes from the metastability set, the escape routes almost always closely resemble these most likely paths. However, the sample-path-level large deviations are typically available only in the form of logarithmic asymptotics, and hence, the asymptotic scale of the exit time can

MSC2020 subject classifications: Primary 00X00, 00X00; secondary 00X00.
Keywords and phrases: First keyword, second keyword.

be determined only up to its exponential rate, requiring different approaches to identify the prefactor. Another breakthrough is *the potential-theoretic approach* initiated in [12, 13, 14] and later summarized in [11]. Instead of relying on large deviations machinery, this approach leverages potential-theoretic tools: the scale of exit times for Markov processes can be expressed in terms of capacity, which, in turn, can be bounded using variational principles. The key advantage of this approach, compared to the pathwise approach, is that it is often possible to find test functions that tightly bound the capacity of the Markov chains, thereby yielding *precise* asymptotics—rather than merely logarithmic asymptotics as in the pathwise approach—of the scales of exit times. Although the potential theoretic approach does not provide extra information—such as the most likely paths—as the pathwise approach beyond the asymptotics of the exit times, its sharpness has inspired extensive research activity. The early works in the potential theoretic approach were focused on reversible Markov processes. However, recent developments have extended the scope of the approach to enable the analysis of non-reversible Markov processes; see, for example, [77, 56, 40, 57].

While these developments provide powerful means to understand rare events and metastability of light-tailed systems, heavy-tailed systems exhibit a fundamentally different large deviations and metastability behaviors and call for a different set of technical tools for successful analysis. For example, early foundational works in heavy-tailed context [47, 48, 49, 50] proved that the exit times of the stochastic processes driven by heavy-tailed noises scale polynomially with respect to the scaling parameter. These papers also reveal that the exit events are almost always driven by a single disproportionately large jump, while the rest of the system's behavior remains close to its nominal behavior. Here, nominal behavior refers to the functional law of large numbers limit of the scaled processes. Note that this is in stark contrast to the light-tailed counterparts, where the exit times scale geometrically, and the exit events are driven by smooth tilting of the entire system from its nominal behavior. One can view this as a manifestation of *the principle of a single big jump*, a well-known folklore in extreme value theory. For stochastic processes with independent increments over a finite time horizon, [45] systematically characterized the principle of a single big jump with an early formulation of heavy-tailed sample-path large deviations.

However, many heavy-tailed rare events in machine learning, finance, operations research, and other disciplines cannot be driven by a single big jump; see e.g. [1, 79, 34, 35, 81]. A notable example arises in the context of deep learning. Stochastic gradient descent (SGD) and its variants are the methods of choice in training deep neural networks (DNNs). Heavy-tailed SGDs have attracted significant attention in the recent past because of their ability to escape local minima with a single big jump, enabling them to explore non-convex loss landscapes within realistic training time horizons. Such ability is widely believed to have fundamental connection to DNNs' remarkable generalization performance on test data. However, the pure form of SGD is rarely employed in practice. In particular, when the gradient noise appears to exhibit heavy-tailed behavior causing SGD to occasionally attempt to travel a long distance in a single step, the step size is truncated at a threshold. This is a common practice known as gradient clipping; see, e.g., [31, 61, 42, 69, 83]. With gradient clipping, the exit event from a large attraction field cannot be solely driven by a single big jump. In general—as we rigorously confirm in this paper—when a single big jump is insufficient to cause the rare event of interest, it is driven by the minimal number of big jumps required to trigger it, while the rest of the system remains close to its nominal dynamics. This portrayal provides a more complete picture than the principle of a single big jump and is referred to as *the catastrophe principle*. More recently, a rigorous mathematical characterization of the catastrophe principle for Lévy processes and random walks was established in the form of heavy-tailed sample-path large deviations [75], leveraging the $\mathbb{M}$-convergence theory originally introduced in [59]. The results in [75] can be viewed as the heavy-tailed counterpart of the Mogulskii's theorem

[60, 62]. Notably, the new large deviations formulation in [75] provides precise asymptotics for heavy-tailed processes, in contrast to the logarithmic asymptotics of the classical large deviation principle; see (1.4). This raises the hope that, for heavy-tailed dynamical systems, it may be possible to simultaneously obtain *both* detailed descriptions of the scenarios leading to phase transitions (as in the pathwise approach [39, 68]) and sharp asymptotics for the exit time (as in the potential-theoretic approach [11]). Successfully implementing this strategy for practical systems requires establishing sufficiently strong sample-path large deviations and developing machinery to translate these results into exit-time analyses tailored for heavy-tailed dynamical systems with transition dynamics potentially modulated by truncation.

In this paper, we propose a new formulation of large deviations, along with systematic tools to establish them and translate them into exit-time asymptotics for heavy-tailed dynamical systems. Using this framework, we derive precise sample-path large deviations and sharp scaling limits of the joint distributions of the exit times and locations for heavy-tailed stochastic difference equations. In particular, we characterize the asymptotics of processes whose jump sizes are modulated by truncation; see (1.2) and (**??**) for the precise definitions. It turns out that such modulation introduces phase transition within phase transition: the polynomial rate of exit time's asymptotic scale changes discontinuously w.r.t. the truncation parameter, changing the way the exit events occur qualitatively; see Theorem 2.8 and the form of $\mathcal{J}_b^I$ in (2.27). This behavior sharply contrasts with the light-tailed counterpart, where truncation does not affect the large deviations behavior. This is another manifestation of the dichotomy between the catastrophe principle and conspiracy principle. In view of these, our results provide comprehensive heavy-tailed counterparts to the Freidlin–Wentzell theory for stochastic dynamical systems. 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 [59, 75]. Our large deviations characterize the catastrophe principle, which reveals a discrete hierarchy governing the causes and probabilities of a wide variety of rare events associated with heavy-tailed stochastic difference and differential equations; see Theorems 2.5, 2.6, **??**, and **??**. We also obtain the precise distributional limit of the scaled sample paths conditional on the rare events in Corollary 2.7 and **??**. In the second half of this paper, we focus on their implications for the exit-time (and exit-location) analysis. However, we emphasize that these results provide general, systematic tools for heavy-tailed rare-event analysis far beyond exit times.
- **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 these asymptotic atoms, we derive sharp asymptotics for the joint distribution of (scaled) exit-times and exit-locations for heavy-tailed processes, as detailed in Theorem 2.8 and Corollary 2.9. Notably, the scaling rate parameter reflects an intricate interplay between the truncation threshold and the geometry of the drift, which is a feature absent in both the principle of a single big jump regime (heavy-tailed systems without truncation) and the conspiracy principle regime (light-tailed systems).

A culmination of metastability analysis is the sharp characterization of the global dynamics of heavy-tailed processes, often established in the form of process-level convergence of scaled processes to simpler ones, such as Markov jump processes on a discrete state space; see, for example, [68, 11, 48, 49, 58, 74]. For unregulated processes such as (**??**), which are governed by the principle of a single big jump, it is well known that the scaling limit is a Markov jump process with a state space consisting of the local minima of the potential function; see e.g., [43, 48, 49]. In a companion paper [82], we demonstrate that the framework developed in the current paper is strong enough to extend the above-mentioned results to the systems *not* governed by the principle of a single big jump—such as (1.2) and (**??**)—within a multi-well potential, by identifying scaling limits and characterizing their global behavior at the process level. In particular, the scaling limit for the truncated heavy-tailed dynamics is a Markov jump process that *only visits the widest minima*. This is in sharp contrast to the untruncated cases [43, 48, 49] where the limiting Markov jump process visits all the local minima with certain fractions. As a result, the fraction of time such processes spend in narrow attraction fields converges to zero as the scaling parameter (often called learning rate in the machine learning literature) tends to zero. Precise characterization of such phenomena is of fundamental importance in understanding and further leveraging the curious effectiveness of the stochastic gradient descent (SGD) algorithms in training deep neural networks.

1.1. *Overview of the Paper.* 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 $\mathbb{R}^d$ such that $\mathbf{E}\boldsymbol{Z}_1 = \boldsymbol{0}$ 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

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

under initial condition $\boldsymbol{X}^\eta_0(\boldsymbol{x}) = \boldsymbol{x}$. 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 consider a truncated variation of $\boldsymbol{X}^\eta_{j+1}(\boldsymbol{x})$. Let $\varphi_b(\cdot)$ be the projection operator from $\mathbb{R}^m$ onto the closed ball centered at the origin with radius b. Define

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

under the initial condition $\boldsymbol{X}^{\eta|b}_0(\boldsymbol{x}) = \boldsymbol{x}$. 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, and the dynamics of $\boldsymbol{X}^\eta_j(\boldsymbol{x})$ is recovered by setting the truncation threshold b as ∞. As mentioned above, such dynamics arise in the training of DNNs, and their global behaviors are closely connected to the performance of the trained models. In particular, if $\boldsymbol{a}$ is the negative gradient of the training loss f, then the argument of $\varphi_b(\cdot)$ in (1.2), $\eta\boldsymbol{a}\big(\boldsymbol{X}^\eta_j(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^\eta_j(\boldsymbol{x})\big)\boldsymbol{Z}_{j+1} = -\eta\big(\nabla f\big(\boldsymbol{X}^\eta_j(\boldsymbol{x})\big) - \boldsymbol{\sigma}\big(\boldsymbol{X}^\eta_j(\boldsymbol{x})\big)\boldsymbol{Z}_{j+1}\big)$, represents the state-dependent stochastic gradient of f at $\boldsymbol{X}^\eta_j(\boldsymbol{x})$, scaled by the negative learning rate $-\eta$, which corresponds to the one-step displacement of SGD. Therefore, (1.1) and (1.2) serve as models for the dynamics of heavy-tailed SGD and its variation with gradient clipping, respectively. See, for example, [81, 69, 83, 54] and the references therein for more details. Note that (1.1) and (1.2) can be interpreted as discretizations of small-noise SDEs driven by Lévy processes. In this paper, we primarily focus on these discrete-time processes, as they provide more accurate models of the stochastic algorithms in deep learning compared to the continuous counterparts. In particular, note that approximating SGDs with SDEs is meaningful only if Z_i's are α-stable. In contrast, (1.1) and (1.2) do not require the Z_i's to be α-stable for modeling SGDs and impose no restrictions on the specific choice of their distribution. We also emphasize that 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, with a straightforward adaptation of the machinery we develop here. We present the results for Lévy-driven SDEs in Appendix **??**. Finally, note also that although (1.1) and (1.2) are probably the most natural scaling regime in many contexts, more general scaling can be considered as well. In Appendix **??**, we present the corresponding results for

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

with some $\gamma > 0$ and under intial conditions $\boldsymbol{X}^{\eta}_0(\boldsymbol{x}) = \boldsymbol{X}^{\eta|b}_0(\boldsymbol{x}) = \boldsymbol{x}$.

At the crux of the problem studied in this paper is a fundamental distinction 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 classical large deviations principle (LDP) framework [24, 26, 33, 80] has been wildly successful in providing systematic tools for studying rare events. In particular, the sample-path large deviation principle rigorously characterize the conspiracy principle. Notable developments include the Mogulskii's theorem [60, 62], the Freidline and Wentzell theory [37, 38, 39], and various extensions for discrete-time processes [64, 53] for finite dimensional processes under relaxed assumptions [22, 28, 27, 2, 29], and for infinite dimensional processes [15, 16, 78, 19, 63].

On the other hand, due to the fundamental differences in the way rare events arise, sample-path large deviations for heavy-tailed processes has been developed much later. Instead, 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 [65, 66], and [45] confirmed the principle of a single big jump systematically at the sample-path level for random walks. The summary of the subsequent developments in the context of processes with independent increments can be found in, for example, [9, 25, 30, 36]. More recently, [75] established a general catastrophe principle for regularly varying Lévy processes and random walks, which goes beyond the principle of a single big jump and characterizes the rare events driven by any number of catastrophes. For example, let $\mathbb{D}\big([0,1],\mathbb{R}\big)$ be the space of real-valued càdlàg functions over $[0,1]$, $S_j \triangleq Z_1 + \cdots + Z_j$ be a mean-zero random walk, and $\boldsymbol{S}^n \triangleq \{S^n_{\lfloor nt \rfloor}/n : t \in [0,1]\}$ be the scaled sample path. Under regularly varying Z_i's, the sample path large deviations established in [75] takes the following form for "general" $B \in \mathbb{D}\big([0,1],\mathbb{R}\big)$,

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

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 supported on the set of step functions with k jumps, and B° and B^{-} are the interior and closure of B, respectively. Here, the index 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 [6] for analogous results 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, it 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 [20]. Moreover, [75, 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, by relaxing the upper bound of the standard LDP, an alternative formulation known as "extended LDP" was proposed in [10], and such a formulation is also feasible for heavy-tailed processes; see, for example, [8, 3, 4]. 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 [20]. We will also see that exact asymptotics are crucial in Section 2.3 and Section 4 of this paper for sharp exit time and exit location analysis. In fact, it demands an even stronger formulation than (1.4), which we will introduce in (1.5) shortly. Below, we describe the main contributions of this paper.

Large Deviations for Heavy-Tailed Dynamical Systems. Our first contribution 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 $\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}) \triangleq \{\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) :\ t \in [0,1]\}$ be the time-scaled version of the sample path of $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ defined in (1.2), embedded in the space of continuous-time processes. Note that $\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})$ is a random element in $\mathbb{D} = \mathbb{D}\big([0,1], \mathbb{R}^m\big)$. As η decreases, $\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})$ 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 $k \geq 0$:

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

The precise statement and the definition of $\mathbf{C}^{(k)|b}$ can be found in Section 2.2.1. Here, we note that $C^{(k)|b}$ is precisely identified, its intuitive meaning is clear, and its computation is straightforward using Monte Carlo simulation.

Additionally, we point out that the index k leading to non-degenerate upper and lower bounds in (1.5) represents the minimal number of jumps (with sizes truncated under b) that must be added to the path of $\boldsymbol{y}_t(\boldsymbol{x})$ for it to enter the set B, given $\boldsymbol{x} \in A$. Such an index 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 a singleton, 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 the sharp characterization of the catastrophe principle for (1.2). Specifically, Section 2.2.2 proves that the conditional distribution of $\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})$ given

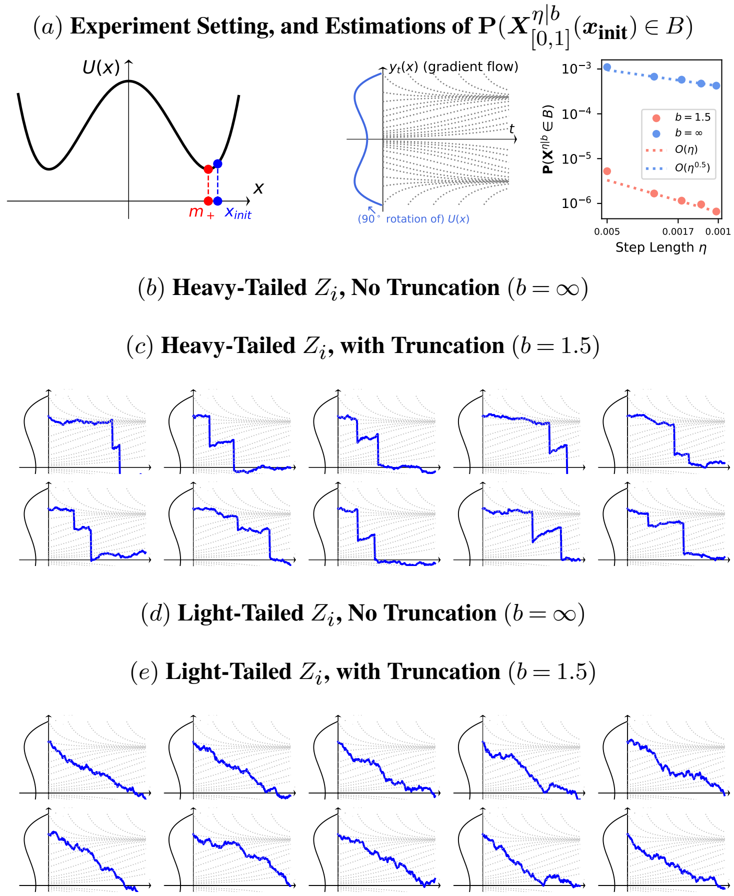

FIG 1. *Numerical examples for large deviations and the catastrophe principle.* ***(a, Left)*** *the potential function $U(\cdot)$ defined in (**??**).* ***(a, Middle)*** *gradient flows under $-U'(\cdot)$.* ***(a, Right)*** *Estimation of $\mathbf{P}(\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}_{init}) \in B)$ through Monte-Carlo simulation; dashed lines are predictions according to our large deviations asymptotics.* ***(b)–(e)****: Samples from $\mathbf{P}(\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}_{init}) \in \cdot\ |\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}_{init}) \in B)$, $\eta = \frac{1}{200}$.*

the rare event of interest converges to the distribution of a piecewise deterministic random function $\boldsymbol{X}^{*|b}_{|B}(\boldsymbol{x})$ with precisely k random jumps whose sizes are bounded from below:

$$\mathscr{L}\big(\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})\big|\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}) \in B\big) \to \mathscr{L}\big(\boldsymbol{X}^{*|b}_{|B}(\boldsymbol{x})\big). \tag{1.6}$$

We give the formal statement of the catastrophe principle in Corollary 2.7. Here, we note that the perturbation associated with $\boldsymbol{Z}_i$ is $\eta\boldsymbol{\sigma}(\boldsymbol{X}^{\eta|b}_{i-1}(\boldsymbol{x}))\boldsymbol{Z}_i$. Hence, the jump size associated with $\boldsymbol{Z}_i$ being bounded from below implies that $\boldsymbol{Z}_i$ is of order $1/\eta$. This means that the rare event $\{\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}) \in B\}$ is driven by k jumps of size $O(1/\eta)$, whereas the rest of the system behaves close to the law-of-large-numbers limit of the system. Figure 1 illustrates the catastrophe principle in a univariate setting where the drift is given by the negative gradient of a potential: $\boldsymbol{b}(\cdot) = -U'(\cdot)$. In (a, Left) of Figure 1, we show the potential function $U : \mathbb{R} \to \mathbb{R}$, while (a, Middle) shows its gradient flows starting from different initial points. By gradient flows, we refer to the solution of the ODE $d\boldsymbol{y}_t(\boldsymbol{x})/dt = -U'\big(\boldsymbol{y}_t(\boldsymbol{x})\big)$ with initial condition $\boldsymbol{y}_0(\boldsymbol{x}) = \boldsymbol{x}$. Given an initial value $\boldsymbol{x}_{\text{init}}$, suppose that we are interested in the conditional law

$$\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}_{\text{init}}) \in \cdot\ \big|\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}_{\text{init}}) \in B\big), \tag{1.7}$$

where $B = \{\xi \in \mathbb{D}([0,1],\mathbb{R}) : \xi(t) \leq 0 \text{ for some } t \leq 1\}$. That is, the behavior of (1.2) when they escape from the attraction field $(0,\infty)$ associated with the local minimum $m_+ = \sqrt{5}$ within $\lfloor 1/\eta \rfloor$ steps. As shown in (b) of Figure 1, when driven-by heavy-tailed Z_i's, the untruncated dynamics $\boldsymbol{X}^\eta_j$ closely resembles the gradient flows and stays close to m_+ until a single large Z_j sends $\boldsymbol{X}^\eta_j$ outside of $(0,\infty)$ in one shot. In comparison, under small enough choices of the truncation threshold b in (1.2), the process $\boldsymbol{X}^{\eta|b}_j$ can no longer exit $(0,\infty)$ from m_+ in one step. Indeed, (c) of Figure 1 depicts a case where the sample paths of $\boldsymbol{X}^{\eta|b}_j$ resembles the gradient flow with two large perturbations truncated at b. This clearly confirms the catastrophe principle (1.6): the rare event $\{\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x}) \in B\}$ arises almost always because of $k = 2$ catastrophically large—i.e., $O(1/\eta)$—perturbations, whereas the rest of the system is indistinguishable from its nominal behavior; here, the index k is the minimum number of jumps required by the nominal path (i.e., gradient flow) to enter the set B. Compare this to (b) of Figure 1, which is governed by the principle of a single big jump, i.e., the catastrophe principle with $k = 1$. Note also that both of these sharply contrast with the light-tailed cases predicted by the classical Freidlin-Wentzell theory, where the rare events arise as the SGD fights against the negative gradient in each step, climbing up the potential hill to transition into the adjacent potential well; see part (d) and (e) of Figure 1. It is also worth noting that, unlike the light-tailed exit scenarios, which closely follow a single deterministic path defined by the solution of a variational problem associated with the rate function, heavy-tailed scenarios exhibit significant stochasticity in the location and size of the big jumps with only the number of jumps being deterministically k. This reflects the fact that the distributional limit of the scaled process described in (1.6) is non-degenerate. See Section **??** for more details of this numerical example.

The notion of $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence, introduced in [59] and further developed in [75], was a key technical tool behind (1.4) in [75]. In this paper, we introduce a uniform version of the $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence and prove an associated Portmanteau theorem in Section 2.1. These developments form the backbone that supports our proofs in Section 2.2 for the uniform sample-path large deviations of the form (1.5).

Metastability Analysis. The second contribution of this paper is the first exit-time analysis for heavy-tailed systems. As described at the beginning of this section, two modern approaches to the analysis of exit times for light-tailed stochastic dynamical systems are the Freidlin-Wentzell theory (or pathwise approach) detailed in the monographs [39, 68] and the potential theoretic approach summarized in the monograph [11]. Despite their success in the light-tailed contexts, neither the pathwise approach nor the potential theoretic approach readily extends to heavy-tailed contexts. In particular, for truncated heavy-tailed dynamics such as $\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})$, the explicit formula for the stationary distribution is rarely available, and its generator lacks the simplicity of the Brownian case, making the adaptation of potential theoretic approach to our context challenging. Meanwhile, the pathwise approach hinges on the large deviation principles at the sample-path level. Historically, however, the heavy-tailed large deviations at the sample-path level have been unavailable and considered to be out of reach until recently.

Successful results in heavy-tailed contexts are relatively recent. For one-dimensional Lévy driven SDEs, [47, 49] proved that the exit times from metastability sets scale at a polynomial rate and the prefactor of the of the scale depend on the width of the potential wells rather than the height of the potential barrier. Similar results have been established in more general settings, such as the multi-dimensional analog in [50], exit times for a global attractor instead of a stable point [46] (see also [43] for its application in characterizing the limiting Markov chain of hyperbolic dynamical systems driven by heavy-tailed perturbations), exit times under multiplicative noises in $\mathbb{R}^d$ [70], extensions to infinite-dimensional spaces [23], and the (discretized) stochastic difference equations driven by α-stable noises [67], to name a few. Such metastability analyses were applied in [76] to study the generaliztion performance of DNNs trained by SGDs with heavy-tailed dynamics and, more recently, in [5] to analyze the sample efficiency of policy gradient algorithms in reinforcement learning. It should be noted that these results focus on the events associated with the principle of a single big jump.

In contrast, this paper develops a systematic tool for analyzing the exit times and locations, even in cases where the principle of a single big jump fails to account for the exit events, and more complex patterns arise during the exit process. The process $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ exemplifies such a scenario, as the truncation operator $\varphi_c(\cdot)$ prevents exit events driven by a single big jump. We reveal phase transitions in the first exit times of $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$, which depend on a notion of the "discretized widths" of the attraction fields. Specifically, we consider (1.1) with drift coefficients $\boldsymbol{a}(\cdot) = -\nabla U(\cdot)$ for some potential function $U \in \mathcal{C}^1(\mathbb{R}^m)$. Without loss of generality, let $I \subseteq \mathbb{R}^m$ be some open and bounded set containing the origin. Suppose that the entire domain I falls within the attraction field of the origin, and the gradient field $-\nabla U(\cdot)$ is locally contractive around the origin. In other words, when initialized within I, the deterministic gradient flow $d\boldsymbol{y}_t(\boldsymbol{x})/dt = -\nabla U(\boldsymbol{y}_t(\boldsymbol{x}))$ (under the initial condition $\boldsymbol{y}_0(\boldsymbol{x}) = \boldsymbol{x}$) will be attracted to and remain trapped near the origin. However, due to the presence of random perturbations, $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ will eventually escape from I after a sufficiently long time. Of particular interest are the asymptotic scale of the first exit times as $\eta \downarrow 0$. Theorem 2.8 proves 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 limit (uniformly for all $\boldsymbol{x}$ bounded away from I^c):

$$\Big(\lambda_b^I(\eta)\cdot\tau^{\eta|b}(\boldsymbol{x}),\ \boldsymbol{X}_\tau^{\eta|b}(\boldsymbol{x})\Big) \Rightarrow (E, V_b) \qquad \text{as } \eta \downarrow 0 \tag{1.8}$$

with some (deterministic) time-scaling function $\lambda_b^I(\eta)$. Here, E is an exponential random variable with the rate parameter 1, V_b is some random element independent of E and supported on I^c, and the scaling function $\lambda_b^I(\eta)$ is regularly varying with index $-[1+\mathcal{J}_b^I(\alpha-1)]$ as $\eta \downarrow 0$, where $\mathcal{J}_b^I$ is the aforementioned discretized width of domain I relative to the truncation threshold b. The precise definition of $\mathcal{J}_b^I$ is provided in (2.27) in Section 2.3.1. However, we note that in the special case $b = \infty$, one can immediately verify that $\mathcal{J}_b^I = 1$, regardless of the geometry of U. Consequently, (1.8) reduces to the principle of a single big jump, as expected. When the drift is contractive so that $\nabla U(\boldsymbol{x}) \cdot \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \in I$, it is also straightforward to see that $\mathcal{J}_b^I = \lceil r/b \rceil$ where r is the distance between $\boldsymbol{0}$ and I^c, and hence, $\mathcal{J}_b^I$ is indeed precisely the discretized width of the attraction field I relative to b. In particular, note that the drift is contractive within any attraction field in the one-dimensional cases. However, in general multi-dimensional spaces, $\mathcal{J}_b^I$ reflects a much more intricate interplay between the geometry of the drift $\boldsymbol{a}(\cdot)$ (or the potential $U(\cdot)$) and the truncation threshold b.

Figure 2 illustrates the key role of the relative width $\mathcal{J}_b^{\mathcal{I}}$ in one dimension. Specifically, we consider a one-dimensional case with a potential function $U : \mathbb{R} \to \mathbb{R}$ depicted in Figure 2 (i), where $I = (s_1, s_2)$ is the attraction field of the local minimum m. Since m is closer to the left boundary s_1, the minimal number of steps required to exit I when starting from m is $\mathcal{J}_b^{\mathcal{I}} = \lceil |s - m_1|/b \rceil$ where $b \in (0, \infty)$. In the untruncated case (1.1) (i.e., with $b = \infty$), we simply have $\mathcal{J}_\infty^{\mathcal{I}} = 1$. Figure 2 (ii) illustrates the discrete structure of phase transitions in (1.8), where 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, with $\alpha = 1.2$ being the index of $\boldsymbol{Z}_i$'s regular variation. This means that the order of the first exit time $\tau^{\eta|b}(\boldsymbol{x})$ does not vary continuously with respect to the truncation threshold b. Instead, it exhibits a

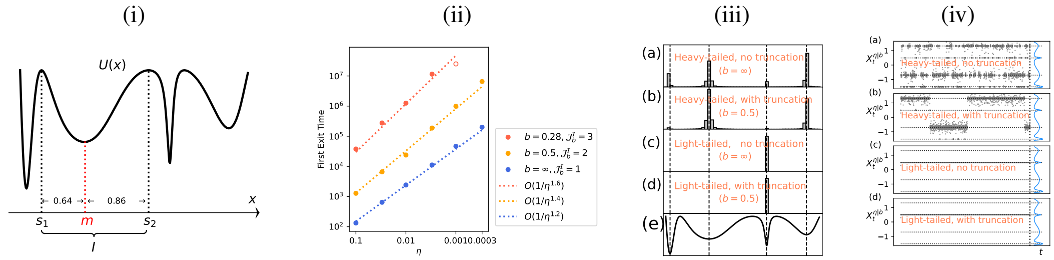

FIG 2. *Numerical examples of the metastability analysis.* ***(i)*** *The univariate potential $U(\cdot)$ defined in (**??**).* ***(ii)*** *First exit times $\tau^{\eta|b}(m)$ from I under different truncation thresholds b and scale parameters η. Dashed lines are predictions from our results in Section 2.3, whereas the dots are the exit times estimated using 20 samples. It can be observed that the predictions and estimates align well.* ***(iii)*** *Histograms of locations within the potential potential $U(\cdot)$ visited by $X_t^{\eta|b}(x)$. Note that in (b), the sharp minima are almost completely eliminated from the trajectory of the SGD.* ***(iv)*** *Sample paths of $X_t^{\eta|b}(x)$. Dashed lines in (iii) and (iv) are added as references for the locations of local minima. Driven by truncated heavy-tailed noise, $X_t^{\eta|b}(x)$ almost completely avoids the sharp minima of $U(\cdot)$ in (b).*

discrete dependence on b through the integer-valued quantity $\mathcal{J}_b^I$. Consequently, the wider the domain I, the *asymptotically longer* the exit time $\tau^{\eta|b}(\boldsymbol{x})$ will be. In the companion paper [82], we build on these phase transitions in exit times to reveal an intriguing global behavior of $\boldsymbol{X}_j^{\eta|b}$ over a multi-well potential: the distribution of $\boldsymbol{X}_j^{\eta|b}$'s sample path closely resembles a Markov chain that **completely avoids narrow local minima**; see Figure 2 (iii) and (iv). More importantly, we demonstrate in [82] that such global dynamics under truncated heavy tails are intimately related to the generalization performance of deep neural networks. See Section **??** for more details of the numerical experiments presented in Figure 2.

Our approach to the metastability analysis hinges on the concept of asymptotic atoms, a general machinery we develop in Section 2.3.2. Asymptotic atoms are 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 characterize the behavior of the stochastic processes uniformly for all initial values over the asymptotic atoms. It should be noted that [51] also investigated the exit events driven by multiple truncated jumps. However, in their context, the mechanism through which multiple jumps arise 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. (See also [52] for the summary of the hierarchy in the asymptotics of the first exit times for heavy-tailed dynamics.) Our results complement the picture and provide a missing piece of the puzzle by unveiling the phase transitions in the exit times under truncated regularly varying perturbations.

Some of the the metastability analysis in Section 2.3 of this paper have been presented in a preliminary form at a conference [81]. The main focus of [81] 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 [81], 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, with numerical examples collected in Section **??**. Section 3 and Section 4 provide the proofs of Sections 2.1, 2.2, and 2.3. Results for stochastic difference equations under more general scaling regimes are presented in Appendix **??**. Results for SDEs driven by Lévy processes with regularly varying increments are collected in Appendix **??**.

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. Section **??** presents numerical examples of our theoretical results. 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 supp(μ)) be the smallest closed set C such that $\mu(\mathbb{S}\setminus C)=0$. For any function $g:\mathbb{S}\to\mathbb{R}$, let 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 $x_n\leq Cy_n\ \forall n\geq 1$. 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 [59, 75] 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 the metric subspace of $\mathbb{S}$ in 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 [59] for equivalent 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 [75], 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 metastability analysis in the later sections.

DEFINITION 2.1 (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 $\theta\in\Theta$. We say that μ^η_θ converges to μ_θ in $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ uniformly in θ on Θ as $\eta\downarrow 0$ 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 $\theta \in \Theta$. Suppose that 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. We provide two additional remarks regarding Theorem 2.2. First, it is generally 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) 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 $\mu^\eta_\theta \triangleq \boldsymbol{\delta}_{\theta-\eta},\ \mu_\theta \triangleq \boldsymbol{\delta}_\theta$, where $\boldsymbol{\delta}_x$ is the Dirac measure at x. For closed set $F = [-1,0]$ and any $\eta \in (0,2)$,

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

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 given $\theta \in \Theta$. To illustrate that (iv) does not imply (i)–(iii), it suffices to examine the case where $\mathbb{C} = \emptyset$, $\mathbb{S} = \mathbb{R}$, $\Theta = [-1,1]$, $\mu^\eta_\theta = \boldsymbol{\delta}_{-\theta}$, and $\mu_\theta = \boldsymbol{\delta}_\theta$.

To conclude this section, we note that the proof of $\mathbb{M}$-convergence (and hence the application of Theorem 2.2) is often facilitated by the notion of asymptotic equivalence between two families of random objects. Specifically, we consider the following version of asymptotic equivalence that generalizes Definition 2.9 in [21], which is particularly useful in the context of Lemma 2.4. The proof of Lemma 2.4 will be provided in Section 2.1.

DEFINITION 2.3 (Asymptotic Equivalence). Let X_n and Y_n^δ be random elements taking values in a complete separable metric space $(\mathbb{S}, \boldsymbol{d})$ and supported on the same probability space. 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^δ in $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$ with respect to ϵ_n as $\delta \downarrow 0$** if the following holds: given $\Delta > 0$ and $B \in \mathscr{S}_{\mathbb{S}}$ that is bounded away from $\mathbb{C}$,

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

LEMMA 2.4. *Let X_n and Y_n^δ be random elements taking values in a complete separable metric space $(\mathbb{S}, \boldsymbol{d})$ and supported on the same probability space. Let $\mu \in \mathbb{M}(\mathbb{S} \setminus \mathbb{C})$. Suppose*

(i) *X_n is asymptotically equivalent to Y_n^δ in $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$ with respect to ϵ_n as $\delta \downarrow 0$,*
(ii) *Given $B \in \mathscr{S}_{\mathbb{S}}$ that is bounded away from $\mathbb{C}$, it holds for all $\delta > 0$ small enough that*

$$\limsup_{\delta \downarrow 0} \limsup_{n \to \infty} \epsilon_n^{-1} \mathbf{P}(Y_n^\delta \in B) \leq \mu(B^-), \qquad \liminf_{\delta \downarrow 0} \liminf_{n \to \infty} \epsilon_n^{-1} \mathbf{P}(Y_n^\delta \in B) \geq \mu(B^\circ).$$

Then $\epsilon_n^{-1} \mathbf{P}(X_n \in \cdot) \to \mu(\cdot)$ in $\mathbb{M}(\mathbb{S} \setminus \mathbb{C})$.

2.2. *Heavy-Tailed Large Deviations.* In Section 2.2.1, we study the sample-path large deviations for stochastic difference equations driven by heavy-tailed dynamics. Section 2.2.2 then characterizes the catastrophe principle of heavy-tailed systems by presenting the conditional limit theorems. The results reveal a discrete hierarchy of the most likely scenarios and probabilities of rare events in heavy-tailed stochastic difference equations. We note that analogous results under more general scaling regimes and for stochastic differential equations are collected in Sections **??** and **??** of the Appendix.

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 unless stated otherwise. 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 $\mathbb{F} = (\mathcal{F}_j)_{j \geq 0}$. Given $b \in (0, \infty)$, the drift coefficient $\boldsymbol{a} : \mathbb{R}^m \to \mathbb{R}^m$, and the diffusion coefficient $\boldsymbol{\sigma} : \mathbb{R}^m \to \mathbb{R}^{m \times d}$, our goal is to study the sample-path large deviations for the discrete-time process $\big\{\boldsymbol{X}_t^{\eta|b}(\boldsymbol{x}) : t \in \mathbb{N}\big\}$ in $\mathbb{R}^m$ driven by the recursion

$$\boldsymbol{X}_t^{\eta|b}(\boldsymbol{x}) = \boldsymbol{X}_{t-1}^{\eta|b}(\boldsymbol{x}) + \varphi_b\Big(\eta \boldsymbol{a}\big(\boldsymbol{X}_{t-1}^{\eta|b}(\boldsymbol{x})\big) + \eta \boldsymbol{\sigma}\big(\boldsymbol{X}_{t-1}^{\eta|b}(\boldsymbol{x})\big) \boldsymbol{Z}_t\Big) \qquad \forall t \geq 1 \tag{2.3}$$

under the initial condition $\boldsymbol{X}_0^{\eta|b}(\boldsymbol{x}) = \boldsymbol{x}$, where

$$\varphi_b(\boldsymbol{w}) \triangleq \Big(\frac{b}{\|\boldsymbol{w}\|} \wedge 1\Big) \cdot \boldsymbol{w} \quad \forall \boldsymbol{w} \neq \boldsymbol{0}, \qquad \varphi_b(\boldsymbol{0}) \triangleq \boldsymbol{0}. \tag{2.4}$$

Here, $u \wedge v = \min\{u, v\}$ and $u \vee v = \max\{u, v\}$. For any $\boldsymbol{w} \neq \boldsymbol{0}$, we have $\varphi_b(\boldsymbol{w}) = (b \wedge \|\boldsymbol{w}\|) \cdot \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$. In other words, the truncation operator $\varphi_b(\boldsymbol{w})$ in (2.3) maintains the direction of the vector $\boldsymbol{w}$ but rescales it to ensure that the norm would not exceed the threshold value b. In particular, we are interested in the case where $\boldsymbol{Z}_i$'s are heavy-tailed. In this paper, we capture the heavy-tailed phenomena 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, [7, 73, 36, 17]. 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

$$H(x)\triangleq \mathbf{P}(\|\boldsymbol{Z}\|>x). \tag{2.5}$$

For any $\alpha>0$, let ν_α be the (Borel) measure on $(0,\infty)$ with

$$\nu_\alpha[x,\infty)=x^{-\alpha}. \tag{2.6}$$

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

$$\Phi(\boldsymbol{x})\triangleq\begin{cases}\Big(\|\boldsymbol{x}\|,\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\Big) & \text{if } \boldsymbol{x}\neq 0,\\ \big(0,(1,0,0,\cdots,0)\big) & \text{otherwise.}\end{cases} \tag{2.7}$$

Note that the origin is included in the domain of Φ simply to lighten the notations in the proofs. However, $\Phi(\boldsymbol{x})$ will not be applied at $\boldsymbol{x}=\mathbf{0}$ in our proofs. Thus, Φ can be interpreted as the polar transform with domain extended to $\mathbf{0}$. 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 $\mathfrak{N}_d$ 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 $x\to\infty$)

$$\frac{\mathbf{P}\Big((x^{-1}R,\boldsymbol{\Theta})\in\ \cdot\ \Big)}{H(x)}\to\nu_\alpha\times\mathbf{S}\qquad\text{in }\mathbb{M}\Big(\big([0,\infty)\times\mathfrak{N}_d\big)\setminus\big(\{0\}\times\mathfrak{N}_d\big)\Big). \tag{2.8}$$

REMARK 2. The multivariate regular variation condition (2.8) is typically stated in terms of vague convergence; see, e.g., [72, 44]. While vague convergence is generally weaker than $\mathbb{M}$-convergence (see Lemma 2.1 of [59]), due to $\alpha>1$ we have $(\nu_\alpha\times\mathbf{S})(A)<\infty$ for any Borel set $A\subseteq(0,\infty)\times\mathfrak{N}_d$ that is bounded away from $\{0\}\times\mathfrak{N}_d$. Therefore, it is easy to verify that the $\mathbb{M}$-convergence stated in (2.8) is equivalent to vague convergence. Furthermore, by the alternative definitions for multivariate regular variation (see [72, 44]), Assumption 1 is equivalent to the vague convergence of $H^{-1}(x)\mathbf{P}(x^{-1}\boldsymbol{Z}\in\ \cdot\ )$ to some Borel measure $\mu(\cdot)$ in $\mathbb{M}(\mathbb{R}^d\setminus\{\mathbf{0}\})$, where $\mu(\cdot)$ exhibits self-similarity in terms of $\mu(\lambda A)=\lambda^{-\alpha}\mu(A)$ for any Borel set $A\subseteq\mathbb{R}^d$ that is bounded away from the origin.

Next, we introduce the assumptions on the drift coefficient $\boldsymbol{a}(\cdot)=\big(a_1(\cdot),\cdots,a_m(\cdot)\big)^T$ and the diffusion coefficient $\boldsymbol{\sigma}(\cdot)=\big(\sigma_{i,j}(\cdot)\big)_{i\in[m],j\in[d]}$. 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 $\mathbf{A}\in\mathbb{R}^{p\times q}$. Obviously, the lower bound for D in Assumption 2 is not necessary, and it is imposed w.l.o.g. for the notational simplicity in the proof.

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

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

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

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

Here, Λ_T is the set of all homeomorphism on $[0,T]$. Throughout this paper, we fix some m and d and consider $\boldsymbol{X}^{\eta|b}_t(\boldsymbol{x})$ taking values in $\mathbb{R}^m$ driven by $\boldsymbol{Z}_t$'s in $\mathbb{R}^d$. Given $A \subseteq \mathbb{R}$, 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 on A with length k. For any $b,\ T \in (0,\infty)$ and $k \in \mathbb{N}$, define the 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 $\boldsymbol{x} \in \mathbb{R}^m$, $\mathbf{W} = (\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$, $\mathbf{V} = (\boldsymbol{v}_1,\cdots,\boldsymbol{v}_k) \in \mathbb{R}^{m\times k}$, 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

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

$$\frac{d\xi_s}{ds} = \boldsymbol{a}(\xi_s) \quad \forall s \in [0,T],\ s \neq t_1,t_2,\cdots,t_k; \tag{2.11}$$

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

Similarly, define the mapping $h^{(k)|b}_{[0,T]}:\mathbb{R}^m \times \mathbb{R}^{d\times k} \times (0,T]^{k\uparrow} \to \mathbb{D}[0,T]$ by

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

In essence, $h^{(k)|b}_{[0,T]}(\boldsymbol{x},\mathbf{W},\boldsymbol{t})$ produces an ODE path perturbed by jumps $\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k$ (with sizes modulated by $\boldsymbol{\sigma}(\cdot)$ and then truncated under threshold b) at times $t_1,\cdots,t_k$, and the mapping $\bar{h}^{(k)|b}_{[0,T]}$ further includes perturbations $\boldsymbol{v}_j$'s right before each jump. For $k=0$, we adopt the convention that $\xi = \bar{h}^{(0)|b}_{[0,T]}(\boldsymbol{x})$ is the solution to the ODE $d\xi_s/ds = \boldsymbol{a}(\xi_s)\ \forall s \in [0,T]$ with the initial condition $\xi_0 = \boldsymbol{x}$. 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 $\boldsymbol{x}$. Given $b,T \in (0,\infty)$, $\epsilon \geq 0$, $A \subseteq \mathbb{R}^m$ and $k \in \mathbb{N}$, let

$$\mathbb{D}^{(k)|b}_A[0,T](\epsilon) \triangleq \bar{h}^{(k)|b}_{[0,T]}\Big(A \times \mathbb{R}^{m\times k} \times \big(\bar{B}_\epsilon(\mathbf{0})\big)^k \times (0,T]^{k\uparrow}\Big) \tag{2.14}$$

be the set that contains all the ODE path with k jumps by time T, i.e., the image of the mapping $\bar{h}^{(k)|b}_{[0,T]}$ defined in (2.10)–(2.12), under small perturbations $\|\boldsymbol{v}_j\| \leq \epsilon$ for all $j \in [k]$. By our definition of $\bar{h}^{(0)|b}_{[0,T]}$ above, $\mathbb{D}^{(0)|b}_A[0,T](\epsilon)$ simply contains all ODE paths under vector field $\boldsymbol{a}(\cdot)$ with initial values over A. For $k=-1$, we adopt the convention that $\mathbb{D}^{(-1)|b}_A[0,T](\epsilon) \triangleq \emptyset$. Also, note that $\mathbb{D}^{(k)|b}_A[0,T](\epsilon) \subseteq \mathbb{D}^{(k)|b}_A[0,T](\epsilon')$ for any $0 \leq \epsilon < \epsilon'$ and $k \geq -1$. We state useful properties of $h^{(k)|b}_{[0,T]}$ and $\mathbb{D}^{(k)|b}_A[0,T](\epsilon)$ in Section **??** of the appendix.

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 $\boldsymbol{x} \in \mathbb{R}^m$, $k \in \mathbb{N}$ and $b,T \in (0,\infty)$, define the Borel measure

$$\mathbf{C}^{(k)|b}_{[0,T]}(\,\cdot\,;\boldsymbol{x}) \triangleq \int \mathbb{I}\Big\{h^{(k)|b}_{[0,T]}\big(\boldsymbol{x},\mathbf{W},\boldsymbol{t}\big) \in\ \cdot\ \Big\}\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^k(d\mathbf{W}) \times \mathcal{L}^{k\uparrow}_T(d\boldsymbol{t}), \tag{2.15}$$

where $\mathbf{S}$ is the probability measure on the unit sphere $\mathfrak{N}_d$ characterized in Assumption 1, ν_α is specified in (2.6), $(\nu_\alpha \times \mathbf{S})\circ\Phi$ is the composition of the product measure $\nu_\alpha \times \mathbf{S}$ with the polar transform Φ, i.e.,

$$\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)(B) \triangleq (\nu_\alpha \times \mathbf{S})\big(\Phi(B)\big) \qquad \forall \text{Borel set } B \subseteq \mathbb{R}^d \setminus \{\mathbf{0}\}, \tag{2.16}$$

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 $\mathbf{W} = (\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k) \in \mathbb{R}^{d\times k}$ with $\boldsymbol{w}_j \neq 0\ \forall j \in [k]$, we have $\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)^k(d\mathbf{W}) = \bigtimes_{j\in[k]}\big((\nu_\alpha \times \mathbf{S})\circ\Phi\big)(d\boldsymbol{w}_j)$.

Note that for any $\boldsymbol{x} \in A$, the measure $\mathbf{C}^{(k)|b}_{[0,T]}(\,\cdot\,;\boldsymbol{x})$ is supported on $\mathbb{D}^{(k)|b}_A[0,T](0)$. Next, define the rate functionc

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

with $H(x) = \mathbf{P}(\|\boldsymbol{Z}\| > x)$ defined in (2.5). By Assumption 1, $\lambda(\eta) \in \mathcal{RV}_{\alpha-1}(\eta)$ as $\eta \downarrow 0$. We write $\lambda^k(\eta) = \big(\lambda(\eta)\big)^k$. For any $T,\ \eta,\ b \in (0,\infty)$, and $\boldsymbol{x} \in \mathbb{R}^m$, 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 $\mathbb{D}[0,T]$, with $\lfloor t \rfloor \triangleq \max\{n \in \mathbb{Z}:\ n \leq t\}$ and $\lceil t \rceil \triangleq \min\{n \in \mathbb{Z}:\ n \geq t\}$. In case that $T = 1$, we suppress the time horizon $[0,1]$ and write $\mathbb{D} \triangleq \mathbb{D}[0,1]$, $\boldsymbol{d}_{J_1} \triangleq \boldsymbol{d}^{[0,1]}_{J_1}$, $\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \triangleq \boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})$, $h^{(k)|b} \triangleq h^{(k)|b}_{[0,1]}$, $\mathbb{D}^{(k)|b}_A(\epsilon) \triangleq \mathbb{D}^{(k)|b}_A[0,1](\epsilon)$, and $\mathbf{C}^{(k)|b} \triangleq \mathbf{C}^{(k)|b}_{[0,1]}$. Now, we are ready to state Theorem 2.5, which establishes the uniform $\mathbb{M}$-convergence for the law of $\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x})$ to $\mathbf{C}^{(k)|b}_{[0,T]}(\,\cdot\,;\boldsymbol{x})$ and a uniform version of the sample path large deviations for $\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x})$.

THEOREM 2.5. *Under Assumptions 1 and 2, it holds for any $k \in \mathbb{N}$, any $b, T, \epsilon \in (0,\infty)$, and any compact $A \subset \mathbb{R}^m$ that $\lambda^{-k}(\eta)\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in \,\cdot\,\big) \to \mathbf{C}^{(k)|b}_{[0,T]}(\,\cdot\,;\boldsymbol{x})$ in $\mathbb{M}\big(\mathbb{D}[0,T] \setminus \mathbb{D}^{(k-1)|b}_A[0,T](\epsilon)\big)$ uniformly in $\boldsymbol{x}$ on A as $\eta \downarrow 0$. In particular, for any $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ that is bounded away from $\mathbb{D}^{(k-1)|b}_A[0,T](\epsilon)$ for some (and hence all) $\epsilon > 0$ small enough,*

(2.17)

$$\inf_{\boldsymbol{x}\in A} \mathbf{C}^{(k)|b}_{[0,T]}\big(B^\circ;\boldsymbol{x}\big) \leq \liminf_{\eta\downarrow 0} \frac{\inf_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta)}$$

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

We provide the proof of Theorem 2.5 in Section 3.3. Furthermore, by sending $b \to \infty$ in Theorem 2.5, we are able to establish uniform sample path large deviations for the process $\big\{\boldsymbol{X}^\eta_t(\boldsymbol{x}) : t \in \mathbb{N}\big\}$ driven by the recursion

$$\boldsymbol{X}^\eta_t(\boldsymbol{x}) = \boldsymbol{X}^\eta_{t-1}(\boldsymbol{x}) + \eta\boldsymbol{a}\big(\boldsymbol{X}^\eta_{t-1}(\boldsymbol{x})\big) + \eta\boldsymbol{\sigma}\big(\boldsymbol{X}^\eta_{t-1}(\boldsymbol{x})\big)\boldsymbol{Z}_t \qquad \forall t \geq 1 \tag{2.18}$$

under initial condition $\boldsymbol{X}^\eta_0(\boldsymbol{x}) = \boldsymbol{x}$. Note that the scalar version of (2.18) is

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

where $\boldsymbol{a}(\cdot) = \big(a_1(\cdot),\cdots,a_m(\cdot)\big)^T$, $\boldsymbol{\sigma}(\cdot) = \big(\sigma_{i,j}(\cdot)\big)_{i\in[m],j\in[d]}$, $\boldsymbol{X}^\eta_t(x) = \big(X^\eta_{t,1}(x),\cdots,X^\eta_{t,m}(x)\big)^T$, and $\boldsymbol{Z}_t = (Z_{t,1},\cdots,Z_{t,d})^T$. By interpreting $\varphi_\infty(\boldsymbol{w}) = \boldsymbol{w}$ as the identity mapping in (2.4), the definition of $\boldsymbol{X}^\eta_t(\boldsymbol{x})$ in (2.18) coincides with that of $\boldsymbol{X}^{\eta|\infty}_t(\boldsymbol{x})$ in (2.3) under the choice of $b = \infty$. Analogously, we adopt the notations $\bar{h}^{(k)}_{[0,T]} \triangleq \bar{h}^{(k)|\infty}_{[0,T]}$, $h^{(k)}_{[0,T]} \triangleq h^{(k)|\infty}_{[0,T]}$, $\mathbf{C}^{(k)}_{[0,T]}(\,\cdot\,;\boldsymbol{x}) \triangleq \mathbf{C}^{(k)|\infty}_{[0,T]}(\,\cdot\,;\boldsymbol{x})$, and $\mathbb{D}^{(k)}_A[0,T](\epsilon) \triangleq \mathbb{D}^{(k)|\infty}_A[0,T](\epsilon)$. For $k = -1$, we again adopt the convention that $\mathbb{D}^{(-1)}_A[0,T](\epsilon) \triangleq \emptyset$. Define the time-scaled version of the sample path as

$$\boldsymbol{X}^\eta_{[0,T]}(\boldsymbol{x}) \triangleq \big\{\boldsymbol{X}^\eta_{\lfloor t/\eta \rfloor}(\boldsymbol{x}):\ t \in [0,T]\big\} \quad \forall T > 0. \tag{2.19}$$

In case that $T = 1$, we suppress the time horizon $[0,1]$ and write $h^{(k)}$, $\mathbf{C}^{(k)}$, $\mathbb{D}_A^{(k)}(\epsilon)$, and $\boldsymbol{X}^{\eta}(x)$ to denote $h^{(k)}_{[0,1]}$, $\mathbf{C}^{(k)}_{[0,1]}$, $\mathbb{D}_A^{(k)}[0,1](\epsilon)$, and $\boldsymbol{X}^{\eta}_{[0,1]}(x)$, respectively. Under Assumption 3, we establish in Theorem 2.6 the uniform $\mathbb{M}$-convergence and sample path large deviations for $\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x})$. Again, the lower bound for C in Assumption 3 is imposed w.l.o.g. simply for the convenience of the proof.

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

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

THEOREM 2.6. *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 $\lambda^{-k}(\eta)\mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in \ \cdot\ \big) \to \mathbf{C}^{(k)}_{[0,T]}(\ \cdot\ ;\boldsymbol{x})$ in $\mathbb{M}\big(\mathbb{D}[0,T] \setminus \mathbb{D}_A^{(k-1)}[0,T](\epsilon)\big)$ uniformly in $\boldsymbol{x}$ on A as $\eta \downarrow 0$. In particular, for any $B \in \mathscr{S}_{\mathbb{D}[0,T]}$ that is bounded away from $\mathbb{D}_A^{(k-1)}[0,T](\epsilon)$ for some (and hence all) $\epsilon > 0$ small enough,*

(2.20)
$$\begin{aligned}\inf_{\boldsymbol{x}\in A} \mathbf{C}^{(k)}_{[0,T]}(B^\circ;\boldsymbol{x}) &\leq \liminf_{\eta\downarrow 0} \frac{\inf_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta)} \\ &\leq \limsup_{\eta\downarrow 0} \frac{\sup_{\boldsymbol{x}\in A}\mathbf{P}\big(\boldsymbol{X}^{\eta}_{[0,T]}(\boldsymbol{x}) \in B\big)}{\lambda^k(\eta)} \leq \sup_{\boldsymbol{x}\in A} \mathbf{C}^{(k)}_{[0,T]}(B^-;\boldsymbol{x}) < \infty.\end{aligned}$$

REMARK 3. We add a remark on the connection between (2.17) (2.20) and the classical LDP framework. Given a measurable set $B \subseteq \mathbb{D}[0,T]$, there is a particular k that plays the role of the rate function. Specifically, let $\mathbb{D}_A^{(k)}[0,T] = \mathbb{D}_A^{(k)}[0,T](0)$ and $\mathcal{J}_A(B) \triangleq \min\{k \in \mathbb{N}:\ B \cap \mathbb{D}_A^{(k)}[0,T] \neq \emptyset\}$. 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 (2.20) is 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 (e.g., $A = \{\mathbf{0}\}$), $T = 1$, $\boldsymbol{a} \equiv 0$, and $\boldsymbol{\sigma} \equiv \mathbf{I}_m$ (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 [75]. Furthermore, the condition of B being bounded away from $\mathbb{D}_A^{(k-1)}(\epsilon)$ (for small $\epsilon > 0$) will reduce to that B is bounded away from the set of step functions (i.e., piecewise constant functions) in $\mathbb{D}$, vanishing at the origin, with at most $k-1$ jumps. This confirms that Theorems 2.5 and 2.6 are proper generalizations of the heavy-tailed large deviations for Lévy processes and random walks in [75].

We provide the proof of Theorem 2.6 in Section 3.3. A high-level description of the proof strategy for Theorem 2.5 is that we first establish the asymptotic equivalence between $\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x})$ and an ODE perturbed by the "large" noises in $(\boldsymbol{Z}_j)_{j\leq T/\eta}$ in terms of $\mathbb{M}$-convergence, and then study the limiting behavior of this large jump approximation. Once we establish Theorem 2.5, by sending $b \to \infty$ and analyzing the limits involved, we obtain the sample path large deviations for $\boldsymbol{X}^{\eta}_j(\boldsymbol{x})$ in Theorem 2.6. 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 $\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x})$. In fact, the conditional limit theorem below follows immediately from the sample-path large deviations established above, i.e., (2.20) and (2.17), and Portmanteau Theorem. The results in Section 2.2.2 can be easily extended to $\mathbb{D}\big([0,T],R^m\big)$ with arbitrary $T \in (0,\infty)$, and we focus on $\mathbb{D} = \mathbb{D}\big([0,1],\mathbb{R}^m\big)$ simply for the clarity of the presentation.

COROLLARY 2.7. *Let Assumptions 1 and 2 hold.*

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

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

(ii) *Furthermore, suppose that Assumption 3 holds. Given $k \in \mathbb{N}$, $\boldsymbol{x} \in \mathbb{R}^m$, and measurable $B \subseteq \mathbb{D}$, suppose that B is bounded away from $\mathbb{D}^{(k-1)}_{\{\boldsymbol{x}\}}(\epsilon)$ for some (and hence all) $\epsilon > 0$ small enough, 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 2.7 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 $\mathbf{C}^{(k)|b}$ in (2.15), its support belongs to the set of paths of the form $h^{(k)|b}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),(t_1,\cdots,t_k)\big)$, where the mapping $h^{(k)|b}$ is defined in (2.10)–(2.12), and the norms $\|\boldsymbol{w}_j\|$'s are bounded from below; see, for instance, Lemma 3.3 and 3.4. 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) of Corollary 2.7.

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. Again, the results for stochastic differential equations and/or under more general scaling regimes are collected in the Appendix.

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

$$\boldsymbol{y}_0(\boldsymbol{x}) = \boldsymbol{x}, \qquad \frac{d\boldsymbol{y}_t(\boldsymbol{x})}{dt} = \boldsymbol{a}\big(\boldsymbol{y}_t(\boldsymbol{x})\big) \;\; \forall t \geq 0. \tag{2.21}$$

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

ASSUMPTION 4. $\boldsymbol{a}(\mathbf{0}) = \mathbf{0}$. The open set $I \subset \mathbb{R}^m$ contains the origin and is bounded, i.e., $\sup_{\boldsymbol{x}\in I}\|\boldsymbol{x}\| < \infty$ and $\mathbf{0} \in I$. 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 $\mathcal{C}^2$ around 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. [71, 50].

Define

$$\tau^\eta(\boldsymbol{x}) \triangleq \min\{j \geq 0: \ \boldsymbol{X}^\eta_j(\boldsymbol{x}) \notin I\}, \ \ \tau^{\eta|b}(\boldsymbol{x}) \triangleq \min\{j \geq 0: \ \boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) \notin I\} \tag{2.22}$$

as the first exit time of $\boldsymbol{X}^\eta_j(\boldsymbol{x})$ and $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ 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:

$$\bar{g}^{(k)|b}\big(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, (t_1, \cdots, t_k)\big) \triangleq \bar{h}^{(k)|b}_{[0,t_k+1]}\Big(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, (t_1, \cdots, t_k)\Big)(t_k), \tag{2.23}$$

where $\bar{h}^{(k)|b}_{[0,T]}$ is the perturbed ODE mapping defined in (2.10)–(2.12). Note that the definition remains the same if, in (2.23), we use mapping $\bar{h}^{(k)|b}_{[0,T]}$ with any $T \in [t_k, \infty)$ instead of $\bar{h}^{(k)|b}_{[0,t_k+1]}$, and we choose the $+1$ only for the consistency and simplicity of the proof. 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

$$\breve{g}^{(k)|b}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t}) \triangleq \bar{g}^{(k)|b}\big(\boldsymbol{x}, \mathbf{W}, (\mathbf{0}, \cdots, \mathbf{0}), \boldsymbol{t}\big) = h^{(k)|b}_{[0,t_k+1]}(\boldsymbol{x}, \mathbf{W}, \boldsymbol{t})(t_k), \tag{2.24}$$

where $\boldsymbol{t} = (t_1, \ldots, t_k) \in (0,\infty)^{k\uparrow}$, and the mapping $h^{(k)|b}_{[0,T]}$ is defined in (2.13). For $k = 0$, we adopt the convention that $\bar{g}^{(0)|b}(\boldsymbol{x}) = \boldsymbol{x}$. With mappings $\bar{g}^{(k)|b}$ defined, we are able to introduce (for any $k \geq 1$, $b > 0$, and $\epsilon \geq 0$)

$$\begin{aligned} \mathcal{G}^{(k)|b}(\epsilon) \triangleq \Big\{ & \bar{g}^{(k-1)|b}\Big(\boldsymbol{v}_1 + \varphi_b\big(\boldsymbol{\sigma}(\boldsymbol{v}_1)\boldsymbol{w}_1\big), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_k), (\boldsymbol{v}_2, \cdots, \boldsymbol{v}_k), \boldsymbol{t}\Big): \\ & \mathbf{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k) \in \mathbb{R}^{d\times k}, \mathbf{V} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k) \in \Big(\bar{B}_\epsilon(\mathbf{0})\Big)^k, \boldsymbol{t} \in (0,\infty)^{k-1\uparrow} \Big\} \end{aligned} \tag{2.25}$$

as the set covered by the k^{th} jump of along ODE path initialized at the origin, with each jump modulated by $\boldsymbol{\sigma}(\cdot)$ and truncated under b (and an ϵ perturbation right before each jump). Here, the truncation operator φ_b is defined in (2.4), and $\bar{B}_r(\mathbf{0})$ is the closed ball with radius r centered at the origin. Under $\epsilon = 0$, we write

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

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

$$\mathcal{G}^{(0)|b}(\epsilon) \triangleq \bar{B}_\epsilon(\mathbf{0}). \tag{2.26}$$

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 2.8) 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 space reachable by ODE paths would expand as more jumps are added to the ODE path. This leads to an intriguing phase transition for the law of the first exit times $\tau^{\eta|b}(\boldsymbol{x})$ (as $\eta \downarrow 0$) in terms of the minimum number of jumps required for exit. More precisely, let

$$\mathcal{J}_b^I \triangleq \min\left\{k \geq 1 :\ \mathcal{G}^{(k)|b} \cap I^c \neq \emptyset\right\} \tag{2.27}$$

be the smallest k such that, under truncation at level b, the k-jump-coverage sets can reach outside the attraction field I. Theorem 2.8 reveals a discrete hierarchy that the order of the first exit time $\tau^{\eta|b}(\boldsymbol{x})$ and the limiting law of the exit location $\boldsymbol{X}^{\eta|b}_{\tau^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x})$ are dictated by this "discretized width" metric $\mathcal{J}_b^I$ of the domain I, relative to the truncation threshold b. Here, the limiting law is characterized by measures

$$\check{\mathbf{C}}^{(k)|b}(\,\cdot\,) \triangleq \int \mathbb{I}\Big\{\check{g}^{(k-1)|b}\Big(\varphi_b\big(\boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{w}_1\big), (\boldsymbol{w}_2, \cdots, \boldsymbol{w}_k), \boldsymbol{t}\Big) \in \,\cdot\,\Big\} \tag{2.28}$$
$$\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)^k(d\mathbf{W}) \times \mathcal{L}_\infty^{k-1\uparrow}(d\boldsymbol{t}),$$

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 (2.16), 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 **??** 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 2.8 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 proof of Theorem 2.8 is based on a general framework developed in Section 2.3.2, and we detail the proof in Section 4.2.

THEOREM 2.8. ***(First Exit Times and Locations: Truncated Case)*** *Let Assumptions 1, 2, and 4 hold. Let $b > 0$. Suppose that $\mathcal{J}_b^I < \infty$, 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 have $C_b^I = 0$, and

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

REMARK 5. Regarding the regularity conditions in Theorem 2.8, conditions of form $\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\partial I) = 0$ are standard even for metastability analyses of untruncated dynamics; see e.g. [43, 46]. Besides, we note that theses conditions hold almost automatically in the non-degenerate one-dimensional settings: suppose that $m = d = 1$ (so $\boldsymbol{Z}_j$'s and $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$'s are random variables in $\mathbb{R}^1$) and for the diffusion coefficient $\sigma : \mathbb{R} \to \mathbb{R}$ we have $\inf_{x \in I} \sigma(x) > 0$; then for (Lebesgue) almost every $b \in (0, \infty)$, I^c is bounded away from $\mathcal{G}^{(\mathcal{J}_b^I - 1)|b}(\epsilon)$ (for small $\epsilon > 0$), $\check{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\partial I) = 0$, and $C_b^I \in (0, \infty)$ with $\mathcal{J}_b^I = \inf_{x \notin I} \lceil |x|/b \rceil$. See Lemmas D.4 and D.5 in the full-length version of the paper (arXiv:2307.03479v4). The full-length version is also included as one of the related files (`full_length_paper.pdf`) of this submission.

REMARK 6. As noted in Section 1.1 and will be confirmed in Corollary 2.9 below, $\mathcal{J}_b^I = 1$ when $b = \infty$, regardless of the geometry of $\boldsymbol{a}(\cdot)$. In this case, Theorem 2.8 reduces to the manifestation of the principle of a single big jump. For $b \neq \infty$ and a contractive drift—i.e., $a(\boldsymbol{x}) \cdot \boldsymbol{x} \leq 0$ for all $\boldsymbol{x} \in I$—note that $\mathcal{J}_b^I = \lceil r/b \rceil$, where $r \triangleq \inf\{\|\boldsymbol{x} - \mathbf{0}\| : \boldsymbol{x} \in I^c\}$. This is because gradient flow will not bring $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ closer to I^c, and hence, the most efficient way to escape from I is through $\lceil r/b \rceil$ consecutive jumps in the direction where I^c is closest. In the general case, however, $\mathcal{J}_b^I$ is determined as the solution to the discrete optimization problem in (2.27), where the geometry of $\boldsymbol{a}(\cdot)$—in particular, gradient flows and their distances from I^c—plays a more sophisticated role.

To conclude this section, we note that the first exit analysis for untruncated heavy-tailed dynamics (see e.g. [47, 49, 71, 50] for analogous results under different scaling in the existing literature) in fact follows directly from Theorem 2.8. Let

$$\check{\mathbf{C}}(\,\cdot\,) \triangleq \int \mathbb{I}\Big\{\boldsymbol{\sigma}(\mathbf{0})\boldsymbol{w} \in \,\cdot\,\Big\}\big((\nu_\alpha \times \mathbf{S}) \circ \Phi\big)(d\boldsymbol{w}). \tag{2.29}$$

The asymptotic analysis for exit times and locations of the untruncated dynamics $\boldsymbol{X}_j^{\eta}(\boldsymbol{x})$ follows from the result for $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ 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 2.9. The proof is straightforward and we collect it in Section **??** for the sake of completeness.

COROLLARY 2.9. ***(First Exit Times and Locations: Untruncated Case)*** *Let Assumptions 1, 2, and 4 hold. Suppose that $\check{\mathbf{C}}(\partial I) = 0$ and $\|\boldsymbol{\sigma}(\mathbf{0})\| > 0$. Then $C_\infty^I \triangleq \check{\mathbf{C}}(I^c) < \infty$. Furthermore, if $C_\infty^I > 0$, then for any $t \geq 0$, $\epsilon > 0$, and measurable set $B \subseteq I^c$,*

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\bigg(C_\infty^I H(\eta^{-1})\tau^\eta(\boldsymbol{x}) > t;\ \boldsymbol{X}^\eta_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg) \leq \frac{\check{\mathbf{C}}(B^-)}{C_\infty^I} \cdot \exp(-t),$$

$$\liminf_{\eta \downarrow 0} \inf_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\bigg(C_\infty^I H(\eta^{-1})\tau^\eta(\boldsymbol{x}) > t;\ \boldsymbol{X}^\eta_{\tau^\eta(\boldsymbol{x})}(\boldsymbol{x}) \in B\bigg) \geq \frac{\check{\mathbf{C}}(B^\circ)}{C_\infty^I} \cdot \exp(-t).$$

Otherwise, we have $C_\infty^I = 0$, and

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in I_\epsilon} \mathbf{P}\bigg(H(\eta^{-1})\tau^\eta(\boldsymbol{x}) \leq t\bigg) = 0 \qquad \forall \epsilon > 0,\ t \geq 0.$$

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 $\big\{\{V_j^\eta(x) : j \geq 0\} : \eta > 0\big\}$ 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 $\{V_j^\eta(x) : j \geq 0\}$ as a $\mathbb{D}[0,T]$-valued random element. For a given set E, let $\tau^\eta_E(x) \triangleq \min\{j \geq 0 : V_j^\eta(x) \in E\}$ denote $\{V_j^\eta(s) : j \geq 0\}$'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 2.11;

(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 2.10. $\big\{\{V_j^\eta(x) : j \geq 0\} : \eta > 0\big\}$ 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

$$C(B^\circ) - \delta_B(\epsilon,T) \leq \liminf_{\eta\downarrow 0} \frac{\inf_{x\in A(\epsilon)} \mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(x) \leq T/\eta;\ V^\eta_{\tau_\epsilon}(x) \in B\big)}{\gamma(\eta)T/\eta} \tag{2.30}$$

$$\leq \limsup_{\eta\downarrow 0} \frac{\sup_{x\in A(\epsilon)} \mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(x) \leq T/\eta;\ V^\eta_{\tau_\epsilon}(x) \in B\big)}{\gamma(\eta)T/\eta} \leq C(B^-) + \delta_B(\epsilon,T) \tag{2.31}$$

$$\limsup_{\eta\downarrow 0} \frac{\sup_{x\in I(\epsilon)} \mathbf{P}\big(\tau^\eta_{(I(\epsilon)\setminus A(\epsilon))^c}(x) > T/\eta\big)}{\gamma(\eta)T/\eta} = 0 \tag{2.32}$$

$$\liminf_{\eta\downarrow 0} \inf_{x\in I(\epsilon)} \mathbf{P}\big(\tau^\eta_{A(\epsilon)}(x) \leq T/\eta\big) = 1 \tag{2.33}$$

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 2.10 asymptotically characterize the atoms in $V^\eta_\cdot(x)$ for the first exit analysis from domain I, note that the condition (2.33) requires the process to efficiently return to the asymptotic atoms $A(\epsilon)$. The conditions (2.30) and (2.31) 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 (2.32) 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 2.8. 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 $\{I(\epsilon) \subseteq I : \epsilon > 0\}$.

Right before Definition 2.10, 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.

THEOREM 2.11. *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 $t \geq 0$,*

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

(ii) *If $C(I^c) = 0$, then $\lim_{\eta \downarrow 0} \sup_{x \in I(\epsilon)} \mathbf{P}\Big(\gamma(\eta)\tau^\eta_{I^c}(x) \leq t\Big) = 0$ holds for each $\epsilon, t > 0$.*

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) and Lemma 2.4. 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 2.5 and 2.6) in Section 3.3.

3.1. *Proof of Theorem 2.2 and Lemma 2.4.*

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 $\limsup_{\eta \downarrow 0} \sup_{\theta \in \Theta} \mu^\eta_\theta(F) - \mu_\theta(F^\epsilon) \leq 0$ 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 $\mu^{\eta_n}_{\theta_n}(F) - \mu_{\theta_n}(F^\epsilon) > \delta\ \forall n \geq 1$. Now, we make two observations. First, using Urysohn's lemma (see, e.g., lemma 2.3 of [59]), 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 $\mu^{\eta_n}_{\theta_n}(F) - \mu_{\theta_n}(F^\epsilon) \leq \mu^{\eta_n}_{\theta_n}(f) - \mu_{\theta_n}(f)$ for each n. Secondly, from statement (ii) we get $\lim_{n \to \infty} \big|\mu^{\eta_n}_{\theta_n}(f) - \mu_{\theta_n}(f)\big| = 0$. In summary, we yield the contradiction

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

Analogously, if the claim $\liminf_{\eta\downarrow 0}\inf_{\theta\in\Theta}\mu^{\eta}_{\theta}(G)-\mu_{\theta}(G^{\epsilon})\geq 0$, 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 $\lim_{\eta\downarrow 0}\sup_{\theta\in\Theta}\left|\mu^{\eta}_{\theta}(g)-\mu_{\theta}(g)\right|=0$ 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

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

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 $\limsup_{n\to\infty}\mu^{\eta_n}_{\theta_n}(F)-\mu_{\theta_n}(F^{\epsilon})\leq 0$ and $\liminf_{n\to\infty}\mu^{\eta_n}_{\theta_n}(G)-\mu_{\theta_n}(G_{\epsilon})\geq 0$ 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 [59]), 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 $\epsilon>0$,

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

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 $\lim_{n\to\infty}\left|\mu^{\eta_n}_{\theta_n}(g)-\mu_{\theta^*}(g)\right|=0$ for the $g\in\mathcal{C}(\mathbb{S}\setminus\mathbb{C})$ fixed in (3.1). However, recall that we have already obtained $\lim_{n\to\infty}\big|\mu_{\theta_n}(g)-\mu_{\theta^*}(g)\big|=0$ using assumption (2.1). We now arrive at the contradiction

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

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 $\limsup_{\eta\downarrow 0}\sup_{\theta\in\Theta}\mu^{\eta}_{\theta}(F)\leq\sup_{\theta\in\Theta}\mu_{\theta}(F)$ 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 $\mu^{\eta_n}_{\theta_n}(F)>\sup_{\theta\in\Theta}\mu_{\theta}(F)+\delta\ \forall n\geq 1$. 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. $\square$

PROOF OF LEMMA 2.4. We arbitrarily pick some Borel measurable $B \subseteq \mathbb{S}$ that is bounded away from $\mathbb{C}$. Henceforth in this proof, we only consider $\Delta > 0$ small enough that $\boldsymbol{d}(B,\mathbb{C}) > \Delta$, and hence B^{Δ} is still bounded away from $\mathbb{C}$. Observe that

$$\begin{aligned}
\mathbf{P}(X_n \in B) &\leq \mathbf{P}\big(X_n \in B;\ \boldsymbol{d}(X_n, Y_n^{\delta}) \leq \Delta\big) + \mathbf{P}\big(X_n \in B;\ \boldsymbol{d}(X_n, Y_n^{\delta}) > \Delta\big) \\
&\leq \mathbf{P}\big(Y_n^{\delta} \in B^{\Delta}\big) + \mathbf{P}\big(X_n \in B \text{ or } Y_n^{\delta} \in B;\ \boldsymbol{d}(X_n, Y_n^{\delta}) > \Delta\big).
\end{aligned}$$

As a result,

$$\begin{aligned}
&\limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \\
&\leq \limsup_{\delta\downarrow 0}\limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}\big(Y_n^{\delta} \in B^{\Delta}\big) \\
&\qquad + \limsup_{\delta\downarrow 0}\limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}\Big(\boldsymbol{d}\big(X_n, Y_n^{\delta}\big)\mathbb{I}\big(X_n \in B \text{ or } Y_n^{\delta} \in B\big) > \Delta\Big) \\
&\leq \mu(B^{\Delta}) \qquad \text{by conditions (i) and (ii) of Lemma 2.4.}
\end{aligned} \tag{3.2}$$

Analogously, observe the lower bound

$$\begin{aligned}
\mathbf{P}(X_n \in B) &\geq \mathbf{P}\big(X_n \in B;\ \boldsymbol{d}(X_n, Y_n^{\delta}) \leq \Delta\big) \\
&\geq \mathbf{P}\big(Y_n^{\delta} \in B_{\Delta};\ \boldsymbol{d}(X_n, Y_n^{\delta}) \leq \Delta\big) \\
&\geq \mathbf{P}(Y_n^{\delta} \in B_{\Delta}) - \mathbf{P}\big(Y_n^{\delta} \in B_{\Delta};\ \boldsymbol{d}(X_n, Y_n^{\delta}) > \Delta\big) \\
&\geq \mathbf{P}(Y_n^{\delta} \in B_{\Delta}) - \mathbf{P}\big(Y_n^{\delta} \in B \text{ or } X_n \in B;\ \boldsymbol{d}(X_n, Y_n^{\delta}) > \Delta\big),
\end{aligned}$$

and hence

$$\begin{aligned}
&\liminf_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \\
&\geq \liminf_{\delta\downarrow 0}\liminf_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(Y_n^{\delta} \in B_{\Delta}) \\
&\qquad - \limsup_{\delta\downarrow 0}\limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}\Big(\boldsymbol{d}\big(X_n, Y_n^{\delta}\big)\mathbb{I}\big(X_n \in B \text{ or } Y_n^{\delta} \in B\big) > \Delta\Big) \\
&\geq \mu(B_{\Delta}) \qquad \text{by conditions (i) and (ii) of Lemma 2.4.}
\end{aligned} \tag{3.3}$$

Since $\mu \in \mathbb{M}(\mathbb{S}\setminus\mathbb{C})$ and B^{Δ} is bounded away from $\mathbb{C}$, we have $\mu(B^{\Delta}) < \infty$. By sending $\Delta \downarrow 0$ in (3.2) and (3.3), it then follows from the continuity of measure μ that

$$\mu(B^{\circ}) \leq \liminf_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \leq \limsup_{n\to\infty} \epsilon_n^{-1}\mathbf{P}(X_n \in B) \leq \mu(B^{-}).$$

Due to the arbitrariness of our choice of B, we conclude the proof using Theorem 2.1 of [59], which is the Portmanteau theorem for the standard $\mathbb{M}(\mathbb{S}\setminus\mathbb{C})$-convergence and, essentially, a special case of Theorem 2.2. □

3.2. *Technical Lemmas for Theorems 2.5 and 2.6.* This section states technical lemmas for the proof of Theorems 2.5 and 2.6. Their proofs can be found in Section 3.2 of the full-length version of this paper (arXiv:2307.03479v4). The full-length version is also included as one of the related files (`full_length_paper.pdf`) of this submission.

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 $\eta\,\|\boldsymbol{Z}_j\|$ 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

$$\tau_i^{>\delta}(\eta) \triangleq \min\{n > \tau_{i-1}^{>\delta}(\eta) :\ \eta\|\boldsymbol{Z}_j\| > \delta\}, \quad \tau_0^{>\delta}(\eta) = 0 \tag{3.4}$$

$$\boldsymbol{W}_i^{>\delta}(\eta) \triangleq \boldsymbol{Z}_{\tau_i^{>\delta}(\eta)}. \tag{3.5}$$

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

$$\begin{aligned}
\mathbf{P}\Big(\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor\Big) &\leq \mathbf{P}\Big(\tau_j^{>\delta}(\eta) - \tau_{j-1}^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor \ \ \forall j \in [k]\Big) \\
&= \Big[\sum_{i=1}^{\lfloor 1/\eta \rfloor} \big(1 - H(\delta/\eta)\big)^{i-1} H(\delta/\eta)\Big]^k \leq \Big[\sum_{i=1}^{\lfloor 1/\eta \rfloor} H(\delta/\eta)\Big]^k \\
&\leq \Big[1/\eta \cdot H(\delta/\eta)\Big]^k. \qquad (3.6)
\end{aligned}$$

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:

$$\boldsymbol{\Gamma}_M \triangleq \Big\{(\mathbf{V}_j)_{j \geq 0} \text{ is adapted to } \mathbb{F} :\ \|\mathbf{V}_j\| \leq M\ \forall j \geq 0 \text{ almost surely}\Big\}. \tag{3.7}$$

LEMMA 3.1. *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*

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

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

$$A_i(\eta, b, \epsilon, \delta, \boldsymbol{x}) \triangleq \Bigg\{\max_{j \in I_i(\eta, \delta)} \eta \Bigg\|\sum_{n = \tau_{i-1}^{>\delta}(\eta) + 1}^{j} \boldsymbol{\sigma}\big(\boldsymbol{X}_{n-1}^{\eta|b}(\boldsymbol{x})\big) \boldsymbol{Z}_n\Bigg\| \leq \epsilon\Bigg\}; \tag{3.8}$$

$$I_i(\eta, \delta) \triangleq \Big\{j \in \mathbb{N} :\ \tau_{i-1}^{>\delta}(\eta) + 1 \leq j \leq \big(\tau_i^{>\delta}(\eta) - 1\big) \wedge \lfloor 1/\eta \rfloor\Big\}. \tag{3.9}$$

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

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

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

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

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

$$E^{\delta}_{c,k}(\eta) \triangleq \Big\{\tau_k^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor < \tau_{k+1}^{>\delta}(\eta);\ \eta\Big\|\boldsymbol{W}_j^{>\delta}(\eta)\Big\| > c\ \ \forall j \in [k]\Big\}. \tag{3.10}$$

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

$$\mathbf{P}\Big(\Big(R^*(c), \boldsymbol{\Theta}^*(c)\Big) \in \ \cdot\ \Big) = \big(\bar{\nu}_\alpha|_{(c,\infty)} \times \mathbf{S}\big)(\cdot). \tag{3.11}$$

Here, recall the definition of the measure ν_α in (2.6) 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 $\bar{\nu}_\alpha|_{(c,\infty)}\big(\ \cdot\ \big) \triangleq c^\alpha \cdot \nu_\alpha\Big(\ \cdot \cap (c,\infty)\Big)$ 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 3.2. *Let Assumption 1 hold. For any $\delta > 0, c \geq \delta$ and $k \in \mathbb{Z}^+$,*

$$\lim_{\eta \downarrow 0} \frac{\mathbf{P}\big(E^{\delta}_{c,k}(\eta)\big)}{\lambda^k(\eta)} = \frac{1/c^{\alpha k}}{k!}, \tag{3.12}$$

and

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

The next few lemmas collect useful properties regarding the mappings and $\bar{h}^{(k)|b}_{[0,T]}$ and $\bar{h}^{(k)}_{[0,T]} = \bar{h}^{(k)|\infty}_{[0,T]}$ defined in (2.10)–(2.13) as well as the images under these mappings.

LEMMA 3.3. *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 $\mathbb{D}_A^{(k-1)}(r)$, 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) $\boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)}(r)\big) > 0.$

LEMMA 3.4. *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 $\mathbb{D}_A^{(k-1)|b}(r)$, then there exist $\bar{\epsilon} > 0$ and $\bar{\delta} > 0$ such that the following claims hold:*

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

LEMMA 3.5. *Let Assumption 2 hold. Given any $b, T \in (0,\infty)$ and any $k \in \mathbb{N}$, the mapping $\bar{h}^{(k)|b}_{[0,T]}$ is continuous on $\mathbb{R}^m \times \mathbb{R}^{d\times k} \times \mathbb{R}^{m\times k} \times (0,T)^{k\uparrow}$.*

In Lemma 3.6, we show that the image of $h^{(1)}$ (resp. $h^{(1)|b}$) provides good approximations of the sample path of $\boldsymbol{X}^{\eta}_j(\boldsymbol{x})$ (resp. $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$) up until $\tau^{>\delta}_1(\eta)$, i.e. the arrival time of the first "large noise"; see (3.4),(3.5) for the definition of $\tau^{>\delta}_i(\eta), \boldsymbol{W}^{>\delta}_i(\eta)$.

LEMMA 3.6. *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*

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

that

$$\sup_{t\in[0,1]:\ t<\eta\tau^{>\delta}_1(\eta)} \Big\| \xi_t - \boldsymbol{X}^{\eta}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) \Big\| \le \rho \cdot \big(\epsilon + \|\boldsymbol{x}-\boldsymbol{y}\| + \eta C\big), \tag{3.13}$$

where

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

(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*

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

that

$$\sup_{t\in[0,1]:\ t<\eta\tau^{>\delta}_1(\eta)} \Big\| \xi_t - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) \Big\| \le \rho \cdot \big(\epsilon + \|\boldsymbol{x}-\boldsymbol{y}\| + \eta C\big), \tag{3.14}$$

$$\sup_{t\in[0,1]:\ t\le\eta\tau^{>\delta}_1(\eta)} \Big\| \xi_t - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) \Big\| \le \rho D \cdot \big(\epsilon + \|\boldsymbol{x}-\boldsymbol{y}\| + 2\eta C\big) \cdot \epsilon^{-\gamma} \tag{3.15}$$

where

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

By applying Lemma 3.6 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 3.7. *Let Assumptions 2 and 3 hold. Let $A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})$ be defined as in (3.8). 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*

$$\bigg(\bigcap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x}) \bigg) \cap \Big\{ \tau^{>\delta}_k(\eta) \le \lfloor 1/\eta \rfloor < \tau^{>\delta}_{k+1}(\eta) \Big\} \cap \Big\{ \eta \Big\| \boldsymbol{W}^{>\delta}_i(\eta) \Big\| \le 1/\epsilon^{\frac{1}{2k}}\ \forall i \in [k] \Big\}$$

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*

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

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

Lemma 3.8 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 3.8. *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* $\lim_{n\to\infty}\boldsymbol{x}_n = \boldsymbol{x}^*$,

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

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

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

Lastly, to facilitate the proof of Theorem 2.5, we consider a "truncated" version of the drift and diffusion coefficients $\boldsymbol{a}(\cdot), \boldsymbol{\sigma}(\cdot)$. Given any $M \geq 1$, let

(3.16)
$$\boldsymbol{a}_M(\boldsymbol{x}) \triangleq \begin{cases} \boldsymbol{a}\Big(M\cdot\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\Big) & \text{if } \|\boldsymbol{x}\| > M, \\ \boldsymbol{a}(\boldsymbol{x}) & \text{otherwise.}\end{cases} \qquad \boldsymbol{\sigma}_M(\boldsymbol{x}) \triangleq \begin{cases} \boldsymbol{\sigma}\Big(M\cdot\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\Big) & \text{if } \|\boldsymbol{x}\| > M, \\ \boldsymbol{\sigma}(\boldsymbol{x}) & \text{otherwise.}\end{cases}$$

That is, we project $\boldsymbol{x}$ onto the closed ball $\{\boldsymbol{x}\in\mathbb{R}^m:\ \|\boldsymbol{x}\| \leq M\}$. 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, for the mapping $\bar{h}^{(k)|b}$ in (2.10)-(2.12), we 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. Recall that $(0,1]^{k\uparrow} = \{(t_1,\ldots,t_k):\ 0 < t_1 < \ldots < t_k \leq 1\}$. 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

$$\xi_0 = \boldsymbol{x}; \tag{3.17}$$

$$\frac{d\xi_t}{dt} = \boldsymbol{a}_M(\xi_t) \qquad \forall t\in[0,1],\ t \neq t_1, t_2, \cdots, t_k; \tag{3.18}$$

$$\xi_t = \xi_{t-} + \boldsymbol{v}_j + \varphi_b\big(\boldsymbol{\sigma}_M(\xi_{t-}+\boldsymbol{v}_j)\boldsymbol{w}_j\big) \qquad \text{if } t = t_j \text{ for some } j\in[k]. \tag{3.19}$$

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

$$h^{(k)|b}_{M\downarrow}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),\boldsymbol{t}\big) \triangleq \bar{h}^{(k)|b}_{M\downarrow}\big(\boldsymbol{x},(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_k),(\mathbf{0},\cdots,\mathbf{0}),\boldsymbol{t}\big). \tag{3.20}$$

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

$$\mathbb{D}^{(k)|b}_{A;M\downarrow}(r) \triangleq \bar{h}^{(k)|b}_{M\downarrow}\Big(A\times\mathbb{R}^{m\times k}\times\big(\bar{B}_r(\mathbf{0})\big)^k\times(0,1]^{k\uparrow}\Big). \tag{3.21}$$

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 $\boldsymbol{\sigma}_M(\cdot)$. Recall the definition of $\mathbb{D}^{(k)|b}_A(r)$ in (2.14). The next lemma establishes the boundedness of paths in sets of form $\mathbb{D}^{(k)|b}_A(r)$ in terms of $\|\xi\| \triangleq \sup_{t\in[0,1]}\|\xi(t)\|$, and is useful for the subsequent analysis.

LEMMA 3.9. *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* $\boldsymbol{x}\in A$,

$$\bar{h}^{(k)|b}(\boldsymbol{x},\boldsymbol{W},\boldsymbol{V},\boldsymbol{t})=\bar{h}_{M\downarrow}^{(k)|b}(\boldsymbol{x},\boldsymbol{W},\boldsymbol{V},\boldsymbol{t}).$$

3.3. *Proofs of Theorem 2.5.* In this section, 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 $\boldsymbol{X}^{\eta|b}(\boldsymbol{x})=\boldsymbol{X}^{\eta|b}_{[0,1]}(\boldsymbol{x})=\{\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}):t\in[0,1]\}$.

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 2.5 is stronger than the standard $\mathbb{M}$-convergence introduced in [59]. Nevertheless, under the conditions stated in Theorem 2.5 regarding the initial values of $\boldsymbol{X}^{\eta|b}$, we can show that it suffices to prove the standard notion of $\mathbb{M}$-convergence, and the proof of Theorem 2.5 hinges on the following proposition.

PROPOSITION 3.10. *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*

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

We defer the proof of Proposition 3.10 to Section 3.3.1, and first demonstrate its application in the proof of Theorem 2.5.

PROOF OF THEOREM 2.5. For the first half of the proof, we impose the boundedness condition in Assumption 3 and prove the claims of Theorem 2.5. In the second half of the proof, we explain how to extend the proof and cover the case where Assumption 3 is dropped.

To establish the uniform $\mathbb{M}$-convergence under Assumption 3, we consider a proof by contradiction. Fix some $r,b>0$ and $k\in\mathbb{N}$, and suppose that there exist some $f\in\mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)$, 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

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

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 3.10 and yield $\lim_{n\to\infty}\big|\mu_n^{(k)}(f)-\mathbf{C}^{(k)|b}(f;\boldsymbol{x}^*)\big|=0$. On the other hand, under Assumption 3 we are able to apply part (b) of Lemma 3.8 and get $\lim_{n\to\infty}\big|\mathbf{C}^{(k)|b}(f;\boldsymbol{x}_n)-\mathbf{C}^{(k)|b}(f;\boldsymbol{x}^*)\big|=0$. We now arrive at the contradiction

$$\begin{aligned}&\lim_{n\to\infty}\big|\mu_n^{(k)}(f)-\mathbf{C}^{(k)|b}(f;\boldsymbol{x}_n)\big|\\&\leq\lim_{n\to\infty}\big|\mu_n^{(k)}(f)-\mathbf{C}^{(k)|b}(f;\boldsymbol{x}^*)\big|+\lim_{n\to\infty}\big|\mathbf{C}^{(k)|b}(f;\boldsymbol{x}^*)-\mathbf{C}^{(k)|b}(f;\boldsymbol{x}_n)\big|=0\end{aligned}$$

and conclude the proof of the uniform $\mathbb{M}$-convergence claim. Next, we prove the uniform sample-path large deviations stated in (2.17) under Assumption 3. Part (b) of Lemma 3.8 verifies the compactness condition (2.1) for the family of measures $\{\mathbf{C}^{(k)|b}(\ \cdot\ ;\boldsymbol{x}):\ \boldsymbol{x}\in A\}$.

The claim (2.17) then follows directly from Theorem 2.2, and it only remains to verify that $\sup_{\boldsymbol{x}\in A}\mathbf{C}^{(k)|b}\big(B^-;\boldsymbol{x}\big)<\infty$. To do so, note that B^- is bounded away from $\mathbb{D}_A^{(k-1)|b}(r)$ for some $r>0$. This allows us to apply Lemma 3.4 and find $\bar{\epsilon}>0$, $\bar{\delta}>0$ such that, for any $\boldsymbol{x}\in A$ and $\boldsymbol{t}\in(0,1]^{k\uparrow}$,

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

Then, by the definition of $\mathbf{C}^{(k)|b}$ in (2.15),

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

This concludes the proof under Assumption 3.

In the remainder of this proof, we extend the proof to 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 exist some $g\in\mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)$, 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

(3.22)
$$\Big|\mu_n^{(k)}(g)-\mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big|>\epsilon\ \forall n\geq 1\qquad\text{with }\mu_n^{(k)}(\cdot)\triangleq\mathbf{P}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\in\ \cdot\ \big)/\lambda^k(\eta_n).$$

Again, due to the compactness of A, we w.l.o.g. assume that $\lim_{n\to\infty}\boldsymbol{x}_n=\boldsymbol{x}^*$ for some $\boldsymbol{x}^*\in A$. Let $B\triangleq\text{supp}(g)$ and note that B is bounded away from $\mathbb{D}_A^{(k-1)|b}(r)$. Applying Lemma 3.9, we can fix some M_0 such that the following claim holds for each $M\geq M_0$: 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 $\boldsymbol{x}\in A$, we have

$$\xi=\bar{h}^{(k)|b}(\boldsymbol{x},\mathbf{W},\mathbf{V},\boldsymbol{t})=\bar{h}_{M\downarrow}^{(k)|b}(\boldsymbol{x},\mathbf{W},\mathbf{V},\boldsymbol{t})\qquad\text{and}\qquad\sup_{t\in[0,1]}\|\xi_t\|\leq M_0.\tag{3.23}$$

Here, recall that the mappings $\bar{h}_{M\downarrow}^{(k)|b}$ and $h_{M\downarrow}^{(k)|b}$ are defined in (3.17)–(3.20). Now, we fix some $M\geq M_0+1$ and recall the definitions of $\boldsymbol{a}_M$, $\boldsymbol{\sigma}_M$ in (3.16). 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

$$\widetilde{\boldsymbol{X}}^{\eta|b}_j(\boldsymbol{x})=\widetilde{\boldsymbol{X}}^{\eta|b}_{j-1}(\boldsymbol{x})+\varphi_b\Big(\eta\boldsymbol{a}_M\big(\widetilde{\boldsymbol{X}}^{\eta|b}_{j-1}(\boldsymbol{x})\big)+\eta\boldsymbol{\sigma}_M\big(\widetilde{\boldsymbol{X}}^{\eta|b}_{j-1}(\boldsymbol{x})\big)\boldsymbol{Z}_j\Big)\ \ \forall j\geq 1\tag{3.24}$$

under initial condition $\widetilde{\boldsymbol{X}}^{\eta|b}_0(\boldsymbol{x})=\boldsymbol{x}$. In particular, by comparing the definition of $\widetilde{\boldsymbol{X}}^{\eta|b}_j(\boldsymbol{x})$ with that of $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$ in (2.3), one can see that (for any $\boldsymbol{x}\in\mathbb{R}^m,\eta>0$)

$$\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\Big\|>M\iff\sup_{t\in[0,1]}\Big\|\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\Big\|>M,\tag{3.25}$$

$$\sup_{t\in[0,1]}\Big\|\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\Big\|\leq M\implies\boldsymbol{X}^{\eta|b}(\boldsymbol{x})=\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x}).\tag{3.26}$$

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

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

Due to (3.23), we must have

$$\widetilde{\mathbf{C}}^{(k)|b}(\,\cdot\,;\boldsymbol{x}) = \mathbf{C}^{(k)|b}(\,\cdot\,;\boldsymbol{x}) \qquad \forall \boldsymbol{x} \in A. \tag{3.27}$$

Next, recall that $\boldsymbol{a}_M$ and $\boldsymbol{\sigma}_M$ satisfy Assumption 3. As has been established in the first half of this proof, we have

$$\lambda^{-k}(\eta)\mathbf{P}\big(\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x}) \in \,\cdot\,\big) \to \widetilde{\mathbf{C}}^{(k)|b}(\,\cdot\,;\boldsymbol{x}) = \mathbf{C}^{(k)|b}(\,\cdot\,;\boldsymbol{x}) \tag{3.28}$$

in $\mathbb{M}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)$ uniformly in $\boldsymbol{x}$ on A as $\eta\downarrow 0$. Therefore, for the function $g\in\mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)$ fixed above we have

$$\lim_{n\to\infty}\big|\tilde{\mu}_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\big| = 0 \qquad \text{with } \tilde{\mu}_n^{(k)}(\cdot) \triangleq \mathbf{P}\big(\widetilde{\boldsymbol{X}}^{\eta_n|b}(\boldsymbol{x}_n)\in\,\cdot\,\big)\big/\lambda^k(\eta_n). \tag{3.29}$$

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

$$\begin{aligned}\mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big] &= \mathbf{E}\Bigg[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\mathbb{I}\Big\{\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\in B;\ \sup_{t\in[0,1]}\Big\|\boldsymbol{X}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\|\leq M\Big\}\Bigg] \\ &\quad + \mathbf{E}\Bigg[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\mathbb{I}\Big\{\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\in B;\ \sup_{t\in[0,1]}\Big\|\boldsymbol{X}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\Big\}\Bigg].\end{aligned} \tag{3.30}$$

The following bound follows from (3.25) and (3.26):

$$\left|\mathbf{E}\Big[g\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big] - \mathbf{E}\Big[g\big(\widetilde{\boldsymbol{X}}^{\eta_n|b}(\boldsymbol{x}_n)\big)\Big]\right| \leq \|g\|\,\mathbf{P}\left(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\right). \tag{3.31}$$

Furthermore, we claim that

$$\lim_{n\to\infty}\lambda^{-k}(\eta_n)\mathbf{P}\left(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\right) = 0. \tag{3.32}$$

Then,

$$\begin{aligned}&\limsup_{n\to\infty}\Big|\mu_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\ &\leq \limsup_{n\to\infty}\Big|\mu_n^{(k)}(g) - \tilde{\mu}_n^{(k)}(g)\Big| + \limsup_{n\to\infty}\Big|\tilde{\mu}_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\ &\leq \|g\|\cdot\limsup_{n\to\infty}\lambda^{-k}(\eta_n)\mathbf{P}\left(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\right) + 0 \qquad \text{due to (3.31) and (3.29)} \\ &= 0 \qquad \text{due to (3.32).}\end{aligned}$$

In summary, we end up with a clear contradiction to (3.22), which allows us to conclude the proof. Now, it only remains to prove the claim (3.32).

Proof of Claim (3.32). 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 for $\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x})$ in (3.28), we get $\limsup_{n\to\infty}\mathbf{P}\Big(\widetilde{\boldsymbol{X}}^{\eta_n|b}(\boldsymbol{x}_n)\in E\Big)\Big/\lambda^{k+1}(\eta_n) < \infty$ and hence (3.32). To see why E is bounded away from $\mathbb{D}_A^{(k)|b}(r)$, note that by (3.23),

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

under our choice of $M \geq M_0 + 1$. This confirms $\boldsymbol{d}_{J_1}\big(\mathbb{D}_A^{(k)|b}(r), E\big) \geq 1$. □

3.3.1. *Proof of Proposition 3.10.* This section is devoted to the proof of Proposition 3.10. Analogous to the proof of Theorem 2.5 above, we show that it suffices to prove the seemingly more restrictive results stated below in Proposition 3.11, where we impose the the boundedness condition in Assumption 3.

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, 2, and 3, it holds for all $k \in \mathbb{N}$ and $b, r > 0$ that*

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

PROOF OF PROPOSITION 3.10. The proof is almost identical to the second half of the proof for Theorem 2.5. Specifically, we fix some $M \geq M_0 + 1$ with M_0 specified in (3.23), and we arbitrarily pick some $g \in \mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)$. Besides, define the stochastic processes $\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x}) \triangleq \big\{\widetilde{\boldsymbol{X}}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}) :\ t \in [0,1]\big\}$ as in (3.24), and note that the claims (3.27) and (3.31) in Theorem 2.5 still hold. Next, by applying Proposition 3.11 onto $\widetilde{\boldsymbol{X}}^{\eta|b}(\boldsymbol{x})$, we again obtain (3.29) and (3.32) (in particular, for the claim (3.32), note that at the end of the proof for Theorem 2.5 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 $\mu_n^{(k)}(\cdot) \triangleq \mathbf{P}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \in \cdot\,\big)\big/\lambda^k(\eta_n)$ and $\tilde{\mu}_n^{(k)}(\cdot) \triangleq \mathbf{P}\big(\widetilde{\boldsymbol{X}}^{\eta_n|b}(\boldsymbol{x}_n) \in \cdot\,\big)\big/\lambda^k(\eta_n)$, observe that

$$\begin{aligned}
&\lim_{n\to\infty}\Big|\mu_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\
&\leq \limsup_{n\to\infty}\Big|\mu_n^{(k)}(g) - \tilde{\mu}_n^{(k)}(g)\Big| + \limsup_{n\to\infty}\Big|\tilde{\mu}_n^{(k)}(g) - \mathbf{C}^{(k)|b}(g;\boldsymbol{x}_n)\Big| \\
&\leq \|g\|\cdot\limsup_{n\to\infty}\lambda^{-k}(\eta_n)\mathbf{P}\left(\sup_{t\in[0,1]}\Big\|\widetilde{\boldsymbol{X}}^{\eta_n|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x}_n)\Big\| > M\right) + 0 \qquad \text{due to (3.31) and (3.29)} \\
&= 0 \qquad \text{due to (3.32)}.
\end{aligned}$$

We conclude the proof by Portmanteau theorem for $\mathbb{M}$-convergence (see theorem 2.1 of [59], a special case of our Theorem 2.2) and the arbitrariness of $g \in \mathcal{C}\big(\mathbb{D}\setminus\mathbb{D}_A^{(k-1)|b}(r)\big)$. □

The rest of Section 3.3.1 is devoted to establishing Proposition 3.11. In light of Lemma 2.4, a natural approach to Proposition 3.11 is to construct some stochastic process whose law is asymptotically equivalent to $\boldsymbol{X}^{\eta|b}$ and converges to the limiting measure $\mathbf{C}^{(k)|b}$ stated in Proposition 3.11. Specifically, recall the definitions of $\tau_i^{>\delta}(\eta)$ and $\boldsymbol{W}_i^{>\delta}(\eta)$ in (3.4)–(3.5). Given $\eta, b, \delta > 0$ and $\boldsymbol{x} \in \mathbb{R}^m$, we define $\hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x}) \triangleq \big\{\hat{\boldsymbol{X}}^{\eta|b;>\delta}_t(\boldsymbol{x}) :\ t \in [0,1]\big\}$ by

(3.33)
$$\frac{d\hat{\boldsymbol{X}}^{\eta|b;>\delta}_t(\boldsymbol{x})}{dt} = \boldsymbol{a}\big(\hat{\boldsymbol{X}}^{\eta|b;>\delta}_t(\boldsymbol{x})\big) \quad \forall t \geq 0,\ t \notin \big\{\eta\tau_i^{>\delta}(\eta) :\ i \geq 1\big\},$$

(3.34)
$$\hat{\boldsymbol{X}}^{\eta|b;>\delta}_t(\boldsymbol{x}) = \hat{\boldsymbol{X}}^{\eta|b;>\delta}_{t-}(\boldsymbol{x}) + \varphi_b\Big(\eta\boldsymbol{\sigma}\big(\hat{\boldsymbol{X}}^{\eta|b;>\delta}_{t-}(\boldsymbol{x})\big)\boldsymbol{W}_i^{>\delta}(\eta)\Big) \quad \text{if } t = \eta\tau_i^{>\delta}(\eta) \text{ for some } i \geq 1,$$

under the initial condition $\hat{\boldsymbol{X}}^{\eta|b;>\delta}_0(\boldsymbol{x}) = \boldsymbol{x}$. By definition of the mapping $h^{(k)|b}$ in (2.10)–(2.13), we have the following property: for any $\eta, b, \delta > 0$, $j \geq 0$, and $\boldsymbol{x} \in \mathbb{R}^m$,

(3.35)
$$\text{on event } \Big\{\tau^{>\delta}_j(\eta) \leq \lfloor 1/\eta \rfloor < \tau^{>\delta}_{j+1}(\eta)\Big\},\ \hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x}) = h^{(j)|b}\big(\boldsymbol{x}, \eta\mathbf{W}^{>\delta}(\eta), \eta\boldsymbol{\tau}^{>\delta}(\eta)\big)$$

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

We first state two results that allow us to apply Lemma 2.4.

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, 3, it holds for all $k \in \mathbb{N}$ and $b, r \in (0,\infty)$ that $\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n)$ is asymptotically equivalent to $\hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x}_n)$ in $\mathbb{M}\big(\mathbb{D} \setminus \mathbb{D}^{(k)|b}_A(r)\big)$ w.r.t. $\lambda^k(\eta_n)$ as $\delta \downarrow 0$.*

PROPOSITION 3.13. *Let η_n be a sequence of strictly positive real numbers with $\lim_{n\to\infty} \eta_n = 0$, $A \subseteq \mathbb{R}^m$ be compact, and $\boldsymbol{x}_n, \boldsymbol{x}^* \in A$ be such that $\lim_{n\to\infty} \boldsymbol{x}_n = \boldsymbol{x}^*$. Let Assumptions 1, 2, and 3 hold. Let $k \geq 0$ and $b, r \in (0,\infty)$. For any $g \in \mathcal{C}\big(\mathbb{D} \setminus \mathbb{D}^{(k-1)|b}_A(r)\big)$,*

$$\lim_{n\to\infty} \mathbf{E}\Big[g\big(\hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x}_n)\big)\Big]\Big/\lambda^k(\eta_n) = \mathbf{C}^{(k)|b}(g\,;\boldsymbol{x}^*) \qquad \forall \delta > 0 \text{ small enough,} \tag{3.36}$$

where $\mathbf{C}^{(k)|b}$ is the measure defined in (2.15), 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 $\mathbb{C}$.

PROOF OF PROPOSITION 3.11. In the context of Lemma 2.4 and under the choice of

$$(\mathbb{S}, \boldsymbol{d}) = (\mathbb{D}, \boldsymbol{d}_{J_1}), \quad \mathbb{C} = \mathbb{D}^{(k-1)|b}_A(r), \quad X_n = \boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n), \quad Y^\delta_n = \hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x}_n), \quad \epsilon_n = \lambda^k(\eta_n),$$

Proposition 3.12 verifies condition (i), and Proposition 3.13 (together with Urysohn's Lemma) verifies condition (ii). Specifically, condition (ii) hold trivially if B is an empty set. In case that B is non-emtpy, given any $\Delta > 0$, one can apply Urysohn's Lemma to fix some continuous function $g : \mathbb{D} \to [0,1]$ such that $\mathbb{I}_B \leq g \leq \mathbb{I}_{B^\Delta}$. By Proposition 3.13,

$$\begin{aligned}
&\limsup_{\delta\downarrow 0}\limsup_{n\to\infty} \mathbf{P}\big(\hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x}_n) \in B\big)\big/\lambda^k(\eta_n) \\
&\leq \limsup_{\delta\downarrow 0}\lim_{n\to\infty} \mathbf{E}\big[g\big(\hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x}_n)\big)\big]\big/\lambda^k(\eta_n) = \mathbf{C}^{(k)|b}(g) \leq \mathbf{C}^{(k)|b}(B^\Delta).
\end{aligned}$$

Sending $\Delta \downarrow 0$, we verify the upper bound in condition (ii) of Lemma 2.4. The lower bound can be verified analogously. Applying Lemma 2.4, we conclude the proof. $\square$

Now, it only remains to prove Propositions 3.12 and 3.13.

PROOF OF PROPOSITION 3.12. 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 2.3, it suffice to show that for any $\delta > 0$ small enough,

(3.37)
$$\lim_{n\to\infty} \lambda^{-k}(\eta_n)\mathbf{P}\Big(\boldsymbol{d}_{J_1}\big(\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n), \hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x}_n)\big)\mathbb{I}\big\{\boldsymbol{X}^{\eta_n|b}(\boldsymbol{x}_n) \text{ or } \hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x}_n) \in B\big\} > \Delta\Big)$$
$$= 0.$$

By Lemma 3.4, there exist some $\bar{\epsilon} \in (0, r)$ and $\bar{\delta} > 0$ such that

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

$$\bar{h}^{(k)|b}\big(\boldsymbol{x}, (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k), (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k), \boldsymbol{t}\big) \in B^{\bar{\epsilon}} \quad \Longrightarrow \quad \|\boldsymbol{w}_i\| > \bar{\delta}\ \forall i \in [k]; \tag{3.38}$$

- furthermore,

$$\boldsymbol{d}_{J_1}\big(B^{\bar{\epsilon}}, \mathbb{D}_A^{(k-1)|b}(r)\big) > \bar{\epsilon}. \tag{3.39}$$

Henceforth in this proof, we only consider $\delta \in (0, \bar{\delta})$. Given $\eta, \delta, \epsilon > 0$ and $\boldsymbol{x} \in A$, let

$$B_0 \triangleq \Big\{\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \in B \text{ or } \hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x}) \in B;\ \boldsymbol{d}_{J_1}\big(\boldsymbol{X}^{\eta|b}(\boldsymbol{x}), \hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x})\big) > \Delta\Big\},$$

$$B_1 \triangleq \Big\{\tau_{k+1}^{>\delta}(\eta) > \lfloor 1/\eta \rfloor\Big\},$$

$$B_2 \triangleq \Big\{\tau_{k}^{>\delta}(\eta) \leq \lfloor 1/\eta \rfloor\Big\},$$

$$B_3 \triangleq \Big\{\eta \Big\|\boldsymbol{W}_i^{>\delta}(\eta)\Big\| > \bar{\delta} \text{ for all } i \in [k]\Big\},$$

$$B_4 \triangleq \Big\{\eta \Big\|\boldsymbol{W}_i^{>\delta}(\eta)\Big\| \leq 1/\epsilon^{\frac{1}{2k}} \text{ for all } i \in [k]\Big\},$$

and note the following decomposition of events:

$$\begin{aligned} B_0 = {} & (B_0 \cap B_1^c) \cup (B_0 \cap B_1 \cap B_2^c) \cup (B_0 \cap B_1 \cap B_2 \cap B_3^c) \\ & \cup (B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4^c) \cup (B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4). \end{aligned} \tag{3.40}$$

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

$$(2\rho D)^{k+1}\sqrt{\epsilon} < \Delta, \qquad 2\rho\epsilon < \bar{\epsilon}, \qquad \epsilon \in (0,1), \tag{3.41}$$

we claim that

$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1^c\Big) \Big/ \lambda^k(\eta) = 0, \tag{3.42}$$

$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2^c\Big) \Big/ \lambda^k(\eta) = 0, \tag{3.43}$$

$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2 \cap B_3^c\Big) \Big/ \lambda^k(\eta) = 0, \tag{3.44}$$

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4^c\Big) \Big/ \lambda^k(\eta) \leq k \cdot \bar{\delta}^{-(k-1)\alpha} \cdot \epsilon^{\frac{\alpha}{2k}}, \tag{3.45}$$

$$\lim_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\Big(B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4\Big) \Big/ \lambda^k(\eta) = 0, \tag{3.46}$$

if we pick $\delta > 0$ sufficiently small. Under such δ, it follows from (3.40) that

$$\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \mathbf{P}\big(B_0\big) \big/ \lambda^k(\eta) \leq k \cdot \bar{\delta}^{-(k-1)\alpha} \cdot \epsilon^{\frac{\alpha}{2k}}$$

for any $\epsilon > 0$ small enough. Note that $\bar{\delta} > 0$ is the constant fixed in (3.38). Driving $\epsilon \downarrow 0$, we conclude the proof of (3.37). Now, it only remains to prove claims (3.42)–(3.46).

Proof of Claim (3.42). For any $\delta > 0$, (3.6) implies that $\sup_{\boldsymbol{x} \in A} \mathbf{P}(B_0 \cap B_1^c) \leq \mathbf{P}(B_1^c) \leq \big(\eta^{-1} H(\delta\eta^{-1})\big)^{k+1} = \boldsymbol{O}\big(\lambda^{k+1}(\eta)\big) = \boldsymbol{o}\big(\lambda^k(\eta)\big)$.

Proof of Claim (3.43). It suffices to show that (for all $\delta > 0$ small enough)

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

By property (3.35), it holds on event $\{\tau_k^{>\delta}(\eta)>\lfloor 1/\eta\rfloor\}$ that $\hat{\boldsymbol{X}}^{\eta_n|b;>\delta}(\boldsymbol{x})\in\mathbb{D}_A^{(k-1)}(0)\subseteq\mathbb{D}_A^{(k-1)}(r)$. By (3.39), we have $\hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x})\notin B$ on event $\{\tau_k^{>\delta}(\eta)>\lfloor 1/\eta\rfloor\}$, and hence

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

Furthermore, let event $A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})$ be defined as in (3.8). We claim that

$$\{\boldsymbol{X}^{\eta|b}(x)\in B\}\cap\{\tau_k^{>\delta}(\eta)>\lfloor 1/\eta\rfloor\}\cap\big(\cap_{i=1}^k A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\big)=\emptyset \tag{3.47}$$

holds for any $\eta>0$ small enough with $\eta<\min\{\frac{b\wedge 1}{2C},\frac{\epsilon}{C}\}$, any $\delta\in(0,\frac{b}{2C})$, and any $\boldsymbol{x}\in A$. Here, $C\geq 1$ being the constant in Assumption 3. Then,

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

To conclude the proof, one only needs to apply Lemma 3.1 (b) with some $N>k(\alpha-1)$, due to $\lambda^k(\eta)\in\mathcal{RV}_{k(\alpha-1)}(\eta)$ as $\eta\downarrow 0$.

Now, we prove claim (3.47) 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

$$\frac{d\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x})}{dt}=\boldsymbol{a}\big(\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x})\big)\qquad\forall t\in[0,\infty)\setminus\{\eta\tau_j^{>\delta}(\eta):\ j\geq 1\}, \tag{3.48}$$

$$\breve{\boldsymbol{X}}^{\eta|b;\delta}_{\eta\tau_j^{>\delta}(\eta)}(\boldsymbol{x})=\boldsymbol{X}^{\eta|b}_{\tau_j^{>\delta}(\eta)}(\boldsymbol{x})\qquad\forall j\geq 1, \tag{3.49}$$

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

$$\begin{aligned}
&\boldsymbol{d}_{J_1}\Big(\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}),\boldsymbol{X}^{\eta|b}(\boldsymbol{x})\Big)\\
&\leq\sup_{t\in[0,1]}\Big\|\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x})-\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\Big\|\\
&\leq\sup_{t\in\big[0,\eta\tau_1^{>\delta}(\eta)\big)\cup\big[\eta\tau_1^{>\delta}(\eta),\eta\tau_2^{>\delta}(\eta)\big)\cup\cdots\cup\big[\eta\tau_{j-1}^{>\delta}(\eta),\eta\tau_j^{>\delta}(\eta)\big)}\Big\|\breve{\boldsymbol{X}}_t^{\eta|b;\delta}(\boldsymbol{x})-\boldsymbol{X}^{\eta|b}_{\lfloor t/\eta\rfloor}(\boldsymbol{x})\Big\|\\
&\leq\rho\cdot\big(\epsilon+\eta C\big)\leq 2\rho\epsilon<\bar{\epsilon}\qquad\text{by (3.14) of Lemma 3.6.}
\end{aligned} \tag{3.50}$$

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 3.6, and (ii) the inequalities follow from the choice of $\eta<\frac{\epsilon}{C}$ above and the choice of $2\rho\epsilon<\bar{\epsilon}$ in (3.41). Moreover, recall that we have fixed $\bar{\epsilon}<r$ at the beginning of the proof, and that (3.50) confirms (under the choice of $j=k$) that claims

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

hold on the event $\big(\cap_{i=1}^k A_i(\eta,b,\epsilon,\delta,\boldsymbol{x})\big)\cap\{\tau_k^{>\delta}(\eta)>\lfloor 1/\eta\rfloor\}$. By (3.39), on the said event we must have $\boldsymbol{X}^{\eta|b}(\boldsymbol{x})\notin B$, which verifies the claim (3.47).

Proof of Claim (3.44). Note that (3.35) holds on the event $B_1 \cap B_2 = \{\tau^{>\delta}_k(\eta) \leq \lfloor 1/\eta \rfloor < \tau^{>\delta}_{k+1}(\eta)\}$, and on the event B_3^c there is some $i \in [k]$ such that $\eta \left\| \boldsymbol{W}^{>\delta}_i(\eta) \right\| \leq \bar{\delta}$. By (3.35) and the choice of $\bar{\delta}$ in (3.38), given $\boldsymbol{x} \in A$ it holds on event $B_1 \cap B_2 \cap B_3^c$ that $\hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x}) \notin B$, and hence

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

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

(3.51)

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

where the event $A_i(\eta, b, \epsilon, \delta, \boldsymbol{x})$ is defined as in (3.8). Then, for any $\delta \in (0, \bar{\delta} \wedge \frac{b}{2C})$,

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

Applying Lemma 3.1 (b) with some $N > k(\alpha - 1)$, we conclude the proof of (3.44).

Now, it only remains to prove the claim (3.51) 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 $\eta \left\| \boldsymbol{W}^{>\delta}_J(\eta) \right\| \leq \bar{\delta}$. Next, recall the definition of the process $\breve{\boldsymbol{X}}^{\eta|b;\delta}_t(\boldsymbol{x})$ in (3.48)–(3.49). Applying (3.50) with $j = k+1$, we get that on event $\big(\cap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x}) \big) \cap \{\tau^{>\delta}_{k+1}(\eta) > \lfloor 1/\eta \rfloor\}$,

$$
\boldsymbol{d}_{J_1}\Big(\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}), \boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \Big) \leq \sup_{t \in [0,1]} \left\| \breve{\boldsymbol{X}}^{\eta|b;\delta}_t(\boldsymbol{x}) - \boldsymbol{X}^{\eta|b}_{\lfloor t/\eta \rfloor}(\boldsymbol{x}) \right\| < 2\rho\epsilon < \bar{\epsilon}. \tag{3.52}
$$

Furthermore, (3.52) implies 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 our choice of $\bar{\epsilon} < r$ at the beginning) such that

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

where the mapping $\bar{h}^{(k)|b}$ is defined in (2.10)–(2.12). Due to $\eta \left\| \boldsymbol{W}^{>\delta}_J(\eta) \right\| \leq \bar{\delta}$, it follows from (3.38) that $\breve{\boldsymbol{X}}^{\eta|b;\delta}(\boldsymbol{x}) \notin B^{\bar{\epsilon}}$. Then by (3.39) and (3.52), we must have $\boldsymbol{X}^{\eta|b}(\boldsymbol{x}) \notin B$ on the event $\big\{ \tau^{>\delta}_k(\eta) \leq \lfloor 1/\eta \rfloor < \tau^{>\delta}_{k+1}(\eta);\ \eta \left\| \boldsymbol{W}^{>\delta}_i(\eta) \right\| \leq \bar{\delta} \text{ for some } i \in [k] \big\} \cap \big(\bigcap_{i=1}^{k+1} A_i(\eta, b, \epsilon, \delta, \boldsymbol{x}) \big)$, thus verifying the claim (3.51).

Proof of Claim (3.45). Recall that $H(x) = \mathbf{P}(\|\boldsymbol{Z}\| > x)$. Due to

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

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

$$
\limsup_{\eta \downarrow 0} \sup_{\boldsymbol{x} \in A} \frac{\mathbf{P}\big(B_0 \cap B_1 \cap B_2 \cap B_3 \cap B_4^c \big)}{\lambda^k(\eta)}
$$

$$
\begin{aligned}
&\leq \lim_{\eta\downarrow 0}\frac{1}{\lambda^k(\eta)}\cdot\Big(\eta^{-1}H(\delta\eta^{-1})\Big)^k\cdot k\cdot\left(\frac{H(\bar\delta\eta^{-1})}{H(\delta\eta^{-1})}\right)^{k-1}\cdot\frac{H(\epsilon^{-\frac{1}{2k}}\eta^{-1})}{H(\delta\eta^{-1})} \qquad \text{by (3.6)}\\
&= \lim_{\eta\downarrow 0}\frac{1}{\lambda^k(\eta)}\cdot\Big(\eta^{-1}H(\eta^{-1})\Big)^k\cdot k\cdot\left(\frac{H(\bar\delta\eta^{-1})}{H(\eta^{-1})}\right)^{k-1}\cdot\frac{H(\epsilon^{-\frac{1}{2k}}\eta^{-1})}{H(\eta^{-1})}\\
&= k\cdot\lim_{\eta\downarrow 0}\left(\frac{H(\bar\delta\eta^{-1})}{H(\eta^{-1})}\right)^{k-1}\cdot\frac{H(\epsilon^{-\frac{1}{2k}}\eta^{-1})}{H(\eta^{-1})} \qquad \text{recall that } \lambda(\eta)=\eta^{-1}H(\eta^{-1})\\
&= k\cdot\bar\delta^{-(k-1)\alpha}\cdot\epsilon^{\frac{\alpha}{2k}} \qquad \text{due to } H(x)\in\mathcal{RV}_{-\alpha}(x) \text{ as } x\to\infty\text{; see Assumption 1.}
\end{aligned}
$$

Proof of Claim (3.46). We only consider $\delta\in(0,\frac{b}{2C})$. On event $B_1\cap B_2=\{\tau_k^{>\delta}(\eta)\leq\lfloor 1/\eta\rfloor<\tau_{k+1}^{>\delta}(\eta)\}$, $\hat{\boldsymbol X}^{\eta|b;>\delta}(\boldsymbol x)$ admits the expression in (3.35) with $j=k$. Then, by Lemma 3.7, we have that for any $\boldsymbol x\in A$ and any $\eta\in(0,\frac{\epsilon\wedge b}{2C})$, the inequality

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

holds on event $\Big(\bigcap_{i=1}^{k+1}A_i(\eta,b,\epsilon,\delta,\boldsymbol x)\Big)\cap B_1\cap B_2\cap B_3\cap B_4$. By our choice of $(2\rho D)^{k+1}\sqrt{\epsilon}<\Delta$ in (3.41), we get $\Big(\bigcap_{i=1}^{k+1}A_i(\eta,b,\epsilon,\delta,\boldsymbol x)\Big)\cap B_1\cap B_2\cap B_3\cap B_4\cap B_0=\emptyset,$ and hence

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

Again, we conclude the proof by applying Lemma 3.1 (b) with some $N>k(\alpha-1)$. □

Recall that $\big(\boldsymbol W_j^*(c)\big)_{j\geq 1}$ is a sequence of iid copies of $\boldsymbol W^*(c)$ defined in (3.11), 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 3.13, we prepare a lemma regarding events of form $E^\delta_{c,k}(\eta)=\big\{\tau_k^{>\delta}(\eta)\leq\lfloor 1/\eta\rfloor<\tau_{k+1}^{>\delta}(\eta);\ \eta\Big\|\boldsymbol W_j^{>\delta}(\eta)\Big\|>c\ \ \forall j\in[k]\big\}$ defined in (3.10).

LEMMA 3.14. *Let Assumption 1 hold. Let $A\subseteq\mathbb R^m$ be a compact set. Suppose that the function $\Psi:\mathbb R^m\times\mathbb R^{d\times k}\times(0,1]^{k\uparrow}\to\mathbb R$ is bounded, and is continuous on $\mathbb R^m\times\mathbb R^{d\times k}\times(0,1)^{k\uparrow}$. Then, for any $\delta>0,\ c>\delta$ and $k\in\mathbb N$,*

$$
\begin{aligned}
&\limsup_{\eta\downarrow 0}\sup_{\boldsymbol x\in A}\left|\frac{\mathbf E\Big[\Psi\Big(\boldsymbol x,\big(\eta\boldsymbol W_1^{>\delta}(\eta),\cdots,\eta\boldsymbol W_k^{>\delta}(\eta)\big),\big(\eta\tau_1^{>\delta}(\eta),\cdots,\eta\tau_k^{>\delta}(\eta)\big)\Big)\mathbb I_{E^\delta_{c,k}(\eta)}\Big]}{\lambda^k(\eta)}-\frac{(1/c^{\alpha k})\psi_{c,k}(\boldsymbol x)}{k!}\right|\\
&\qquad=0,
\end{aligned}
$$

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

(3.53)
$$
\Big|\lambda^{-k}(\eta_n)\mathbf E\Big[\Psi\big(\boldsymbol x_n,\eta_n\mathbf W^{\eta_n},\eta_n\boldsymbol\tau^{\eta_n}\big)\mathbb I_{E^\delta_{c,k}(\eta_n)}\Big]-(1/k!)\cdot c^{-\alpha k}\cdot\psi_{c,k}(\boldsymbol x_n)\Big|>\epsilon\quad\forall n\geq 1,
$$

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 we can assume w.l.o.g. that $\boldsymbol{x}_n \to \boldsymbol{x}^*$ for some $\boldsymbol{x}^* \in A$. Then,

$$
\begin{aligned}
&\lim_{n\to\infty} \lambda^{-k}(\eta_n)\mathbf{E}\Big[\Psi\big(\boldsymbol{x}_n, \eta_n \mathbf{W}^{\eta_n}, \eta_n \boldsymbol{\tau}^{\eta_n}\big)\mathbb{I}_{E^\delta_{c,k}(\eta_n)}\Big] \\
&= \Big[\lim_{n\to\infty} \lambda^{-k}(\eta_n)\mathbf{P}\Big(E^\delta_{c,k}(\eta_n)\Big)\Big] \cdot \lim_{n\to\infty} \mathbf{E}\Big[\Psi\big(\boldsymbol{x}_n, \eta_n \mathbf{W}^{\eta_n}, \eta_n \boldsymbol{\tau}^{\eta_n}\big)\Big| E^\delta_{c,k}(\eta_n)\Big] \\
&= (1/k!) \cdot c^{-\alpha k} \cdot \psi_{c,k}(\boldsymbol{x}^*) \text{ by Lemma 3.2, } \boldsymbol{x}_n \to \boldsymbol{x}^*\text{, and continuous mapping theorem.}
\end{aligned}
$$

Meanwhile, bounded convergence theorem implies $\psi_{c,k}(\boldsymbol{x}_n) \to \psi_{c,k}(\boldsymbol{x}^*)$. We now get $\lim_{n\to\infty}\Big|\lambda^{-k}(\eta_n)\mathbf{E}\Big[\Psi\big(\boldsymbol{x}_n, \eta_n \mathbf{W}^{\eta_n}, \eta_n \boldsymbol{\tau}^{\eta_n}\big)\mathbb{I}_{E^\delta_{c,k}(\eta_n)}\Big] - (1/k!) \cdot c^{-\alpha k} \cdot \psi_{c,k}(\boldsymbol{x}_n)\Big| = 0$, which contradicts (3.53) and allows us to conclude the proof. □

We are now ready to prove Proposition 3.13.

PROOF OF PROPOSITION 3.13. 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 non-negative, continuous, and bounded, whose support $B \triangleq \text{supp}(g)$ bounded away from $\mathbb{D}_A^{(k-1)|b}(r)$. By Lemma 3.4, we can fix some $\bar\epsilon \in (0, r)$ and $\bar\delta > 0$ such that

- for any $\boldsymbol{x} \in A$,

$$
h^{(k)|b}\big(\boldsymbol{x}, (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_k), \boldsymbol{t}\big) \in B^{\bar\epsilon} \quad \Longrightarrow \quad \|\boldsymbol{w}_j\| > \bar\delta \ \ \forall j \in [k]; \tag{3.54}
$$

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

Let $E^\delta_{c,k}(\eta)$ be defined as in (3.10). For any $\eta > 0$ and $\boldsymbol{x} \in A$, note that

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

The claim (3.36) then follows from the analyses below for each term.

Term $I_1(\eta, \boldsymbol{x})$. It follows from (3.6) that $\sup_{\boldsymbol{x}\in\mathbb{R}^m} \mathbf{E}[I_1(\eta, \boldsymbol{x})] \le \|g\| \cdot \big[\eta^{-1} \cdot H(\delta/\eta)\big]^{k+1}$. Therefore, $\lim_{\eta\downarrow 0} \sup_{\boldsymbol{x}\in A} \mathbf{E}[I_1(\eta, \boldsymbol{x})]\Big/\big(\eta^{-1}H(\eta^{-1})\big)^k \le \frac{\|g\|}{\delta^{\alpha(k+1)}} \cdot \lim_{n\to\infty} \eta^{-1} \cdot H(1/\eta) = 0$ due to $H(x) \in \mathcal{RV}_{-\alpha}(x)$ and $\alpha > 1$.

Term $I_2(\eta, \boldsymbol{x})$. As has been shown in the proof of claim (3.43) in Proposition 3.12 that, for any $\delta > 0$ and $\boldsymbol{x} \in A$, it holds on event $\{\tau^{>\delta}_k(\eta) > \lfloor 1/\eta \rfloor\}$ that $\hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x}) \notin B$, and hence $I_2(\eta, \boldsymbol{x}) = 0$.

Term $I_3(\eta,\boldsymbol{x})$. On the event $\{\tau^{>\delta}_k(\eta) \le \lfloor 1/\eta \rfloor < \tau^{>\delta}_{k+1}(\eta)\}$ the process $\hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x})$ admits the expression in (3.35) with $j=k$. Since there is some $i \in [k]$ such that $\eta \left\|\boldsymbol{W}^{>\delta}_i(\eta)\right\| \le \bar{\delta}$, by (3.54) we must have $\hat{\boldsymbol{X}}^{\eta|b;>\delta}(\boldsymbol{x}) \notin B$, and hence $I_3(\eta,\boldsymbol{x}) = 0$.

Term $I_4(\eta,\boldsymbol{x})$. On the event $E^{\delta}_{\bar{\delta},k}(\eta)$, the process $\hat{\boldsymbol{X}}^{\eta|b;(k)}(\boldsymbol{x})$ admits the expression in (3.35). As a result, for any $\eta > 0$ and $\boldsymbol{x} \in A$, we have $\mathbf{E}[I_4(\eta,\boldsymbol{x})] = \mathbf{E}\left[\Psi\big(\boldsymbol{x},\eta\mathbf{W}^{\eta},\eta\boldsymbol{\tau}^{\eta}\big)\mathbb{I}_{E^{\delta}_{\bar{\delta},k}(\eta)}\right]$, 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 $\Psi(\boldsymbol{x},\mathbf{W},\boldsymbol{t}) \triangleq g\big(h^{(k)|b}(\boldsymbol{x},\mathbf{W},\boldsymbol{t})\big)$. Also, let $\psi(\boldsymbol{x}) \triangleq \mathbf{E}\Big[\Psi\Big(\boldsymbol{x},\big(\boldsymbol{W}^*_1(\bar{\delta}),\cdots,\boldsymbol{W}^*_k(\bar{\delta})\big),\big(U_{(1;k)},\cdots,U_{(k;k)}\big)\Big)\Big]$, where $\big(\boldsymbol{W}^*_j(c)\big)_{j\ge 1}$ are iid copies of $\boldsymbol{W}^*(c)$ defined in (3.11), and $\big(U_{(j:k)}\big)_{j\in[k]}$ are the order statistics of k samples of Unif$(0,1)$. First, the continuity of mapping Ψ on $\mathbb{R}^m \times \mathbb{R}^{d\times k} \times (0,1)^{k\uparrow}$ follows from the continuity of g and $h^{(k)|b}$ (see Lemma 3.5). Besides, by bounded convergence theorem, one can see that $\psi(\cdot)$ is also continuous, and Ψ and ψ are bounded due to $\|\Psi\| \le \|g\| < \infty$. By Lemma 3.14,

$$\lim_{\eta\downarrow 0}\sup_{\boldsymbol{x}\in A}\left|\lambda^{-k}(\eta)\mathbf{E}\big[I_4(\eta,\boldsymbol{x})\big] - (1/k!)\cdot\bar{\delta}^{-\alpha k}\cdot\psi(\boldsymbol{x})\right| = 0.$$

By 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 $\lim_{n\to\infty}\psi(\boldsymbol{x}_n) = \psi(\boldsymbol{x}^*)$. Now, to conclude the proof for claim (3.36), it only remains to show that

$$\frac{(1/\bar{\delta}^{\alpha k})\psi(\boldsymbol{x}^*)}{k!} = \mathbf{C}^{(k)|b}(g;\boldsymbol{x}^*). \tag{3.55}$$

To do so, recall the law of $\boldsymbol{W}^*(c)$ in (3.11). By the definition of $\psi(\cdot)$,

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

where ν_α is defined in (2.6), the measure $\mathbf{S}$ in Assumption 1, and $\alpha > 1$ is the heavy-tail index in Assumption 1. By our choice of $\bar{\delta}$ in (3.54),

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

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

$$\begin{aligned}&\psi(\boldsymbol{x}^*)\\ &= k!\cdot\bar{\delta}^{\alpha k}\int g\Big(h^{(k)|b}\big(\boldsymbol{x}^*,(w_1\boldsymbol{\theta}_1,\cdots,w_k\boldsymbol{\theta}_k),\boldsymbol{t}\big)\Big)\left(\bigtimes_{j=1}^{k}\Big(\nu_\alpha(dw_j)\times\mathbf{S}(d\boldsymbol{\theta}_j)\Big)\right)\times\mathcal{L}^{k\uparrow}_1(d\boldsymbol{t})\\ &= k!\cdot\bar{\delta}^{\alpha k}\cdot\mathbf{C}^{(k)|b}\big(g;\boldsymbol{x}^*\big) \qquad \text{by the definitions in (2.15).}\end{aligned}$$

We thus verify (3.55) and conclude the proof. $\square$

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 2.11. Section 4.2 then applies the framework in the context of heavy-tailed stochastic difference equations and proves Theorem 2.8.

4.1. *Proof of Theorem 2.11.* The proof 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*

$$\liminf_{\eta\downarrow 0}\inf_{x\in A(\epsilon)}\mathbf{P}\Big(\gamma(\eta)\tau^{\eta}_{I(\epsilon)^c}(x)>t;\ V^{\eta}_{\tau_\epsilon}(x)\in B\Big)\geq C(B^\circ)\cdot e^{-t}-\delta_{t,B}(\epsilon),$$

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

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*

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

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 $0<r<R$,

$$\begin{aligned}\mathbf{P}\Big(\tau^{\eta}_{I(\epsilon)^c}(y)\in\big(RT/\eta,\,\rho^{\eta}_1(y)\big]\Big)&\leq\mathbf{P}\Big(\tau^{\eta}_{I(\epsilon)^c}(y)\wedge\rho^{\eta}_1(y)>RT/\eta\Big)\\&\leq\mathbf{P}\Big(V^{\eta}_j(y)\in I(\epsilon)\setminus A(\epsilon)\quad\forall j\in\big[\lfloor rT/\eta\rfloor,\,RT/\eta\big]\Big)\\&\leq\sup_{z\in I(\epsilon)\setminus A(\epsilon)}\mathbf{P}\Big(\tau^{\eta}_{(I(\epsilon)\setminus A(\epsilon))^c}(z)>\frac{R-r}{2}T/\eta\Big)\\&=\gamma(\eta)T/\eta\cdot o(1),\end{aligned}\tag{4.1}$$

where the last equality is from (2.32) of Condition 1, and

$$\begin{aligned}&\sup_{y\in A(\epsilon)}\mathbf{P}\Big(V^{\eta}_{\tau_\epsilon}(y)\in B;\ \tau^{\eta}_{I(\epsilon)^c}(y)\leq\rho^{\eta}_1(y)\Big)\\&\leq\sup_{y\in A(\epsilon)}\mathbf{P}\Big(V^{\eta}_{\tau_\epsilon}(y)\in B;\ \tau^{\eta}_{I(\epsilon)^c}(y)\leq RT/\eta\Big)+\sup_{y\in A(\epsilon)}\mathbf{P}\Big(\tau^{\eta}_{I(\epsilon)^c}(y)\in\big(RT/\eta,\,\rho^{\eta}_1(y)\big]\Big)\\&\leq\sup_{y\in A(\epsilon)}\mathbf{P}\Big(V^{\eta}_{\tau_\epsilon}(y)\in B;\ \tau^{\eta}_{I(\epsilon)^c}(y)\leq RT/\eta\Big)+\gamma(\eta)T/\eta\cdot o(1)\\&\leq\big(C(B^-)+\delta_B(\epsilon,RT)+o(1)\big)\cdot\gamma(\eta)RT/\eta,\end{aligned}\tag{4.2}$$

where the second inequaility is from (4.1) and the last equality is from (2.31) 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 $\mathcal{F}_{\rho^\eta_j(x)}$,

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

$$
\geq \inf_{y \in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(y) \leq T/\eta;\ V^\eta_{\tau_\epsilon}(y) \in B\Big) \cdot \sum_{j=K}^{\infty} \inf_{x \in A(\epsilon)} \mathbf{P}\Big(\tau^\eta_{I(\epsilon)^c}(x) > \rho^\eta_j(x)\Big). \tag{4.3}
$$

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

(4.4)

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

where the last inequality is from (4.2) with $B = \mathbb{S}$. From (4.3), (4.4), and (2.30) 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 $k = \left\lfloor \frac{t/\gamma(\eta)}{T/\eta} \right\rfloor$ 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

$$
\begin{aligned}
\text{(II)} &\leq \sum_{j=1}^k \sup_{x \in A(\epsilon)} \mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(x) \wedge \rho^\eta_j(x) - \rho^\eta_{j-1}(x) \geq T/\eta\big) \\
&= \sum_{j=1}^k \sup_{x \in A(\epsilon)} \mathbf{E}\Big[\mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(x) \wedge \rho^\eta_j(x) - \rho^\eta_{j-1}(x) \geq T/\eta \big| \mathcal{F}_{\rho^\eta_{j-1}(x)}\big)\Big] \\
&\leq \sum_{j=1}^k \sup_{y \in A(\epsilon)} \mathbf{P}\big(\tau^\eta_{I(\epsilon)^c}(y) \wedge \rho^\eta_1(y) \geq T/\eta\big) \\
&\leq \frac{t}{\gamma(\eta)T/\eta} \cdot \gamma(\eta)T/\eta \cdot o(1) = o(1)
\end{aligned}
$$

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.

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

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

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

whereas the second term is bounded via (2.30) of Condition 1 as follows:

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

Therefore,

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

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 $\mathcal{F}_{\rho^\eta_j(x)}$,

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

Lastly, by Condition 1 (specifically, $\lim_{\epsilon\downarrow0}\lim_{T\uparrow\infty}\delta_{I^c}(\epsilon,T)=0$ in Definition 2.10), we verify the upper bounds in Case (ii) and conclude the proof. □

Now, we are ready to prove Theorem 2.11.

PROOF OF THEOREM 2.11. 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 $\epsilon,\epsilon'>0$, $t\geq0$, and measurable $B\subseteq\mathbb{S}$,

$$\begin{aligned}
&\liminf_{\eta\downarrow0}\inf_{x\in I(\epsilon')}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_{I(\epsilon)^c}(x)>t,\ V^\eta_{\tau_\epsilon}(x)\in B\Big)\geq C(B^\circ)\cdot e^{-t}-\delta_{t,B}(\epsilon)\\
&\limsup_{\eta\downarrow0}\sup_{x\in I(\epsilon')}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_{I(\epsilon)^c}(x)>t,\ V^\eta_{\tau_\epsilon}(x)\in B\Big)\leq C(B^-)\cdot e^{-t}+\delta_{t,B}(\epsilon)
\end{aligned}\tag{4.5}$$

where $\delta_{t,B}(\epsilon)$ is characterized in part (i) of Proposition 4.1 such that $\delta_{t,B}(\epsilon)\to0$ as $\epsilon\to0$. 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

$$\liminf_{\eta\downarrow0}\inf_{x\in I(\epsilon')}\mathbf{P}\Big(\gamma(\eta)\cdot\tau^\eta_{I^c}(x)>t,\ V^\eta_\tau(x)\in B\Big)$$

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

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

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

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 $x\in I$,

(4.6)

$$\mathbf{P}\Big(\gamma(\eta)\cdot\tau^{\eta}_{\epsilon}(x)>t,\,V^{\eta}_{\tau_\epsilon}(x)\in B\Big)$$
$$=\mathbf{E}\Bigg[\mathbf{P}\Big(\gamma(\eta)\cdot\tau^{\eta}_{\epsilon}(x)>t,\,V^{\eta}_{\tau_\epsilon}(x)\in B\Big|\mathcal{F}_{\tau^{\eta}_{A(\epsilon)}(x)}\Big)\cdot\Big(\mathbb{I}\big\{\tau^{\eta}_{A(\epsilon)}(x)\leq T/\eta\big\}+\mathbb{I}\big\{\tau^{\eta}_{A(\epsilon)}(x)>T/\eta\big\}\Big)\Bigg].$$

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

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

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,

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

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

4.2. *Proof of Theorem 2.8.* In this section, we apply the framework developed in Section 2.3.2 and prove Theorem 2.8. Throughout this section, we impose Assumptions 1, 2, and 4.

We start by fixing a few constants. Recall the definition of the discretized width metric $\mathcal{J}_b^I$ defined in (2.27). To prove Theorem 2.8, in this section we fix some $b > 0$ such that the conditions in Theorem 2.8 hold. This allows us to fix some $\breve{\epsilon} > 0$ small enough such that

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

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

$$\boldsymbol{x} \in \bar{B}_r(\mathbf{0}) \quad \Longrightarrow \quad \boldsymbol{y}_t(\boldsymbol{x}) \in \bar{B}_r(\mathbf{0}) \ \forall t \geq 0, \tag{4.8}$$

with the ODE $\boldsymbol{y}_t(x)$ defined in (2.21). Next, for any $\epsilon \in (0, \breve{\epsilon})$, let

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

By Gronwall's inequality, it is easy to see that $\breve{I}(\epsilon)$ is an open set. Meanwhile, by Assumption 4, given $\boldsymbol{x} \in I$ we must have $\boldsymbol{x} \in \breve{I}(\epsilon)$ for any $\epsilon > 0$ 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 $(I_\epsilon)^-$ is still bounded away from I^c, i.e., $\|\boldsymbol{x} - \boldsymbol{y}\| \geq \epsilon$ for all $\boldsymbol{x} \in (I_\epsilon)^-$, $\boldsymbol{y} \in I^c$. Besides, the continuity of $\boldsymbol{a}(\cdot)$ (see Assumption 2) implies the continuity (w.r.t. $\boldsymbol{x}$) of $\boldsymbol{y}_t(\boldsymbol{x})$'s hitting time to $\bar{B}_{\breve{\epsilon}}(\mathbf{0})$. Then, by 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 any $T > 0$ large enough. This confirms that given $\epsilon > 0$, it holds for any $\epsilon' > 0$ 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 (2.25). 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 (2.21). 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 any $\epsilon' > 0$ 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.,

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

This can be argued with a proof by contradiction. Suppose that 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 each $n \geq 1$. Besides, we have shown that $\mathcal{G}^{(\mathcal{J}_b^I-1)|b}(\breve{\epsilon}) \subseteq \check{I}(\epsilon)$ holds for any $\epsilon > 0$ small enough. Then under such $\epsilon > 0$, by (4.8)(4.9), we must have $t_n \leq 1/\epsilon$ for each n; otherwise, we will have $\boldsymbol{x}'_n = \boldsymbol{y}_{t_n}(\boldsymbol{x}_n) \in \bar{B}_{\breve{\epsilon}}(\mathbf{0})$, which prevents $\|\boldsymbol{x}'_n - \boldsymbol{z}_n\| \leq 1/n$ to hold for the sequence $\boldsymbol{z}_n \in I^c$ due to $\bar{B}_{\breve{\epsilon}}(\mathbf{0}) \subseteq I_{\breve{\epsilon}}$; see (4.7). Now, in light of the boundedness for the sequence $\boldsymbol{x}_n \in \bar{\mathcal{G}}^{(\mathcal{J}_b^I-1)|b}(\breve{\epsilon}) \subset I$ (due to the boundedness of I) and the boundedness of the sequence t_n, by picking a sub-sequence if necessary, we can w.l.o.g. assume that $\boldsymbol{x}_n \to \boldsymbol{x}_*$ for some $\boldsymbol{x}_* \in \left(\mathcal{G}^{(\mathcal{J}_b^I-1)|b}(\breve{\epsilon})\right)^- \subset I$ and $t_n \to t_*$ for some $t_* \in [0,\infty)$. Let $\boldsymbol{x}'_* \triangleq \boldsymbol{y}_{t_*}(\boldsymbol{x}_*)$. Since $\boldsymbol{x}_* \in I$, by Assumption 4 we must have $\boldsymbol{x}'_* \in I$. However, the continuity of the flow (specifically, using Gronwall's inequality) implies $\boldsymbol{x}'_n \to \boldsymbol{x}'_*$. Under the condition $\|\boldsymbol{x}'_n - \boldsymbol{z}_n\| \leq 1/n$ and $\boldsymbol{z}_n \in I^c$ for each n, we arrive at the contradiction that $\boldsymbol{z}_n \to \boldsymbol{x}'_* \in I^c$ due to I being an open set and I^c being closed. This verifies (4.12). Summarizing (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 to 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 $\boldsymbol{x} \in (I_\epsilon)^-$. By 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 $(I_\epsilon)^-$, we have $\boldsymbol{t}(\epsilon) < \infty$ for any $\epsilon > 0$. Besides, by the definition of $\boldsymbol{t}(\cdot)$, we have (for any $\epsilon \in (0,\bar{\epsilon})$)

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

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

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

Specifically, as a direct consequence of (4.18), there is some $\bar{c} \in (0,1)$ such that

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

where $\bar{\epsilon} > 0$ is the constant fixed in (4.13)–(4.15).

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 of the full-length version of the paper (arXiv:2307.03479v4; see also `full_length_paper.pdf` in the related files of this submission), together with the proofs of other useful properties regarding measures $\check{\mathbf{C}}^{(k)|b}$.

LEMMA 4.2. *Let $\bar{t}, \bar{\delta} \in (0,\infty)$ be the constants characterized in part (b) of Lemma* **??**. *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 $B \subseteq (I_\epsilon)^c$,*

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

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

where $\mathbf{C}^{(k)|b}_{[0,T]}$ *is defined in* (2.15), $\breve{\mathbf{C}}^{(k)|b}$ *is defined in* (2.28), $\mathcal{J}^I_b$ *is defined in* (2.27),

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

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

To see how we apply the framework developed in Section 2.3.2, let us specialize Condition 1 to a setting where $\mathbb{S}=\mathbb{R}$, $A(\epsilon)=B_\epsilon(\mathbf{0})=\{\boldsymbol{x}\in\mathbb{R}^m:\ \|\boldsymbol{x}\|<\epsilon\}$, and the covering $I(\epsilon)=I_\epsilon$. Let $V^\eta_j(x)=\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x})$. Meanwhile, for $C^I_b=\breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(I^c)$, it is shown in Lemma **??** that $C^I_b<\infty$. Now, recall that $H(\cdot)=\mathbf{P}(\|\boldsymbol{Z}_1\|>\cdot)$ and $\lambda(\eta)=\eta^{-1}H(\eta^{-1})$. Recall that in Theorem 2.8, we consider two cases: (i) $C^I_b\in(0,\infty)$, and (ii) $C^I_b=0$. We first discuss our choices in Case (i). When $C^I_b>0$, we set the location measure and scale as

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

The regularity conditions in Theorem 2.8 dictate that $\breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}(\partial I)=0$, and hence $C(\partial I)=0$. Besides, note that $C(\cdot)$ is a probability measure and $\gamma(\eta)T/\eta=C^I_bT\cdot\big(\lambda(\eta)\big)^{\mathcal{J}^I_b}$.

The application of the framework developed in Section 2.3.2 (specifically, Theorem 2.11) hinges on the verification of (2.30)–(2.33). We start by verifying (2.30) and (2.31). 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 $\lim_{\Delta\downarrow 0}\breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}\Big((B^\Delta\cap I^c)\setminus(B^-\cap I^c)\Big)=0$ and $\lim_{\Delta\downarrow 0}\breve{\mathbf{C}}^{(\mathcal{J}^I_b)|b}\Big((B^\circ\cap I^c)\setminus(B_\Delta\cap I^c)\Big)=0$. This allows us to fix a sequence $(\Delta^{(n)})_{n\geq 1}$ such that $\Delta^{(n+1)}\in(0,\Delta^{(n)}/2)$ and

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

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 $(\epsilon^{(n)})_{n\geq 1}$ with $\epsilon^{(n+1)}\in(0,\epsilon^{(n)}/2)$ and $\epsilon^{(n)}\in(0,\bar{\epsilon}]$ $\forall n\geq 1$, such that for any $n\geq 1$, $\epsilon\in(0,\epsilon^{(n)}]$,

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

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

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 (under $n=n_\epsilon$)

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

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

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

where $n = n_\epsilon$ 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 implies $\lim_{r\downarrow 0} \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}\Big((\partial I)^r\Big) = \breve{\mathbf{C}}^{(\mathcal{J}_b^I)|b}(\partial I) = 0$. 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 (ii) (i.e., $C_b^I = 0$), we set

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

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

To verify conditions (2.30) and (2.31), we introduce stopping times

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

LEMMA 4.3 (Verifying conditions (2.30) and (2.31)). *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 $T > \bar{t}$,*

$$\liminf_{\eta\downarrow 0} \inf_{\boldsymbol{x}:\ \|\boldsymbol{x}\|\leq\epsilon} \frac{\mathbf{P}\Big(\tau_\epsilon^{\eta|b}(\boldsymbol{x}) \leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau_\epsilon^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\Big)}{\gamma(\eta)T/\eta} \geq C(B^\circ) - \delta_B(\epsilon,T),$$

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

PROOF. Recall that

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

(ii) in case that $C_b^I = 0$, we have $\gamma(\eta)T/\eta = T\cdot\big(\lambda(\eta)\big)^{\mathcal{J}_b^I}$, $C(\cdot)\equiv 0$, and $\delta_B(\epsilon,T) = \breve{\delta}_B(\epsilon,T)/T$.

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

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

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

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

$$\begin{aligned}\Big\{\tau_\epsilon^{\eta|b}(\boldsymbol{x}) \leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau_\epsilon^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in B\Big\} &= \Big\{\tau_\epsilon^{\eta|b}(\boldsymbol{x}) \leq T/\eta;\ \boldsymbol{X}^{\eta|b}_{\tau_\epsilon^{\eta|b}(\boldsymbol{x})}(\boldsymbol{x}) \in \widetilde{B}(\epsilon)\Big\} \\ &= \Big\{\boldsymbol{X}^{\eta|b}_{[0,T]}(\boldsymbol{x}) \in \breve{E}\big(\epsilon,\widetilde{B}(\epsilon),T\big)\Big\}.\end{aligned}$$

For any $\epsilon \in (0,\bar{\epsilon})$ and $\xi \in \breve{E}(\epsilon, \widetilde{B}(\epsilon), T)$, 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 **??**, given $\epsilon \in (0,\bar{\epsilon}]$, it holds for any $\xi \in \mathbb{D}^{(\mathcal{J}_b^I - 1)|b}_{\bar{B}_\epsilon(\mathbf{0})}[0,T](\epsilon)$ that $\xi_t \in (I_{2\bar{\epsilon}})^- \ \forall t \in [0,T]$. Therefore, the claim

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

holds for any $\epsilon \in (0,\bar{\epsilon}]$. Next, recall the sequence $(\epsilon^{(n)})_{n\geq 1}$ specified in (4.24)–(4.25). For any $\epsilon > 0$ small enough we have $\epsilon \in (0, \epsilon^{(1)}]$, and for such ϵ we set $n = n_\epsilon$ as the unique positive integer such that $\epsilon \in (\epsilon^{(n+1)}, \epsilon^{(n)}]$. It then follows from Theorem 2.5 that

(4.30)

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

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

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

Plugging this bound back into (4.30), the upper bound in the claim (4.28) then follows from the definition of $\breve{\delta}_B$ in (4.26) and our choice of $C(\cdot)$ in (4.22). Similarly, by Theorem 2.5 and the property (4.25), we obtain (for any $\epsilon > 0$ small enough)

(4.31)

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

where $n = n_\epsilon$ is the unique positive integer such that $\epsilon \in (\epsilon^{(n+1)}, \epsilon^{(n)}]$. Furthermore, by the preliminary bound $(E \cap F)_\Delta \supseteq E_\Delta \cap F_\Delta$,

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

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

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

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

The next result verifies condition (2.32). 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 (2.32) means that $\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) \in I_\epsilon \setminus \{\boldsymbol{x} : \|\boldsymbol{x}\| < \epsilon\}$ for all $j \leq T/\eta$. Also, recall our choice of $\gamma(\eta)T/\eta = C_b^I T \cdot \big(\lambda(\eta)\big)^{\mathcal{J}_b^I}$ if $C_b^I > 0$, or $\gamma(\eta)T/\eta = T \cdot \big(\lambda(\eta)\big)^{\mathcal{J}_b^I}$ if $C_b^I = 0$. To verify condition (2.32), it suffices to prove the following result.

LEMMA 4.4 (Verifying condition (2.32)). *Let $\bar\epsilon > 0$ be the constant in (4.13)–(4.15), and $\boldsymbol{t}(\cdot)$ be defined as in (4.16). Given $k \geq 1$ and $\epsilon \in (0, \bar\epsilon)$, it holds for any $T \geq k \cdot \boldsymbol{t}(\epsilon/2)$ that*

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

PROOF. In this proof, we write $\xi(t) = \xi_t$ for any $\xi \in \mathbb{D}[0,T]$, and write $B_\epsilon(\mathbf{0}) = \{\boldsymbol{x} \in \mathbb{R}^m : \|\boldsymbol{x}\| < \epsilon\}$. Note that $\big\{\boldsymbol{X}^{\eta|b}_j(\boldsymbol{x}) \in I_\epsilon \setminus B_\epsilon(\mathbf{0})\ \ \forall j \leq T/\eta\big\} = \big\{\boldsymbol{X}^{\eta|b}_{[0,T]}(x) \in E(\epsilon)\big\}$ 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 $\mathbb{D}_A^{(k)|b}[0,T](\epsilon)$ in (2.14). 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 2.5 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 follows if we show the existence of some $\delta > 0$ such that

$$\boldsymbol{d}^{[0,T]}_{J_1}(\xi, \xi') \geq \delta > 0 \qquad \forall \xi \in \mathbb{D}^{(k-1)|b}_{(I_\epsilon)^-}[0,T](\epsilon),\ \xi' \in E(\epsilon). \tag{4.32}$$

First, by definition of $E(\epsilon)$, we have $\xi'(t) \in I_\epsilon\ \forall t \in [0,T]$ for any $\xi' \in E(\epsilon)$. Also, note that $\inf\{\|\boldsymbol{x} - \boldsymbol{y}\| :\ \boldsymbol{x} \in I_\epsilon,\ \boldsymbol{y} \notin I_{\epsilon/2}\} \geq \epsilon/2$. Therefore, for any $\xi \in \mathbb{D}^{(k-1)|b}_{(I_\epsilon)^-}[0,T](\epsilon)$, if $\xi(t) \notin I_{\epsilon/2}$ for some $t \in [0,T]$, then $\boldsymbol{d}^{[0,T]}_{J_1}(\xi, \xi') \geq \epsilon/2 > 0$. Next, given $\xi \in \mathbb{D}^{(k-1)|b}_{(I_\epsilon)^-}[0,T](\epsilon)$, we consider the case where $\xi(t) \in I_{\epsilon/2}$ for any $t \in [0,T]$. By definition of $\mathbb{D}^{(k-1)|b}_{(I_\epsilon)^-}[0,T](\epsilon)$, 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 $\xi = \bar{h}^{(k-1)|b}_{[0,T]}\big(\boldsymbol{x}, \mathbf{W}, \mathbf{V}, (t_1, \cdots, t_{k-1})\big)$. Under the convention that $t_0 = 0$ and $t_k = T$, we have (for each $j \in [k]$)

$$\xi(t) = \boldsymbol{y}_{t - t_{j-1}}\big(\xi(t_{j-1})\big) \qquad \forall t \in [t_{j-1}, t_j). \tag{4.33}$$

Here, $\boldsymbol{y}_\cdot(x)$ is the ODE defined in (2.21). By the running assumption in this lemma that $T \geq k \cdot \boldsymbol{t}(\epsilon/2)$, there must be some $j \in [k]$ such that $t_j - t_{j-1} \geq \boldsymbol{t}(\epsilon/2)$. Since $\xi(t) \in I_{\epsilon/2}$ $\forall t \in [0,T]$, we have $\xi(t_{j-1}) \in I_{\epsilon/2}$. Combining (4.33) and the property (4.17), we get $\lim_{t \uparrow t_j} \xi(t) \in \bar{B}_{\epsilon/2}(\boldsymbol{0}) \subset B_\epsilon(\boldsymbol{0})$. On the other hand, by definition of $E(\epsilon)$, we have $\xi'(t) \notin B_\epsilon(\boldsymbol{0})$ for all $t \in [0,T]$, which implies $\boldsymbol{d}_{J_1}^{[0,T]}(\xi,\xi') \geq \frac{\epsilon}{2}$. This concludes the proof. $\square$

Let

$$R_\epsilon^{\eta|b}(\boldsymbol{x}) \triangleq \min\left\{ j \geq 0 : \ \left\| \boldsymbol{X}_j^{\eta|b}(\boldsymbol{x}) \right\| < \epsilon \right\} \tag{4.34}$$

be the first time $\boldsymbol{X}_j^{\eta|b}(\boldsymbol{x})$ returns to the ϵ-neighborhood of the origin. Lastly, we establish condition (2.33) under our choice of $A(\epsilon) = \{\boldsymbol{x} \in \mathbb{R}^m : \ \|\boldsymbol{x}\| < \epsilon\}$. Note that the first visit time $\tau_{A(\epsilon)}^\eta(x)$ therein coincides with $R_\epsilon^{\eta|b}(x)$.

LEMMA 4.5 (Verifying condition (2.33)). *Let $\boldsymbol{t}(\cdot)$ be defined as in (4.16) and*

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

For any $\epsilon \in (0,\bar{\epsilon})$, $\lim_{\eta\downarrow 0} \sup_{\boldsymbol{x} \in (I_\epsilon)^-} \mathbf{P}\Big(\big(E(\eta,\epsilon,\boldsymbol{x})\big)^c \Big) = 0$.

PROOF. In this proof, we write $B_\epsilon(\boldsymbol{0}) = \{\boldsymbol{x} \in \mathbb{R}^m : \ \|\boldsymbol{x}\| < \epsilon\}$. Note that $\big(E(\eta,\epsilon,\boldsymbol{x})\big)^c \subseteq \big\{\boldsymbol{X}_{[0,\boldsymbol{t}(\epsilon/2)]}^{\eta|b}(\boldsymbol{x}) \in E_1^*(\epsilon) \cup E_2^*(\epsilon)\big\}$, where

$$E_1^*(\epsilon) \triangleq \left\{ \xi \in \mathbb{D}[0,\boldsymbol{t}(\epsilon/2)] : \ \xi(t) \notin B_\epsilon(\boldsymbol{0}) \ \forall t \in [0,\boldsymbol{t}(\epsilon/2)] \right\},$$

$$E_2^*(\epsilon) \triangleq \left\{ \xi \in \mathbb{D}[0,\boldsymbol{t}(\epsilon/2)] : \ \exists 0 \leq s < t \leq \boldsymbol{t}(\epsilon/2) \ s.t. \ \xi(t) \in B_\epsilon(\boldsymbol{0}), \ \xi(s) \notin I \right\}.$$

Recall the definition of $\mathbb{D}_A^{(k)|b}[0,T](\epsilon)$ in (2.14). We claim that both $E_1^*(\epsilon)$ and $E_2^*(\epsilon)$ are bounded away from

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

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

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

Meanwhile, by property (4.18),

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

This allows us to apply Theorem 2.5 and obtain

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

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 $\lim_{\eta\downarrow 0} \lambda(\eta) = 0$. $\square$

We conclude this section with the proof of Theorem 2.8.

PROOF OF THEOREM 2.8. As noted above, the claim $C^I_b < \infty$ is verified by Lemma **??**. Next, since Lemmas 4.3–4.5 verify Condition 1, Theorem 2.8 follows immediately from Theorem 2.11. □

Acknowledgments. The authors would like to thank the anonymous referees, an Associate Editor and the Editor for their constructive comments that improved the quality of this paper.

Funding. The authors gratefully acknowledge the support of the National Science Foundation (NSF) under CMMI-2146530.

SUPPLEMENTARY MATERIAL

Title of Supplement A
Short description of Supplement A.

Title of Supplement B
Short description of Supplement B.

REFERENCES

[1] ALBRECHER, H., CHEN, B., VATAMIDOU, E. and ZWART, B. (2020). Finite-time ruin probabilities under large-claim reinsurance treaties for heavy-tailed claim sizes. *Journal of Applied Probability* **57** 513–530.

[2] BALDI, P. and CARAMELLINO, L. (2011). General Freidlin–Wentzell Large Deviations and positive diffusions. *Statistics & Probability Letters* **81** 1218-1229.

[3] BAZHBA, M., BLANCHET, J., RHEE, C.-H. and ZWART, B. (2020). Sample path large deviations for Lévy processes and random walks with Weibull increments. *The Annals of Applied Probability* **30** 2695 – 2739.

[4] BAZHBA, M., RHEE, C.-H. and ZWART, B. (2022). Large deviations for stochastic fluid networks with Weibullian tails. *Queueing Systems* **102** 25–52.

[5] BEDI, A. S., PARAYIL, A., ZHANG, J., WANG, M. and KOPPEL, A. (2024). On the Sample Complexity and Metastability of Heavy-tailed Policy Search in Continuous Control. *Journal of Machine Learning Research* **25** 1–58.

[6] BERNHARD, H., DAS, B. et al. (2020). Heavy-tailed random walks, buffered queues and hidden large deviations. *Bernoulli* **26** 61–92.

[7] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular variation* **27**. Cambridge university press, Cambridge.

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

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

[10] BOROVKOV, A. A. and MOGUL'SKII, A. A. (2010). On large deviation principles in metric spaces. *Siberian Mathematical Journal* **51** 989–1003.

[11] BOVIER, A. and DEN HOLLANDER, F. (2016). *Metastability: a potential-theoretic approach* **351**. Springer, New York, NY.

[12] BOVIER, A., ECKHOFF, M., GAYRARD, V. and KLEIN, M. (2001). Metastability in stochastic dynamics of disordered mean-field models. *Probability Theory and Related Fields* **119** 99–161.

[13] BOVIER, A., ECKHOFF, M., GAYRARD, V. and KLEIN, M. (2004). Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc.(JEMS)* **6** 399–424.

[14] BOVIER, A., ECKHOFF, M., GAYRARD, V. and KLEIN, M. (2005). Metastability in reversible diffusion processes II: Precise asymptotics for small eigenvalues. *Journal of the European Mathematical Society* **7** 69–99.

[15] BUDHIRAJA, A. and DUPUIS, P. (2000). A variational representation for positive functionals of infinite dimensional Brownian motion. *Probability and mathematical statistics-Wroclaw University* **20** 39–61.

[16] BUDHIRAJA, A., DUPUIS, P. and MAROULAS, V. (2008). Large deviations for infinite dimensional stochastic dynamical systems. *The Annals of Probability* **36** 1390 – 1420.

[17] BURACZEWSKI, D., DAMEK, E., MIKOSCH, T. et al. (2016). Stochastic models with power-law tails. *The equation X= AX+ B. Cham: Springer*.

[18] CASSANDRO, M., GALVES, A., OLIVIERI, E. and VARES, M. E. (1984). Metastable behavior of stochastic dynamics: a pathwise approach. *Journal of statistical physics* **35** 603–634.

[19] CERRAI, S. and RÖCKNER, M. (2004). Large deviations for stochastic reaction-diffusion systems with multiplicative noise and non-Lipshitz reaction term. *The Annals of Probability* **32** 1100 – 1139.

[20] CHEN, B., BLANCHET, J., RHEE, C.-H. and ZWART, B. (2019). Efficient rare-event simulation for multiple jump events in regularly varying random walks and compound Poisson processes. *Mathematics of Operations Research* **44** 919–942.

[21] CHEN, B., RHEE, C.-H. and ZWART, B. (2024). Sample-path large deviations for a class of heavy-tailed Markov-additive processes. *Electronic Journal of Probability* **29** 1 – 44.

[22] CHIGANSKY, P. and LIPTSER, R. (2007). *The Freidlin-Wentzell LDP with Rapidly Growing Coeffcients* In *Stochastic Differential Equations: Theory And Applications: A Volume in Honor of Professor Boris L Rozovskii* 197-219. World Scientific, Hackensack, NJ.

[23] DEBUSSCHE, A., HÖGELE, M. and IMKELLER, P. (2013). *The dynamics of nonlinear reaction-diffusion equations with small Lévy noise* **2085**. Springer, New York, NY.

[24] DEMBO, A. and ZEITOUNI, O. (2010). *Large deviations techniques and applications*. *Stochastic Modelling and Applied Probability* **38**. Springer-Verlag, Berlin, New York, NY. Corrected reprint of the second (1998) edition. MR2571413

[25] DENISOV, D., DIEKER, A. B. and SHNEER, V. (2008). Large deviations for random walks under subexponentiality: The big-jump domain. *The Annals of Probability* **36** 1946 – 1991.

[26] DEUSCHEL, J.-D. and STROOCK, D. W. (1989). *Large deviations*. *Pure and Applied Mathematics* **137**. Academic Press, Inc., Boston, MA. MR997938

[27] DONATI-MARTIN, C. (2008). Large deviations for Wishart processes. *Probability and Mathematical Statistics* **28**.

[28] DONATI-MARTIN, C., ROUAULT, A., YOR, M. and ZANI, M. (2004). Large deviations for squares of Bessel and Ornstein-Uhlenbeck processes. *Probability Theory and Related Fields* **129** 261–289.

[29] DUPUIS, P. and ELLIS, R. S. (2011). *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, Hoboken, NJ.

[30] EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (2013). *Modelling extremal events: for insurance and finance* **33**. Springer Science & Business Media, New York, NY.

[31] ENGSTROM, L., ILYAS, A., SANTURKAR, S., TSIPRAS, D., JANOOS, F., RUDOLPH, L. and MADRY, A. (2020). Implementation Matters in Deep RL: A Case Study on PPO and TRPO. In *International Conference on Learning Representations*.

[32] EYRING, H. (1935). The Activated Complex in Chemical Reactions. *The Journal of Chemical Physics* **3** 107-115.

[33] FENG, J. and KURTZ, T. G. (2006). *Large deviations for stochastic processes*. *Mathematical Surveys and Monographs* **131**. American Mathematical Society, Providence, RI. MR2260560

[34] FOSS, S. and KORSHUNOV, D. (2006). Heavy tails in multi-server queue. *Queueing Systems* **52** 31–48.

[35] FOSS, S. and KORSHUNOV, D. (2012). On Large Delays in Multi-Server Queues with Heavy Tails. *Mathematics of Operations Research* **37** 201-218.

[36] FOSS, S., KORSHUNOV, D., ZACHARY, S. et al. (2011). *An introduction to heavy-tailed and subexponential distributions* **6**. Springer, New York, NY.

[37] FREIDLIN, M. I. and WENTZELL, A. D. (1970). ON SMALL RANDOM PERTURBATIONS OF DYNAMICAL SYSTEMS. *Russian Mathematical Surveys* **25** 1.

[38] FREIDLIN, M. I. and WENTZELL, A. D. (1973). Some problems concerning stability under small random perturbations. *Theory of Probability & Its Applications* **17** 269–283.

[39] FREIDLIN, M. I. and WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York, NY.

[40] GAUDILLIERE, A. and LANDIM, C. (2014). A Dirichlet principle for non reversible Markov chains and some recurrence theorems. *Probability Theory and Related Fields* **158** 55–89.

[41] GLASSTONE, S., LAIDLER, K. J. and EYRING, H. (1941). *The theory of rate processes: the kinetics of chemical reactions, viscosity, diffusion and electrochemical phenomena*. McGraw-Hill, New York.

[42] GRAVES, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.

[43] HÖGELE, M. and PAVLYUKEVICH, I. (2015). Metastability in a class of hyperbolic dynamical systems perturbed by heavy-tailed Lévy type noise. *Stochastics and Dynamics* **15** 1550019.

[44] HULT, H. and LINDSKOG, F. (2006). Regular variation for measures on metric spaces. *Publications de l'Institut Mathématique* **80** 121–140.

[45] HULT, H., LINDSKOG, F., MIKOSCH, T. and SAMORODNITSKY, G. (2005). Functional large deviations for multivariate regularly varying random walks. *The Annals of Applied Probability* **15** 2651 – 2680.

[46] HÖGELE, M. and PAVLYUKEVICH, I. (2014). The Exit Problem from a Neighborhood of the Global Attractor for Dynamical Systems Perturbed by Heavy-Tailed Lévy Processes. *Stochastic Analysis and Applications* **32** 163–190.

[47] IMKELLER, P. and PAVLYUKEVICH, I. (2006). First exit times of SDEs driven by stable Lévy processes. *Stochastic Processes and their Applications* **116** 611-642.

[48] IMKELLER, P. and PAVLYUKEVICH, I. (2008). Lévy flights: transitions and meta-stability. *Journal of Physics A: Mathematical and General* **39** L237.

[49] IMKELLER, P. and PAVLYUKEVICH, I. (2008). Metastable behaviour of small noise Lévy-Driven diffusions. *ESAIM: PS* **12** 412-437.

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

[51] IMKELLER, P., PAVLYUKEVICH, I. and WETZEL, T. (2009). First exit times for Lévy-driven diffusions with exponentially light jumps. *The Annals of Probability* **37** 530 – 564.

[52] IMKELLER, P., PAVLYUKEVICH, I. and WETZEL, T. (2010). The hierarchy of exit times of Lévy-driven Langevin equations. *The European Physical Journal Special Topics* **191** 211–222.

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

[54] KOLOSKOVA, A., HENDRIKX, H. and STICH, S. U. (2023). Revisiting Gradient Clipping: Stochastic bias and tight convergence guarantees. In *Proceedings of the 40th International Conference on Machine Learning* (A. KRAUSE, E. BRUNSKILL, K. CHO, B. ENGELHARDT, S. SABATO and J. SCARLETT, eds.). *Proceedings of Machine Learning Research* **202** 17343–17363. PMLR, Honolulu, HI.

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

[56] LANDIM, C. (2014). Metastability for a non-reversible dynamics: the evolution of the condensate in totally asymmetric zero range processes. *Communications in Mathematical Physics* **330** 1–32.

[57] LEE, J. and SEO, I. (2022). Non-reversible metastable diffusions with Gibbs invariant measure I: Eyring–Kramers formula. *Probability Theory and Related Fields* **182** 849–903.

[58] LEE, J. and SEO, I. (2022). Non-reversible metastable diffusions with Gibbs invariant measure II: Markov chain convergence. *Journal of Statistical Physics* **189** 25.

[59] LINDSKOG, F., RESNICK, S. I., ROY, J. et al. (2014). Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps. *Probability Surveys* **11** 270–314.

[60] LYNCH, J. and SETHURAMAN, J. (1987). Large deviations for processes with independent increments. *The annals of probability* **15** 610–627.

[61] MERITY, S., KESKAR, N. S. and SOCHER, R. (2018). Regularizing and Optimizing LSTM Language Models. In *International Conference on Learning Representations*.

[62] MOGULSKII, A. (1993). Large deviations for processes with independent increments. *The annals of probability* **12** 202–215.

[63] MOHAN, M. T. (2021). Wentzell–Freidlin Large Deviation Principle for Stochastic Convective Brinkman–Forchheimer Equations. *Journal of Mathematical Fluid Mechanics* **23** 62.

[64] MORROW, G. J. and SAWYER, S. (1989). Large Deviation Results for a Class of Markov Chains Arising from Population Genetics. *The Annals of Probability* **17** 1124–1146.

[65] NAGAEV, A. (1969). Limit theorems for large deviations where Cramér's conditions are violated. *Fiz-Mat. Nauk* **7** 17–22.

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

[67] NGUYEN, T. H., SIMSEKLI, U., GURBUZBALABAN, M. and RICHARD, G. (2019). First Exit Time Analysis of Stochastic Gradient Descent Under Heavy-Tailed Gradient Noise. In *Advances in Neural Information Processing Systems* (H. WALLACH, H. LAROCHELLE, A. BEYGELZIMER, F. D' ALCHÉ-BUC, E. FOX and R. GARNETT, eds.) **32**. Curran Associates, Inc., Vancouver.

[68] OLIVIERI, E. and VARES, M. E. (2005). *Large Deviations and Metastability*. *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge.

[69] PASCANU, R. (2013). On the difficulty of training recurrent neural networks. *arXiv preprint arXiv:1211.5063*.

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

[71] PAVLYUKEVICH, I. (2011). First Exit Times of Solutions of Stochastic Differential Equations Driven by Multiplicative Lévy Noise with Heavy Tails. *Stochastics and Dynamics* **11** 495-519.

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

[73] RESNICK, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media, New York, NY.

[74] REZAKHANLOU, F. and SEO, I. (2023). Scaling limit of small random perturbation of dynamical systems. In *Annales de l'Institut Henri Poincare (B) Probabilites et statistiques* **59** 867–903. Institut Henri Poincaré.

[75] RHEE, C.-H., BLANCHET, J., ZWART, B. et al. (2019). Sample path large deviations for Lévy processes and random walks with regularly varying increments. *The Annals of Probability* **47** 3551–3605.

[76] SIMSEKLI, U., SAGUN, L. and GURBUZBALABAN, M. (2019). A tail-index analysis of stochastic gradient noise in deep neural networks. In *International Conference on Machine Learning* 5827–5837. PMLR.

[77] SLOWIK, M. (2012). A note on variational representations of capacities for reversible and nonreversible Markov chains. *Unpublished, Technische Universität Berlin*.

[78] SOWERS, R. B. (1992). Large Deviations for a Reaction-Diffusion Equation with Non-Gaussian Perturbations. *The Annals of Probability* **20** 504–537.

[79] TANKOV, P. (2003). *Financial modelling with jump processes*. Chapman and Hall/CRC, Boca Raton, FL.

[80] VARADHAN, S. R. S. (1984). *Large deviations and applications*. *CBMS-NSF Regional Conference Series in Applied Mathematics* **46**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. MR758258

[81] WANG, X., OH, S. and RHEE, C.-H. (2022). Eliminating Sharp Minima from SGD with Truncated Heavy-tailed Noise. In *International Conference on Learning Representations*.

[82] WANG, X. and RHEE, C.-H. Global Dynamics of Heavy-Tailed SGD in Nonconvex Loss Landscape: characterization and Control. https://chrhee.github.io/papers/WangRhee24b.pdf. *Working Paper*.

[83] ZHANG, J., HE, T., SRA, S. and JADBABAIE, A. (2020). Why Gradient Clipping Accelerates Training: A Theoretical Justification for Adaptivity. In *International Conference on Learning Representations*.