Large Deviation Properties of the Empirical Measure of a Metastable Small Noise Diffusion

Paul Dupuis[1] · Guo-Jhen Wu[2]

Abstract

The aim of this paper is to develop tractable large deviation approximations for the empirical measure of a small noise diffusion. The starting point is the Freidlin–Wentzell theory, which shows how to approximate via a large deviation principle the invariant distribution of such a diffusion. The rate function of the invariant measure is formulated in terms of quasipotentials, quantities that measure the difficulty of a transition from the neighborhood of one metastable set to another. The theory provides an intuitive and useful approximation for the invariant measure, and along the way many useful related results (e.g., transition rates between metastable states) are also developed. With the specific goal of design of Monte Carlo schemes in mind, we prove large deviation limits for integrals with respect to the empirical measure, where the process is considered over a time interval whose length grows as the noise decreases to zero. In particular, we show how the first and second moments of these integrals can be expressed in terms of quasipotentials. When the dynamics of the process depend on parameters, these approximations can be used for algorithm design, and applications of this sort will appear elsewhere. The use of a small noise limit is well motivated, since in this limit good sampling of the state space becomes most challenging. The proof exploits a regenerative structure, and a number of new techniques are needed to turn large deviation estimates over a regenerative cycle into estimates for the empirical measure and its moments.

Keywords Large deviations · Freidlin–Wentzell theory · Small noise diffusion · Empirical measure · Quasipotential · Monte Carlo method

✉ Guo-Jhen Wu
   gjwu@kth.se

[1]   Division of Applied Mathematics, Brown University, Providence, USA

[2]   Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden

Mathematics Subject Classification 60F10 · 60J60 · 65C05 · 37A50 · 60K10

1 Introduction

Among the many interesting results proved by Freidlin and Wentzell in the 70's and 80's concerning small random perturbations of dynamical systems, one of particular note is the large deviation principle for the invariant measure of such a system. Consider the small noise diffusion

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dW_t, \quad X_0^\varepsilon = x,$$

where $X_t^\varepsilon \in \mathbb{R}^d$, $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^k$ (the $d \times k$ matrices) and $W_t \in \mathbb{R}^k$ is a standard Brownian motion. Under mild regularity conditions on b and σ, one has that for any $T \in (0, \infty)$ the processes $\{X_\cdot^\varepsilon\}_{\varepsilon>0}$ satisfy a large deviation principle on $C([0, T] : \mathbb{R}^d)$ with rate function

$$I_T(\phi) \doteq \int_0^T \sup_{\alpha \in \mathbb{R}^d} \left[\langle \dot{\phi}_t, \alpha \rangle - \langle b(\phi_t), \alpha \rangle - \frac{1}{2} \|\sigma(\phi_t)\alpha\|^2 \right] dt$$

when ϕ is absolutely continuous and $\phi(0) = x$, and $I_T(\phi) = \infty$ otherwise. If $\sigma(x)\sigma(x)' > 0$ (in the sense of symmetric square matrices) for all $x \in \mathbb{R}^d$, then one can evaluate the supremum and find

$$I_T(\phi) = \int_0^T \frac{1}{2} \left\langle \dot{\phi}_t - b(\phi_t), \left[\sigma(\phi_t)\sigma(\phi_t)' \right]^{-1} (\dot{\phi}_t - b(\phi_t)) \right\rangle dt. \qquad (1.1)$$

To simplify the discussion, we will assume this non-degeneracy condition. It is also assumed by Freidlin and Wentzell in [12], but can be weakened.

Define the **quasipotential** $V(x, y)$ for $x, y \in \mathbb{R}^d$ by

$$V(x, y) \doteq \inf \{ I_T(\phi) : \phi(0) = x, \phi(T) = y, T < \infty \}.$$

Suppose that $\{X^\varepsilon\}$ is ergodic on a compact manifold $M \subset \mathbb{R}^d$ with invariant measure $\mu^\varepsilon \in \mathcal{P}(M)$. Then, under a number of additional assumptions, including assumptions on the structure of the dynamical system $\dot{X}_t^0 = b(X_t^0)$, Freidlin and Wentzell [12, Chapter 6] show how to construct a function $J : M \to [0, \infty]$ in terms of V, such that J is the large deviation rate function for $\{\mu^\varepsilon\}_{\varepsilon>0}$: J has compact level sets, and

$$\liminf_{\varepsilon \to 0} \varepsilon \log \mu^\varepsilon(G) \geq - \inf_{y \in G} J(y) \text{ for open } G \subset M,$$
$$\limsup_{\varepsilon \to 0} \varepsilon \log \mu^\varepsilon(F) \leq - \inf_{y \in F} J(y) \text{ for closed } F \subset M.$$

This gives a very useful approximation to μ^ε, and along the way many interesting related results (e.g., transition rates between metastable states) are also developed.

The aim of this paper is to develop large deviation type estimates for a quantity that is closely related to μ^ε, which is the empirical measure over an interval $[0, T^\varepsilon]$. This is defined by

$$\rho^\varepsilon(A) \doteq \frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} 1_A(X_s^\varepsilon) ds \qquad (1.2)$$

for $A \in \mathcal{B}(M)$. For reasons that will be made precise later on, we will assume $T^\varepsilon \to \infty$ as $\varepsilon \to 0$, and typically T^ε will grow exponentially in the form $e^{c/\varepsilon}$ for some $c > 0$.

There is of course a large deviation theory for the empirical measure when $\varepsilon > 0$ is held fixed and the length of the time interval tends to infinity (see, e.g., [7,8]). However, it can be hard to extract information from the corresponding rate function. Our interest in proving large deviations estimates when $\varepsilon \to 0$ and $T^\varepsilon \to \infty$ is in the hope that one will find it easier to extract information in this double limit, analogous to the simplified approximation to μ^ε just mentioned. These results will be applied in [10] to analyze and optimize a Monte Carlo method known as infinite swapping [9,15] when the noise is small. Small noise models are common in applications and are also the setting in which Monte Carlo methods can have the greatest difficulty. We expect that the general set of results will be useful for other purposes as well.

We note that while developed in the context of small noise diffusions, the collection of results due to Freidlin and Wentzell that are discussed in [12] also hold for other classes of processes, such as scaled stochastic networks, when appropriate conditions are assumed and the finite time sample path large deviation results are available (see, e.g., [19]). We expect that such generalizations are possible for the results we prove as well.

The outline of the paper is as follows. In Sect. 2, we explain our motivation and the relevance for studying the particular quantities that are the topic of the paper. In Sect. 3, we provide definitions and assumptions that are used throughout the paper, and Sect. 4 states the main asymptotic results as well as a related conjecture. Examples that illustrate the results are given in Sect. 5. In Sect. 6, we introduce an important tool for our analysis—the regenerative structure, and with this concept, we decompose the original asymptotic problem into two sub-problems that require very different forms of analysis. These two types of asymptotic problems are then analyzed separately in Sects. 7 and 8. In Sect. 9, we combine the partial asymptotic results from Sects. 7 and 8 to prove the main large deviation type results that are stated in Section 4. Section 10 gives the proof of a key theorem from Section 8, which asserts an approximately exponential distribution for return times that arise in the decomposition based on regenerative structure, as well as a tail bound needed for some integrability arguments. The last section of the paper, Sect. 11, presents the proof of an upper bound for the rate of decay of the variance per unit time in the context of a special case, thereby showing for the case that the lower bounds of Sect. 4 are in a sense tight. To focus on the main discussion, proofs of some lemmas are collected in an Appendix.

Remark 1.1 There are certain time-scaling parameters that play key roles throughout this paper. For the reader's convenience, we record here where they are first described: h_1 and w are defined in (4.1) and (4.2); c is introduced and its relation to h_1 and w is given in Theorem 4.3; m is introduced at the beginning of Sect. 6.2.

2 Quantities of Interest

The quantities we are interested in are the higher order moments, and in particular second moments, of an integral of a risk-sensitive functional with respect to the empirical measure ρ^ε defined in (1.2). To be more precise, the integral is of the form

$$\int_M e^{-\frac{1}{\varepsilon} f(x)} 1_A(x) \, \rho^\varepsilon(dx) \tag{2.1}$$

for some nice (e.g., bounded and continuous) function $f : M \to \mathbb{R}$ and a closed set $A \in \mathcal{B}(M)$. Note that this integral can also be expressed as

$$\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt. \tag{2.2}$$

In order to understand the large deviation behavior of moments of such an integral, we must identify the correct scaling to extract meaningful information. Moreover, as will be shown, there is an important difference between centered moments and ordinary (non-centered) moments.

By the use of the regenerative structure of $\{X_t^\varepsilon\}_{t \geq 0}$, we can decompose (2.2) [equivalently (2.1)] into the sum of a random number of independent and identically distributed (iid) random variables, plus a residual term which here we will ignore. To simplify the notation, we temporarily drop the ε, and without being precise about how the regenerative structure is introduced, let Y_j denote the integral of $e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon)$ over a regenerative cycle. (The specific regenerative structure we use will be identified later on.)

Thus, we consider a sequence $\{Y_j\}_{j \in \mathbb{N}}$ of iid random variables with finite second moments and want to compare the scaling properties of, for example, the second moment and the second centered moment of $\frac{1}{n} \sum_{j=1}^n Y_j$. When used for the small noise system, both n and moments of Y_i will scale exponentially in $1/\varepsilon$, and n will be random, but for now we assume n is deterministic. The second moment is

$$E\left(\frac{1}{n}\sum_{k=1}^n Y_k\right)^2 = \frac{1}{n^2}\sum_{k=1}^n E(Y_k)^2 + \frac{1}{n^2}\sum_{i,j:i \neq j} E(Y_i Y_j) = (EY_1)^2 + \frac{1}{n}\mathrm{Var}(Y_1),$$

and the second centered moment is

$$E\left(\frac{1}{n}\sum_{k=1}^n (Y_k - EY_1)\right)^2 = \mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^n Y_k\right) = \frac{1}{n}\mathrm{Var}(Y_1).$$

When analyzing the performance of the Monte Carlo schemes, one is concerned of course with both bias and variance, but in situations where we would like to apply the results of this paper one assumes T^ε is large enough that the bias term is unimportant, so that all we are concerned with is the variance. However, some care will be needed to determine a suitable measure of quality of the algorithm, since as noted Y_i could scale exponentially with in $1/\varepsilon$ with a negative coefficient (exponentially small), while n will be exponentially large.

In the analysis of unbiased accelerated Monte Carlo methods for small noise systems over bounded time intervals (e.g., to estimate escape probabilities), it is standard to use the second moment, which is often easier to analyze, in lieu of the variance [3, Chapter VI], [4, Chapter 14]. This situation corresponds to $n = 1$. The alternative criterion is more convenient since by Jensen's inequality one can easily establish a best possible rate of decay of the second moment, and estimators are deemed efficient if they possess the optimal rate of decay [3,4]. However, with n exponentially large this is no longer true. Using the previous calculations, we see that the second moment of $\frac{1}{n} \sum_{j=1}^{n} Y_j$ can be completely dominated by $(E Y_1)^2$, and therefore using this quantity to compare algorithms may be misleading, since our true concern is the variance of $\frac{1}{n} \sum_{j=1}^{n} Y_j$.

This observation suggests that our study of moments of the empirical measure we should consider only centered moments, and in particular quantities like

$$T^\varepsilon \mathrm{Var} \left(\int_M e^{-\frac{1}{\varepsilon} f(x)} 1_A(x) \, \rho^\varepsilon(dx) \right) = T^\varepsilon \mathrm{Var} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt \right),$$

which is the variance per unit time. For Monte Carlo, one wants to minimize the variance per unit time, and to make the problem more tractable we instead try to maximize the decay rate of the variance per unit time. Assuming the limit exists, this is defined by

$$\lim_{\varepsilon \to 0} -\varepsilon \log \left[T^\varepsilon \mathrm{Var} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt \right) \right]$$

and so we are especially interested in lower bounds on this decay rate.

Thus, our goal is to develop methods that allow the approximation of at least first and second moments of (2.2). In fact, the methods we introduce can be developed further to obtain large deviation estimates of higher moments if that were needed or desired.

3 Setting of the Problem, Assumptions and Definitions

The process model we would like to consider is an $\mathbb{R}^d$-valued solution to an Itô stochastic differential equation (SDE), where the drift so strongly returns the process to some compact set that events involving exit of the process from some larger compact set are so rare that they can effectively be ignored when analyzing the empirical measure. However, to simplify the analysis we follow the convention of [12, Chapter

6], and work with a small noise diffusion that takes values in a compact and connected manifold $M \subset \mathbb{R}^d$ of dimension r and with smooth boundary. The precise regularity assumptions for M are given on [12, page 135]. With this convention in mind, we consider a family of diffusion processes $\{X^\varepsilon\}_{\varepsilon \in (0,\infty)}$, $X^\varepsilon \in C([0,\infty) : M)$, that satisfy the following condition.

Condition 3.1 *Consider continuous $b : M \to \mathbb{R}^d$ and $\sigma : M \to \mathbb{R}^d \times \mathbb{R}^d$ (the $d \times d$ matrices) and assume that σ is uniformly nondegenerate, in that there is $c > 0$ such that for any x and any v in the tangent space of M at x, $\langle v, \sigma(x)\sigma(x)'v \rangle \geq c\langle v, v \rangle$. For absolutely continuous $\phi \in C([0,T] : M)$ define $I_T(\phi)$ by (1.1), where the inverse $\left[\sigma(x)\sigma(x)'\right]^{-1}$ is relative to the tangent space of M at x. Let $I_T(\phi) = \infty$ for all other $\phi \in C([0,T] : M)$. Then, we assume that for each $T < \infty$, $\{X_t^\varepsilon\}_{0 \leq t \leq T}$ satisfies the large deviation principle with rate function I_T, uniformly with respect to the initial condition [4, Definition 1.13].*

We note that for such diffusion processes nondegeneracy of the diffusion matrix implies there is a unique invariant measure $\mu^\varepsilon \in \mathcal{P}(M)$. A discussion of weak sufficient conditions under which Condition 3.1 holds appears in [12, Sect. 3, Chapter 5].

Remark 3.2 There are several ways one can approximate a diffusion of the sort described at the beginning of this section by a diffusion on a smooth compact manifold. One such "compactification" of the state space can be obtained by assuming that for some bounded but large enough rectangle trajectories that exit the rectangle do not affect the large deviation behavior of quantities of interest and then extend the coefficients of the process periodically and smoothly off an even larger rectangle to all of $\mathbb{R}^d$ (a technique sometimes used to bound the state space for purposes of numerical approximation). One can then map $\mathbb{R}^d$ to a manifold that is topologically equivalent to a torus, and even arrange that the metric structure on the part of the manifold corresponding to the smaller rectangle coincides with a Euclidean metric.

Define the **quasipotential** $V(x, y) : M \times M \to [0, \infty)$ by

$$V(x, y) \doteq \inf \{I_T(\phi) : \phi(0) = x, \phi(T) = y, T < \infty\}. \tag{3.1}$$

For a given set $A \subset M$, define $V(x, A) \doteq \inf_{y \in A} V(x, y)$ and $V(A, y) \doteq \inf_{x \in A} V(x, y)$.

Remark 3.3 For any fixed y and set A, $V(x, y)$ and $V(x, A)$ are both continuous functions of x. Similarly, for any given x and any set A, $V(x, y)$ and $V(A, y)$ are also continuous in y.

Definition 3.4 We say that a set $N \subset M$ is **stable** if for any $x \in N$, $y \notin N$ we have $V(x, y) > 0$. A set which is not stable is called **unstable**.

Definition 3.5 We say that $O \in M$ is an **equilibrium point** of the ordinary differential equation (ODE) $\dot{x}_t = b(x_t)$ if $b(O) = 0$. Moreover, we say that this equilibrium point O is **asymptotically stable** if for every neighborhood $\mathcal{E}_1$ of O (relative to M) there exists a smaller neighborhood $\mathcal{E}_2$ such that the trajectories of system $\dot{x}_t = b(x_t)$ starting in $\mathcal{E}_2$ converge to O without leaving $\mathcal{E}_1$ as $t \to \infty$.

Remark 3.6 An asymptotically stable equilibrium point is a stable set, but a stable set might contain no asymptotically stable equilibrium point.

The following restrictions on the structure of the dynamical system in M will be used. These restrictions include the assumption that the equilibrium points are a finite collection. This is a more restrictive framework than that of [12], which allows, e.g., limit cycles. In a remark at the end of this section, we comment on what would be needed to extend to the general setup of [12].

Condition 3.7 *There exists a finite number of points $\{O_j\}_{j\in L} \subset M$ with $L \doteq \{1, 2, \ldots, l\}$ for some $l \in \mathbb{N}$, such that $\cup_{j\in L}\{O_j\}$ coincides with the ω-limit set of the ODE $\dot{x}_t = b(x_t)$.*

Without loss of generality, we may assume that O_j is stable if and only if $j \in L_s$ where $L_s \doteq \{1, \ldots, l_s\}$ for some $l_s \leq l$.

Next, we give a definition from graph theory which will be used in the statement of the main results.

Definition 3.8 Given a subset $W \subset L = \{1, \ldots, l\}$, a directed graph consisting of arrows $i \rightarrow j$ ($i \in L \setminus W, j \in L, i \neq j$) is called a W**-graph on** L if it satisfies the following conditions.

1. Every point $i \in L \setminus W$ is the initial point of exactly one arrow.
2. For any point $i \in L \setminus W$, there exists a sequence of arrows leading from i to some point in W.

We note that we could replace the second condition by the requirement that there are no closed cycles in the graph. We denote by $G(W)$ the set of W-graphs; we shall use the letter g to denote graphs. Moreover, if p_{ij} ($i, j \in L, j \neq i$) are numbers, then $\prod_{(i\rightarrow j)\in g} p_{ij}$ will be denoted by $\pi(g)$.

Remark 3.9 We mostly consider the set of $\{i\}$-graphs, i.e., $G(\{i\})$ for some $i \in L$, and also use $G(i)$ to denote $G(\{i\})$. We occasionally consider the set of $\{i, j\}$-graphs, i.e., $G(\{i, j\})$ for some $i, j \in L$ with $i \neq j$. Again, we also use $G(i, j)$ to denote $G(\{i, j\})$.

Definition 3.10 For all $j \in L$, define

$$W\left(O_j\right) \doteq \min_{g\in G(j)} \left[\sum_{(m\rightarrow n)\in g} V\left(O_m, O_n\right)\right] \tag{3.2}$$

and

$$W\left(O_1 \cup O_j\right) \doteq \min_{g\in G(1,j)} \left[\sum_{(m\rightarrow n)\in g} V\left(O_m, O_n\right)\right]. \tag{3.3}$$

Remark 3.11 Heuristically, if we interpret $V\left(O_m, O_n\right)$ as the "cost" of moving from O_m to O_n, then $W\left(O_j\right)$ is the "least total cost" of reaching O_j from every O_i with $i \in L \setminus \{j\}$. According to [12, Theorem 4.1, Chapter 6], one can interpret $W(O_i) - \min_{j\in L} W(O_j)$ as the decay rate of $\mu^\varepsilon(B_\delta(O_i))$, where $B_\delta(O_i)$ is a small open neighborhood of O_i.

Definition 3.12 We use $G_s(W)$ to denote the collection of all W-graphs on $L_s = \{1, \dots, l_s\}$ with $W \subset L_s$.

We make the following technical assumptions on the structure of the SDE. Let $B_\delta(K)$ denote the δ-neighborhood of a set $K \subset M$. Recall that μ^ε is the unique invariant probability measure of the diffusion process $\{X_t^\varepsilon\}_t$. The existence of the limits appearing in the first part of the condition is ensured by Theorem 4.1 in [12, Chapter 6].

Condition 3.13 *1. There exists a unique asymptotically stable equilibrium point O_1 of the system $\dot{x}_t = b(x_t)$ such that*

$$\lim_{\delta \to 0} \lim_{\varepsilon \to 0} -\varepsilon \log \mu^\varepsilon(B_\delta(O_1)) = 0, \ and$$
$$\lim_{\delta \to 0} \lim_{\varepsilon \to 0} -\varepsilon \log \mu^\varepsilon(B_\delta(O_j)) > 0 \ for \ any \ j \in L \setminus \{1\}.$$

2. All of the eigenvalues of the matrix of partial derivatives of b at O_ℓ relative to M have negative real parts for $\ell \in L_s$.
3. $b : M \to \mathbb{R}^d$ and $\sigma : M \to \mathbb{R}^d \times \mathbb{R}^d$ are C^1.

Remark 3.14 According to [12, Theorem 4.1, Chapter 6] and the first part of Condition 3.13, we know that $W(O_j) > W(O_1)$ for all $j \in L \setminus \{1\}$.

Remark 3.15 We comment on the use of the various parts of the condition. Part 1 means that neighborhoods of O_1 capture more of the mass as $\varepsilon \to 0$ than neighborhoods of any other equilibrium point. It simplifies the analysis greatly, but we expect it could be weakened if desired. Parts 2 and 3 are assumed in [6], which gives an explicit exponential bound on the tail probability of the exit time from the domain of attraction. It is largely because of our reliance on the results of [6] that we must assume that equilibrium sets are points in Condition 3.7, rather than the more general compacta as considered in [12]. Both Condition 3.7 and 3.13 could be weakened if the corresponding versions of the results we use from [6] were available.

Remark 3.16 The quantities $V(O_i, O_j)$ determine various key transition probabilities and time scales in the analysis of the empirical measure. The more general framework of [12], as well as the one-dimensional case (i.e., $r = 1$) in the present setting, requires some closely related but slightly more complicated quantities. These are essentially the analogues of $V(O_i, O_j)$ under the assumption that trajectories used in the definition are not allowed to pass through equilibrium compacta (such as a limit cycle) when traveling from O_i to O_j. The related quantities, which are designated using notation of the form $\tilde{V}(O_i, O_j)$ in [12], are needed since the probability of a direct transition from O_i to O_j without passing though another equilibrium structure may be zero, which means that transitions from O_i to O_j must be decomposed according to these intermediate transitions. To simplify the presentation, we do not provide the details of the one-dimensional case in our setup, but simply note that it can be handled by the introduction of these additional quantities.

Consider the filtration $\{\mathcal{F}_t\}_{t \geq 0}$ defined by $\mathcal{F}_t \doteq \sigma(X_s^\varepsilon, s \leq t)$ for any $t \geq 0$. For any $\delta > 0$ smaller than a quarter of the minimum of the distances between O_i and O_j for all $i \neq j$, we consider two types of stopping times with respect to the filtration $\{\mathcal{F}_t\}_t$. The first type are the hitting times of $\{X_t^\varepsilon\}_t$ at the δ-neighborhood of all equilibrium points $\{O_j\}_{j \in L}$ after traveling a reasonable distance away from those neighborhoods. More precisely, we define stopping times by $\tau_0 \doteq 0$,

$$\sigma_n \doteq \inf\{t > \tau_n : X_t^\varepsilon \in \cup_{j \in L} \partial B_{2\delta}(O_j)\} \text{ and } \tau_n \doteq \inf\{t > \sigma_{n-1} : X_t^\varepsilon \in \cup_{j \in L} \partial B_\delta(O_j)\}.$$

The second type of stopping times is the return times of $\{X_t^\varepsilon\}_t$ to the δ-neighborhood of O_1, where as noted previously O_1 is in some sense the most important equilibrium point. The exact definitions are $\tau_0^\varepsilon \doteq 0$,

$$\sigma_n^\varepsilon \doteq \inf\{t > \tau_n^\varepsilon : X_t^\varepsilon \in \cup_{j \in L \setminus \{1\}} \partial B_\delta(O_j)\} \text{ and } \tau_n^\varepsilon \doteq \inf\left\{t > \sigma_{n-1}^\varepsilon : X_t^\varepsilon \in \partial B_\delta(O_1)\right\}. \tag{3.4}$$

We then define two embedded Markov chains $\{Z_n\}_{n \in \mathbb{N}_0} \doteq \{X_{\tau_n}^\varepsilon\}_{n \in \mathbb{N}_0}$ with state space $\cup_{j \in L} \partial B_\delta(O_j)$, and $\{Z_n^\varepsilon\}_{n \in \mathbb{N}_0} \doteq \{X_{\tau_n^\varepsilon}^\varepsilon\}_{n \in \mathbb{N}_0}$ with state space $\partial B_\delta(O_1)$.

Let $p(x, \partial B_\delta(O_j))$ denote the one-step transition probabilities of $\{Z_n\}_{n \in \mathbb{N}_0}$ starting from a point $x \in \cup_{i \in L} \partial B_\delta(O_i)$, namely,

$$p(x, \partial B_\delta(O_j)) \doteq P_x(Z_1 \in \partial B_\delta(O_j)).$$

We have the following estimates on $p(x, \partial B_\delta(O_j))$ in terms of V. The lemma is a consequence of [12, Lemma 2.1, Chapter 6] and the fact that under our conditions $V(O_i, O_j)$ and $\tilde{V}(O_i, O_j)$ as defined in [12] coincide.

Lemma 3.17 *For any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$ and $\varepsilon \in (0, \varepsilon_0)$, for all $x \in \partial B_\delta(O_i)$, the one-step transition probability of the Markov chain $\{Z_n\}_{n \in \mathbb{N}}$ on $\partial B_\delta(O_j)$ satisfies*

$$e^{-\frac{1}{\varepsilon}(V(O_i, O_j) + \eta)} \leq p(x, \partial B_\delta(O_j)) \leq e^{-\frac{1}{\varepsilon}(V(O_i, O_j) - \eta)}$$

for any $i, j \in L$.

Remark 3.18 According to Lemma 4.6 in [16], Condition 3.1 guarantees the existence and uniqueness of invariant measures for $\{Z_n\}_n$ and $\{Z_n^\varepsilon\}_n$. We use $\nu^\varepsilon \in \mathcal{P}(\cup_{i \in L} \partial B_\delta(O_i))$ and $\lambda^\varepsilon \in \mathcal{P}(\partial B_\delta(O_1))$ to denote the associated invariant measures.

4 Results and a Conjecture

The following main results of this paper assume Conditions 3.1, 3.7 and 3.13. Although moments higher than the second moment are not considered in this paper, as noted previously one can use arguments such as those used here to identify and prove the analogous results.

Recall that $\{O_j\}_{j \in L}$ are the equilibrium points and that they satisfy Conditions 3.7 and 3.13. In addition, O_j is stable if and only if $j \in L_s$, where $L_s \doteq \{1, \ldots, l_s\}$ for some $l_s \leq l = |L|$, and τ_1^ε is the first return time to the δ-neighborhood of O_1 after having first visited the δ-neighborhood of any other equilibrium point.

Lemma 4.1 *For any $\delta \in (0, 1)$ smaller than a quarter of the minimum of the distances between O_i and O_j for all $i \neq j$, any $\varepsilon > 0$ and any nonnegative measurable function $g : M \to \mathbb{R}$*

$$E_{\lambda^\varepsilon} \left(\int_0^{\tau_1^\varepsilon} g\left(X_s^\varepsilon\right) ds \right) = E_{\lambda^\varepsilon} \tau_1^\varepsilon \cdot \int_M g(x) \mu^\varepsilon (dx),$$

where $\lambda^\varepsilon \in \mathcal{P}(\partial B_\delta(O_1))$ is the unique invariant measure of $\{Z_n^\varepsilon\}_n = \{X_{\tau_n^\varepsilon}^\varepsilon\}_n$ and $\mu^\varepsilon \in \mathcal{P}(M)$ is the unique invariant measure of $\{X_t^\varepsilon\}_t$.

Proof We define a measure on M by

$$\hat{\mu}^\varepsilon (B) \doteq E_{\lambda^\varepsilon} \left(\int_0^{\tau_1^\varepsilon} 1_B \left(X_t^\varepsilon\right) dt \right)$$

for $B \in \mathcal{B}(M)$, so that for any nonnegative measurable function $g : M \to \mathbb{R}$

$$\int_M g(x) \hat{\mu}^\varepsilon (dx) = E_{\lambda^\varepsilon} \left(\int_0^{\tau_1^\varepsilon} g\left(X_t^\varepsilon\right) dt \right).$$

According to the proof of Theorem 4.1 in [16], the measure given by $\hat{\mu}^\varepsilon (B) / \hat{\mu}^\varepsilon (M)$ is an invariant measure of $\{X_t^\varepsilon\}_t$. Since we already know that μ^ε is the unique invariant measure of $\{X_t^\varepsilon\}_t$, this means that $\mu^\varepsilon(B) = \hat{\mu}^\varepsilon (B) / \hat{\mu}^\varepsilon (M)$ for any $B \in \mathcal{B}(M)$. Therefore, for any nonnegative measurable function $g : M \to \mathbb{R}$

$$E_{\lambda^\varepsilon} \left(\int_0^{\tau_1^\varepsilon} g\left(X_t^\varepsilon\right) dt \right) = \int_M g(x) \mu^\varepsilon (dx) \cdot \hat{\mu}^\varepsilon (M) = \int_M g(x) \mu^\varepsilon (dx) \cdot E_{\lambda^\varepsilon} \tau_1^\varepsilon.$$

$\square$

Recall the definitions of $W(O_j)$ and $W(O_1 \cup O_j)$ in Definition 3.10, as well as the definition of the quasipotential $V(x, y)$ in (3.1). For any $k \in L$, we define

$$h_k \doteq \min_{j \in L \setminus \{k\}} V(O_k, O_j). \tag{4.1}$$

In addition, define

$$w \doteq W(O_1) - \min_{j \in L \setminus \{1\}} W(O_1 \cup O_j). \tag{4.2}$$

Remark 4.2 The quantity h_k is related to the time that it takes for the process to leave a neighborhood of O_k, and $W(O_1) - W(O_1 \cup O_j)$ is related to the transition time from a neighborhood of O_j to one of O_1. It turns out that our results and arguments depend on which of h_1 or w is larger. Throughout the paper, the constructions used in the case when $h_1 > w$ will be in terms of what we call a **single cycle**, and those for the case when $h_1 \leq w$ in terms of a **multicycle**.

Theorem 4.3 *Let* $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ *for some* $c > h_1 \vee w$. *Given* $\eta > 0$, *a continuous function* $f : M \to \mathbb{R}$ *and any compact set* $A \subset M$, *there exists* $\delta_0 \in (0, 1)$ *such that for any* $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left| E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon \right) dt \right) - \int_M e^{-\frac{1}{\varepsilon}f(x)} 1_A (x) \, \mu^\varepsilon (dx) \right|$$

$$\geq \inf_{x \in A} \left[f(x) + W(x) \right] - W(O_1) + c - (h_1 \vee w) - \eta,$$

where $W(x) \doteq \min_{j \in L}[W(O_j) + V(O_j, x)]$.

Remark 4.4 Since $W(x) = \min_{j \in L}[W(O_j) + V(O_j, x)]$, the lower bound appearing in Theorem 4.3 is equivalent to

$$\min_{j \in L} \left(\inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) - W(O_1) \right) + c - (h_1 \vee w) - \eta.$$

The next result gives an upper bound on the variance per unit time, or equivalently a lower bound on its rate of decay. In the design of a Markov chain Monte Carlo method, one would maximize this rate of decay to improve the method's performance.

Theorem 4.5 *Let* $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ *for some* $c > h_1 \vee w$. *Given* $\eta > 0$, *a continuous function* $f : M \to \mathbb{R}$ *and any compact set* $A \subset M$, *there exists* $\delta_0 \in (0, 1)$ *such that for any* $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(T^\varepsilon \cdot \mathrm{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon \right) dt \right) \right)$$

$$\geq \begin{cases} \min_{j \in L} \left(R_j^{(1)} \wedge R_j^{(2)} \right) - \eta, & \text{if } h_1 > w \\ \min_{j \in L} \left(R_j^{(1)} \wedge R_j^{(2)} \wedge R_j^{(3)} \right) - \eta, & \text{otherwise} \end{cases},$$

where

$$R_j^{(1)} \doteq \inf_{x \in A} \left[2f(x) + V(O_j, x) \right] + W(O_j) - W(O_1),$$

$$R_1^{(2)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_1, x) \right] - h_1,$$

and for $j \in L \setminus \{1\}$

$$R_j^{(2)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) - 2W(O_1) + W(O_1 \cup O_j),$$

$$R_j^{(3)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + 2W(O_j) - 2W(O_1) - w.$$

Remark 4.6 If one mistakenly treated a single cycle case as a multicycle case in the application of Theorem 4.5, then the result is the same since with $h_1 > w$, (4.2) implies that $R_j^{(3)} \geq R_j^{(2)}$ for any $j \in L$.

Remark 4.7 Although Theorems 4.3 and 4.5 as stated assume the starting distribution λ^ε, they can be extended to general initial distributions by using results from Sect. 10, which show that the process essentially forgets the initial distribution before leaving the neighborhood of O_1.

Remark 4.8 In this remark, we interpret the use of Theorems 4.3 and 4.5 in the context of Monte Carlo and also explain the role of the time scaling T^ε.

There is a minimum amount of time that must elapse before the process can visit all stable equilibrium points often enough that good estimation of risk-sensitive integrals is possible. As is well known, this time scales exponentially in the form of $T^\varepsilon = e^{c/\varepsilon}$, and the issue is the selection of the constant $c > 0$, which motivates the assumptions on T^ε for the two cases. However, when designing a scheme there typically will be parameters available for selection. The growth constant in T^ε will then depend on these parameters, which will then be chosen to (either directly or indirectly, depending on the criteria used) reduce the size of T^ε. For a compelling example, we refer to [10], which shows how for a system with fixed well depths a scheme known as infinite swapping can be designed so that given any $a > 0$ one can design a scheme so that an interval of length $e^{a/\varepsilon}$ suffices.

Theorem 4.3 is concerned with bias, and for T^ε as above will give a negligible contribution to the total error in comparison with the variance. Thus, it is Theorem 4.5 that determines the performance of the scheme and serves as a criteria for optimization. Of particular note is that the value of c does *not* appear in the variational problem appearing in Theorem 4.5.

Theorem 4.5 gives a lower bound on the rate of decay of variance per unit time. For applications to the design of Monte Carlo schemes as in [10], there is an a priori bound on the best possible performance, and so this lower bound (which yields an upper bound on variances) is sufficient to determine if a scheme is nearly optimal. However, for other purposes an upper bound on the decay rate could be useful, and we expect the other direction holds as well.

The proofs of Theorems 4.3 and 4.5 for single cycles and multicycles are almost identical with a few key differences. We focus on providing proofs in the single cycle case, and then point out the required modifications in the proofs for the multicycle case.

Theorem 4.9 *The bound in Theorem 4.3 can be calculated using only stable equilibrium points. Specifically,*

1. $W(x) = \min_{j \in L_s}[W(O_j) + V(O_j, x)]$
2. $W(O_j) = \min_{g \in G_s(j)} \left[\sum_{(m \to n) \in g} V(O_m, O_n) \right]$
3. $W(O_1 \cup O_j) = \min_{g \in G_s(1,j)} \left[\sum_{(m \to n) \in g} V(O_m, O_n) \right]$
4. $\min_{j \in L}(\inf_{x \in A}[f(x) + V(O_j, x)] + W(O_j)) = \min_{j \in L_s}(\inf_{x \in A}[f(x) + V(O_j, x)] + W(O_j))$.

Remark 4.10 Theorem 4.9 says that the bound appearing in Theorem 4.3 depends on the set of indices of only stable equilibrium points. This is not surprising, since in [12, Chapter 6], it has been shown that the logarithmic asymptotics of the invariant measure of a Markov process in this framework can be characterized in terms of graphs on the set of indices of just stable equilibrium points. It is natural to ask if the same property holds for the lower bound appearing in Theorem 4.5. Notice that part 4 of Theorem 4.9 implies $\min_{j \in L} R_j^{(1)} = \min_{j \in L_s} R_j^{(1)}$, so if one can prove (possibly under extra conditions, for example, by considering a double-well model as in Sect. 11) that $\min_{j \in L} R_j^{(2)} = \min_{j \in L_s} R_j^{(2)}$, then these two equations assert the property we want for the single cycle case, namely, $\min_{j \in L}(R_j^{(1)} \wedge R_j^{(2)}) = \min_{j \in L_s}(R_j^{(1)} \wedge R_j^{(2)})$. An analogous comment applies for the multicycle case.

Conjecture 4.11 *Let $T^\varepsilon = e^{\frac{1}{\varepsilon} c}$ for some $c > h_1 \vee w$. Let f be continuous and suppose that A is the closure of its interior. Then for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(T^\varepsilon \cdot \mathrm{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon \right) dt \right) \right)$$

$$\leq \begin{cases} \min_{j \in L} \left(R_j^{(1)} \wedge R_j^{(2)} \right) + \eta, & \text{if } h_1 > w \\ \min_{j \in L} \left(R_j^{(1)} \wedge R_j^{(2)} \wedge R_j^{(3)} \right) + \eta, & \text{otherwise} \end{cases}.$$

In Section 11, we outline the proof of Conjecture 4.11 for a special case.

5 Examples

Example 5.1 We first consider the situation depicted in Fig. 1. Values of $W(O_j)$ are given in the figure. If one interprets the figure as a potential with minimum zero, then the corresponding heights of the equilibrium points are given by $W(O_j) - W(O_1)$. We take $f = 0$ and A to be a small closed interval about O_5. As we will see and should be clear from the figure, this example can be analyzed using single regenerative cycles.

Recall that

$$R_j^{(1)} \doteq \inf_{x \in A} [2f(x) + V(O_j, x)] + W(O_j) - W(O_1)$$

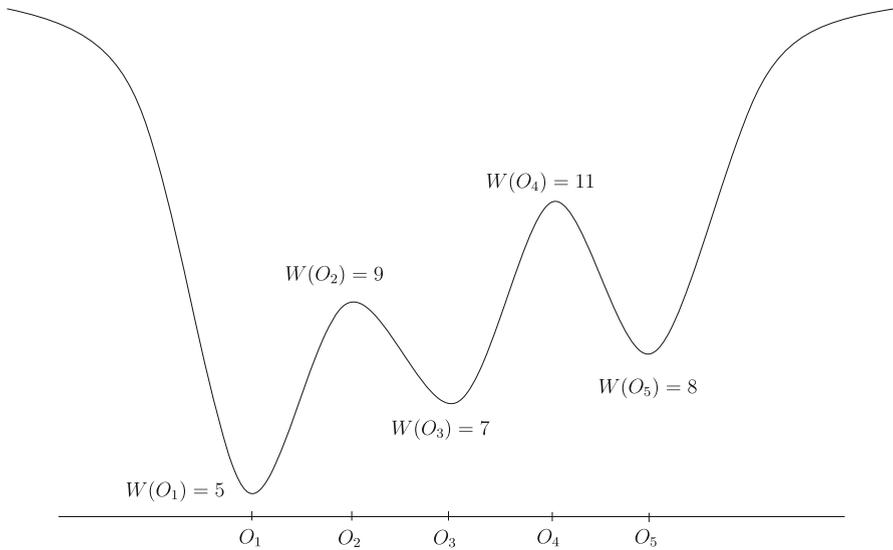$$R_1^{(2)} \doteq 2 \inf_{x \in A} [f(x) + V(O_1, x)] - h_1$$

Fig. 1 Single cycle example

and for $j > 1$

$$R_j^{(2)} \doteq 2 \inf_{x \in A} [f(x) + V(O_j, x)] + \left(W(O_j) - W(O_1)\right) - W(O_1) + W(O_1 \cup O_j)$$

If one traces through the proof of Theorem 4.5 for the case of a single cycle, then one finds that the constraining bound is given in Lemma 7.23, which is in turn based on Lemma 7.9. As we will see, in the minimization problem $\min_{j \in L}(R_j^{(1)} \wedge R_j^{(2)})$ the min on j turns out to be achieved at $j = 5$. This is of course not surprising, since A is an interval about O_5. It is then the minimum of $R_5^{(1)}$ and $R_5^{(2)}$ which determines the dominant source of the variance of the estimator.

We recall that τ_1^ε is the time for a full regenerative cycle, and that τ_1 is the time to first reach the 2δ neighborhood of an equilibrium point and then reach a δ neighborhood of a (perhaps the same) equilibrium point. The quantities that are relevant in Lemma 7.9 are

$$\sup_{y \in \partial B_\delta(O_5)} E_y \left(\int_0^{\tau_1} 1_A(X_t^\varepsilon) dt \right)^2 \quad \text{and} \quad E_x N_5$$

for $R_j^{(1)}$ and

$$\left[\sup_{y \in \partial B_\delta(O_5)} E_y \int_0^{\tau_1} 1_A(X_t^\varepsilon) dt \right]^2, \quad E_x N_5, \quad \text{and essentially} \quad \sup_{y \in \partial B_\delta(O_5)} E_y N_5$$

for $R_j^{(2)}$. Decay rates are in turn determined by (see the proof of Lemma 7.23)

$$0 \text{ and } W(O_1) - W(O_5) + h_1$$

and

$$0, \ W(O_1) - W(O_5) + h_1 \text{ and } W(O_1) - W(O_1 \cup O_5),$$

respectively. Thus, for this example it is only the term $W(O_1) - W(O_1 \cup O_5)$ that distinguishes between the two. Since this is always greater than zero and it appears in $R_j^{(2)}$ in the form $-(W(O_1) - W(O_1 \cup O_5))$, it must be the case that $R_5^{(2)} < R_5^{(1)}$.

The numerical values for the example are

$$(W(O_1 \cup O_j), j = 2, \ldots, 5) = (5, 3, 5, 2)$$
$$(V(O_j, O_5), j = 1, \ldots, 5) = (8, 4, 4, 0, 0)$$
$$(W(O_j) - W(O_1), j = 1, \ldots, 5) = (0, 4, 2, 6, 3)$$
$$(R_j^{(1)}, j = 1, \ldots, 5) = (8, 8, 6, 6, 3)$$
$$(R_j^{(2)}, j = 2, \ldots, 5) = (12, 8, 6, 0)$$

and $R_1^{(2)} = 16 - 4 = 12$, $h_1 = 4$ and $w = 5 - 2 = 3$. Since $w < h_1$, it falls into the single cycle case. We therefore find $\min_j R_j^{(1)} \wedge R_j^{(2)}$ equals to 0 and occurs with superscript 2 and at $j = 5$.

For an example where the dominant contribution to the variance is through the quantities associated with $R_j^{(1)}$, we move the set A further to the right of O_5. All other quantities are unchanged save

$$\sup_{y \in \partial B_\delta(O_5)} E_y \left(\int_0^{\tau_1} 1_A(X_t^\varepsilon) dt \right)^2 \text{ and } \left[\sup_{y \in \partial B_\delta(O_5)} E_y \int_0^{\tau_1} 1_A(X_t^\varepsilon) dt \right]^2,$$

whose decay rates are governed (for $j = 5$) by $\inf_{x \in A}[V(O_5, x)]$ and $2 \inf_{x \in A}[V(O_5, x)]$, respectively. Choosing A so that $\inf_{x \in A}[V(O_5, x)] > 3$, it is now the case that $R_5^{(1)} < R_5^{(2)}$.

Example 5.2 We consider the situation depicted in Fig. 2. In this example, we again take $f = 0$ and A to be a small closed interval about O_3. Since the well at O_5 is deeper than that at O_1, we expect that multicycles will be needed, and so recall

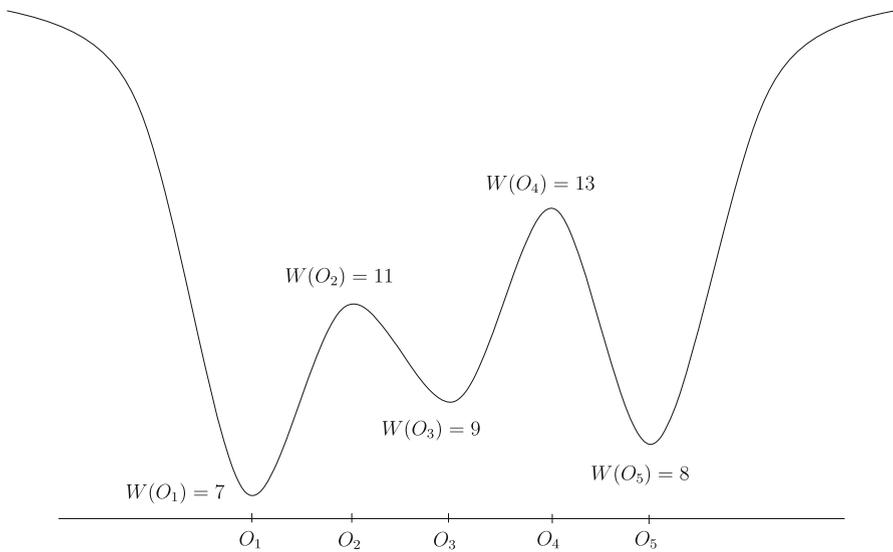$$R_j^{(3)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + 2W(O_j) - 2W(O_1) - w.$$

Fig. 2 Multicycle example

The needed values are

$$(W(O_1 \cup O_j), j = 2, \ldots, 5) = (7, 5, 7, 2)$$
$$(V(O_j, O_3), j = 1, \ldots, 5) = (4, 0, 0, 0, 5)$$
$$(W(O_j) - W(O_1), j = 1, \ldots, 5) = (0, 4, 2, 6, 1)$$
$$(R_j^{(1)}, j = 1, \ldots, 5) = (4, 4, 2, 6, 6)$$
$$(R_j^{(2)}, j = 2, \ldots, 5) = (4, 0, 6, 6)$$
$$(R_j^{(3)}, j = 1, \ldots, 5) = (3, 3, -1, 7, 7)$$

and $R_1^{(2)} = 8 - 4 = 4$, $h_1 = 4$ and $w = 7 - 2 = 5$. Since $w > h_1$ a single cycle cannot be used for the analysis of the variance, and we need to use multicycles. We find $\min_j R_j^{(1)} \wedge R_j^{(2)} \wedge R_j^{(3)}$ is equal to -1 and occurs with superscript 3 and $j = 3$.

6 Wald's Identities and Regenerative Structure

To prove Theorems 4.3 and 4.5, we will use the regenerative structure to analyze the system over the interval $[0, T^\varepsilon]$. Since the number of regenerative cycles will be random, Wald's identities will be useful.

Recall that τ_n^ε is the n-th return time to $\partial B_\delta(O_1)$ after having visited the neighborhood of a different equilibrium point, and $\lambda^\varepsilon \in \mathcal{P}(\partial B_\delta(O_1))$ is the invariant measure of the Markov process $\{X_{\tau_n^\varepsilon}^\varepsilon\}_{n \in \mathbb{N}_0}$ with state space $\partial B_\delta(O_1)$. If we let the process $\{X_t^\varepsilon\}_t$ start with λ^ε at time 0, that is, assume the distribution of X_0^ε is λ^ε, then by the strong

Markov property of $\{X_t^\varepsilon\}_t$, we find that $\{X_t^\varepsilon\}_t$ is a regenerative process and the cycles $\{\{X_{\tau_{n-1}^\varepsilon + t}^\varepsilon : 0 \le t < \tau_n^\varepsilon - \tau_{n-1}^\varepsilon\}, \tau_n^\varepsilon - \tau_{n-1}^\varepsilon\}$ are iid objects. Moreover, $\{\tau_n^\varepsilon\}_{n \in \mathbb{N}_0}$ is a sequence of renewal times under λ^ε.

6.1 Single cycle

Define the filtration $\{\mathcal{H}_n\}_{n \in \mathbb{N}}$, where $\mathcal{H}_n \doteq \mathcal{F}_{\tau_n^\varepsilon}$ and $\mathcal{F}_t \doteq \sigma(\{X_s^\varepsilon; s \le t\})$. With respect to this filtration, in the single cycle case (i.e., when $h_1 > w$), we consider the stopping times $N^\varepsilon(T) \doteq \inf\{n \in \mathbb{N} : \tau_n^\varepsilon > T\}$. Note that $N^\varepsilon(T) - 1$ is the number of complete single renewal intervals contained in $[0, T]$.

With this notation, we can bound $\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt$ from above and below by

$$\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)-1} S_n^\varepsilon \le \frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt \le \frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon, \quad (6.1)$$

where

$$S_n^\varepsilon \doteq \int_{\tau_{n-1}^\varepsilon}^{\tau_n^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt.$$

Applying Wald's first identity shows

$$E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) = \frac{1}{T^\varepsilon} E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon)) E_{\lambda^\varepsilon} S_1^\varepsilon. \quad (6.2)$$

Therefore, the logarithmic asymptotics of $E_{\lambda^\varepsilon}(\int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt / T^\varepsilon)$ are determined by those of $E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))/T^\varepsilon$ and $E_{\lambda^\varepsilon} S_1^\varepsilon$. Likewise, to understand the logarithmic asymptotics of $T^\varepsilon \cdot \mathrm{Var}_{\lambda^\varepsilon}(\int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A(X_t^\varepsilon) \, dt / T^\varepsilon)$, it is sufficient to identify the corresponding logarithmic asymptotics of $\mathrm{Var}_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))/T^\varepsilon$, $\mathrm{Var}_{\lambda^\varepsilon}(S_1^\varepsilon)$, $E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))/T^\varepsilon$ and $E_{\lambda^\varepsilon} S_1^\varepsilon$. This can be done with the help of Wald's second identity, since

$$T^\varepsilon \cdot \mathrm{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right)$$

$$\le 2T^\varepsilon \cdot E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon - \frac{1}{T^\varepsilon} N^\varepsilon(T^\varepsilon) E_{\lambda^\varepsilon} S_1^\varepsilon \right)^2$$

$$+ 2T^\varepsilon \cdot E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} N^\varepsilon(T^\varepsilon) E_{\lambda^\varepsilon} S_1^\varepsilon - \frac{1}{T^\varepsilon} E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon)) E_{\lambda^\varepsilon} S_1^\varepsilon \right)^2$$

$$= 2 \frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} \mathrm{Var}_{\lambda^\varepsilon} S_1^\varepsilon + 2 \frac{\mathrm{Var}_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} (E_{\lambda^\varepsilon} S_1^\varepsilon)^2. \quad (6.3)$$

In the next two sections, we derive bounds on $E_{\lambda^\varepsilon} S_1^\varepsilon$, $\mathrm{Var}_{\lambda^\varepsilon} (S_1^\varepsilon)$ and $E_{\lambda^\varepsilon} (N^\varepsilon (T^\varepsilon))$, $\mathrm{Var}_{\lambda^\varepsilon} (N^\varepsilon (T^\varepsilon))$, respectively.

6.2 Multicycle

Recall that in the case of a multicycle, we have $w \geq h_1$. For any $m > 0$ such that $h_1 + m > w$ and for any $\varepsilon > 0$, on the same probability space as $\{\tau_n^\varepsilon\}$, one can define a sequence of independent and geometrically distributed random variables $\{\mathbf{M}_i^\varepsilon\}_{i \in \mathbb{N}}$ with parameter $e^{-m/\varepsilon}$ that are independent of $\{\tau_n^\varepsilon\}$. We then define multicycles according to

$$\mathbf{K}_i^\varepsilon \doteq \sum_{j=1}^{i} \mathbf{M}_j^\varepsilon, \quad \hat{\tau}_i^\varepsilon \doteq \sum_{n=\mathbf{K}_{i-1}^\varepsilon+1}^{\mathbf{K}_i^\varepsilon} \tau_n^\varepsilon, \quad i \in \mathbb{N}. \tag{6.4}$$

Consider the stopping times $\hat{N}^\varepsilon (T) \doteq \inf \{n \in \mathbb{N} : \hat{\tau}_n^\varepsilon > T\}$. Note that $\hat{N}^\varepsilon (T) - 1$ is the number of complete multicycles contained in $[0, T]$. With this notation and by following the same idea as in the single cycle case, we can bound $\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A (X_t^\varepsilon) \, dt$ from above and below by

$$\frac{1}{T^\varepsilon} \sum_{n=1}^{\hat{N}^\varepsilon (T^\varepsilon)-1} \hat{S}_n^\varepsilon \leq \frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A (X_t^\varepsilon) \, dt \leq \frac{1}{T^\varepsilon} \sum_{n=1}^{\hat{N}^\varepsilon (T^\varepsilon)} \hat{S}_n^\varepsilon, \tag{6.5}$$

where

$$\hat{S}_n^\varepsilon \doteq \int_{\hat{\tau}_{n-1}^\varepsilon}^{\hat{\tau}_n^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A (X_t^\varepsilon) \, dt.$$

Therefore, by applying Wald's first and second identities, we know that the logarithmic asymptotics of $E_{\lambda^\varepsilon} (\int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A (X_t^\varepsilon) \, dt / T^\varepsilon)$ are determined by those of $E_{\lambda^\varepsilon} (\hat{N}^\varepsilon (T^\varepsilon))/T^\varepsilon$ and $E_{\lambda^\varepsilon} \hat{S}_1^\varepsilon$, and the asymptotics of $T^\varepsilon \cdot \mathrm{Var}_{\lambda^\varepsilon} (\int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A (X_t^\varepsilon) \, dt / T^\varepsilon)$ by those of $\mathrm{Var}_{\lambda^\varepsilon} (\hat{N}^\varepsilon (T^\varepsilon))/T^\varepsilon$, $\mathrm{Var}_{\lambda^\varepsilon} (\hat{S}_1^\varepsilon)$, $E_{\lambda^\varepsilon} (\hat{N}^\varepsilon (T^\varepsilon))/T^\varepsilon$ and $E_{\lambda^\varepsilon} \hat{S}_1^\varepsilon$. In particular, we have

$$E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{\hat{N}^\varepsilon (T^\varepsilon)} \hat{S}_n^\varepsilon \right) = \frac{1}{T^\varepsilon} E_{\lambda^\varepsilon} \left(\hat{N}^\varepsilon (T^\varepsilon) \right) E_{\lambda^\varepsilon} \hat{S}_1^\varepsilon \tag{6.6}$$

and

$$T^\varepsilon \cdot \mathrm{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{\hat{N}^\varepsilon (T^\varepsilon)} \hat{S}_n^\varepsilon \right)$$

$$\leq 2 \frac{E_{\lambda^\varepsilon} (\hat{N}^\varepsilon (T^\varepsilon))}{T^\varepsilon} \mathrm{Var}_{\lambda^\varepsilon} \hat{S}_1^\varepsilon + 2 \frac{\mathrm{Var}_{\lambda^\varepsilon} (\hat{N}^\varepsilon (T^\varepsilon))}{T^\varepsilon} \left(E_{\lambda^\varepsilon} \hat{S}_1^\varepsilon \right)^2. \tag{6.7}$$

In the next two sections, we derive bounds on $E_{\lambda^\varepsilon}\hat{S}_1^\varepsilon$, $\mathrm{Var}_{\lambda^\varepsilon}(\hat{S}_1^\varepsilon)$ and $E_{\lambda^\varepsilon}(\hat{N}^\varepsilon(T^\varepsilon))$, $\mathrm{Var}_{\lambda^\varepsilon}(\hat{N}^\varepsilon(T^\varepsilon))$, respectively.

Remark 6.1 It should be kept in mind that $\hat{\tau}_n^\varepsilon$, $\hat{N}^\varepsilon(T^\varepsilon)$ and $\hat{S}_n^\varepsilon$ all depend on m, although this dependence is not explicit in the notation.

Remark 6.2 In general, for any quantity in the single cycle case, we use analogous notation with a "hat" on it to represent the corresponding quantity in the multicycle version. For instance, we use τ_n^ε for a single regenerative cycle, and $\hat{\tau}_n^\varepsilon$ for a multi-regenerative cycle.

7 Asymptotics of Moments of S_1^ε and $\hat{S}_1^\varepsilon$

In this section, we will first introduce the elementary theory of an irreducible finite state Markov chain $\{Z_n\}_{n \in \mathbb{N}_0}$ with state space L, and then state and prove bounds for the asymptotics of moments of S_1^ε and $\hat{S}_1^\varepsilon$.

For the asymptotic analysis, the following useful facts will be used repeatedly.

Lemma 7.1 *For any nonnegative sequences $\{a_\varepsilon\}_{\varepsilon>0}$ and $\{b_\varepsilon\}_{\varepsilon>0}$, we have*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log(a_\varepsilon b_\varepsilon) \geq \liminf_{\varepsilon \to 0} -\varepsilon \log a_\varepsilon + \liminf_{\varepsilon \to 0} -\varepsilon \log b_\varepsilon, \qquad (7.1)$$

$$\limsup_{\varepsilon \to 0} -\varepsilon \log(a_\varepsilon + b_\varepsilon) \leq \min\left\{\limsup_{\varepsilon \to 0} -\varepsilon \log a_\varepsilon, \limsup_{\varepsilon \to 0} -\varepsilon \log b_\varepsilon\right\},$$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log(a_\varepsilon + b_\varepsilon) = \min\left\{\liminf_{\varepsilon \to 0} -\varepsilon \log a_\varepsilon, \liminf_{\varepsilon \to 0} -\varepsilon \log b_\varepsilon\right\}. \quad (7.2)$$

7.1 Markov Chains and Graph Theory

In this subsection, we state some elementary theory for finite state Markov chains taken from [1, Chapter 2]. For a finite state Markov chain, the invariant measure, the mean exit time, etc., can be expressed explicitly as the ratio of certain determinants, i.e., sums of products consisting of transition probabilities, and these sums only contain terms with a plus sign. Which products should appear in the various sums can be described conveniently by means of graphs on the set of states of the chain. This method of linking graphs and quantities associated with a finite state Markov chain was introduced by Freidlin and Wentzell in [12, Chapter 6].

Consider an irreducible finite state Markov chain $\{Z_n\}_{n \in \mathbb{N}_0}$ with state space L. For any $i, j \in L$, let p_{ij} be the one-step transition probability of $\{Z_n\}_n$ from state i to state j. Write $P_i(\cdot)$ and $E_i(\cdot)$ for probabilities and expectations of the chain started at state i at time 0. Recall the notation $\pi(g) \doteq \prod_{(i \to j) \in g} p_{ij}$.

Lemma 7.2 *The unique invariant measure of $\{Z_n\}_{n \in \mathbb{N}}$ can be expressed*

$$\lambda_i = \frac{\sum_{g \in G(i)} \pi(g)}{\sum_{j \in L}\left(\sum_{g \in G(j)} \pi(g)\right)}.$$

Proof See Lemma 3.1, Chapter 6 in [12]. $\qquad\square$

To analyze the empirical measure, we will need additional results, including representations for the number of visits to a state during a regenerative cycle. Write

$$T_i \doteq \inf\{n \geq 0 : Z_n = i\}$$

for the *first hitting time* of state i, and write

$$T_i^+ \doteq \inf\{n \geq 1 : Z_n = i\}.$$

Observe that $T_i^+ = T_i$ unless $Z_0 = i$, in which case we call T_i^+ the *first return time* to state i.

Let $\hat{N} \doteq \inf\{n \in \mathbb{N}_0 : Z_n \in L \setminus \{1\}\}$ and $N \doteq \inf\{n \in \mathbb{N} : Z_n = 1, n \geq \hat{N}\}$. $\hat{N}$ is the first time of visiting a state other than state 1, and N is the first time of visiting state 1 after $\hat{N}$. For any $j \in L$, let N_j be the number of visits (including time 0) of state j before N, i.e., $N_j = |\{n \in \mathbb{N}_0 : n < N \text{ and } Z_n = j\}|$. We would like to understand $E_1 N_j$ and $E_j N_j$ for any $j \in L$. These quantities will appear later on in Subsection 7.2. The next lemma shows how they can be related to the invariant measure of $\{Z_n\}_n$.

Lemma 7.3 *1. For any $j \in L \setminus \{1\}$*

$$E_j N_j = \frac{\sum_{g \in G(1,j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)} \text{ and } E_j N_j = \lambda_j \left(E_j T_1 + E_1 T_j\right).$$

2. For any $i, j \in L$, $j \neq i$

$$P_i\left(T_j < T_i^+\right) = \frac{1}{\lambda_i \left(E_j T_i + E_i T_j\right)}.$$

3. For any $j \in L$

$$E_1 N_j = \frac{1}{1 - p_{11}} \frac{\lambda_j}{\lambda_1}.$$

Proof See Lemma 3.4 in [12, Chapter 6] for the first assertion of part 1 and see Lemma 2.7 in [1, Chapter 2] for the second assertion of part 1. For part 2, see Corollary 2.8 in [1, Chapter 2]. For part 3, since $E_1 N_j = \sum_{\ell=1}^{\infty} P_1\left(N_j \geq \ell\right)$, we need to understand $P_1\left(N_j \geq \ell\right)$, which means we need to know how to count all the ways to get $N_j \geq \ell$ before returning to state 1.

We first have to move away from state 1, so the types of sequences are of the form

$$\underbrace{1, 1, \ldots, 1}_{i \text{ times}}, k_1, k_2, \ldots, k_q, 1$$

for some $i, q \in \mathbb{N}$ and $k_1 \neq 1, \cdots, k_q \neq 1$. When $j = 1$, we do not care about $k_1, k_2, \ldots, k_q$, and therefore

$$P_1(N_1 \geq i) = p_{11}^{i-1} \text{ and } E_1 N_1 = \sum_{i=1}^{\infty} P_1(N_1 \geq i) = \frac{1}{1 - p_{11}}.$$

For $j \in L \setminus \{1\}$, the event $\{N_j \geq \ell\}$ requires that within $k_1, k_2, \ldots, k_q$, we

1. first visit state j before returning to state 1, which has corresponding probability $P_1(T_j < T_1^+)$,
2. then start from state j and again visit state j before returning to state 1, which has corresponding probability $P_j(T_j^+ < T_1)$.

Step 2 needs to happen at least $\ell - 1$ times in a row, and after that we do not care. Thus,

$$P_1(N_j \geq \ell) = \sum_{i=1}^{\infty} (p_{11})^{i-1} P_1(T_j < T_1^+) (P_j(T_j^+ < T_1))^{\ell-1}$$

$$= \frac{1}{1 - p_{11}} P_1(T_j < T_1^+) (P_j(T_j^+ < T_1))^{\ell-1}$$

and

$$\sum_{\ell=1}^{\infty} P_1(N_j \geq \ell) = \frac{1}{1 - p_{11}} \frac{P_1(T_j < T_1^+)}{P_j(T_1 < T_j^+)} = \frac{1}{1 - p_{11}} \frac{\lambda_j (E_1 T_j + E_j T_1)}{\lambda_1 (E_1 T_j + E_j T_1)}$$

$$= \frac{1}{1 - p_{11}} \frac{\lambda_j}{\lambda_1}.$$

The third equality comes from part 2.   □

To apply the preceding results using the machinery developed by Freidlin and Wentzell, one must have analogues that allow for small perturbations of the transition probabilities due to the fact that initial conditions are to be taken in small neighborhoods of the equilibrium points. The addition of a tilde will be used to identify the corresponding objects, such as hitting and return times. Take as given a Markov chain $\{\tilde{Z}_n\}_{n \in \mathbb{N}_0}$ on a state space $\mathcal{X} = \cup_{i \in L} \mathcal{X}_i$, with $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset$ ($i \neq j$), and assume there is $a \in [1, \infty)$ such that for any $i, j \in L$ and $j \neq i$, the transition probability of the chain from $x \in \mathcal{X}_i$ to $\mathcal{X}_j$ (denoted by $p(x, \mathcal{X}_j)$) satisfies the inequalities

$$a^{-1} p_{ij} \leq p(x, \mathcal{X}_j) \leq a p_{ij} \tag{7.3}$$

for any $x \in \mathcal{X}_i$. Write $P_x(\cdot)$ and $E_x(\cdot)$ for probabilities and expectations of the chain started at $x \in \mathcal{X}$ at time 0. Write

$$\tilde{T}_i \doteq \inf\{n \geq 0 : \tilde{Z}_n \in \mathcal{X}_i\}$$

for the *first hitting time* of $\mathcal{X}_i$, and write

$$\tilde{T}_i^+ \doteq \inf\{n \geq 1 : \tilde{Z}_n \in \mathcal{X}_i\}.$$

Observe that $\tilde{T}_i^+ = \tilde{T}_i$ unless $\tilde{Z}_0 \in \mathcal{X}_i$, in which case we call $\tilde{T}_i^+$ the *first return time* to $\mathcal{X}_i$. Recall that $l = |L|$.

Remark 7.4 Observe that given $j \in L$ and for any $x \in \mathcal{X}_j$, $1 - p(x, \mathcal{X}_j) = \sum_{k \in L \setminus \{j\}} p(x, \mathcal{X}_k)$. Therefore, we can apply (7.3) to obtain

$$a^{-1} \sum_{k \in L \setminus \{j\}} p_{jk} \leq 1 - p(x, \mathcal{X}_j) \leq a \sum_{k \in L \setminus \{j\}} p_{jk}.$$

Lemma 7.5 *1. Consider distinct $i, j, k \in L$. Then, for $x \in \mathcal{X}_k$,*

$$a^{-4^{l-2}} P_k(T_j < T_i) \leq P_x(\tilde{T}_j < \tilde{T}_i) \leq a^{4^{l-2}} P_k(T_j < T_i).$$

2. For any $i \in L$, $j \in L \setminus \{i\}$ and $x \in \mathcal{X}_i$,

$$a^{-4^{l-2}-1} P_i(T_j < T_i^+) \leq P_x(\tilde{T}_j < \tilde{T}_i^+) \leq a^{4^{l-2}+1} P_i(T_j < T_i^+).$$

Proof For part 1, see Lemma 3.3 in [12, Chapter 6]. We only need to prove part 2. Note that by a first step analysis on $\{\tilde{Z}_n\}_{n \in \mathbb{N}_0}$, for any $i \in L$, $j \in L \setminus \{i\}$ and $x \in \mathcal{X}_i$,

$$
\begin{aligned}
P_x(\tilde{T}_j < \tilde{T}_i^+) &= p(x, \mathcal{X}_j) + \sum_{k \in L \setminus \{i,j\}} \int_{\mathcal{X}_k} P_y(\tilde{T}_j < \tilde{T}_i) p(x, dy) \\
&\leq a p_{ij} + \sum_{k \in L \setminus \{i,j\}} \left(a^{4^{l-2}} P_k(T_j < T_i)\right) (a p_{ik}) \\
&\leq a^{4^{l-2}+1} \left(p_{ij} + \sum_{k \in L \setminus \{i,j\}} P_k(T_j < T_i) p_{ik}\right) \\
&= a^{4^{l-2}+1} P_i(T_j < T_i^+).
\end{aligned}
$$

The first inequality comes from the use of (7.3) and part 1; the last equality holds since we can do a first step analysis on $\{Z_n\}_n$. Similarly, we can show the lower bound. □

Let $\check{N} \doteq \inf\{n \in \mathbb{N}_0 : \tilde{Z}_n \in \cup_{j \in L \setminus \{1\}} \mathcal{X}_j\}$ and $\tilde{N} \doteq \inf\{n \in \mathbb{N} : Z_n \in \mathcal{X}_1, n \geq \check{N}\}$. For any $j \in L$, let $\tilde{N}_j$ be the number of visits (including time 0) of state $\mathcal{X}_j$ before $\tilde{N}$, i.e., $\tilde{N}_j = |\{n \in \mathbb{N}_0 : n < \tilde{N} \text{ and } Z_n \in \mathcal{X}_j\}|$. We would like to understand $E_x \tilde{N}_j$ for any $j \in L$ and $x \in \mathcal{X}_1$ or $\mathcal{X}_j$.

Lemma 7.6 *For any $j \in L$ and $x \in \mathcal{X}_1$*

$$E_x \tilde{N}_j \leq \frac{a^{4^{l-1}}}{\sum_{\ell \in L \setminus \{1\}} p_{1\ell}} \frac{\sum_{g \in G(j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)}.$$

Moreover, for any $j \in L \setminus \{1\}$

$$\sum_{\ell=1}^{\infty} \sup_{x \in \mathcal{X}_j} P_x\left(\tilde{N}_j \geq \ell\right) \leq a^{4^{l-1}} \frac{\sum_{g \in G(1,j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)}$$

and

$$\sum_{\ell=1}^{\infty} \sup_{x \in \mathcal{X}_1} P_x\left(\tilde{N}_1 \geq \ell\right) \leq \frac{a}{\sum_{\ell \in L \setminus \{1\}} p_{1\ell}}.$$

Proof For any $x \in \mathcal{X}_1$, note that for any $\ell \in \mathbb{N}$, by a conditioning argument as in the proof of Lemma 7.3 (3), we find that for $j \in L \setminus \{1\}$

$$P_x(\tilde{N}_j \geq \ell) \leq \frac{\sup_{y \in \mathcal{X}_1} P_y(\tilde{T}_j < \tilde{T}_1^+)}{1 - \sup_{y \in \mathcal{X}_1} p(y, \mathcal{X}_1)} \left(\sup_{y \in \mathcal{X}_j} P_y(\tilde{T}_j^+ < \tilde{T}_1)\right)^{\ell-1}$$

and

$$P_x(\tilde{N}_1 \geq \ell) \leq \left(\sup_{y \in \mathcal{X}_1} p(y, \mathcal{X}_1)\right)^{\ell-1}.$$

Thus, for any $x \in \mathcal{X}_1$ and for $j \in L \setminus \{1\}$

$$\begin{aligned}
E_x \tilde{N}_j &= \sum_{\ell=1}^{\infty} P_x(\tilde{N}_j \geq \ell) \leq \frac{\sup_{y \in \mathcal{X}_1} P_y(\tilde{T}_j < \tilde{T}_1^+)}{1 - \sup_{y \in \mathcal{X}_1} p(y, \mathcal{X}_1)} \cdot \frac{1}{1 - \sup_{y \in \mathcal{X}_j} P_y(\tilde{T}_j^+ < \tilde{T}_1)} \\
&= \frac{\sup_{y \in \mathcal{X}_1} P_y(\tilde{T}_j < \tilde{T}_1^+)}{\left(\inf_{y \in \mathcal{X}_j}(1 - p(y, \mathcal{X}_1))\right)\left(\inf_{y \in \mathcal{X}_j} P_y(\tilde{T}_1 < \tilde{T}_j^+)\right)} \\
&\leq a^{4^{l-1}} \frac{P_1(T_j < T_1^+)}{(\sum_{\ell \in L \setminus \{1\}} p_{1\ell}) P_j(T_1 < T_j^+)} \\
&= \frac{a^{4^{l-1}}}{\sum_{\ell \in L \setminus \{1\}} p_{1\ell}} \frac{\lambda_j}{\lambda_1} = \frac{a^{4^{l-1}}}{\sum_{\ell \in L \setminus \{1\}} p_{1\ell}} \frac{\sum_{g \in G(j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)}.
\end{aligned}$$

The second inequality is from Remark 7.4 and Lemma 7.5 (2); the third equality comes from Lemma 7.3 (2); the last equality holds due to Lemma 7.2. Also,

$$\begin{aligned}
E_x \tilde{N}_1 &= \sum_{\ell=1}^{\infty} P_x(\tilde{N}_1 \geq \ell) \leq \frac{1}{1 - \sup_{y \in \mathcal{X}_1} p(y, \mathcal{X}_1)} = \frac{1}{\inf_{y \in \mathcal{X}_1}(1 - p(y, \mathcal{X}_1))} \\
&\leq \frac{a}{\sum_{\ell \in L \setminus \{1\}} p_{1\ell}}.
\end{aligned}$$

The last inequality is from Remark 7.4. This completes the proof of part 1.

Turning to part 2, since for any $\ell \in \mathbb{N}$

$$\sup_{x \in \mathcal{X}_1} P_x(\tilde{N}_1 \geq \ell) \leq \left(\sup_{y \in \mathcal{X}_1} p(y, \mathcal{X}_1)\right)^{\ell-1},$$

we have

$$\sum_{\ell=1}^{\infty} \sup_{x \in \mathcal{X}_1} P_x(\tilde{N}_1 \geq \ell) \leq \frac{1}{1 - \sup_{y \in \mathcal{X}_1} p(y, \mathcal{X}_1)} \leq \frac{a}{\sum_{\ell \in L \setminus \{1\}} p_{1\ell}}.$$

Furthermore, we use the conditioning argument again to find that for any $j \in L \setminus \{1\}$ and $\ell \in \mathbb{N}$

$$\sup_{x \in \mathcal{X}_j} P_x(\tilde{N}_j \geq \ell) \leq (\sup_{y \in \mathcal{X}_j} P_y(\tilde{T}_j^+ < \tilde{T}_1))^{\ell-1}.$$

This implies that

$$\sum_{\ell=1}^{\infty} \sup_{x \in \mathcal{X}_j} P_x(\tilde{N}_j \geq \ell)$$

$$\leq \sum_{\ell=1}^{\infty} (\sup_{y \in \mathcal{X}_j} P_y(\tilde{T}_j^+ < \tilde{T}_1))^{\ell-1} = \frac{1}{1 - \sup_{y \in \mathcal{X}_j} P_y(\tilde{T}_j^+ < \tilde{T}_1)}$$

$$= \frac{1}{\inf_{y \in \mathcal{X}_j} P_y(\tilde{T}_1 < \tilde{T}_j^+)} \leq a^{4^{l-1}} \frac{1}{P_j(T_1 < T_j^+)}$$

$$= a^{4^{l-1}} \lambda_j (E_1 T_j + E_j T_1) = a^{4^{l-1}} \frac{\sum_{g \in G(1,j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)}.$$

We use Lemma 7.5 (2) to obtain the second inequality and Lemma 7.3, parts (2) and (1), for the penultimate and last equalities. $\qquad\square$

7.2 Asymptotics of Moments of S_1^{ε}

Recall that $\{X^{\varepsilon}\}_{\varepsilon \in (0,\infty)} \subset C([0,\infty) : M)$ is a sequence of stochastic processes satisfying Conditions 3.1, 3.7 and 3.13. Moreover, recall that S_1^{ε} is defined by

$$S_1^{\varepsilon} \doteq \int_0^{\tau_1^{\varepsilon}} e^{-\frac{1}{\varepsilon} f(X_t^{\varepsilon})} 1_A(X_t^{\varepsilon}) \, dt. \tag{7.4}$$

As mentioned in Section 6, we are interested in the logarithmic asymptotics of $E_{\lambda^{\varepsilon}} S_1^{\varepsilon}$ and $E_{\lambda^{\varepsilon}} (S_1^{\varepsilon})^2$. To find these asymptotics, the main tool we will use is Freidlin–Wentzell theory [12]. In fact, we will generalize the results of Freidlin–Wentzell to the following: For any given continuous function $f : M \to \mathbb{R}$ and any compact set $A \subset M$, we will provide lower bounds for

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z \left(\int_0^{\tau_1^{\varepsilon}} e^{-\frac{1}{\varepsilon} f(X_s^{\varepsilon})} 1_A(X_s^{\varepsilon}) \, ds \right) \right) \tag{7.5}$$

and

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z \left(\int_0^{\tau_1^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_s^\varepsilon)} 1_A \left(X_s^\varepsilon \right) ds \right)^2 \right). \tag{7.6}$$

As will be shown, these two bounds can be expressed in terms of the quasipotentials $V(O_i, O_j)$ and $V(O_i, x)$.

Remark 7.7 In the Freidlin–Wentzell theory as presented in [12], they only consider bounds for

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z \tau_1^\varepsilon \right).$$

Thus, their result is a special case of (7.5) with $f \equiv 0$ and $A = M$. Moreover, we generalize their result further by considering the logarithmic asymptotics of higher moment quantities such as (7.6).

Before proceeding, we recall that $L = \{1, \ldots, l\}$ and for any $\delta > 0$, we define $\tau_0 \doteq 0$,

$$\sigma_n \doteq \inf\{t > \tau_n : X_t^\varepsilon \in \bigcup_{j \in L} \partial B_{2\delta}(O_j)\} \text{ and } \tau_n \doteq \inf\{t > \sigma_{n-1} : X_t^\varepsilon \in \bigcup_{j \in L} \partial B_\delta(O_j)\}.$$

Moreover, $\tau_0^\varepsilon \doteq 0$,

$$\sigma_n^\varepsilon \doteq \inf\{t > \tau_n^\varepsilon : X_t^\varepsilon \in \bigcup_{j \in L \setminus \{1\}} \partial B_\delta(O_j)\} \text{ and } \tau_n^\varepsilon \doteq \inf \left\{ t > \sigma_{n-1}^\varepsilon : X_t^\varepsilon \in \partial B_\delta(O_1) \right\}.$$

In addition, $\{Z_n\}_{n \in \mathbb{N}_0} \doteq \{X_{\tau_n}^\varepsilon\}_{n \in \mathbb{N}_0}$ is a Markov chain on $\bigcup_{j \in L} \partial B_\delta(O_j)$ and $\{Z_n^\varepsilon\}_{n \in \mathbb{N}_0} \doteq \{X_{\tau_n^\varepsilon}^\varepsilon\}_{n \in \mathbb{N}_0}$ is a Markov chain on $\partial B_\delta(O_1)$. It is essential to keep the distinction clear: when there is an ε superscript the chain makes transitions between neighborhoods of distinct equilibria, while if absent such transitions are possible, but for stable equilibria there will be many more transitions between the δ and 2δ neighborhoods.

Following the notation of Subsect. 7.1, let $\hat{N} \doteq \inf\{n \in \mathbb{N}_0 : Z_n \in \bigcup_{j \in L \setminus \{1\}} \partial B_\delta(O_j)\}$, $N \doteq \inf\{n \geq \hat{N} : Z_n \in \partial B_\delta(O_1)\}$, and recall $\mathcal{F}_t \doteq \sigma(\{X_s^\varepsilon; s \leq t\})$. Then, since $\{\tau_n\}_{n \in \mathbb{N}_0}$ are stopping times with respect to the filtration $\{\mathcal{F}_t\}_{t \geq 0}$, $\mathcal{F}_{\tau_n}$ are well-defined for any $n \in \mathbb{N}_0$ and we use $\mathcal{G}_n$ to denote $\mathcal{F}_{\tau_n}$. One can prove that $\hat{N}$ and N are stopping times with respect to $\{\mathcal{G}_n\}_{n \in \mathbb{N}}$. For any $j \in L$, let N_j be the number of visits of $\{Z_n\}_{n \in \mathbb{N}_0}$ to $\partial B_\delta(O_j)$ (including time 0) before N.

The proofs of the following two lemmas are given in Appendix.

Lemma 7.8 *Given $\delta > 0$ sufficiently small, for any $x \in \partial B_\delta(O_1)$ and any nonnegative measurable function $g : M \to \mathbb{R}$,*

$$E_x \left(\int_0^{\tau_1^\varepsilon} g \left(X_s^\varepsilon \right) ds \right) \leq \sum_{j \in L} \left[\sup_{y \in \partial B_\delta(O_j)} E_y \left(\int_0^{\tau_1} g \left(X_s^\varepsilon \right) ds \right) \right] \cdot E_x N_j.$$

Lemma 7.9 *Given $\delta > 0$ sufficiently small, for any $x \in \partial B_\delta(O_1)$ and any nonnegative measurable function $g : M \to \mathbb{R}$,*

$$
E_x \left(\int_0^{\tau_1^\varepsilon} g\left(X_s^\varepsilon\right) ds \right)^2 \le l \sum_{j \in L} \left[\sup_{y \in \partial B_\delta(O_j)} E_y \left(\int_0^{\tau_1} g\left(X_s^\varepsilon\right) ds \right)^2 \right] \cdot E_x N_j
$$

$$
+ 2l \sum_{j \in L} \left[\sup_{y \in \partial B_\delta(O_j)} E_y \left(\int_0^{\tau_1} g\left(X_s^\varepsilon\right) ds \right) \right]^2 \cdot E_x N_j
$$

$$
\cdot \sum_{k=1}^\infty \sup_{y \in \partial B_\delta(O_j)} P_y \left(k \le N_j \right). \tag{7.7}
$$

Although as noted the proofs are given in Appendix, these results follow in a straightforward way by decomposing the excursion away from O_1 during $[0, \tau_1^\varepsilon]$, which only stops when returning to a neighborhood of O_1, into excursions between any pair of equilibrium points, counting the number of such excursions that start near a particular equilibrium point, and using the strong Markov property.

Remark 7.10 Following an analogous argument as in the proof of Lemmas 7.8 and 7.9, we can prove the following: Given $\delta > 0$ sufficiently small, for any $x \in \partial B_\delta(O_1)$ and any nonnegative measurable function $g : M \to \mathbb{R}$,

$$
E_x \left(\int_{\sigma_0^\varepsilon}^{\tau_1^\varepsilon} g\left(X_s^\varepsilon\right) ds \right) \le \sum_{j \in L \setminus \{1\}} \left[\sup_{y \in \partial B_\delta(O_j)} E_y \left(\int_0^{\tau_1} g\left(X_s^\varepsilon\right) ds \right) \right] \cdot E_x N_j
$$

and

$$
E_x \left(\int_{\sigma_0^\varepsilon}^{\tau_1^\varepsilon} g\left(X_s^\varepsilon\right) ds \right)^2 \le l \sum_{j \in L \setminus \{1\}} \left[\sup_{y \in \partial B_\delta(O_j)} E_y \left(\int_0^{\tau_1} g\left(X_s^\varepsilon\right) ds \right)^2 \right] \cdot E_x N_j
$$

$$
+ 2l \sum_{j \in L \setminus \{1\}} \left[\sup_{y \in \partial B_\delta(O_j)} E_y \left(\int_0^{\tau_1} g\left(X_s^\varepsilon\right) ds \right) \right]^2 \cdot E_x N_j
$$

$$
\cdot \sum_{\ell=1}^\infty \sup_{y \in \partial B_\delta(O_j)} P_y \left(k \le N_j \right).
$$

The main difference is that if the integration starts from σ_0^ε (the first visiting time of $\bigcup_{j \in L \setminus \{1\}} \partial B_\delta(O_j)$), then any summation appearing in the upper bounds should sum over all indices in $L \setminus \{1\}$ instead of L.

Owing to its frequent appearance but with varying arguments, we introduce the notation

$$I^\varepsilon(t_1, t_2; f, A) \doteq \int_{t_1}^{t_2} e^{-\frac{1}{\varepsilon} f(X_s^\varepsilon)} 1_A(X_s^\varepsilon) ds, \tag{7.8}$$

and write $I^\varepsilon(t; f, A)$ if $t_1 = 0$ and $t_2 = t$ so that, e.g., $S_1^\varepsilon = I^\varepsilon(\tau_1^\varepsilon; f, A)$.

Corollary 7.11 *Given any measurable set $A \subset M$, a measurable function $f : M \to \mathbb{R}$, $j \in L$ and $\delta > 0$, we have*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z I^\varepsilon(\tau_1^\varepsilon; f, A) \right)$$

$$\geq \min_{j \in L} \left\{ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) \right\},$$

and

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z I^\varepsilon(\tau_1; f, A)^2 \right) \geq \min_{j \in L} \left(\hat{R}_j^{(1)} \wedge \hat{R}_j^{(2)} \right),$$

where

$$\hat{R}_j^{(1)} \doteq \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A)^2 \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right)$$

and

$$\hat{R}_j^{(2)} \doteq 2 \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right)$$

$$+ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^{\infty} \sup_{z \in \partial B_\delta(O_j)} P_z \left(\ell \leq N_j \right) \right).$$

Proof For the first part, applying Lemma 7.8 with $g(x) = e^{-\frac{1}{\varepsilon} f(x)} 1_A(x)$ and using (7.1) and (7.2) completes the proof. For the second part, using Lemma 7.9 with $g(x) = e^{-\frac{1}{\varepsilon} f(x)} 1_A(x)$ and using (7.1) and (7.2) again completes the proof. □

Remark 7.12 Owing to Remark 7.10, we can modify the proof of Corollary 7.11 and show that given any set $A \subset M$, a measurable function $f : M \to \mathbb{R}$, $j \in L$ and $\delta > 0$,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z I^\varepsilon(\sigma_0^\varepsilon, \tau_1^\varepsilon; f, A) \right)$$

$$\geq \min_{j \in L \setminus \{1\}} \left\{ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) \right\}.$$

Moreover,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z I^\varepsilon(\sigma_0^\varepsilon, \tau_1^\varepsilon; f, A)^2 \right) \geq \min_{j \in L \setminus \{1\}} \left(\hat{R}_j^{(1)} \wedge \hat{R}_j^{(2)} \right),$$

where the definitions of $\hat{R}_j^{(1)}$ and $\hat{R}_j^{(2)}$ can be found in Corollary 7.11.

We next consider lower bounds on

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) \quad \text{and}$$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A)^2 \right)$$

for $j \in L$. We state some useful results before studying the lower bounds. Recall also that τ_1 is the time to reach the δ-neighborhood of any of the equilibrium points after leaving the 2δ-neighborhood of one of the equilibrium points.

Lemma 7.13 *For any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that for all $\delta \in (0, \delta_0)$ and $\varepsilon \in (0, \varepsilon_0)$*

$$\sup_{x \in M} E_x \tau_1 \leq e^{\frac{\eta}{\varepsilon}} \text{ and } \sup_{x \in M} E_x (\tau_1)^2 \leq e^{\frac{\eta}{\varepsilon}}.$$

Proof If x is not in $\cup_{j \in L} B_{2\delta}(O_j)$, then a uniform (in x and small ε) upper bound on these expected values follows from the corollary to [12, Lemma 1.9, Chapter 6].

If $x \in \cup_{j \in L} B_{2\delta}(O_j)$, then we must wait till the process reaches $\cup_{j \in L} \partial B_{2\delta}(O_j)$, after which we can use the uniform bound (and the strong Markov property). Since there exists $\delta > 0$ such the lower bound $P_x(\inf\{t \geq 0 : X_t^\varepsilon \in \cup_{j \in L} \partial B_{2\delta}(O_j)\} \leq 1) \geq e^{-\eta/2\varepsilon}$ is valid for all $x \in \cup_{j \in L} B_{2\delta}(O_j)$ and small $\varepsilon > 0$, upper bounds of the desired form follow from the Markov property and standard calculations. $\square$

For any compact set $A \subset M$, we use ϑ_A to denote the first hitting time

$$\vartheta_A \doteq \inf \left\{ t \geq 0 : X_t^\varepsilon \in A \right\}.$$

Note that ϑ_A is a stopping time with respect to filtration $\{\mathcal{F}_t\}_{t \geq 0}$. The following result is relatively straightforward given the just discussed bound on the distribution of τ_1, and follows by partitioning according to $\tau_1 \geq T$ and $\tau_1 < T$ for large but fixed T.

Lemma 7.14 *For any compact set $A \subset M$, $j \in L$ and any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that for all $\varepsilon \in (0, \varepsilon_0)$ and $\delta \in (0, \delta_0)$*

$$\sup_{z \in \partial B_\delta(O_j)} P_z (\vartheta_A \leq \tau_1) \leq e^{-\frac{1}{\varepsilon}(\inf_{x \in A}[V(O_j, x)] - \eta)}.$$

Lemma 7.15 *Given a compact set $A \subset M$, any $j \in L$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z \left[\int_0^{\tau_1} 1_A \left(X_s^\varepsilon \right) ds \right] \right) \geq \inf_{x \in A} V \left(O_j, x \right) - \eta$$

and

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z \left(\int_0^{\tau_1} 1_A \left(X_s^\varepsilon \right) ds \right)^2 \right) \geq \inf_{x \in A} V \left(O_j, x \right) - \eta.$$

Proof The idea of this proof follows from the proof of Theorem 4.3 in [12, Chapter 4]. Since $I^\varepsilon(\tau_1; 0, A) = \int_0^{\tau_1} 1_A \left(X_s^\varepsilon \right) ds$, for any $x \in \partial B_\delta(O_j)$,

$$E_x I^\varepsilon (\tau_1; 0, A)$$
$$= E_x \left[I^\varepsilon(\tau_1; 0, A) 1_{\{\vartheta_A \leq \tau_1\}} \right] = E_x \left[E_x \left[I^\varepsilon(\tau_1; 0, A) \middle| \mathcal{F}_{\vartheta_A} \right] 1_{\{\vartheta_A \leq \tau_1\}} \right]$$
$$= E_x \left[(E_{X_{\vartheta_A}^\varepsilon} I^\varepsilon(\tau_1; 0, A)) 1_{\{\vartheta_A \leq \tau_1\}} \right] \leq \sup_{y \in \partial A} E_y \tau_1 \cdot \sup_{z \in \partial B_\delta(O_j)} P_z \left(\vartheta_A \leq \tau_1 \right).$$

The inequality is due to $E_{X_{\vartheta_A}^\varepsilon} I^\varepsilon(\tau_1; 0, A) \leq E_{X_{\vartheta_A}^\varepsilon} \tau_1 \leq \sup_{y \in \partial A} E_y \tau_1$. We then apply Lemmas 7.13 and 7.14 to find that for the given $\eta > 0$, there exists $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that for all $\varepsilon \in (0, \varepsilon_0)$ and $\delta \in (0, \delta_0)$,

$$E_x I^\varepsilon(\tau_1; 0, A) \leq \sup_{y \in \partial A} E_y \tau_1 \cdot \sup_{z \in \partial B_\delta(O_j)} P_z \left(\vartheta_A \leq \tau_1 \right) \leq e^{\frac{\eta/2}{\varepsilon}} e^{-\frac{1}{\varepsilon} \left(\inf_{y \in A} V(O_j, y) - \eta/2 \right)}.$$

Thus,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; 0, A) \right) \geq \inf_{x \in A} V \left(O_j, x \right) - \eta.$$

This completes the proof of part 1.

For part 2, following the same conditioning argument as for part 1 with the use of Lemmas 7.13 and 7.14 gives that for the given $\eta > 0$, there exists $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that for all $\varepsilon \in (0, \varepsilon_0)$ and $\delta \in (0, \delta_0)$,

$$E_x I^\varepsilon(\tau_1; 0, A)^2 \leq \sup_{y \in \partial A} E_y (\tau_1)^2 \cdot \sup_{z \in \partial B_\delta(O_j)} P_z \left(\vartheta_A \leq \tau_1 \right) \leq e^{\frac{\eta/2}{\varepsilon}} e^{-\frac{1}{\varepsilon} \left(\inf_{x \in A} V(O_j, x) - \eta/2 \right)}.$$

Therefore,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; 0, A)^2 \right) \geq \inf_{x \in A} V \left(O_j, x \right) - \eta.$$

□

Lemma 7.16 *Given compact sets $A_1, A_2 \subset M$, $j \in L$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z \left[\left(\int_0^{\tau_1} 1_{A_1}(X_s^\varepsilon) \, ds \right) \left(\int_0^{\tau_1} 1_{A_2}(X_s^\varepsilon) \, ds \right) \right] \right)$$

$$\geq \max \left\{ \inf_{x \in A_1} V(O_j, x), \inf_{x \in A_2} V(O_j, x) \right\} - \eta.$$

Proof We set $\vartheta_{A_i} \doteq \inf \{ t \geq 0 : X_t^\varepsilon \in A_i \}$ for $i = 1, 2$. For any $x \in \partial B_\delta(O_j)$, using a conditioning argument as in the proof of Lemma 7.15 we obtain that for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that for all $\varepsilon \in (0, \varepsilon_0)$ and $\delta \in (0, \delta_0)$,

$$E_x \left[\left(\int_0^{\tau_1} 1_{A_1}(X_s^\varepsilon) \, ds \right) \left(\int_0^{\tau_1} 1_{A_2}(X_s^\varepsilon) \, ds \right) \right]$$

$$= E_x \left[\left(\int_0^{\tau_1} \int_0^{\tau_1} 1_{A_1}(X_s^\varepsilon) 1_{A_2}(X_t^\varepsilon) \, ds \, dt \right) 1_{\{\vartheta_{A_1} \vee \vartheta_{A_2} \leq \tau_1\}} \right]$$

$$= E_x \left[\left(E_{X_{\vartheta_{A_1} \vee \vartheta_{A_2}}^\varepsilon} \left[\int_0^{\tau_1} \int_0^{\tau_1} 1_{A_1}(X_s^\varepsilon) 1_{A_2}(X_t^\varepsilon) \, ds \, dt \right] \right) 1_{\{\vartheta_{A_1} \vee \vartheta_{A_2} \leq \tau_1\}} \right]$$

$$\leq \sup_{y \in \partial A_1 \cup \partial A_2} E_y(\tau_1)^2 \cdot \sup_{z \in \partial B_\delta(O_j)} P_z(\vartheta_{A_1} \leq \tau_1, \vartheta_{A_2} \leq \tau_1)$$

$$\leq e^{\frac{\eta/2}{\varepsilon}} \cdot \min \left\{ \sup_{z \in \partial B_\delta(O_j)} P_z(\vartheta_{A_1} \leq \tau_1), \sup_{z \in \partial B_\delta(O_j)} P_z(\vartheta_{A_2} \leq \tau_1) \right\}, \quad (7.9)$$

The last inequality holds since for $i = 1, 2$

$$\sup_{z \in \partial B_\delta(O_j)} P_z(\vartheta_{A_1} \leq \tau_1, \vartheta_{A_2} \leq \tau_1) \leq \sup_{z \in \partial B_\delta(O_j)} P_z(\vartheta_{A_i} \leq \tau_1)$$

and owing to Lemma 7.13, for all $\varepsilon \in (0, \varepsilon_0)$

$$\sup_{y \in \partial A_1} E_y(\tau_1)^2 \leq e^{\frac{\eta/2}{\varepsilon}} \text{ and } \sup_{y \in \partial A_2} E_y(\tau_1)^2 \leq e^{\frac{\eta/2}{\varepsilon}}.$$

Furthermore, for the given $\eta > 0$, by Lemma 7.14, there exists $\delta_i \in (0, 1)$ such that for any $\delta \in (0, \delta_i)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} P_z(\vartheta_{A_i} \leq \tau_1) \right) \geq \inf_{x \in A_i} V(O_j, x) - \eta/2$$

for $i = 1, 2$. Hence, letting $\delta_0 = \delta_1 \wedge \delta_2$, for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z \left[\left(\int_0^{\tau_1} 1_{A_1} (X_s^\varepsilon) \, ds \right) \left(\int_0^{\tau_1} 1_{A_2} (X_s^\varepsilon) \, ds \right) \right] \right)$$

$$\geq \liminf_{\varepsilon \to 0} -\varepsilon \log \left(e^{\frac{\eta}{2\varepsilon}} \min \left\{ \sup_{z \in \partial B_\delta(O_j)} P_z \left(\vartheta_{A_1} \leq \tau_1 \right), \sup_{z \in \partial B_\delta(O_j)} P_z \left(\vartheta_{A_2} \leq \tau_1 \right) \right\} \right)$$

$$\geq -\eta/2 + \max \left\{ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} P_z \left(\vartheta_{A_1} \leq \tau_1 \right) \right), \right.$$

$$\left. \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} P_z \left(\vartheta_{A_2} \leq \tau_1 \right) \right) \right\}$$

$$\geq \max \left\{ \inf_{x \in A_1} V (O_j, x), \inf_{x \in A_2} V (O_j, x) \right\} - \eta.$$

The first inequality is from (7.9). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

Remark 7.17 The next lemma considers asymptotics of the first and second moments of a certain integral that will appear in a decomposition of S_1^ε. It is important to note that the variational bounds for both moments have the same structure as an infimum over $x \in A$. While one might consider it possible that the variational problem for the second moment could require a pair of parameters (e.g., infimum over $x, y \in A$), the infimum is in fact achieved on the "diagonal" $x = y$. This means that the biggest contribution to the second moment is likewise due to mass along the "diagonal."

Lemma 7.18 *Given a compact set $A \subset M$, a continuous function $f : M \to \mathbb{R}$, $j \in L$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon (\tau_1; f, A) \right) \geq \inf_{x \in A} \left[f(x) + V (O_j, x) \right] - \eta$$

and

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon (\tau_1; f, A)^2 \right) \geq \inf_{x \in A} \left[2f(x) + V (O_j, x) \right] - \eta.$$

Proof Since a continuous function is bounded on a compact set, there exists $m \in (0, \infty)$ such that $-m \leq f(x) \leq m$ for all $x \in A$. For $n \in \mathbb{N}$ and $k \in \{1, 2, \ldots, n\}$, consider the sets

$$A_{n,k} \doteq \left\{ x \in A : f(x) \in \left[-m + \frac{2(k-1)m}{n}, -m + \frac{2km}{n} \right] \right\}.$$

Note that $A_{n,k}$ is a compact set for any n, k. In addition, for any n fixed, $\bigcup_{k=1}^n A_{n,k} = A$. With this expression, for any $x \in \partial B_\delta(O_j)$ and $n \in \mathbb{N}$

$$E_x I^\varepsilon (\tau_1; f, A) \leq \sum_{k=1}^n E_x I^\varepsilon (\tau_1; f, A_{n,k})$$

$$\leq \sum_{k=1}^n E_x I^\varepsilon (\tau_1; 0, A_{n,k}) e^{-\frac{1}{\varepsilon} (F_{n,k} - 2m/n)}.$$

The second inequality holds because by definition of $A_{n,k}$, for any $x \in A_{n,k}$, $f(x) \geq F_{n,k} - 2m/n$ with $F_{n,k} \doteq \sup_{y \in A_{n,k}} f(y)$.

Next, we first apply (7.2) and then Lemma 7.15 with compact sets $A_{n,k}$ for $k \in \{1, 2, \ldots, n\}$ to get

$$
\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right)
$$

$$
\geq \min_{k \in \{1,\ldots,n\}} \left\{ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; 0, A_{n,k}) e^{-\frac{1}{\varepsilon}\left(F_{n,k} - \frac{2m}{n}\right)} \right) \right\}
$$

$$
= \min_{k \in \{1,\ldots,n\}} \left\{ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; 0, A_{n,k}) \right) + F_{n,k} \right\} - \frac{2m}{n}
$$

$$
\geq \min_{k \in \{1,\ldots,n\}} \left\{ \sup_{x \in A_{n,k}} f(x) + \inf_{x \in A_{n,k}} V(O_j, x) \right\} - \eta - \frac{2m}{n}.
$$

Finally, we know that $V(O_j, x)$ is bounded below by 0, and then we use the fact that for any two functions $f, g : \mathbb{R}^d \to \mathbb{R}$ with g being bounded below (to ensure that the right hand side is well defined) and any set $A \subset \mathbb{R}^d$, $\inf_{x \in A}(f(x) + g(x)) \leq \sup_{x \in A} f(x) + \inf_{x \in A} g(x)$ to find that the last minimum in the previous display is greater than or equal to

$$
\min_{k \in \{1,\ldots,n\}} \left\{ \inf_{x \in A_{n,k}} [f(x) + V(O_j, x)] \right\} = \inf_{x \in A} [f(x) + V(O_j, x)].
$$

Therefore,

$$
\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) \geq \inf_{x \in A} [f(x) + V(O_j, x)] - \eta - \frac{2m}{n}.
$$

Since n is arbitrary, sending $n \to \infty$ completes the proof for the first part.

Turning to part 2, we follow the same argument as for part 1. For any $n \in \mathbb{N}$, we use the decomposition of A into $\bigcup_{k=1}^n A_{n,k}$ to have that for any $x \in \partial B_\delta(O_j)$,

$$
E_x I^\varepsilon(\tau_1; f, A)^2 \leq E_x \left(\sum_{k=1}^n I^\varepsilon(\tau_1; f, A_{n,k}) \right)^2
$$

$$
= \sum_{k=1}^n \sum_{\ell=1}^n E_x \left[I^\varepsilon(\tau_1; f, A_{n,k}) I^\varepsilon(\tau_1; f, A_{n,\ell}) \right].
$$

Recall that $F_{n,k}$ is used to denote $\sup_{y \in A_{n,k}} f(y)$. Using the definition of $A_{n,k}$ gives that for any $k, \ell \in \{1, \ldots, n\}$

$$E_x\left[I^\varepsilon(\tau_1; f, A_{n,k})I^\varepsilon(\tau_1; f, A_{n,\ell})\right]$$

$$\leq \sup_{z\in\partial B_\delta(O_j)} E_z\left[I^\varepsilon(\tau_1; 0, A_{n,k})I^\varepsilon(\tau_1; 0, A_{n,\ell})\right] e^{-\frac{1}{\varepsilon}\left(F_{n,k}+F_{n,\ell}-\frac{4m}{n}\right)}.$$

Applying (7.2) first and then Lemma 7.16 with compact sets $A_{n,k}$ and $A_{n,\ell}$ pairwise for all $k, \ell \in \{1, 2, \ldots, n\}$ gives that

$$\liminf_{\varepsilon\to0} -\varepsilon\log\left(\sup_{z\in\partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A)^2\right)$$

$$\geq \min_{k,\ell\in\{1,\ldots,n\}} \liminf_{\varepsilon\to0} -\varepsilon\log \sup_{z\in\partial B_\delta(O_j)} E_z\left[I^\varepsilon(\tau_1; f, A_{n,k})I^\varepsilon(\tau_1; f, A_{n,\ell})\right]$$

$$\geq \min_{k,\ell\in\{1,\ldots,n\}} \left\{\max\left\{\inf_{x\in A_{n,k}} V(O_j, x), \inf_{x\in A_{n,\ell}} V(O_j, x)\right\} + F_{n,k} + F_{n,\ell}\right\} - \eta - \frac{4m}{n}$$

$$\geq \min_{k\in\{1,\ldots,n\}} \left\{\sup_{x\in A_{n,k}} [2f(x)] + \inf_{x\in A_{n,k}} V(O_j, x)\right\} - \eta - \frac{4m}{n}$$

$$\geq \min_{k\in\{1,\ldots,n\}} \left\{\inf_{x\in A_{n,k}} [2f(x) + V(O_j, x)]\right\} - \eta - \frac{4m}{n}$$

$$= \inf_{x\in A} [2f(x) + V(O_j, x)] - \eta - \frac{4m}{n}.$$

Sending $n \to \infty$ completes the proof for the second part. $\square$

Our next interest is to find lower bounds for

$$\liminf_{\varepsilon\to0} -\varepsilon\log\left(\sup_{z\in\partial B_\delta(O_1)} E_z N_j\right) \quad\text{and}\quad \liminf_{\varepsilon\to0} -\varepsilon\log\left(\sum_{\ell=1}^\infty \sup_{z\in\partial B_\delta(O_j)} P_z\left(\ell \leq N_j\right)\right).$$

We first recall that N_j is the number of visits of the embedded Markov chain $\{Z_n\}_n = \{X^\varepsilon_{\tau_n}\}_n$ to $\partial B_\delta(O_j)$ within one loop of regenerative cycle. Also, the definitions of $G(i)$ and $G(i, j)$ for any $i, j \in L$ with $i \neq j$ are given in Definition 3.8 and Remark 3.9.

Lemma 7.19 *For any $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$ and for any $j \in L$*

$$\liminf_{\varepsilon\to0} -\varepsilon\log\left(\sup_{z\in\partial B_\delta(O_1)} E_z N_j\right) \geq -\min_{\ell\in L\setminus\{1\}} V(O_1, O_\ell) + W(O_j) - W(O_1) - \eta,$$

where

$$W(O_j) \doteq \min_{g\in G(j)} \left[\sum_{(m\to n)\in g} V(O_m, O_n)\right].$$

Proof According to Lemma 3.17, we know that for any $\eta > 0$, there exist $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$ and $\varepsilon \in (0, \varepsilon_0)$, for all $x \in \partial B_\delta(O_i)$,

the one-step transition probability of the Markov chain $\{Z_n\}_n$ on $\partial B_\delta(O_j)$ satisfies the inequalities

$$e^{-\frac{1}{\varepsilon}(V(O_i,O_j)+\eta/4^{l-1})} \leq p(x, \partial B_\delta(O_j)) \leq e^{-\frac{1}{\varepsilon}(V(O_i,O_j)-\eta/4^{l-1})}. \quad (7.10)$$

We can then apply Lemma 7.6 with $p_{ij} = e^{-\frac{1}{\varepsilon}V(O_i,O_j)}$ and $a = e^{\frac{1}{\varepsilon}\eta/4^{l-1}}$ to obtain that

$$\sup_{z\in\partial B_\delta(O_1)} E_z N_j \leq \frac{e^{\frac{1}{\varepsilon}\eta}}{\sum_{\ell\in L\setminus\{1\}} e^{-\frac{1}{\varepsilon}V(O_1,O_\ell)}} \frac{\sum_{g\in G(j)}\pi(g)}{\sum_{g\in G(1)}\pi(g)}$$

$$\leq \frac{e^{\frac{1}{\varepsilon}\eta}}{e^{-\frac{1}{\varepsilon}\min_{\ell\in L\setminus\{1\}} V(O_1,O_\ell)}} \frac{\sum_{g\in G(j)}\pi(g)}{\sum_{g\in G(1)}\pi(g)}.$$

Thus,

$$\liminf_{\varepsilon\to 0} -\varepsilon\log\left(\sup_{z\in\partial B_\delta(O_1)} E_z N_j\right)$$

$$\geq -\min_{\ell\in L\setminus\{1\}} V(O_1, O_\ell) - \eta + \liminf_{\varepsilon\to 0} -\varepsilon\log\left(\frac{\sum_{g\in G(j)}\pi(g)}{\sum_{g\in G(1)}\pi(g)}\right).$$

Hence, it suffices to show that

$$\liminf_{\varepsilon\to 0} -\varepsilon\log\left(\frac{\sum_{g\in G(j)}\pi(g)}{\sum_{g\in G(1)}\pi(g)}\right) \geq W(O_j) - W(O_1).$$

Observe that by definition for any $j\in L$ and $g\in G(j)$

$$\pi(g) = \prod_{(m\to n)\in g} p_{mn} = \prod_{(m\to n)\in g} e^{-\frac{1}{\varepsilon}V(O_m,O_n)} = \exp\left\{-\frac{1}{\varepsilon}\sum_{(m\to n)\in g} V(O_m, O_n)\right\},$$

which implies that

$$\liminf_{\varepsilon\to 0} -\varepsilon\log\left(\frac{\sum_{g\in G(j)}\pi(g)}{\sum_{g\in G(1)}\pi(g)}\right)$$

$$\geq \min_{g\in G(j)}\left[\liminf_{\varepsilon\to 0} -\varepsilon\log\left(\exp\left\{-\frac{1}{\varepsilon}\sum_{(m\to n)\in g} V(O_m, O_n)\right\}\right)\right]$$

$$\quad - \min_{g\in G(1)}\left[\limsup_{\varepsilon\to 0} -\varepsilon\log\left(\exp\left\{-\frac{1}{\varepsilon}\sum_{(m\to n)\in g} V(O_m, O_n)\right\}\right)\right]$$

$$= \min_{g\in G(j)}\left[\sum_{(m\to n)\in g} V(O_m, O_n)\right] - \min_{g\in G(1)}\left[\sum_{(m\to n)\in g} V(O_m, O_n)\right]$$

$$= W(O_j) - W(O_1).$$

The inequality is from Lemma 7.1; the last equality holds due the definition of $W\left(O_j\right)$.

$\square$

Recall the definition of $W(O_1 \cup O_j)$ in (3.3). In the next result, we obtain bounds on, for example, a quantity close to the expected number of visits to $B_\delta(O_j)$ before visiting a neighborhood of O_1, after starting near O_j.

Lemma 7.20 *For any $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^{\infty} \sup_{z \in \partial B_\delta(O_1)} P_z \left(\ell \leq N_1 \right) \right) \geq - \min_{\ell \in L \backslash \{1\}} V\left(O_1, O_\ell\right) - \eta$$

and for any $j \in L \backslash \{1\}$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^{\infty} \sup_{z \in \partial B_\delta(O_j)} P_z \left(\ell \leq N_j \right) \right) \geq W(O_1 \cup O_j) - W\left(O_1\right) - \eta.$$

Proof We again use that by Lemma 3.17, for any $\eta > 0$ there exist $\delta_0 \in (0, 1)$ and $\varepsilon_0 \in (0, 1)$, such that (7.10) holds for any $\delta \in (0, \delta_0)$, $\varepsilon \in (0, \varepsilon_0)$ and all $x \in \partial B_\delta(O_i)$. Then, by Lemma 7.6 with $p_{ij} = e^{-\frac{1}{\varepsilon}V(O_i, O_j)}$ and $a = e^{\frac{1}{\varepsilon}\eta/4^{l-1}}$

$$\sum_{\ell=1}^{\infty} \sup_{x \in \partial B_\delta(O_j)} P_x \left(N_1 \geq \ell \right) \leq \frac{e^{\frac{1}{\varepsilon}\eta}}{\sum_{\ell \in L \backslash \{1\}} e^{-\frac{1}{\varepsilon}V(O_1, O_\ell)}}$$

and for any $j \in L \backslash \{1\}$

$$\sum_{\ell=1}^{\infty} \sup_{x \in \partial B_\delta(O_j)} P_x \left(N_j \geq \ell \right) \leq e^{\frac{1}{\varepsilon}\eta} \frac{\sum_{g \in G(1,j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)}.$$

Thus,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^{\infty} \sup_{z \in \partial B_\delta(O_1)} P_z \left(\ell \leq N_1 \right) \right)$$
$$\geq - \limsup_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell \in L \backslash \{1\}} e^{-\frac{1}{\varepsilon}V(O_1, O_\ell)} \right) - \eta$$

and

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=0}^{\infty} \sup_{z \in \partial B_\delta(O_j)} P_z \left(\ell \leq N_j \right) \right)$$
$$\geq \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\frac{\sum_{g \in G(1,j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)} \right) - \eta.$$

Following the same argument as for the proof of Lemma 7.19, we can use Lemma 7.1 to obtain that

$$
-\limsup_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell \in L \setminus \{1\}} e^{-\frac{1}{\varepsilon} V(O_1, O_\ell)} \right) \geq - \min_{\ell \in L \setminus \{1\}} V(O_1, O_\ell)
$$

and

$$
\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\frac{\sum_{g \in G(1,j)} \pi(g)}{\sum_{g \in G(1)} \pi(g)} \right)
$$
$$
\geq \min_{g \in G(1,j)} \left[\sum_{(m \to n) \in g} V(O_m, O_n) \right] - \min_{g \in G(1)} \left[\sum_{(m \to n) \in g} V(O_m, O_n) \right].
$$

Recalling (3.2) and (3.3), we are done.   □

As mentioned at the beginning of this subsection, our main goal is to provide lower bounds for

$$
\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z \left(\int_0^{\tau_1^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_s^\varepsilon)} 1_A (X_s^\varepsilon) \, ds \right) \right)
$$

and

$$
\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z \left(\int_0^{\tau_1^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_s^\varepsilon)} 1_A (X_s^\varepsilon) \, ds \right)^2 \right)
$$

for a given continuous function $f : M \to \mathbb{R}$ and compact set $A \subset M$. We now state the main results of the subsection. Recall that $h_1 = \min_{\ell \in L \setminus \{1\}} V(O_1, O_\ell)$, $S_1^\varepsilon \doteq \int_0^{\tau_1^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_s^\varepsilon)} 1_A (X_s^\varepsilon) \, ds$ and $W(O_j) \doteq \min_{g \in G(j)} [\sum_{(m \to n) \in g} V(O_m, O_n)]$ and the definitions (7.8).

Lemma 7.21 *Given a compact set $A \subset M$, a continuous function $f : M \to \mathbb{R}$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$
\liminf_{\varepsilon \to 0} -\varepsilon \log \left[\sup_{z \in \partial B_\delta(O_1)} E_z S_1^\varepsilon \right] \geq \min_{j \in L} \left\{ \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) \right\}
$$
$$
- W(O_1) - h_1 - \eta.
$$

Proof Recall that by Lemma 7.18, we have shown that for the given η, there exists $\delta_1 \in (0, 1)$, such that for any $\delta \in (0, \delta_1)$ and $j \in L$

$$
\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) \geq \inf_{x \in A} \left[f(x) + V(O_j, x) \right] - \frac{\eta}{2}.
$$

In addition, by Lemma 7.19, we know that for the same η, there exists $\delta_2 \in (0, 1)$, such that for any $\delta \in (0, \delta_2)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right) \geq - \min_{\ell \in L \setminus \{1\}} V\left(O_1, O_\ell\right) + W\left(O_j\right) - W\left(O_1\right) - \eta/2.$$

Hence, for any $\delta \in (0, \delta_0)$ with $\delta_0 = \delta_1 \wedge \delta_2$, we apply Corollary 7.11 to get

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left[E_x I^\varepsilon(\tau_1^\varepsilon; f, A) \right]$$

$$\geq \min_{j \in L} \left\{ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z\left(N_j\right) \right) \right\}$$

$$\geq \min_{j \in L} \left\{ \inf_{x \in A} \left[f(x) + V\left(O_j, x\right) \right] + W\left(O_j\right) \right\} - W\left(O_1\right) - h_1 - \eta,$$

where τ_1^ε is the time for a regenerative cycle and τ_1 is the first visit time of neighborhoods of equilibrium points after being a certain distance away from them. $\qquad \square$

Remark 7.22 According to Remark 7.12 and using the same argument as in Lemma 7.21, we can find that given a compact set $A \subset M$, a continuous function $f : M \to \mathbb{R}$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left[\sup_{z \in \partial B_\delta(O_1)} E_z I^\varepsilon(\sigma_0^\varepsilon, \tau_1^\varepsilon; f, A) \right]$$

$$\geq \min_{j \in L \setminus \{1\}} \left\{ \inf_{x \in A} \left[f(x) + V\left(O_j, x\right) \right] + W\left(O_j\right) \right\} - W\left(O_1\right) - h_1 - \eta.$$

Lemma 7.23 *Given a compact set $A \subset M$, a continuous function $f : M \to \mathbb{R}$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left[\sup_{z \in \partial B_\delta(O_1)} E_z(S_1^\varepsilon)^2 \right] \geq \min_{j \in L} \left(R_j^{(1)} \wedge R_j^{(2)} \right) - h_1 - \eta,$$

where $S_1^\varepsilon \doteq \int_0^{\tau_1^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_s^\varepsilon)} 1_A\left(X_s^\varepsilon\right) ds$ and $h_1 = \min_{\ell \in L \setminus \{1\}} V\left(O_1, O_\ell\right)$, and

$$R_j^{(1)} \doteq \inf_{x \in A} \left[2f(x) + V\left(O_j, x\right) \right] + W\left(O_j\right) - W\left(O_1\right)$$

$$R_1^{(2)} \doteq 2 \inf_{x \in A} \left[f(x) + V\left(O_1, x\right) \right] - h_1$$

and for $j \in L \setminus \{1\}$

$$R_j^{(2)} \doteq 2 \inf_{x \in A} \left[f(x) + V\left(O_j, x\right) \right] + W\left(O_j\right) - 2W\left(O_1\right) + W(O_1 \cup O_j).$$

Proof Following a similar argument as for the proof of Lemma 7.21, given any $\eta > 0$, owing to Lemmas 7.18, 7.19 and 7.20, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$ and for any $j \in L$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) \geq \inf_{x \in A} \left[f(x) + V(O_j, x) \right] - \frac{\eta}{4},$$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A)^2 \right) \geq \inf_{x \in A} \left[2f(x) + V(O_j, x) \right] - \frac{\eta}{4},$$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right) \geq -h_1 + W(O_j) - W(O_1) - \frac{\eta}{4},$$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^{\infty} \sup_{z \in \partial B_\delta(O_1)} P_z(\ell \leq N_1) \right) \geq -h_1 - \frac{\eta}{4},$$

and for any $j \in L \setminus \{1\}$,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^{\infty} \sup_{z \in \partial B_\delta(O_j)} P_z(\ell \leq N_j) \right) \geq W(O_1 \cup O_j) - W(O_1) - \frac{\eta}{4}.$$

Hence, for any $\delta \in (0, \delta_0)$ we apply Corollary 7.11 to get

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z \left(S_1^\varepsilon \right)^2 \right) \geq \min_{j \in L} \left(\hat{R}_j^{(1)} \wedge \hat{R}_j^{(2)} \right),$$

where

$$\hat{R}_j^{(1)} \doteq \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A)^2 \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right)$$

$$\geq \inf_{x \in A} \left[2f(x) + V(O_j, x) \right] + W(O_j) - W(O_1) - h_1 - \eta = R_j^{(1)} - h_1 - \eta$$

and

$$\hat{R}_1^{(2)} \doteq 2 \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z I^\varepsilon(\tau_1; f, A) \right)$$

$$+ \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_1 \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^{\infty} \sup_{z \in \partial B_\delta(O_1)} P_z(\ell \leq N_1) \right)$$

$$\geq 2 \left(\inf_{x \in A} \left[f(x) + V(O_1, x) \right] - \frac{\eta}{4} \right) + \left(-h_1 - \frac{\eta}{4} \right) + \left(-h_1 - \frac{\eta}{4} \right)$$

$$= 2 \inf_{x \in A} \left[f(x) + V(O_1, x) \right] - 2h_1 - \eta = R_1^{(2)} - h_1 - \eta$$

and for $j \in L \setminus \{1\}$

$$
\begin{aligned}
\hat{R}_j^{(2)} &\doteq 2 \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_j)} E_z I^\varepsilon(\tau_1; f, A) \right) \\
&\quad + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z N_j \right) + \liminf_{\varepsilon \to 0} -\varepsilon \log \left(\sum_{\ell=1}^\infty \sup_{z \in \partial B_\delta(O_j)} P_z \left(\ell \leq N_j \right) \right) \\
&\geq 2 \left(\inf_{x \in A} \left[f(x) + V(O_j, x) \right] - \frac{\eta}{4} \right) + \left(-h_1 + W(O_j) - W(O_1) - \frac{\eta}{4} \right) \\
&\quad + \left(W(O_1 \cup O_j) - W(O_1) - \frac{\eta}{4} \right) \\
&= 2 \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) - 2W(O_1) + W(O_1 \cup O_j) - h_1 - \eta \\
&= R_j^{(2)} - h_1 - \eta.
\end{aligned}
$$

$\square$

7.3 Asymptotics of Moments of $\hat{S}_1^\varepsilon$

Recall that

$$
\hat{S}_n^\varepsilon \doteq \int_{\hat{\tau}_{n-1}^\varepsilon}^{\hat{\tau}_n^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon \right) dt,
$$

where $\hat{\tau}_i^\varepsilon$ is a multicycle defined according to (6.4) and with $\{\mathbf{M}_i^\varepsilon\}_{i \in \mathbb{N}}$ being a sequence of independent and geometrically distributed random variables with parameter $e^{-m/\varepsilon}$ for some $m > 0$ such that $m + h_1 > w$. Moreover, $\{\mathbf{M}_i^\varepsilon\}$ is also independent of $\{\tau_n^\varepsilon\}$. Using the independence of $\{\mathbf{M}_i^\varepsilon\}$ and $\{\tau_n^\varepsilon\}$, and the fact that $\{\tau_n^\varepsilon\}$ and $\{S_n^\varepsilon\}$ are both iid under P_{λ^ε}, we find that $\{\hat{S}_n^\varepsilon\}$ is also iid under P_{λ^ε} and

$$
E_{\lambda^\varepsilon} \hat{S}_1^\varepsilon = E_{\lambda^\varepsilon} \mathbf{M}_1^\varepsilon \cdot E_{\lambda^\varepsilon} S_1^\varepsilon \tag{7.11}
$$

and

$$
\begin{aligned}
\operatorname{Var}_{\lambda^\varepsilon} \hat{S}_1^\varepsilon &= E_{\lambda^\varepsilon} \mathbf{M}_1^\varepsilon \cdot \operatorname{Var}_{\lambda^\varepsilon}(S_1^\varepsilon) + \operatorname{Var}_{\lambda^\varepsilon}(\mathbf{M}_1^\varepsilon) \cdot (E_{\lambda^\varepsilon} S_1^\varepsilon)^2 \\
&\leq E_{\lambda^\varepsilon} \mathbf{M}_1^\varepsilon \cdot E_{\lambda^\varepsilon}(S_1^\varepsilon)^2 + \operatorname{Var}_{\lambda^\varepsilon}(\mathbf{M}_1^\varepsilon) \cdot (E_{\lambda^\varepsilon} S_1^\varepsilon)^2
\end{aligned} \tag{7.12}
$$

On the other hand, since $\mathbf{M}_1^\varepsilon$ is geometrically distributed with parameter $e^{-m/\varepsilon}$, this gives that

$$
E_{\lambda^\varepsilon} \mathbf{M}_1^\varepsilon = e^{\frac{m}{\varepsilon}} \text{ and } \operatorname{Var}_{\lambda^\varepsilon}(\mathbf{M}_1^\varepsilon) = e^{\frac{2m}{\varepsilon}} (1 - e^{\frac{-m}{\varepsilon}}). \tag{7.13}
$$

Therefore, by combining (7.11), (7.12) and (7.13) with Lemma 7.21 and Lemma 7.23, we have the following two lemmas.

Lemma 7.24 *Given a compact set $A \subset M$, a continuous function $f : M \to \mathbb{R}$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log E_{\lambda^\varepsilon} \hat{S}_1^\varepsilon$$

$$\geq \min_{j \in L} \left\{ \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) \right\} - W(O_1) - (m + h_1) - \eta.$$

Lemma 7.25 *Given a compact set $A \subset M$, a continuous function $f : M \to \mathbb{R}$ and $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \mathrm{Var}_{\lambda^\varepsilon}(\hat{S}_1^\varepsilon) \geq \min_{j \in L} \left(R_j^{(1)} \wedge R_j^{(2)} \wedge R_j^{(3,m)} \right) - (m + h_1) - \eta,$$

where $R_j^{(1)}$ and $R_j^{(2)}$ are defined as in Lemma 7.23, and

$$R_j^{(3,m)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + 2W(O_j) - 2W(O_1) - (m + h_1).$$

Later on, we will optimize on m to obtain the largest bound from below. This will require that we consider first $m > w - h_1$, so that as shown in the next section $N^\varepsilon(T^\varepsilon)$ can be suitably approximated in terms of a Poisson distribution, and then sending $m \downarrow w - h_1$.

8 Asymptotics of Moments of $N^\varepsilon(T^\varepsilon)$ and $\hat{N}^\varepsilon(T^\varepsilon)$

Recall that the number of single cycles in the time interval $[0, T^\varepsilon]$ plus one is defined as

$$N^\varepsilon(T^\varepsilon) \doteq \inf \left\{ n \in \mathbb{N} : \tau_n^\varepsilon > T^\varepsilon \right\},$$

where the τ_n^ε are the return times to $B_\delta(O_1)$ after ever visiting one of the δ-neighborhood of other equilibrium points than O_1. In addition, λ^ε is the unique invariant measure of $\{Z_n^\varepsilon\}_n = \{X_{\tau_n^\varepsilon}^\varepsilon\}_n$. The number of multicycles in the time interval $[0, T^\varepsilon]$ plus one is defined as

$$\hat{N}^\varepsilon(T^\varepsilon) \doteq \inf \left\{ n \in \mathbb{N} : \hat{\tau}_n^\varepsilon > T^\varepsilon \right\},$$

where $\hat{\tau}_i^\varepsilon$ are defined as in (6.4).

In this section, we will find the logarithmic asymptotics of the expected value and the variance of $N^\varepsilon(T^\varepsilon)$ with $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > h_1$ in Lemmas 8.2 and 8.4 under the assumption that $h_1 > w$ (i.e., single cycle case), and the analogous quantities for $\hat{N}^\varepsilon(T^\varepsilon)$ with $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > w$ in Lemmas 8.19 and 8.21 under the assumption that $w \geq h_1$ (i.e., multicycle case).

Remark 8.1 While the proofs of these asymptotic results are quite detailed, it is essential that we obtain estimates good enough for a relatively precise comparison of the expected value and the variance of $N^\varepsilon (T^\varepsilon)$, and likewise for $\hat{N}^\varepsilon (T^\varepsilon)$. For this, the key result needed is the characterization of $N^\varepsilon (T^\varepsilon)$ (and $\hat{N}^\varepsilon (T^\varepsilon)$) as having an approximately Poisson distribution. These follow by exploiting the asymptotically exponential character of τ_n^ε (and $\hat{\tau}_n^\varepsilon$), together with some uniform integrability properties.

Lemmas 8.2 and 8.4 below are proved in Sect. 8.3.

Lemma 8.2 If $h_1 > w$ and $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > h_1$, then there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left| \frac{E_{\lambda^\varepsilon} (N^\varepsilon (T^\varepsilon))}{T^\varepsilon} - \frac{1}{E_{\lambda^\varepsilon} \tau_1^\varepsilon} \right| \geq c.$$

Corollary 8.3 If $h_1 > w$ and $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > h_1$, then there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon} (N^\varepsilon (T^\varepsilon))}{T^\varepsilon} \geq \varkappa_\delta,$$

where $\varkappa_\delta \doteq \min_{y \in \cup_{k \in L \setminus \{1\}} \partial B_\delta (O_k)} V (O_1, y)$.

Lemma 8.4 If $h_1 > w$ and $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > h_1$, then for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{\mathrm{Var}_{\lambda^\varepsilon} (N^\varepsilon (T^\varepsilon))}{T^\varepsilon} \geq h_1 - \eta.$$

Before proceeding, we mention a result from [11] and define some notation which will be used in this section. Results in Section 5 and Section 10 of [11, Chapter XI] say that for any $t > 0$, the first and second moment of $N^\varepsilon (t)$ can be represented as

$$E_{\lambda^\varepsilon} \left(N^\varepsilon (t) \right) = \sum_{n=0}^\infty P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \leq t \right) \text{ and } E_{\lambda^\varepsilon} \left(N^\varepsilon (t) \right)^2 = \sum_{n=0}^\infty (2n + 1) P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \leq t \right). \tag{8.1}$$

Let $\Gamma^\varepsilon \doteq T^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon$ and $\gamma^\varepsilon \doteq (\Gamma^\varepsilon)^{-\ell}$ with some $\ell \in (0, 1)$ which will be chosen later. Intuitively, Γ^ε is the typical number of regenerative cycles in $[0, T^\varepsilon]$ since $E_{\lambda^\varepsilon} \tau_1^\varepsilon$ is the expected length of one regenerative cycle. To simplify notation, we pretend that $(1 + 2\gamma^\varepsilon) \Gamma^\varepsilon$ and $(1 - 2\gamma^\varepsilon) \Gamma^\varepsilon$ are positive integers so that we can divide $E_{\lambda^\varepsilon} (N^\varepsilon (T^\varepsilon))$ into three partial sums which are

$$\mathfrak{P}_1 \doteq \sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^\infty P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \leq T^\varepsilon \right), \ \mathfrak{P}_2 \doteq \sum_{n=(1-2\gamma^\varepsilon)\Gamma^\varepsilon}^{(1+2\gamma^\varepsilon)\Gamma^\varepsilon} P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \leq T^\varepsilon \right)$$

and

$$\mathfrak{P}_3 \doteq \sum_{n=0}^{(1-2\gamma^\varepsilon)\Gamma^\varepsilon-1} P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \leq T^\varepsilon \right). \tag{8.2}$$

Similarly, we divide $E_{\lambda^\varepsilon} \left(N^\varepsilon \left(T^\varepsilon \right) \right)^2$ into

$$\mathfrak{R}_1 \doteq \sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^{\infty} (2n+1) \, P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \le T^\varepsilon \right), \quad \mathfrak{R}_2 \doteq \sum_{n=(1-2\gamma^\varepsilon)\Gamma^\varepsilon}^{(1+2\gamma^\varepsilon)\Gamma^\varepsilon} (2n+1) \, P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \le T^\varepsilon \right)$$

and

$$\mathfrak{R}_3 \doteq \sum_{n=0}^{(1-2\gamma^\varepsilon)\Gamma^\varepsilon-1} (2n+1) \, P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \le T^\varepsilon \right). \tag{8.3}$$

The next step is to find upper bounds for these partial sums, and these bounds will help us to find suitable lower bounds for the logarithmic asymptotics of $E_{\lambda^\varepsilon} \left(N^\varepsilon \left(T^\varepsilon \right) \right)$ and $\mathrm{Var}_{\lambda^\varepsilon} \left(N^\varepsilon \left(T^\varepsilon \right) \right)$. Before looking into the upper bound for partial sums, we establish some properties.

Theorem 8.5 *If $h_1 > w$, then for any $\delta > 0$ sufficiently small,*

$$\lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \tau_1^\varepsilon = \varkappa_\delta \text{ and } \tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon \xrightarrow{d} \mathrm{Exp}(1).$$

Moreover, there exists $\varepsilon_0 \in (0, 1)$ and a constant $\tilde{c} > 0$ such that

$$P_{\lambda^\varepsilon} \left(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon > t \right) \le e^{-\tilde{c}t}$$

for any $t > 0$ and any $\varepsilon \in (0, \varepsilon_0)$.

Remark 8.6 For any $\delta > 0$, $\varkappa_\delta \le h_1$.

The proof of Theorem 8.5 will be given in Section 10. In that section, we will first prove an analogous result for the exit time (or first visiting time to other equilibrium points to be more precise) and then show how one can extend those results to the return time. The proof of the following lemma is straightforward and hence omitted.

Lemma 8.7 *If $h_1 > w$ and $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > h_1$, then for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$,*

$$h_1 - \eta \ge \lim_{\varepsilon \to 0} -\varepsilon \log \Gamma^\varepsilon \ge h_1 - c - \eta.$$

Lemma 8.8 *Define $\mathcal{Z}_1^\varepsilon \doteq \tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon$. Then, for any $\delta > 0$ sufficiently small,*

- *there exists some $\varepsilon_0 \in (0, 1)$ such that $\sup_{\varepsilon \in (0,\varepsilon_0)} E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^3 < \infty$,*
- *there exists some $\varepsilon_0 \in (0, 1)$ such that $\inf_{\varepsilon \in (0,\varepsilon_0)} \mathrm{Var}_{\lambda^\varepsilon}(\mathcal{Z}_1^\varepsilon) > 0$ and $E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^2 = E_{\lambda^\varepsilon} \left(\tau_1^\varepsilon \right)^2 / \left(E_{\lambda^\varepsilon} \tau_1^\varepsilon \right)^2 \to 2$ as $\varepsilon \to 0$.*

Proof For the first part, we use Theorem 8.5 to find that there exists $\varepsilon_0 \in (0, 1)$ and a constant $\tilde{c} > 0$ such that

$$P_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon > t \right) = P_{\lambda^\varepsilon} \left(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon > t \right) \le e^{-\tilde{c}t}$$

for any $t > 0$ and any $\varepsilon \in (0, \varepsilon_0)$. Therefore, for $\varepsilon \in (0, \varepsilon_0)$

$$E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^3 = 3 \int_0^\infty t^2 P_{\lambda^\varepsilon} (\mathcal{Z}_1^\varepsilon > t) dt \le 3 \int_0^\infty t^2 e^{-\tilde{c}t} dt < \infty.$$

For the second assertion, since $\sup_{0 < \varepsilon < \varepsilon_0} E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^3 < \infty$, it implies that $\{ \left(\mathcal{Z}_1^\varepsilon \right)^2 \}_{0 < \varepsilon < \varepsilon_0}$ and $\{ \mathcal{Z}_1^\varepsilon \}_{0 < \varepsilon < \varepsilon_0}$ are both uniformly integrable. Moreover, because $\mathcal{Z}_1^\varepsilon \overset{d}{\to} \mathrm{Exp}(1)$ as $\varepsilon \to 0$ from Theorem 8.5 and since for $X \overset{d}{=} \mathrm{Exp}(1)$, $EX = 1$ and $EX^2 = 2$, we obtain

$$E_{\lambda^\varepsilon} \left(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon \right)^2 = E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^2 \to 2 \text{ and } E_{\lambda^\varepsilon} \mathcal{Z}_1^\varepsilon \to 1.$$

as $\varepsilon \to 0$. This implies $\mathrm{Var}_{\lambda^\varepsilon} (\mathcal{Z}_1^\varepsilon) \to 1$ as $\varepsilon \to 0$. Obviously, there exists some $\varepsilon_0 \in (0, 1)$ such that $\inf_{\varepsilon \in (0, \varepsilon_0)} \mathrm{Var}_{\lambda^\varepsilon} (\mathcal{Z}_1^\varepsilon) \ge 1/2 > 0$. This completes the proof. □

Remark 8.9 Throughout the rest of this section, we will use C to denote a constant in $(0, \infty)$ which is independent of ε but whose value may change from use to use.

8.1 Chernoff Bound

In this subsection, we will provide upper bounds for

$$\mathfrak{P}_1 \doteq \sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^\infty P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \le T^\varepsilon \right) \text{ and } \mathfrak{R}_1 \doteq \sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^\infty (2n+1) P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \le T^\varepsilon \right)$$

via a Chernoff bound. The following result is well known, and its proof is standard.

Lemma 8.10 *(Chernoff bound) Let $X_1, \dots, X_n$ be an iid sequence of random variables. For any $a \in \mathbb{R}$ and for any $t \in (0, \infty)$*

$$P \left(X_1 + \cdots + X_n \le a \right) \le \left(E e^{-tX_1} \right)^n e^{ta}.$$

Recall that $\Gamma^\varepsilon \doteq T^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon$ and $\gamma^\varepsilon \doteq (\Gamma^\varepsilon)^{-\ell}$ with some $\ell \in (0, 1)$ which will be chosen later.

Lemma 8.11 *Given any $\delta > 0$ and any $\ell > 0$, there exists $\varepsilon_0 \in (0, 1)$ such that for any $\varepsilon \in (0, \varepsilon_0)$*

$$P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \le T^\varepsilon \right) \le e^{-n(\Gamma^\varepsilon)^{-2\ell}}$$

for any $n \ge (1 + 2\gamma^\varepsilon) \Gamma^\varepsilon$. In addition,

$$\mathfrak{P}_1 \le C \left(\Gamma^\varepsilon \right)^{2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}} \text{ and } \mathfrak{R}_1 \le C \left(\Gamma^\varepsilon \right)^{1+2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}} + C \left(\Gamma^\varepsilon \right)^{4\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}}.$$

Proof Given $\delta > 0$, $\ell > 0$ and $\varepsilon \in (0, 1)$, we find that for $n \geq (1 + 2\gamma^\varepsilon) \Gamma^\varepsilon$

$$P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \leq T^\varepsilon \right) = P_{\lambda^\varepsilon} \left(\frac{\tau_1^\varepsilon + \left(\tau_2^\varepsilon - \tau_1^\varepsilon \right) + \cdots + \left(\tau_n^\varepsilon - \tau_{n-1}^\varepsilon \right)}{E_{\lambda^\varepsilon} \tau_1^\varepsilon} \leq \Gamma^\varepsilon \right)$$

$$\leq P_{\lambda^\varepsilon} \left(\frac{\tau_1^\varepsilon + \left(\tau_2^\varepsilon - \tau_1^\varepsilon \right) + \cdots + \left(\tau_n^\varepsilon - \tau_{n-1}^\varepsilon \right)}{E_{\lambda^\varepsilon} \tau_1^\varepsilon} \leq \frac{n}{1 + 2\gamma^\varepsilon} \right)$$

$$\leq \left(E_{\lambda^\varepsilon} e^{-\gamma^\varepsilon \mathcal{Z}_1^\varepsilon} \right) e^{\frac{n\gamma^\varepsilon}{1+2\gamma^\varepsilon}},$$

where $\mathcal{Z}_1^\varepsilon = \tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon$. We use the fact that $\{ \tau_n^\varepsilon - \tau_{n-1}^\varepsilon \}_{n \in \mathbb{N}}$ are iid and apply Lemma 8.10 (Chernoff bound) with $a = n/ (1 + 2\gamma^\varepsilon)$ and $t = \gamma^\varepsilon$ for the last inequality. Therefore, in order to verify the first claim, it suffices to show that

$$\left(E_{\lambda^\varepsilon} e^{-\gamma^\varepsilon \mathcal{Z}_1^\varepsilon} \right) e^{\frac{\gamma^\varepsilon}{1+2\gamma^\varepsilon}} \leq e^{-(\gamma^\varepsilon)^2} = e^{-(\Gamma^\varepsilon)^{-2\ell}}.$$

We observe that for any $x \geq 0$, $e^{-x} \leq 1 - x + x^2/2$, and this gives

$$E_{\lambda^\varepsilon} e^{-\gamma^\varepsilon \mathcal{Z}_1^\varepsilon} \leq 1 - E_{\lambda^\varepsilon} \left(\gamma^\varepsilon \mathcal{Z}_1^\varepsilon \right) + E_{\lambda^\varepsilon} \left(\gamma^\varepsilon \mathcal{Z}_1^\varepsilon \right)^2 /2 = 1 - \gamma^\varepsilon + \left(\gamma^\varepsilon \right)^2 E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^2 /2.$$

Moreover, since we can apply Lemma 8.8 to find $E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^2 \to 2$ as $\varepsilon \to 0$, there exists $\varepsilon_0 \in (0, 1)$ such that for any $\varepsilon \in (0, \varepsilon_0)$, $E_{\lambda^\varepsilon} \left(\mathcal{Z}_1^\varepsilon \right)^2 \leq 9/4$. Thus, for any $\varepsilon \in (0, \varepsilon_0)$

$$\left(E_{\lambda^\varepsilon} e^{-\gamma^\varepsilon \mathcal{Z}_1^\varepsilon} \right) e^{\frac{\gamma^\varepsilon}{1+2\gamma^\varepsilon}} \leq \exp \left\{ \gamma^\varepsilon /(1 + 2\gamma^\varepsilon) + \log(1 - \gamma^\varepsilon + (9/8) \left(\gamma^\varepsilon \right)^2) \right\}.$$

Using a Taylor series expansion, we find that for all $|x| < 1$

$$1/(1 + x) = 1 - x + O \left(x^2 \right) \text{ and } \log (1 + x) = x - x^2/2 + O \left(x^3 \right),$$

which gives

$$\gamma^\varepsilon /(1 + 2\gamma^\varepsilon) + \log(1 - \gamma^\varepsilon + (9/8) \left(\gamma^\varepsilon \right)^2)$$
$$= \gamma^\varepsilon - 2 \left(\gamma^\varepsilon \right)^2 + [(-\gamma^\varepsilon + (9/8) \left(\gamma^\varepsilon \right)^2] - [-\gamma^\varepsilon + (9/8) \left(\gamma^\varepsilon \right)^2]^2/2 + O((\gamma^\varepsilon)^3)$$
$$= -(11/8) \left(\gamma^\varepsilon \right)^2 + O((\gamma^\varepsilon)^3) \leq - \left(\gamma^\varepsilon \right)^2,$$

for all $\varepsilon \in (0, \varepsilon_0)$. We are done for part 1.

For part 2, we use the estimate from part 1 and find

$$\sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^{\infty} P_{\lambda^\varepsilon} \left(\tau_n^\varepsilon \leq T^\varepsilon \right) \leq \sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^{\infty} e^{-n(\gamma^\varepsilon)^2} \leq \frac{e^{-(1+2\gamma^\varepsilon)\Gamma^\varepsilon(\gamma^\varepsilon)^2}}{1 - e^{-(\gamma^\varepsilon)^2}}.$$

Since $e^{-x} \le 1-x+x^2/2$ for any $x \in \mathbb{R}$, we have $1-e^{-x} \ge x-x^2/2 \ge x-x/2 = x/2$ for all $x \in (0,1)$, and thus $1/(1-e^{-x}) \le 2/x$ for all $x \in (0,1)$. As a result,

$$\sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^{\infty} P_{\lambda^\varepsilon}\left(\tau_n^\varepsilon \le T^\varepsilon\right) \le \frac{e^{-(1+2\gamma^\varepsilon)\Gamma^\varepsilon(\gamma^\varepsilon)^2}}{1-e^{-(\gamma^\varepsilon)^2}} \le \frac{2}{(\gamma^\varepsilon)^2}e^{-(1+2\gamma^\varepsilon)\Gamma^\varepsilon(\gamma^\varepsilon)^2}$$

$$\le 2\left(\Gamma^\varepsilon\right)^{2\ell}e^{-(\Gamma^\varepsilon)^{1-2\ell}}.$$

This completes the proof of part 2.

Finally, for part 3, we use the fact that for $x \in (0,1)$, and for any $k \in \mathbb{N}$,

$$\sum_{n=k}^{\infty} nx^n = kx^k(1-x)^{-1} + x^{k+1}(1-x)^{-2} \le (k(1-x)^{-1} + (1-x)^{-2})x^k.$$

Using the estimate from part 1 once again, we have

$$\sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon+1}^{\infty} nP_{\lambda^\varepsilon}\left(\tau_n^\varepsilon \le T^\varepsilon\right) \le \sum_{n=(1+2\gamma^\varepsilon)\Gamma^\varepsilon}^{\infty} ne^{-n(\gamma^\varepsilon)^2}$$

$$\le \left(\frac{(1+2\gamma^\varepsilon)\Gamma^\varepsilon}{1-e^{-(\gamma^\varepsilon)^2}} + \left(1-e^{-(\gamma^\varepsilon)^2}\right)^{-2}\right)e^{-(1+2\gamma^\varepsilon)\Gamma^\varepsilon(\gamma^\varepsilon)^2}$$

$$\le \left(4\left(\Gamma^\varepsilon\right)^{1+2\ell} + 4\left(\Gamma^\varepsilon\right)^{4\ell}\right)e^{-(\Gamma^\varepsilon)^{1-2\ell}}.$$

We are done.   □

Remark 8.12 If $0 < \ell < 1/2$, then $\mathfrak{P}_1$ and $\mathfrak{R}_1$ converge to 0 doubly exponentially fast as $\varepsilon \to 0$ in the sense that for any $k \in (0,\infty)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log\left[\left(\Gamma^\varepsilon\right)^k e^{-(\Gamma^\varepsilon)^{1-2\ell}}\right] = \infty.$$

8.2 Berry–Esseen Bound

In this subsection, we will provide upper bounds for

$$\mathfrak{P}_2 \doteq \sum_{n=(1-2\gamma^\varepsilon)\Gamma^\varepsilon}^{(1+2\gamma^\varepsilon)\Gamma^\varepsilon} P_{\lambda^\varepsilon}\left(\tau_n^\varepsilon \le T^\varepsilon\right) \text{ and } \mathfrak{R}_2 \doteq \sum_{n=(1-2\gamma^\varepsilon)\Gamma^\varepsilon}^{(1+2\gamma^\varepsilon)\Gamma^\varepsilon} (2n+1)P_{\lambda^\varepsilon}\left(\tau_n^\varepsilon \le T^\varepsilon\right)$$

via the Berry–Esseen bound.

We first recall that $\Gamma^\varepsilon = T^\varepsilon/E_{\lambda^\varepsilon}^\varepsilon\tau_1^\varepsilon$. The following is Theorem 1 in [11, Chapter XVI.5].

Theorem 8.13 *(Berry–Esseen) Let $\{X_n\}_{n \in \mathbb{N}}$ be independent real-valued random variables with a common distribution such that*

$$E(X_1) = 0, \ \sigma^2 \doteq E(X_1)^2 > 0, \ \rho \doteq E|X_1|^3 < \infty.$$

Then, for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$\left| P \left(\frac{X_1 + \cdots + X_n}{\sigma \sqrt{n}} \leq x \right) - \Phi(x) \right| \leq \frac{3\rho}{\sigma^3 \sqrt{n}},$$

where $\Phi(\cdot)$ is the distribution function of $N(0, 1)$.

Corollary 8.14 *For any $\varepsilon > 0$, let $\{X_n^\varepsilon\}_{n \in \mathbb{N}}$ be independent real-valued random variables with a common distribution such that*

$$E\left(X_1^\varepsilon\right) = 0, \quad \left(\sigma^\varepsilon\right)^2 \doteq E\left(X_1^\varepsilon\right)^2 > 0, \quad \rho^\varepsilon \doteq E\left|X_1^\varepsilon\right|^3 < \infty.$$

Assume that there exists $\varepsilon_0 \in (0, 1)$ such that

$$\hat{\rho} \doteq \sup_{\varepsilon \in (0, \varepsilon_0)} \rho^\varepsilon < \infty \text{ and } \hat{\sigma}^2 \doteq \inf_{\varepsilon \in (0, \varepsilon_0)} \left(\sigma^\varepsilon\right)^2 > 0.$$

Then for all $x \in \mathbb{R}$, $n \in \mathbb{N}$ and $\varepsilon \in (0, \varepsilon_0)$,

$$\left| P \left(\frac{X_1^\varepsilon + \cdots + X_n^\varepsilon}{\sigma^\varepsilon \sqrt{n}} \leq x \right) - \Phi(x) \right| \leq \frac{3\rho^\varepsilon}{(\sigma^\varepsilon)^3 \sqrt{n}} \leq \frac{3\hat{\rho}}{\hat{\sigma}^3 \sqrt{n}}.$$

Lemma 8.15 *Given any $\delta > 0$ and any $\ell > 0$, there exists $\varepsilon_0 \in (0, 1)$ such that for any $\varepsilon \in (0, \varepsilon_0)$ and $k \in \mathbb{N}_0$, $0 \leq k \leq 2\gamma^\varepsilon \Gamma^\varepsilon$*

$$P_{\lambda^\varepsilon} \left(\tau_{\Gamma^\varepsilon + k}^\varepsilon \leq T^\varepsilon \right) \leq 1 - \Phi \left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon + k}} \right) + \frac{3\hat{\rho}}{\hat{\sigma}^3 \sqrt{\Gamma^\varepsilon + k}}$$

and

$$P_{\lambda^\varepsilon} \left(\tau_{\Gamma^\varepsilon - k}^\varepsilon \leq T^\varepsilon \right) \leq \Phi \left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon - k}} \right) + \frac{3\hat{\rho}}{\hat{\sigma}^3 \sqrt{\Gamma^\varepsilon - k}},$$

where $(\sigma^\varepsilon)^2 \doteq E_{\lambda^\varepsilon}\left(\mathfrak{X}_1^\varepsilon\right)^2$, $\hat{\rho} \doteq \sup_{\varepsilon \in (0, \varepsilon_0)} E_{\lambda^\varepsilon}\left|\mathfrak{X}_1^\varepsilon\right|^3 < \infty$ and $\hat{\sigma}^2 \doteq \inf_{\varepsilon \in (0, \varepsilon_0)}(\sigma^\varepsilon)^2 > 0$ with $\mathfrak{X}_1^\varepsilon \doteq \tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon - 1$.

Proof For any $n \in \mathbb{N}$, we define $\mathfrak{X}_n^\varepsilon \doteq \mathcal{Z}_n^\varepsilon - E_{\lambda^\varepsilon} \mathcal{Z}_1^\varepsilon$ with $\mathcal{Z}_n^\varepsilon \doteq (\tau_n^\varepsilon - \tau_{n-1}^\varepsilon)/E_{\lambda^\varepsilon}\tau_1^\varepsilon$. Obviously, $E_{\lambda^\varepsilon} \mathcal{Z}_n^\varepsilon = 1$ and $E_{\lambda^\varepsilon} \mathfrak{X}_n^\varepsilon = 0$ and if we apply Lemma 8.8, then we find that there exists some $\varepsilon_0 \in (0, 1)$ such that $\sup_{\varepsilon \in (0, \varepsilon_0)} E_{\lambda^\varepsilon}\left(\mathcal{Z}_1^\varepsilon\right)^3 < \infty$ and $\inf_{\varepsilon \in (0, \varepsilon_0)} \text{Var}_{\lambda^\varepsilon}(\mathcal{Z}_1^\varepsilon) > 0$. Since $\mathcal{Z}_1^\varepsilon \geq 0$, Jensen's inequality implies $\left(E_{\lambda^\varepsilon} \mathcal{Z}_1^\varepsilon\right)^3 \leq E_{\lambda^\varepsilon}\left(\mathcal{Z}_1^\varepsilon\right)^3$, and therefore

$$\hat{\rho} \leq 4 \sup_{\varepsilon \in (0, \varepsilon_0)} \left(E_{\lambda^\varepsilon}\left(\mathcal{Z}_1^\varepsilon\right)^3 + \left(E_{\lambda^\varepsilon} \mathcal{Z}_1^\varepsilon\right)^3 \right) \leq 8 \sup_{\varepsilon \in (0, \varepsilon_0)} E_{\lambda^\varepsilon}\left(\mathcal{Z}_1^\varepsilon\right)^3 < \infty,$$

and

$$\hat{\sigma}^2 = \inf_{\varepsilon \in (0, \varepsilon_0)} E_{\lambda^\varepsilon}\left(\mathfrak{X}_1^\varepsilon\right)^2 = \inf_{\varepsilon \in (0, \varepsilon_0)} \text{Var}_{\lambda^\varepsilon}\left(\mathcal{Z}_1^\varepsilon\right) > 0.$$

Therefore, we can use Corollary 8.14 with the iid sequence $\left\{\mathcal{X}_n^\varepsilon\right\}_{n\in\mathbb{N}}$ to find that for any $k \in \mathbb{N}_0$ and $0 \leq k \leq 2\gamma^\varepsilon\Gamma^\varepsilon$

$$
\begin{aligned}
P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon+k}^\varepsilon \leq T^\varepsilon\right) &= P_{\lambda^\varepsilon}\left(\mathcal{Z}_1^\varepsilon + \cdots + \mathcal{Z}_{\Gamma^\varepsilon+k}^\varepsilon \leq \Gamma^\varepsilon\right) \\
&= P_{\lambda^\varepsilon}\left(\frac{\mathcal{X}_1^\varepsilon + \cdots + \mathcal{X}_{\Gamma^\varepsilon+k}^\varepsilon}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}} \leq \frac{-k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right) \\
&\leq 1 - \Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right) + \frac{3\hat{\rho}}{\hat{\sigma}^3\sqrt{\Gamma^\varepsilon+k}},
\end{aligned}
$$

and similarly

$$
P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon-k}^\varepsilon \leq T^\varepsilon\right) \leq \Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}}\right) + \frac{3\hat{\rho}}{\hat{\sigma}^3\sqrt{\Gamma^\varepsilon-k}}.
$$

$\square$

Lemma 8.16 *Given any* $\delta > 0$ *and any* $\ell \in (0, 1/2)$, *there exists* $\varepsilon_0 \in (0, 1)$ *such that for any* $\varepsilon \in (0, \varepsilon_0)$, $\mathfrak{P}_2 \leq C\left(\Gamma^\varepsilon\right)^{\frac{1}{2}-\ell} + 2\left(\Gamma^\varepsilon\right)^{1-\ell}$.

Proof We rewrite $\mathfrak{P}_2$ as

$$
\mathfrak{P}_2 = \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon} P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon-k}^\varepsilon \leq T^\varepsilon\right) + P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon}^\varepsilon \leq T^\varepsilon\right) + \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon} P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon+k}^\varepsilon \leq T^\varepsilon\right).
$$

Then, we use the upper bounds from Lemma 8.15 to get

$$
\begin{aligned}
\mathfrak{P}_2 &\leq \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\left[\Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}}\right) + \frac{3\hat{\rho}}{\hat{\sigma}^3\sqrt{\Gamma^\varepsilon-k}}\right] \\
&\quad + 1 + \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\left[1 - \Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right) + \frac{3\hat{\rho}}{\hat{\sigma}^3\sqrt{\Gamma^\varepsilon+k}}\right] \\
&\leq \frac{24\hat{\rho}}{\hat{\sigma}^3}\gamma^\varepsilon\sqrt{\Gamma^\varepsilon} + 1 + 2\gamma^\varepsilon\Gamma^\varepsilon + \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\left[\Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}}\right) - \Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right)\right].
\end{aligned}
$$

The sum of the first three terms is easily bounded above by $C\left(\Gamma^\varepsilon\right)^{\frac{1}{2}-\ell} + 2\left(\Gamma^\varepsilon\right)^{1-\ell}$. We will show that the last term is bounded above by a constant to complete the proof.

To prove this, we observe that for any $k \leq 2\gamma^\varepsilon\Gamma^\varepsilon$, we may assume $k \leq \Gamma^\varepsilon/2$ by taking ε sufficiently small. Then, we apply the Mean Value Theorem and find

$$
\begin{aligned}
&\left|\Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}}\right) - \Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right)\right| \\
&\quad \leq \sup_{x\in\left[\frac{\sqrt{2/3}k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon}}, \frac{\sqrt{2}k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon}}\right]} \phi(x) \cdot \left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}} - \frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right),
\end{aligned}
$$

where $\phi(x) \doteq e^{-\frac{x^2}{2}}/\sqrt{2\pi}$ and since $0 \le k \le \Gamma^\varepsilon/2$, we have

$$\left[\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}, \frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}}\right] \subset \left[\frac{\sqrt{2/3}k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon}}, \frac{\sqrt{2}k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon}}\right].$$

Additionally, because $\phi(x) = e^{-\frac{x^2}{2}}/\sqrt{2\pi}$ is a monotone decreasing function on $[0, \infty)$, we find that

$$x \in [(\sqrt{2/3}k)(\sigma^\varepsilon\sqrt{\Gamma^\varepsilon}), (\sqrt{2}k)(\sigma^\varepsilon\sqrt{\Gamma^\varepsilon})] \quad \text{implies} \quad \phi(x) \le e^{-\frac{k^2}{3(\sigma^\varepsilon)^2\Gamma^\varepsilon}}/\sqrt{2\pi}.$$

Also, $\sqrt{1+x} - \sqrt{1-x} \le 2x$ for all $x \in [0, 1]$ and $k \le \Gamma^\varepsilon/2$ and a little algebra give $k/\sqrt{\Gamma^\varepsilon-k} - k/\sqrt{\Gamma^\varepsilon+k} \le 4k^2/\Gamma^\varepsilon\sqrt{\Gamma^\varepsilon}$. Therefore, we find

$$\sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\left[\Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}}\right) - \Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right)\right]$$

$$\le \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\frac{1}{\sqrt{2\pi}}e^{-\frac{k^2}{3(\sigma^\varepsilon)^2\Gamma^\varepsilon}}\frac{4k^2}{\sigma^\varepsilon\Gamma^\varepsilon\sqrt{\Gamma^\varepsilon}} \le \frac{4}{\sigma^\varepsilon\Gamma^\varepsilon}\sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\int_{k-1}^k \frac{(1+x)^2}{\sqrt{2\pi\Gamma^\varepsilon}}e^{-\frac{x^2}{3(\sigma^\varepsilon)^2\Gamma^\varepsilon}}\,dx$$

$$\le \frac{4}{\Gamma^\varepsilon}\sqrt{\frac{3}{2}}\int_0^\infty \frac{(1+x)^2}{\sqrt{3\pi\,(\sigma^\varepsilon)^2\,\Gamma^\varepsilon}}e^{-\frac{x^2}{3(\sigma^\varepsilon)^2\Gamma^\varepsilon}}\,dx \le \frac{6}{\Gamma^\varepsilon}E\left(1+X^+\right)^2,$$

where $X \sim N(0, 3\,(\sigma^\varepsilon)^2\,\Gamma^\varepsilon/2)$. Finally, since $E\left(1+X^+\right)^2 \le 2 + 2E\left(X^2\right) = 2 + 3\,(\sigma^\varepsilon)^2\,\Gamma^\varepsilon$, this implies that

$$\sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\left[\Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon-k}}\right) - \Phi\left(\frac{k}{\sigma^\varepsilon\sqrt{\Gamma^\varepsilon+k}}\right)\right] \le \frac{6}{\Gamma^\varepsilon}\left(2 + 3\,(\sigma^\varepsilon)^2\,\Gamma^\varepsilon\right) \le 12 + 18\hat{\rho}^{2/3},$$

$$(8.4)$$

where the last inequality is from

$$\sup_{\varepsilon\in(0,\varepsilon_0)}\sigma^\varepsilon = \sup_{\varepsilon\in(0,\varepsilon_0)}\left(E_{\lambda^\varepsilon}(\mathfrak{X}_1^\varepsilon)^2\right)^{1/2} \le \sup_{\varepsilon\in(0,\varepsilon_0)}(E_{\lambda^\varepsilon}\left|\mathfrak{X}_1^\varepsilon\right|^3)^{1/3} = \hat{\rho}^{1/3}.$$

Since according to Lemma 8.15 $\hat{\rho}^{1/3}$ is finite, we are done. □

Lemma 8.17 *Given any $\delta > 0$ and any $\ell \in (0, 1/2)$, there exists $\varepsilon_0 \in (0, 1)$ and a constant $C < \infty$ such that for any $\varepsilon \in (0, \varepsilon_0)$, $\mathfrak{R}_2 \le 4\,(\Gamma^\varepsilon)^{2-\ell} + C\,(\Gamma^\varepsilon)^{2(1-\ell)}$.*

Proof The proof of this lemma is similar to the proof of Lemma 8.16. We rewrite $\mathfrak{R}_2$ as

$$\mathfrak{R}_2 = \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\left(2\Gamma^\varepsilon - 2k + 1\right)P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon-k}^\varepsilon \le T^\varepsilon\right) + \left(2\Gamma^\varepsilon + 1\right)P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon}^\varepsilon \le T^\varepsilon\right)$$

$$+ \sum_{k=1}^{2\gamma^\varepsilon\Gamma^\varepsilon}\left(2\Gamma^\varepsilon + 2k + 1\right)P_{\lambda^\varepsilon}\left(\tau_{\Gamma^\varepsilon+k}^\varepsilon \le T^\varepsilon\right).$$

Then, we use the upper bounds from Lemma 8.15 to get

$$
\mathfrak{R}_2 \le \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(2\Gamma^\varepsilon - 2k + 1\right) \left[\Phi\left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon - k}} \right) + \frac{3\hat\rho}{\hat\sigma^3 \sqrt{\Gamma^\varepsilon - k}} \right] + \left(2\Gamma^\varepsilon + 1\right)
$$
$$
+ \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(2\Gamma^\varepsilon + 2k + 1\right) \left[1 - \Phi\left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon + k}} \right) + \frac{3\hat\rho}{\hat\sigma^3 \sqrt{\Gamma^\varepsilon + k}} \right].
$$

The next thing is to pair all the terms carefully and bound these pairs separately. We start with

$$
\sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(2\Gamma^\varepsilon - 2k + 1\right) \Phi\left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon - k}} \right) - \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(2\Gamma^\varepsilon + 2k + 1\right) \Phi\left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon + k}} \right)
$$
$$
\le \left(2\Gamma^\varepsilon + 1\right) \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left[\Phi\left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon - k}} \right) - \Phi\left(\frac{k}{\sigma^\varepsilon \sqrt{\Gamma^\varepsilon + k}} \right) \right] \le C\Gamma^\varepsilon.
$$

We use (8.4) for the last inequality. The second pair is

$$
\sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(2\Gamma^\varepsilon - 2k + 1\right) \frac{3\hat\rho}{\hat\sigma^3 \sqrt{\Gamma^\varepsilon - k}} + \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(2\Gamma^\varepsilon + 2k + 1\right) \frac{3\hat\rho}{\hat\sigma^3 \sqrt{\Gamma^\varepsilon + k}}
$$
$$
= \frac{6\hat\rho}{\hat\sigma^3} \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(\sqrt{\Gamma^\varepsilon - k} + \sqrt{\Gamma^\varepsilon + k} \right) + \frac{3\hat\rho}{\hat\sigma^3} \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(\frac{1}{\sqrt{\Gamma^\varepsilon - k}} + \frac{1}{\sqrt{\Gamma^\varepsilon + k}} \right)
$$
$$
\le \frac{6\hat\rho}{\hat\sigma^3} \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} 2\sqrt{2\Gamma^\varepsilon} + \frac{3\hat\rho}{\hat\sigma^3} \sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} 2 \le C\gamma^\varepsilon \Gamma^\varepsilon \sqrt{\Gamma^\varepsilon} + C\gamma^\varepsilon \Gamma^\varepsilon \le C\left(\Gamma^\varepsilon\right)^{\frac{3}{2}-\ell},
$$

where the first inequality holds due to $k \le \Gamma^\varepsilon/2$. The third term is

$$
\sum_{k=1}^{2\gamma^\varepsilon \Gamma^\varepsilon} \left(2\Gamma^\varepsilon + 2k + 1\right) + \left(2\Gamma^\varepsilon + 1\right)
$$
$$
= 4\gamma^\varepsilon \left(\Gamma^\varepsilon\right)^2 + 2\gamma^\varepsilon \Gamma^\varepsilon + 4\left(\gamma^\varepsilon \Gamma^\varepsilon\right)^2 + 2\gamma^\varepsilon \Gamma^\varepsilon + \left(2\Gamma^\varepsilon + 1\right)
$$
$$
\le 4\gamma^\varepsilon \left(\Gamma^\varepsilon\right)^2 + C\left(\gamma^\varepsilon \Gamma^\varepsilon\right)^2 = 4\left(\Gamma^\varepsilon\right)^{2-\ell} + C\left(\Gamma^\varepsilon\right)^{2(1-\ell)},
$$

where the inequality holds since for $\ell \in (0, 1/2)$, $2 - 2\ell \ge 1$ and this implies that $(2\Gamma^\varepsilon + 1) \le C\left(\gamma^\varepsilon \Gamma^\varepsilon\right)^2$. Lastly, combining all the pairs and the corresponding upper bounds, we find that for any $\ell \in (0, 1/2)$,

$$
\mathfrak{R}_2 \le [4\left(\Gamma^\varepsilon\right)^{2-\ell} + C\left(\Gamma^\varepsilon\right)^{2(1-\ell)}] + C\Gamma^\varepsilon + C\left(\Gamma^\varepsilon\right)^{\frac{3}{2}-\ell} \le 4\left(\Gamma^\varepsilon\right)^{2-\ell} + C\left(\Gamma^\varepsilon\right)^{2(1-\ell)},
$$

where C is a constant which depends on ℓ only (and in particular is independent of ε). $\qquad\square$

8.3 Asymptotics of Moments of $N^\varepsilon(T^\varepsilon)$

In this subsection, we prove Lemmas 8.2 and 8.4.

Proof of Lemma 8.2 First, recall that

$$E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right) = \sum\nolimits_{n=0}^{\infty} P_{\lambda^\varepsilon}\left(\tau_n^\varepsilon \leq T^\varepsilon\right) = \mathfrak{P}_1 + \mathfrak{P}_2 + \mathfrak{P}_3,$$

where the $\mathfrak{P}_i$ are defined in (8.2). We can simply bound $\mathfrak{P}_3$ from above by $(1 - 2\gamma^\varepsilon)\Gamma^\varepsilon$. Applying Lemma 8.11 and Lemma 8.16 for the other terms, we have for any $\ell \in (0, 1/2)$ that

$$E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right) \leq C\left(\Gamma^\varepsilon\right)^{2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}} + (C\left(\Gamma^\varepsilon\right)^{\frac{1}{2}-\ell} + 2\left(\Gamma^\varepsilon\right)^{1-\ell}) + \left(1 - 2\gamma^\varepsilon\right)\Gamma^\varepsilon$$

$$= T^\varepsilon / E_{\lambda^\varepsilon}\tau_1^\varepsilon + C\left(\Gamma^\varepsilon\right)^{\frac{1}{2}-\ell} + C\left(\Gamma^\varepsilon\right)^{2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}}.$$

On the other hand, from the definition of $N^\varepsilon\left(T^\varepsilon\right)$, $E_{\lambda^\varepsilon}\tau_{N^\varepsilon(T^\varepsilon)}^\varepsilon \geq T^\varepsilon$. Using Wald's first identity, we find

$$E_{\lambda^\varepsilon}\tau_{N^\varepsilon(T^\varepsilon)}^\varepsilon = E_{\lambda^\varepsilon}\sum\nolimits_{n=1}^{N^\varepsilon(T^\varepsilon)}\left(\tau_n^\varepsilon - \tau_{n-1}^\varepsilon\right) = E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right) \cdot E_{\lambda^\varepsilon}\tau_1^\varepsilon.$$

Hence,

$$0 \leq \frac{E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right)}{T^\varepsilon} - \frac{1}{E_{\lambda^\varepsilon}\tau_1^\varepsilon} \leq \frac{1}{T^\varepsilon}[C\left(\Gamma^\varepsilon\right)^{\frac{1}{2}-\ell} + C\left(\Gamma^\varepsilon\right)^{2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}}].$$

Therefore,

$$\liminf_{\varepsilon\to 0} -\varepsilon\log\left|\frac{E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right)}{T^\varepsilon} - \frac{1}{E_{\lambda^\varepsilon}\tau_1^\varepsilon}\right|$$

$$\geq \liminf_{\varepsilon\to 0} -\varepsilon\log\left[\frac{1}{T^\varepsilon}\left(C\left(\Gamma^\varepsilon\right)^{\frac{1}{2}-\ell} + \left(\Gamma^\varepsilon\right)^{2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}}\right)\right].$$

It remains to find an appropriate lower bound for the liminf.

We use (7.2), Lemma 8.7 and Remark 8.12 to find that for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$ and any $\ell \in (0, 1/2)$

$$\liminf_{\varepsilon\to 0} -\varepsilon\log\left[\frac{1}{T^\varepsilon}\left(C\left(\Gamma^\varepsilon\right)^{\frac{1}{2}-\ell} + \left(\Gamma^\varepsilon\right)^{2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}}\right)\right] \geq \liminf_{\varepsilon\to 0}\varepsilon\log T^\varepsilon$$

$$+ \min\left\{\liminf_{\varepsilon\to 0} -\varepsilon\log\left(\Gamma^\varepsilon\right)^{1/2-\ell}, \liminf_{\varepsilon\to 0} -\varepsilon\log\left(\left(\Gamma^\varepsilon\right)^{2\ell} e^{-(\Gamma^\varepsilon)^{1-2\ell}}\right)\right\}$$

$$\geq c + \min\left\{(1/2 - \ell)\left(h_1 - c - \eta\right), \infty\right\} = c + (1/2 - \ell)\left(h_1 - c - \eta\right).$$

We complete the proof by sending ℓ to $1/2$. $\qquad\square$

Proof of Lemma 8.4 Recall that

$$E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right)^2 = \sum\nolimits_{n=0}^{\infty}(2n + 1) P_{\lambda^\varepsilon}\left(\tau_n^\varepsilon \leq t\right) = \mathfrak{R}_1 + \mathfrak{R}_2 + \mathfrak{R}_3$$

where the $\mathfrak{R}_i$ are defined in (8.3). We can bound $\mathfrak{R}_3$ from above by

$$\sum_{n=0}^{(1-2\gamma^\varepsilon)\Gamma^\varepsilon-1} (2n+1) = \left(1 - 4\gamma^\varepsilon + 4\left(\gamma^\varepsilon\right)^2\right)\left(\Gamma^\varepsilon\right)^2.$$

Applying Lemma 8.11 and Lemma 8.17, we have for any $\ell \in (0, 1/2)$ that

$$
\begin{aligned}
E_{\lambda^\varepsilon}(N^\varepsilon\left(T^\varepsilon\right))^2 &\leq C\left(\Gamma^\varepsilon\right)^{1+2\ell} e^{-\left(\Gamma^\varepsilon\right)^{1-2\ell}} \\
&\quad + 4\left(\Gamma^\varepsilon\right)^{2-\ell} + C\left(\Gamma^\varepsilon\right)^{2(1-\ell)} + \left(1 - 4\gamma^\varepsilon + 4\left(\gamma^\varepsilon\right)^2\right)\left(\Gamma^\varepsilon\right)^2 \\
&\leq \left(\Gamma^\varepsilon\right)^2 + C\left(\Gamma^\varepsilon\right)^{2(1-\ell)} + C\left(\Gamma^\varepsilon\right)^{1+2\ell} e^{-\left(\Gamma^\varepsilon\right)^{1-2\ell}}.
\end{aligned}
$$

As in the proof of Lemma 8.2 $E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right) \geq \Gamma^\varepsilon$. Thus, for any $\ell \in (0, 1/2)$

$$
\begin{aligned}
\mathrm{Var}_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right) &\leq E_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right)^2 - \left(\Gamma^\varepsilon\right)^2 \\
&\leq [\left(\Gamma^\varepsilon\right)^2 + C\left(\Gamma^\varepsilon\right)^{2(1-\ell)} + C\left(\Gamma^\varepsilon\right)^{1+2\ell} e^{-\left(\Gamma^\varepsilon\right)^{1-2\ell}}] - \left(\Gamma^\varepsilon\right)^2 \\
&= C\left(\Gamma^\varepsilon\right)^{2(1-\ell)} + C\left(\Gamma^\varepsilon\right)^{1+2\ell} e^{-\left(\Gamma^\varepsilon\right)^{1-2\ell}}.
\end{aligned}
$$

Again we use (7.2), Lemma 8.7 and Remark 8.12 to find that for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$ and for any $\ell \in (0, 1/2)$,

$$
\begin{aligned}
&\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{\mathrm{Var}_{\lambda^\varepsilon}\left(N^\varepsilon\left(T^\varepsilon\right)\right)}{T^\varepsilon} \\
&\geq \liminf_{\varepsilon \to 0} \varepsilon \log T^\varepsilon + \min\left\{\liminf_{\varepsilon \to 0} -\varepsilon \log\left(\Gamma^\varepsilon\right)^{2(1-\ell)}, \liminf_{\varepsilon \to 0} -\varepsilon \log\left(\left(\Gamma^\varepsilon\right)^{1+2\ell} e^{-\left(\Gamma^\varepsilon\right)^{1-2\ell}}\right)\right\} \\
&\geq c + \min\left\{2\left(1-\ell\right)\left(h_1 - c - \eta\right), \infty\right\} = 2\left(1-\ell\right)\left(h_1 - \eta\right) + \left(2\ell - 1\right)c.
\end{aligned}
$$

We complete the proof by sending ℓ to $1/2$. $\square$

8.4 Asymptotics of Moments of $\hat{N}^\varepsilon(T^\varepsilon)$

The proof of the following result is given in Section 10.

Theorem 8.18 *If $w \geq h_1$, then given any $m > 0$ such that $m + h_1 > w$ and for any $\delta > 0$ sufficiently small,*

$$\lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \hat{\tau}_1^\varepsilon = m + \varkappa_\delta \text{ and } \hat{\tau}_1^\varepsilon / E_{\lambda^\varepsilon} \hat{\tau}_1^\varepsilon \overset{d}{\to} \mathrm{Exp}(1).$$

Moreover, there exists $\varepsilon_0 \in (0, 1)$ and a constant $\tilde{c} > 0$ such that

$$P_{\lambda^\varepsilon}\left(\hat{\tau}_1^\varepsilon / E_{\lambda^\varepsilon} \hat{\tau}_1^\varepsilon > t\right) \leq e^{-\tilde{c}t}$$

for any $t > 0$ and any $\varepsilon \in (0, \varepsilon_0)$.

Notice that Theorem 8.18 is a multicycle version of Theorem 8.5, which is the key to the proofs of the asymptotics of moments of $N^\varepsilon(T^\varepsilon)$, namely, Lemma 8.2 and Lemma 8.4. Given Theorem 8.18, the proofs of the following analogous results follow from essentially the same arguments as those of Lemma 8.2 and Lemma 8.4.

Lemma 8.19 *Suppose that $w \geq h_1$, $m + h_1 > w$, and that $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > w$. Then, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left| \frac{E_{\lambda^\varepsilon}(\hat{N}^\varepsilon(T^\varepsilon))}{T^\varepsilon} - \frac{1}{E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right| \geq c.$$

Corollary 8.20 *Suppose that $w \geq h_1$, $m + h_1 > w$ and that $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > w$. Then, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon}(\hat{N}^\varepsilon(T^\varepsilon))}{T^\varepsilon} \geq m + \varkappa_\delta.$$

Lemma 8.21 *Suppose that $w \geq h_1$, $m + h_1 > w$ and that $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > w$. Then, for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{\mathrm{Var}_{\lambda^\varepsilon}(\hat{N}^\varepsilon(T^\varepsilon))}{T^\varepsilon} \geq m + h_1 - \eta.$$

9 Large Deviation Type Upper Bounds

In this section, we collect results from the previous sections to prove the main results of the paper, Theorems 4.3 and 4.5 , which give large deviation upper bounds on the bias under the empirical measure and the variance per unit time. We also give the proof of Theorem 4.9, which shows how to simplify some expressions appearing in the large deviation bounds. Before giving the proof of the first result, we establish Lemmas 9.1 and 9.2 for the single cycle case, and Lemmas 9.3 and 9.4 for the multicycle case, which are needed in the proof of Theorem 4.3. Recall that for any $n \in \mathbb{N}$

$$S_n^\varepsilon \doteq \int_{\tau_{n-1}^\varepsilon}^{\tau_n^\varepsilon} e^{-\frac{1}{\varepsilon}f(X_t^\varepsilon)} 1_A(X_t^\varepsilon)\,dt. \tag{9.1}$$

Lemma 9.1 *If $h_1 > w$, $A \subset M$ is compact and $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > h_1$, then for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left| \frac{E_{\lambda^\varepsilon}N^\varepsilon(T^\varepsilon)}{T^\varepsilon} E_{\lambda^\varepsilon}S_1^\varepsilon - \int_M e^{-\frac{1}{\varepsilon}f(x)} 1_A(x)\,\mu^\varepsilon(dx) \right|$$
$$\geq \inf_{x \in A}[f(x) + W(x)] - W(O_1) + c - h_1 - \eta.$$

Proof To begin, by Lemma 4.1 with $g(x) = e^{-\frac{1}{\varepsilon}f(x)}1_A(x)$, we know that for any δ sufficiently small and $\varepsilon > 0$,

$$E_{\lambda^\varepsilon} S_1^\varepsilon = E_{\lambda^\varepsilon} \left(\int_0^{\tau_1^\varepsilon} e^{-\frac{1}{\varepsilon}f(X_s^\varepsilon)}1_A\left(X_s^\varepsilon\right) ds \right) = E_{\lambda^\varepsilon} \tau_1^\varepsilon \cdot \int_M e^{-\frac{1}{\varepsilon}f(x)}1_A(x)\,\mu^\varepsilon(dx).$$

This implies that

$$\left| \frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} E_{\lambda^\varepsilon} S_1^\varepsilon - \int_M e^{-\frac{1}{\varepsilon}f(x)}1_A(x)\,\mu^\varepsilon(dx) \right|$$

$$= E_{\lambda^\varepsilon} S_1^\varepsilon \cdot \left| \frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} - \frac{1}{E_{\lambda^\varepsilon}\tau_1^\varepsilon} \right|.$$

Hence, by (7.1), Lemmas 7.21 and 8.2, we find that given $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left| \frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} E_{\lambda^\varepsilon} S_1^\varepsilon - \int_M e^{-\frac{1}{\varepsilon}f(x)}1_A(x)\,\mu^\varepsilon(dx) \right|$$

$$\geq \liminf_{\varepsilon \to 0} -\varepsilon \log E_{\lambda^\varepsilon} S_1^\varepsilon + \liminf_{\varepsilon \to 0} -\varepsilon \log \left| \frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} - \frac{1}{E_{\lambda^\varepsilon}\tau_1^\varepsilon} \right|$$

$$\geq \inf_{x \in A} [f(x) + W(x)] - W(O_1) + c - h_1 - \eta.$$

$\square$

In the application of Wald's identity, a difficulty arises in that, owing to the randomness of $N^\varepsilon(T^\varepsilon)$, $S_{N^\varepsilon(T^\varepsilon)}^\varepsilon$ need not have the same distribution as S_1^ε. Nevertheless, such minor term can be dealt with by using technique in, for example, [18, Theorem 3.16]. The proof of the following lemma can be found in Appendix.

Lemma 9.2 *If $h_1 > w$, $A \subset M$ is compact and $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > h_1$, then for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon} S_{N^\varepsilon(T^\varepsilon)}^\varepsilon}{T^\varepsilon} \geq \inf_{x \in A} [f(x) + W(x)] - W(O_1) + c - h_1 - \eta.$$

We have similar results for multicycles. To be specific, we have the following two lemmas.

Lemma 9.3 *Suppose that $w \geq h_1$, $m + h_1 > w$, $A \subset M$ is compact and that $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > w$. Then, for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left| \frac{E_{\lambda^\varepsilon} \hat{N}^\varepsilon(T^\varepsilon)}{T^\varepsilon} E_{\lambda^\varepsilon} \hat{S}_1^\varepsilon - \int_M e^{-\frac{1}{\varepsilon}f(x)}1_A(x)\,\mu^\varepsilon(dx) \right|$$

$$\geq \inf_{x \in A} [f(x) + W(x)] - W(O_1) + c - (m + h_1) - \eta.$$

Lemma 9.4 *Suppose that $w \geq h_1$, $m + h_1 > w$, $A \subset M$ is compact and that $T^\varepsilon = e^{\frac{1}{\varepsilon}c}$ for some $c > w$. Then, for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$*

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon} \hat{S}^\varepsilon_{\hat{N}^\varepsilon(T^\varepsilon)}}{T^\varepsilon} \geq \inf_{x \in A} [f(x) + W(x)] - W(O_1) + c - (m + h_1) - \eta.$$

Proof of Theorem 4.3 If $h_1 > w$, then recall that

$$\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)-1} S^\varepsilon_n \leq \frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X^\varepsilon_t)} 1_A (X^\varepsilon_t) \, dt \leq \frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S^\varepsilon_n,$$

where S^ε_n is defined in (9.1). Then, we apply Wald's first identity to obtain

$$
\begin{aligned}
E_{\lambda^\varepsilon} \left(\sum_{n=1}^{N^\varepsilon(T^\varepsilon)-1} S^\varepsilon_n \right) &= E_{\lambda^\varepsilon} \left(\sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S^\varepsilon_n \right) - E_{\lambda^\varepsilon} S^\varepsilon_{N^\varepsilon(T^\varepsilon)} \\
&= E_{\lambda^\varepsilon} \left(N^\varepsilon(T^\varepsilon) \right) E_{\lambda^\varepsilon} S^\varepsilon_1 - E_{\lambda^\varepsilon} S^\varepsilon_{N^\varepsilon(T^\varepsilon)}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\left| E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X^\varepsilon_t)} 1_A (X^\varepsilon_t) \, dt \right) - \int_M e^{-\frac{1}{\varepsilon}f(x)} 1_A(x) \mu^\varepsilon(dx) \right| \\
&\leq \left| \frac{E_{\lambda^\varepsilon} (N^\varepsilon(T^\varepsilon))}{T^\varepsilon} E_{\lambda^\varepsilon} S^\varepsilon_1 - \int_M e^{-\frac{1}{\varepsilon}f(x)} 1_A(x) \mu^\varepsilon(dx) \right| + \frac{E_{\lambda^\varepsilon} S^\varepsilon_{N^\varepsilon(T^\varepsilon)}}{T^\varepsilon}.
\end{aligned}
$$

Therefore, by Lemmas 9.1 and 9.2 we have that for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$,

$$
\begin{aligned}
&\liminf_{\varepsilon \to 0} -\varepsilon \log \left| E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X^\varepsilon_t)} 1_A (X^\varepsilon_t) \, dt \right) - \int_M e^{-\frac{1}{\varepsilon}f(x)} 1_A(x) \mu^\varepsilon(dx) \right| \\
&\geq \inf_{x \in A} [f(x) + W(x)] - W(O_1) + c - h_1 - \eta.
\end{aligned}
$$

The argument for $h_1 \leq w$ is entirely analogous but uses by Lemmas 9.3 and 9.4. □

The following lemma bounds quantities that will arise in the proof of Theorem 4.5. Its proof is given in Appendix.

Lemma 9.5 *Recall the definitions $R^{(2)}_1 \doteq 2\inf_{x \in A} [f(x) + V(O_1, x)] - h_1$, and for $j \in L \setminus \{1\}$, $R^{(2)}_j \doteq 2\inf_{x \in A} [f(x) + V(O_j, x)] + W(O_j) - 2W(O_1) + W(O_1 \cup O_j)$ with $h_1 \doteq \min_{\ell \in L \setminus \{1\}} V(O_1, O_\ell)$. Then, $2\inf_{x \in A} [f(x) + W(x)] - 2W(O_1) - h_1 \geq \min_{j \in L} R^{(2)}_j$.*

Proof of Theorem 4.5 We begin with the observation that is for any random variables X, Y and Z satisfying $0 \leq Y - Z \leq X \leq Y$,

$$\text{Var}(X) = EX^2 - (EX)^2 \leq EY^2 - (E(Y-Z))^2$$
$$= \text{Var}(Y) + 2EY \cdot EZ - (EZ)^2 \leq \text{Var}(Y) + 2EY \cdot EZ.$$

When $h_1 > w$, since

$$0 \leq \frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon - \frac{1}{T^\varepsilon} S_{N^\varepsilon(T^\varepsilon)}^\varepsilon \leq \frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon\right) dt \leq \frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon,$$

we have

$$\text{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon\right) dt \right)$$
$$\leq \text{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) + 2 E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) \frac{E_{\lambda^\varepsilon} S_{N^\varepsilon(T^\varepsilon)}^\varepsilon}{T^\varepsilon},$$

and with the help of (7.2)

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\text{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon}f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon\right) dt \right) T^\varepsilon \right)$$
$$\geq \min \left\{ \liminf_{\varepsilon \to 0} -\varepsilon \log \left[\text{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) T^\varepsilon \right], \right.$$
$$\left. \liminf_{\varepsilon \to 0} -\varepsilon \log \left[E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) \frac{E_{\lambda^\varepsilon} S_{N^\varepsilon(T^\varepsilon)}^\varepsilon}{T^\varepsilon} T^\varepsilon \right] \right\}.$$

We complete the proof in the case of single cycle by showing both terms are bounded below by $\min_{j \in L} (R_j^{(1)} \wedge R_j^{(2)}) - \eta$, where we recall

$$R_j^{(1)} \doteq \inf_{x \in A} \left[2f(x) + V(O_j, x) \right] + W(O_j) - W(O_1),$$
$$R_1^{(2)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_1, x) \right] - h_1,$$

and for $j \in L \setminus \{1\}$

$$R_j^{(2)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) - 2W(O_1) + W(O_1 \cup O_j).$$

For the second term, we apply Wald's first identity, Lemma 7.21, Corollary 8.3 and Lemma 9.2 to find that given $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left[E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) \frac{E_{\lambda^\varepsilon} S_{N^\varepsilon(T^\varepsilon)}^\varepsilon}{T^\varepsilon} T^\varepsilon \right]$$

$$\geq \liminf_{\varepsilon \to 0} -\varepsilon \log T^\varepsilon + \liminf_{\varepsilon \to 0} -\varepsilon \log E_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right)$$

$$+ \liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon} S_{N^\varepsilon(T^\varepsilon)}^\varepsilon}{T^\varepsilon}$$

$$\geq -c + (\inf_{x \in A}[f(x) + W(x)] - W(O_1) - h_1 - \eta/3) + \varkappa_\delta$$

$$+ (\inf_{x \in A}[f(x) + W(x)] - W(O_1) + (c - h_1) - \eta/3)$$

$$\geq 2\inf_{x \in A}[f(x) + W(x)] - 2W(O_1) - h_1 - \eta$$

$$\geq \min_{j \in L} R_j^{(2)} - \eta \geq \min_{j \in L}(R_j^{(1)} \wedge R_j^{(2)}) - \eta.$$

The third inequality holds by choosing δ sufficiently small $h_\delta \geq h_1 - \eta/3$. The fourth inequality is from Lemma 9.5.

Turning to the first term, we can bound the variance by (6.3):

$$\mathrm{Var}_{\lambda^\varepsilon}\left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) T^\varepsilon \leq 2\frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} \mathrm{Var}_{\lambda^\varepsilon} S_1^\varepsilon + 2\frac{\mathrm{Var}_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon}(E_{\lambda^\varepsilon} S_1^\varepsilon)^2$$

$$\leq 2\frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} E_{\lambda^\varepsilon}(S_1^\varepsilon)^2 + 2\frac{\mathrm{Var}_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon}(E_{\lambda^\varepsilon} S_1^\varepsilon)^2.$$

If we use Corollary 8.3 and Lemma 7.23, then we know that given $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left[\frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} E_{\lambda^\varepsilon}(S_1^\varepsilon)^2 \right]$$

$$\geq \liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} + \liminf_{\varepsilon \to 0} -\varepsilon \log E_{\lambda^\varepsilon}(S_1^\varepsilon)^2$$

$$\geq \min_{j \in L}(R_j^{(1)} \wedge R_j^{(2)}) - \eta.$$

In addition, we can apply Lemmas 7.21 and 8.4 to show that given $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left[\frac{\mathrm{Var}_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon}(E_{\lambda^\varepsilon} S_1^\varepsilon)^2 \right]$$

$$\geq \liminf_{\varepsilon \to 0} -\varepsilon \log \frac{\mathrm{Var}_{\lambda^\varepsilon}(N^\varepsilon(T^\varepsilon))}{T^\varepsilon} + 2\liminf_{\varepsilon \to 0} -\varepsilon \log E_{\lambda^\varepsilon} S_1^\varepsilon$$

$$\geq 2\inf_{x \in A}[f(x) + W(x)] - 2W(O_1) - h_1 - \eta$$

$$\geq \min_{j \in L} R_j^{(2)} - \eta \geq \min_{j \in L}(R_j^{(1)} \wedge R_j^{(2)}) - \eta.$$

The second last inequality comes from Lemma 9.5.

Hence, we find that given $\eta > 0$, there exists $\delta_0 \in (0, 1)$, such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\mathrm{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \sum_{n=1}^{N^\varepsilon(T^\varepsilon)} S_n^\varepsilon \right) T^\varepsilon \right) \geq \min_{j \in L}(R_j^{(1)} \wedge R_j^{(2)}) - \eta,$$

and we are done for the single cycle case.

For multicycle case, by using a similar argument and applying Lemmas 7.24, 7.25, 8.21, 9.4 and Corollary 8.20, we find that

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \left(\mathrm{Var}_{\lambda^\varepsilon} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} e^{-\frac{1}{\varepsilon} f(X_t^\varepsilon)} 1_A \left(X_t^\varepsilon \right) dt \right) T^\varepsilon \right)$$

$$\geq \min_{j \in L}(R_j^{(1)} \wedge R_j^{(2)} \wedge R_j^{(3,m)}) - \eta,$$

with

$$R_j^{(3,m)} \doteq 2 \inf_{x \in A} \left[f(x) + V(O_j, x) \right] + 2W(O_j) - 2W(O_1) - (m + h_1).$$

We complete the proof by sending $m \downarrow w - h_1$. $\qquad \square$

Proof of Theorem 4.9 Parts 1, 2 and 3 are from Theorem 4.3, Lemma 4.3 (b) and Theorem 6.1 in [12, Chapter 6], respectively.

We now turn to part 4. Before giving the proof, we state a result from [12]. The result is Lemma 4.3 (c) in [12, Chapter 6], which says that for any unstable equilibrium point O_j, there exists a stable equilibrium point O_i such that $W(O_j) = W(O_i) + V(O_i, O_j)$.

Now, suppose that $\min_{j \in L}(\inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j))$ is attained at some $\ell \in L$ such that O_ℓ is unstable (i.e., $\ell \in L \setminus L_s$). Then, since there exists a stable equilibrium point O_i (i.e., $i \in L_s$) such that $W(O_\ell) = W(O_i) + V(O_i, O_\ell)$ we find

$$\min_{j \in L} \left(\inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) \right)$$

$$= \inf_{x \in A} [f(x) + V(O_\ell, x)] + W(O_\ell) = \inf_{x \in A} [f(x) + V(O_\ell, x)] + V(O_i, O_\ell) + W(O_i)$$

$$\geq \inf_{x \in A} [f(x) + V(O_i, x)] + W(O_i) \geq \min_{j \in L_s} \left(\inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) \right)$$

$$\geq \min_{j \in L} \left(\inf_{x \in A} \left[f(x) + V(O_j, x) \right] + W(O_j) \right).$$

The first inequality is from a dynamic programming inequality. Therefore, the minimum is also attained at $i \in L_s$ and $\min_{j \in L} R_j^{(1)} = \min_{j \in L_s} R_j^{(1)}$. $\qquad \square$

10 Exponential Return Law and Tail Behavior

In this section, we give the proof of Theorem 8.5, which was the key fact needed to obtain bounds on the distribution of $N^\varepsilon(T^\varepsilon)$, and the related multicycle analogy. A

result of this type first appears in [6], which asserts that the time needed to escape from an open subset of the domain of attraction of a stable equilibrium point that contains the equilibrium point has an asymptotically exponential distribution. [6] also proves a nonasymptotic bound on the tail of the probability of escape before a certain time that is also of exponential form. Theorem 8.5 is a more complicated statement, in that it asserts the asymptotically exponential form for the return time to the neighborhood of O_1. To prove this, we build on the results of [6] and decompose the return time into times of transitions between equilibrium points. This in turn will require the proof of a number of related results, such as establishing the independence of certain estimates with respect to initial distributions.

The existence of an exponentially distributed first hitting time is a central topic in the theory of quasistationary distributions. For a recent book length treatment of the topic, we refer to [5]. However, so far as we can tell the types of situations, we encounter are not covered by existing results, and so as noted we develop what is needed using [6] as the starting point. See Remark 3.15.

For any $j \in L$, define υ_j^ε as the hitting time of $\partial B_\delta(O_k)$ for some $k \in L \setminus \{j\}$, i.e.,

$$\upsilon_j^\varepsilon \doteq \inf \left\{ t > 0 : X_t^\varepsilon \in \cup_{k \in L \setminus \{j\}} \partial B_\delta(O_k) \right\}. \tag{10.1}$$

We will prove the following result for first hitting times of another equilibrium point, and later extend to return times.

Lemma 10.1 *For any $j \in L_s$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$ and any distribution λ^ε on $\partial B_\delta(O_j)$,*

$$\lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \upsilon_j^\varepsilon = \min_{y \in \cup_{k \in L \setminus \{j\}} \partial B_\delta(O_k)} V(O_j, y) \ \text{and} \ \upsilon_j^\varepsilon / E_{\lambda^\varepsilon} \upsilon_j^\varepsilon \xrightarrow{d} \mathrm{Exp}(1).$$

Moreover, there exists $\varepsilon_0 \in (0, 1)$ and a constant $\tilde{c} > 0$ such that

$$P_{\lambda^\varepsilon} \left(\upsilon_j^\varepsilon / E_{\lambda^\varepsilon} \upsilon_j^\varepsilon > t \right) \le e^{-\tilde{c}t}$$

for any $t > 0$ and any $\varepsilon \in (0, \varepsilon_0)$.

The organization of this section is as follows. The first part of Lemma 10.1 that is concerned with mean first hitting times is proved in Section 10.1, while the second part that is concerned with an asymptotically exponential distribution but when starting with a special distribution is proved in Section 10.2. The last part of the lemma, which focuses on bounds on the tail of the hitting time of another equilibrium point but when starting with a special distribution, is proved in Section 10.3. We then extend the second and third parts of Lemma 10.1 to general initial distributions in Section 10.4 and Section 10.5. The last two subsections then extend all of Lemma 10.1 to return times for single cycles and multicycles, respectively.

10.1 Mean First Hitting Time

Lemma 10.2 *For any $\delta > 0$ sufficiently small and $x \in \partial B_\delta(O_j)$ with $j \in L_s$*

$$\lim_{\varepsilon \to 0} \varepsilon \log E_x \upsilon_j^\varepsilon = \min_{y \in \cup_{k \in L \setminus \{j\}} \partial B_\delta(O_k)} V(O_j, y). \tag{10.2}$$

Proof For the given $j \in L_s$, let D_j denote the corresponding domain of attraction. We claim there is $k \in L \setminus L_s$ such that

$$q_j \doteq \inf_{y \in \partial D_j} V(O_j, y) = V(O_j, O_k).$$

Since $V(O_j, \cdot)$ is continuous and ∂D_j is compact, there is a point $y^* \in \partial D_j$ such that $q_j = V(O_j, y^*)$. If $y^* \in \cup_{k \in L \setminus L_s} O_k$, then we are done. If this is not true, then since $y^* \notin (\cup_{k \in L_s} D_k) \cup (\cup_{k \in L \setminus L_s} O_k)$, and since the solution to $\dot\phi = b(\phi)$, $\phi(0) = y^*$ must converge to $\cup_{k \in L} O_k$ as $t \to \infty$, it must in fact converge to a point in $\cup_{k \in L \setminus L_s} O_k$, say O_k. Since such trajectories have zero cost, by a standard argument for any $\varepsilon > 0$ we can construct by concatenation a trajectory that connects O_j to O_k in finite time and with cost less than $q_j + \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we have $q_j = V(O_j, O_k)$.

There may be more than one $l \in L \setminus L_s$ such that $O_l \in \partial D_j$ and $q_j = V(O_j, O_l)$, but we can assume that for some $k \in L \setminus L_s$ and $\bar y \in \partial B_\delta(O_k)$ we attain the min in (10.2). Then, $\bar q_j \doteq V(O_j, \bar y) \leq q_j$, and we need to show $\lim_{\varepsilon \to 0} \varepsilon \log E_x \upsilon_j^\varepsilon = \bar q_j$.

Given $s < \bar q_j$, let $D_j(s) = \{x : V(O_j, x) \leq s\}$ and assume s is large enough that $B_\delta(O_j) \subset D_j(s)^\circ$. Then, $D_j(s) \subset D_j^\circ$ is closed and contained in the open set $D_j \setminus \cup_{l \in L \setminus \{j\}} B_\delta(O_l)$, and thus the time to reach $\partial D_j(s)$ is never greater than υ_j^ε. Given $\eta > 0$, we can find a set $D_j^\eta(s)$ that is contained in $D_j(s)$ and satisfies the conditions of [12, Theorem 4.1, Chapter 4], and also $\inf_{z \in \partial D_j^\eta(s)} V(O_j, z) \geq s - \eta$. This theorem gives the equality in the following display:

$$\liminf_{\varepsilon \to 0} \varepsilon \log E_x \upsilon_j^\varepsilon \geq \liminf_{\varepsilon \to 0} \varepsilon \log E_x \inf\{t \geq 0 : X_t^\varepsilon \in \partial D_j^\eta(s)\}$$
$$= \inf_{z \in \partial D_j^\eta(s)} V(O_j, z) \geq s - \eta.$$

Letting $\eta \downarrow 0$ and then $s \uparrow \bar q_j$ gives $\liminf_{\varepsilon \to 0} \varepsilon \log E_x \upsilon_j^\varepsilon \geq \bar q_j$.

For the reverse inequality, we also adapt an argument from the proof of [12, Theorem 4.1, Chapter 4]. One can find $T_1 < \infty$ such that the probability for X_t^ε to reach $\cup_{l \in L} B_\delta(O_l)$ by time T_1 from any $x \in M \setminus \cup_{l \in L} B_\delta(O_l)$ is bounded below by $1/2$. (This follows easily from the law of large numbers and that all trajectories of the noiseless system reach $\cup_{l \in L} B_{\delta/2}(O_l)$ in some finite time that is bounded uniformly in $x \in M \setminus \cup_{l \in L} B_\delta(O_l)$.) Also, given $\eta > 0$ there is $T_2 < \infty$ and $\varepsilon_0 > 0$ such that $P_x\{X_t^\varepsilon \text{ reaches } \cup_{k \in L \setminus \{j\}} \partial B_\delta(O_k) \text{ before } T_2\} \geq \exp -(\bar q_j + \eta)/\varepsilon$ for all $x \in \partial B_\delta(O_j)$. It then follows from the strong Markov property that for any $x \in M \setminus \cup_{l \in L} B_\delta(O_l)$

$$P_x\{\upsilon_j^\varepsilon \leq T_1 + T_2\} \geq e^{-\frac{1}{\varepsilon}(\bar q_j + \eta)}/2.$$

Using the ordinary Markov property, we have

$$
\begin{aligned}
E_x \upsilon_j^\varepsilon &\le \sum_{n=0}^{\infty} (n+1)(T_1 + T_2) P_x \{ n(T_1 + T_2) < \upsilon_j^\varepsilon \le (n+1)(T_1 + T_2) \} \\
&= (T_1 + T_2) \sum_{n=0}^{\infty} P_x \{ \upsilon_j^\varepsilon > n(T_1 + T_2) \} \\
&\le (T_1 + T_2) \sum_{n=0}^{\infty} \left[1 - \inf_{x \in M \setminus \cup_{l \in L} B_\delta(O_l)} P_x \{ \upsilon_j^\varepsilon \le T_1 + T_2 \} \right]^n \\
&= (T_1 + T_2) \left(\inf_{x \in M \setminus \cup_{l \in L} B_\delta(O_l)} P_x \{ \upsilon_j^\varepsilon \le T_1 + T_2 \} \right)^{-1} \\
&\le 2(T_1 + T_2) e^{\frac{1}{\varepsilon}(\bar{q}_j + \eta)}.
\end{aligned}
$$

Thus, $\limsup_{\varepsilon \to 0} \varepsilon \log E_x \upsilon_j^\varepsilon \le \bar{q}_j + \eta$, and letting $\eta \downarrow 0$ completes the proof. $\quad\square$

Remark 10.3 By the standard Freidlin–Wentzell theory, the convergence asserted in Lemma 10.2 is uniform on $\partial B_\delta(O_j)$. Therefore, we have the first part of Lemma 10.1.

10.2 Asymptotically Exponential Distribution

Lemma 10.4 *For each* $j \in L_s$ *there is a distribution* u^ε *on* $\partial B_{2\delta}(O_j)$ *such that* $\upsilon_j^\varepsilon / E_{u^\varepsilon} \upsilon_j^\varepsilon \xrightarrow{d} \mathrm{Exp}(1)$.

Proof To simplify notation and since it plays no role, we write $j = 1$ throughout the proof. We call $\partial B_\delta(O_1)$ and $\partial B_{2\delta}(O_1)$ the inner and outer rings of O_1. We can then decompose the hitting time as

$$
\upsilon_1^\varepsilon = \sum_{k=1}^{\mathcal{N}^\varepsilon - 1} \theta_k^\varepsilon + \zeta^\varepsilon, \tag{10.3}
$$

where θ_k^ε is the k-th amount of time that the process travels from the outer ring to the inner ring and back without visiting $\cup_{j \in L \setminus \{1\}} \partial B_\delta(O_j)$, ζ^ε is the amount of time that the process travels from the outer ring directly to $\cup_{j \in L \setminus \{1\}} \partial B_\delta(O_j)$ without visiting the inner ring, and $\mathcal{N}^\varepsilon - 1$ is the number of times that the process goes back and forth between the inner ring and outer ring. (It is assumed that $\delta > 0$ is small enough that $B_{2\delta}(O_1) \subset M \setminus \cup_{j \in L \setminus \{1\}} B_{2\delta}(O_j)$.) Note that θ_k^ε grows exponentially of the order δ, due to the time taken to travel from the inner ring to the outer ring, and ζ^ε is uniformly bounded in expected value.

For any set A, define the first hitting time by $\tau(A) \doteq \inf \{ t > 0 : X_t^\varepsilon \in A \}$. Consider the conditional transition probability from $x \in \partial B_{2\delta}(O_1)$ to $y \in \partial B_\delta(O_1)$ given by

$$
\psi_1^\varepsilon(dy|x) \doteq P\left(X_{\tau(\partial B_\delta(O_1))}^\varepsilon \in dy | X_0^\varepsilon = x, \ X_t^\varepsilon \notin \cup_{j \in L \setminus \{1\}} \partial B_\delta(O_j), t \in [0, \tau(\partial B_\delta(O_1)))] \right),
$$

and the transition probability from $y \in \partial B_\delta(O_1)$ to $x \in \partial B_{2\delta}(O_1)$ given by

$$
\psi_2^\varepsilon(dx|y) \doteq P\left(X_{\tau(\partial B_{2\delta}(O_1))}^\varepsilon \in dx | X_0^\varepsilon = y \right). \tag{10.4}
$$

Then, we can create a transition probability from $x \in \partial B_{2\delta}(O_1)$ to $y \in \partial B_{2\delta}(O_1)$ by

$$\psi^\varepsilon(dy|x) \doteq \int_{\partial B_\delta(O_1)} \psi_2^\varepsilon(dy|z)\,\psi_1^\varepsilon(dz|x). \tag{10.5}$$

Since $\partial B_{2\delta}(O_1)$ is compact and $\{X_t^\varepsilon\}_t$ is non-degenerate and Feller, there exists an invariant measure $u^\varepsilon \in \mathcal{P}(\partial B_{2\delta}(O_1))$ with respect to the transition probability $\psi^\varepsilon(dy|x)$. If we start with the distribution u^ε on $\partial B_{2\delta}(O_1)$, then it follows from the definition of u^ε and the strong Markov property that the $\{\theta_k^\varepsilon\}_{k<\mathcal{N}^\varepsilon}$ are iid. Moreover, the indicators of escape (i.e., $1_{\{\tau(\cup_{j\in L\setminus\{1\}}\partial B_\delta(O_j))=\tau(\cup_{j\in L}\partial B_\delta(O_j))\}}$) are iid Bernoulli, and we write them as Y_k^ε with $P_{u^\varepsilon}(Y_k^\varepsilon = 1) = e^{-h_1^\varepsilon(\delta)/\varepsilon}$, where $\delta > 0$ is from the construction, $h_1^\varepsilon(\delta) \to h_1(\delta)$ as $\varepsilon \to 0$ and $h_1(\delta) \uparrow h_1$ as $\delta \downarrow 0$ with $h_1 = \min_{j\in L\setminus\{1\}} V(O_1, O_j)$. Note that $\mathcal{N}^\varepsilon = \inf\{k \in \mathbb{N} : Y_k^\varepsilon = 1\}$. We therefore have

$$P_{u^\varepsilon}(\mathcal{N}^\varepsilon = k) = (1 - e^{-h_1^\varepsilon(\delta)/\varepsilon})^{k-1} e^{-h_1^\varepsilon(\delta)/\varepsilon},$$

and thus

$$E_{u^\varepsilon} \upsilon_1^\varepsilon = E_{u^\varepsilon}\left[\sum_{j=1}^{\mathcal{N}^\varepsilon-1} \theta_j^\varepsilon\right] + E_{u^\varepsilon}\zeta^\varepsilon = E_{u^\varepsilon}(\mathcal{N}^\varepsilon - 1)E_{u^\varepsilon}\theta_1^\varepsilon + E_{u^\varepsilon}\zeta^\varepsilon,$$

where the second equality comes from Wald's identity. Using $\sum_{k=1}^\infty k a^{k-1} = 1/(1-a)^2$ for $a \in [0, 1)$, we also have

$$E_{u^\varepsilon}\mathcal{N}^\varepsilon = \sum_{k=1}^\infty k(1 - e^{-h_1^\varepsilon(\delta)/\varepsilon})^{k-1} e^{-h_1^\varepsilon(\delta)/\varepsilon} = e^{-h_1^\varepsilon(\delta)/\varepsilon} e^{2h_1^\varepsilon(\delta)/\varepsilon} = e^{h_1^\varepsilon(\delta)/\varepsilon},$$

and therefore

$$E_{u^\varepsilon}\upsilon_1^\varepsilon = e^{h_1^\varepsilon(\delta)/\varepsilon} E_{u^\varepsilon}\theta_1^\varepsilon + (E_{u^\varepsilon}\zeta^\varepsilon - E_{u^\varepsilon}\theta_1^\varepsilon). \tag{10.6}$$

Next consider the characteristic function of $\upsilon_1^\varepsilon / E_{u^\varepsilon}\upsilon_1^\varepsilon$

$$\phi^\varepsilon(t) = E_{u^\varepsilon} e^{it\upsilon_1^\varepsilon / E_{u^\varepsilon}\upsilon_1^\varepsilon} = \phi_\upsilon^\varepsilon(t/E_{u^\varepsilon}\upsilon_1^\varepsilon),$$

where $\phi_\upsilon^\varepsilon$ is the characteristic function of υ_1^ε. By (10.3), we have

$$\begin{aligned}
\phi_\upsilon^\varepsilon(s) &= E_{u^\varepsilon} e^{is\left(\sum_{k=1}^{\mathcal{N}^\varepsilon-1}\theta_k^\varepsilon + \zeta^\varepsilon\right)} = E_{u^\varepsilon} e^{is\zeta^\varepsilon} E_{u^\varepsilon} e^{is\left(\sum_{k=1}^{\mathcal{N}^\varepsilon-1}\theta_k^\varepsilon\right)} \\
&= \phi_\zeta^\varepsilon(s) \sum_{k=1}^\infty (1 - e^{-h_1^\varepsilon(\delta)/\varepsilon})^{k-1} e^{-h_1^\varepsilon(\delta)/\varepsilon} \phi_\theta^\varepsilon(s)^{k-1} \\
&= \phi_\zeta^\varepsilon(s) e^{-h_1^\varepsilon(\delta)/\varepsilon} (1 - [(1 - e^{-h_1^\varepsilon(\delta)/\varepsilon})\phi_\theta^\varepsilon(s)])^{-1},
\end{aligned}$$

where ϕ_θ^ε and ϕ_ζ^ε are the characteristic functions of θ_1^ε and ζ^ε, respectively. We want to show that for any $t \in \mathbb{R}$

$$\phi^\varepsilon(t) = \phi_\upsilon^\varepsilon(t/E_{u^\varepsilon}\upsilon_1^\varepsilon) \to 1/(1 - it) \text{ as } \varepsilon \to 0.$$

We first show that $\phi_\zeta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon) \to 1$. By definition, $\phi_\zeta^\varepsilon\left(t/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)$
$= E_{u^\varepsilon} \cos\left(t\zeta^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right) + i E_{u^\varepsilon} \sin\left(t\zeta^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)$. According to [12, Lemma 1.9, Chapter 6], we know that there exist $T_0 \in (0, \infty)$ and $\beta > 0$ such that for any $T \in (T_0, \infty)$ and for all ε sufficiently small

$$P_{u^\varepsilon}\left(\zeta^\varepsilon > T\right) \le e^{-\frac{1}{\varepsilon}\beta(T-T_0)}, \tag{10.7}$$

and therefore for any bounded and continuous function $f : \mathbb{R} \to \mathbb{R}$

$$\left| E_{u^\varepsilon} f\left(t\zeta^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right) - f(0) \right| \le 2 \|f\|_\infty P_{u^\varepsilon}\left(\zeta^\varepsilon > T\right)$$
$$+ E_{u^\varepsilon}\left[\left| f\left(t\zeta^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right) - f(0) \right| 1_{\{\zeta^\varepsilon \le T\}} \right].$$

The first term in the last display goes to 0 as $\varepsilon \to 0$. For any fixed t, $t/E_{u^\varepsilon} \upsilon_1^\varepsilon \to 0$ as $\varepsilon \to 0$. Since f is continuous, the second term in the last display also converges to 0 as $\varepsilon \to 0$. $\phi_\zeta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon) \to 1$ follows by taking f to be $\sin x$ and $\cos x$.

It remains to show that for any $t \in \mathbb{R}$

$$e^{-h_1^\varepsilon(\delta)/\varepsilon}\left(1 - \left[(1 - e^{-h_1^\varepsilon(\delta)/\varepsilon})\phi_\theta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon)\right]\right)^{-1} \to 1/(1 - it)$$

as $\varepsilon \to 0$. Observe that

$$e^{-h_1^\varepsilon(\delta)/\varepsilon}\left(1 - \left[(1 - e^{-h_1^\varepsilon(\delta)/\varepsilon})\phi_\theta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon)\right]\right)^{-1} = \left(\frac{1 - \phi_\theta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} + \phi_\theta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon)\right)^{-1},$$

so it suffices to show that $\phi_\theta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon) \to 1$ and $[1 - \phi_\theta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon)]/e^{-h_1^\varepsilon(\delta)/\varepsilon} \to -it$ as $\varepsilon \to 0$.

For the former, note that by (10.6)

$$0 \le E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right) \le \frac{t E_{u^\varepsilon}\theta_1^\varepsilon}{\left(e^{h_1^\varepsilon(\delta)/\varepsilon} - 1\right) E_{u^\varepsilon}\theta_1^\varepsilon} \to 0$$

as $\varepsilon \to 0$, and so $t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon$ converges to 0 in distribution. Moreover, since e^{ix} is bounded and continuous, we find $\phi_\theta^\varepsilon(t/E_{u^\varepsilon} \upsilon_1^\varepsilon) \to 1$. For the second part, using

$$x - x^3/3! \le \sin x \le x \text{ and } 1 - x^2/2 \le \cos x \le 1$$

for $x \in \mathbb{R}$ we find that

$$0 \le \frac{1 - E_{u^\varepsilon} \cos\left(t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} \le \frac{E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)^2}{2e^{-h_1^\varepsilon(\delta)/\varepsilon}}$$

and

$$\frac{E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} - \frac{E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)^3}{3!e^{-h_1^\varepsilon(\delta)/\varepsilon}} \le \frac{E_{u^\varepsilon} \sin\left(t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} \le \frac{E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon} \upsilon_1^\varepsilon\right)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}}.$$

From our previous observation regarding the distribution of ζ^ε and (10.6)

$$\frac{E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon}\upsilon_1^\varepsilon\right)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} \to t \text{ as } \varepsilon \to 0.$$

In addition, since θ_1^ε can be viewed as the time from the outer ring to the inner ring without visiting $\cup_{j\in L\setminus\{1\}}\partial B_\delta(O_j)$ plus the time from the inner ring to the outer ring, by applying (10.7) to the former and using [6, Theorem 4 and Corollary 1] under Condition 3.13 to the later, we find that

$$P_{u^\varepsilon}\left(\theta_1^\varepsilon/E_{u^\varepsilon}\theta_1^\varepsilon > t\right) \le 2e^{-t} \tag{10.8}$$

for all $t \in [0, \infty)$ and ε sufficiently small. This implies that

$$E_{u^\varepsilon}\left(\theta_1^\varepsilon/E_{u^\varepsilon}\theta_1^\varepsilon\right)^2 = 2\int_0^\infty t^2 P_{u^\varepsilon}\left(\theta_1^\varepsilon/E_{u^\varepsilon}\theta_1^\varepsilon > t\right)dt \le 4\int_0^\infty t^2 e^{-t}dt = 8$$

and similarly $E_{u^\varepsilon}\left(\theta_1^\varepsilon/E_{u^\varepsilon}\theta_1^\varepsilon\right)^3 = 3\int_0^\infty t^3 \le 36$. Then combined with (10.6), we have

$$0 \le \frac{E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon}\upsilon_1^\varepsilon\right)^2}{2e^{-h_1^\varepsilon(\delta)/\varepsilon}} \le \frac{t^2 E_{u^\varepsilon}\left(\theta_1^\varepsilon/E_{u^\varepsilon}\theta_1^\varepsilon\right)^2}{2e^{-h_1^\varepsilon(\delta)/\varepsilon}(e^{h_1^\varepsilon(\delta)/\varepsilon}-1)^2} \to 0$$

and

$$0 \le \frac{E_{u^\varepsilon}\left(t\theta_1^\varepsilon/E_{u^\varepsilon}\upsilon_1^\varepsilon\right)^3}{3!e^{-h_1^\varepsilon(\delta)/\varepsilon}} \le \frac{t^3 E_{u^\varepsilon}\left(\theta_1^\varepsilon/E_{u^\varepsilon}\theta_1^\varepsilon\right)^3}{3!e^{-h_1^\varepsilon(\delta)/\varepsilon}(e^{h_1^\varepsilon(\delta)/\varepsilon}-1)^3} \to 0.$$

Therefore, we have shown that for any $t \in \mathbb{R}$

$$\frac{1-\phi_\theta^\varepsilon(t/E_{u^\varepsilon}\upsilon_1^\varepsilon)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} = \frac{1-E_{u^\varepsilon}\cos\left(t\theta_1^\varepsilon/E_{u^\varepsilon}\upsilon_1^\varepsilon\right)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} - i\frac{E_{u^\varepsilon}\sin\left(t\theta_1^\varepsilon/E_{u^\varepsilon}\upsilon_1^\varepsilon\right)}{e^{-h_1^\varepsilon(\delta)/\varepsilon}} \to -it.$$

$\square$

Remark 10.5 From the proof of Lemma 10.4, we actually know that

$$\phi_\upsilon^\varepsilon(t/E_{u^\varepsilon}\upsilon_1^\varepsilon) \to 1/(1-it)$$

uniformly on any compact set in $\mathbb{R}$ as $\varepsilon \to 0$.

10.3 Tail Probability

The goal of this subsection is to prove the following.

Lemma 10.6 *For each $j \in L_s$ there is a distribution u^ε on $\partial B_{2\delta}(O_j)$ and $\tilde{c} > 0$ such that for any $t \in [0, \infty)$, $P_{u^\varepsilon}(v_j^\varepsilon/E_{u^\varepsilon}v_j^\varepsilon > t) \le e^{-\tilde{c}t}$ (here, v_j^ε and u^ε are defined as in the last subsection).*

Proof As in the last subsection, we give the proof for the case $j = 1$. To begin, we note that for any $\alpha > 0$ Chebyshev's inequality implies

$$P_{u^\varepsilon}\left(v_1^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon > t\right) = P_{u^\varepsilon}(e^{\alpha v_1^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon} > e^{\alpha t}) \le e^{-\alpha t} \cdot E_{u^\varepsilon}e^{\alpha v_1^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon}.$$

By picking $\alpha = \alpha^* \doteq 1/8$, it suffices to show that $E_{u^\varepsilon}e^{\alpha^* v_1^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon}$ is bounded by a constant. We will do this by showing how the finiteness of $E_{u^\varepsilon}e^{\alpha^* v_1^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon}$ is implied by the finiteness of $E_{u^\varepsilon}e^{\alpha^* \theta_1^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon}$ and $E_{u^\varepsilon}e^{\alpha^* \zeta^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon}$.

Using (10.8), we find that for any $\alpha > 0$

$$P_{u^\varepsilon}(e^{\alpha \theta_1^\varepsilon/E_{u^\varepsilon}\theta_1^\varepsilon} > t) \le 2e^{-\frac{1}{\alpha}\log t} = 2t^{-\frac{1}{\alpha}}$$

for all $t \in [1, \infty)$ and ε sufficiently small. Then, (10.6) implies $E_{u^\varepsilon}v_1^\varepsilon \ge \left(e^{h_1^\varepsilon(\delta)/\varepsilon} - 1\right)E_{u^\varepsilon}\theta_1^\varepsilon$ and therefore

$$\begin{aligned}
E_{u^\varepsilon}e^{\alpha^*/E_{u^\varepsilon}v_1^\varepsilon \theta_1^\varepsilon} &\le \int_0^1 P_{u^\varepsilon}\left(\exp\left(\alpha^*\theta_1^\varepsilon/[(e^{h_1^\varepsilon(\delta)/\varepsilon} - 1)E_{u^\varepsilon}\theta_1^\varepsilon]\right) > t\right)dt \\
&\quad + \int_1^\infty P_{u^\varepsilon}\left(\exp\left(\alpha^*\theta_1^\varepsilon/[(e^{h_1^\varepsilon(\delta)/\varepsilon} - 1)E_{u^\varepsilon}\theta_1^\varepsilon]\right) > t\right)dt \\
&\le 1 + 2\int_1^\infty t^{-(e^{h_1^\varepsilon(\delta)/\varepsilon}-1)/\alpha^*}dt \\
&= 1 + 2[(e^{h_1^\varepsilon(\delta)/\varepsilon} - 1)/\alpha^* - 1]^{-1} = 1 + 2\alpha^*[e^{h_1^\varepsilon(\delta)/\varepsilon} - \alpha^* - 1]^{-1}.
\end{aligned}$$

To estimate ζ^ε, we use that by (10.7) there are $T_0 \in (0, \infty)$ and $\beta > 0$ such that for any $t \in (T_0, \infty)$ and for all ε sufficiently small $P_{u^\varepsilon}(\zeta^\varepsilon > t) \le e^{-\frac{1}{\varepsilon}\beta(t-T_0)}$, so that for any $\alpha > 0$ $P_{u^\varepsilon}\left(e^{\alpha \zeta^\varepsilon} > t\right) \le e^{-\frac{1}{\varepsilon}\beta\left(\frac{1}{\alpha}\log t - T_0\right)}$ for any $t \ge e^{\alpha T_0}$. Given $n \in \mathbb{N}$, for all sufficiently small ε we have $\alpha^*/E_{u^\varepsilon}v_1^\varepsilon \le 1/n$, and thus

$$P_{u^\varepsilon}\left(e^{\alpha^* \zeta^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon} > t\right) \le P_{u^\varepsilon}\left(e^{\zeta^\varepsilon/n} > t\right) \le e^{-\frac{1}{\varepsilon}\beta(n\log t - T_0)}.$$

Hence for any n such that $e^{T_0/n} \le 3/2$ and $(-\beta n + 1)\log(3/2) + \beta T_0 < 0$, and for ε small enough that $\alpha^*/E_{u^\varepsilon}v_1^\varepsilon \le 1/n$, we have

$$\begin{aligned}
E_{u^\varepsilon}e^{\alpha^* \zeta^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon} &\le \int_0^\infty P_{u^\varepsilon}\left(e^{\alpha^* \zeta^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon} > t\right)dt \le 3/2 + \int_{\frac{3}{2}}^\infty P_{u^\varepsilon}\left(e^{\alpha^* \zeta^\varepsilon/E_{u^\varepsilon}v_1^\varepsilon} > t\right)dt \\
&\le 3/2 + \int_{\frac{3}{2}}^\infty e^{-\frac{1}{\varepsilon}\beta(n\log t - T_0)}dt = 3/2 + e^{\frac{1}{\varepsilon}\beta T_0}(\beta n/\varepsilon - 1)^{-1}(3/2)^{\frac{1}{\varepsilon}(-\beta n + \varepsilon)} \\
&= 3/2 + (\beta n/\varepsilon - 1)^{-1}e^{\frac{1}{\varepsilon}[(-\beta n + \varepsilon)\log(3/2) + \beta T_0]} \le 3/2 + (\beta n/\varepsilon - 1)^{-1} \le 2.
\end{aligned}$$

We have shown that for such α^*, $E_{u^\varepsilon} e^{\alpha^* \zeta^\varepsilon / E_{u^\varepsilon} v_1^\varepsilon}$ and $E_{u^\varepsilon} e^{\alpha^* \theta_1^\varepsilon / E_{u^\varepsilon} v_1^\varepsilon}$ are uniformly bounded for all ε sufficiently small. Lastly, using the same calculation as used for the characteristic function

$$
\begin{aligned}
E_{u^\varepsilon} e^{\alpha^* v_1^\varepsilon / E_{u^\varepsilon} v_1^\varepsilon} &= E_{u^\varepsilon} e^{\alpha^* \zeta^\varepsilon / E_{u^\varepsilon} v_1^\varepsilon} \cdot e^{-h_1^\varepsilon(\delta)/\varepsilon} \left(1 - \left[(1 - e^{-h_1^\varepsilon(\delta)/\varepsilon}) E_{u^\varepsilon} e^{\alpha^* \theta_1^\varepsilon / E_{u^\varepsilon} v_1^\varepsilon}\right]\right)^{-1} \\
&\le 2 e^{-h_1^\varepsilon(\delta)/\varepsilon} \left(1 - \left[(1 - e^{-h_1^\varepsilon(\delta)/\varepsilon}) \left(1 + \frac{2\alpha^*}{e^{h_1^\varepsilon(\delta)/\varepsilon} - \alpha^* - 1}\right)\right]\right)^{-1} \\
&= 2 e^{-h_1^\varepsilon(\delta)/\varepsilon} \left(e^{-h_1^\varepsilon(\delta)/\varepsilon} - \frac{2\alpha^*}{e^{h_1^\varepsilon(\delta)/\varepsilon} - \alpha^* - 1} + \frac{2\alpha^*}{e^{h_1^\varepsilon(\delta)/\varepsilon} - \alpha^* - 1} e^{-h_1^\varepsilon(\delta)/\varepsilon}\right)^{-1} \\
&= 2 \left(1 - 2\alpha^* \frac{e^{h_1^\varepsilon(\delta)/\varepsilon} - 1}{e^{h_1^\varepsilon(\delta)/\varepsilon} - \alpha^* - 1}\right)^{-1} \le 2/(1 - 4\alpha^*) = 4.
\end{aligned}
$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ □

10.4 General Initial Condition

This subsection presents results that will allow us to extend the results in the previous two subsections to arbitrary initial distribution $\lambda^\varepsilon \in \mathcal{P}(\partial B_\delta(O_1))$. Under our assumptions, for any $j \in L_s$ we observe that the process model

$$
dX_t^\varepsilon = b\left(X_t^\varepsilon\right) dt + \sqrt{\varepsilon} \sigma\left(X_t^\varepsilon\right) dW_t \tag{10.9}
$$

has the property that $b(x) = A(x - O_j)[1 + o(1)]$ and $\sigma(x) = \bar{\sigma}[1 + o(1)]$, where $o(1) \to 0$ as $\|x - O_j\| \to 0$, A is stable and $\bar{\sigma}$ is invertible. By an invertible change of variable, we can arrange so that $O_j = 0$ and $\bar{\sigma} = I$, and to simplify we assume this in the rest of the section.

Since A is stable, there exists a positive definite and symmetric solution M to the matrix equation $AM + MA^T = -I$ (we can in fact exhibit the solution in the form $M = \int_0^\infty e^{At} e^{A^T t} dt$). To prove the ergodicity, we introduce some additional notation: $U(x) \doteq \langle x, Mx \rangle$, $B_i \doteq \{x : U(x) < b_i^2\}$ and $\mathcal{S}_i(\varepsilon) \doteq \{x : U(x) < a_i^2 \varepsilon\}$, for $i = 1, 2$, where $0 < a_1 < a_2$, $0 < b_0 < b_1 < b_2$. If $\varepsilon_0 = (b_0^2/a_2^2)/2$, then with cl denoting closure, $\mathrm{cl}(\mathcal{S}_2(\varepsilon_0)) \subset B_0$, and we will assume $\varepsilon \in (0, \varepsilon_0)$ henceforth. For a use later on, we will also assume that $a_1^2 = 2 \sup_{x \in B_2} \mathrm{tr}[\sigma(x) \sigma(x)^T M]$.

Remark 10.7 The sets B_1 and B_2 will play the roles that $B_\delta(O_1)$ and $B_{2\delta}(O_1)$ played previously in this section. Although elsewhere in this paper as well as in the reference [12] these sets are taken to be balls with respect to the Euclidean norm, in this subsection we take them to be level sets of $U(x)$. The shape of these sets and the choice of the factor of 2 relating the radii play no role in the analysis of [12] or in our prior use in this paper. However, in this subsection it is notationally convenient for the sets to be level sets of U, since U is a Lyapunov function for the noiseless dynamics near 0. After this subsection, we will revert to the $B_\delta(O_1)$ and $B_{2\delta}(O_1)$ notation.

In addition to the restrictions $a_1 < a_2$ and $a_2^2 \varepsilon_0 \le b_0^2$, we also assume that a_1, a_2 and $\varepsilon_0 > 0$ are such that if ϕ^x is the solution to the noiseless dynamics $\dot{\phi} = b(\phi)$

with initial condition x, then: (i) for all $x \in \partial \mathcal{S}_2(\varepsilon)$, ϕ^x never crosses ∂B_1; (i) for all $x \in \partial \mathcal{S}_1(\varepsilon)$, ϕ^x never exits $\mathcal{S}_2(\varepsilon)$.

The idea that will be used to establish asymptotic independence from the starting distribution is the following. We start the process on ∂B_1. With some small probability, it will hit ∂B_2 before hitting $\partial \mathcal{S}_2(\varepsilon)$. This gives a contribution to $\psi_2^\varepsilon(dz|x)$ defined in (10.4) that will be relatively unimportant. If instead it hits $\partial \mathcal{S}_2(\varepsilon)$ first, then we do a Freidlin–Wentzell type analysis and decompose the trajectory into excursions between $\partial \mathcal{S}_2(\varepsilon)$ and $\partial \mathcal{S}_1(\varepsilon)$, before a final excursion from $\partial \mathcal{S}_2(\varepsilon)$ to ∂B_2.

To exhibit the asymptotic independence from ε, we introduce the scaled process $Y_t^\varepsilon = X_t^\varepsilon/\sqrt{\varepsilon}$, which solves the SDE

$$dY_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}} b(\sqrt{\varepsilon} Y_t^\varepsilon) dt + \sigma(\sqrt{\varepsilon} Y_t^\varepsilon) dW_t.$$

Let $\bar{\mathcal{S}}_1 = \partial \mathcal{S}_1(1)$ and $\bar{\mathcal{S}}_2 = \partial \mathcal{S}_2(1)$. Let $\omega^\varepsilon(w|x)$ denote the density of the hitting location on $\bar{\mathcal{S}}_2$ by the process Y^ε, given $Y_0^\varepsilon = x \in \bar{\mathcal{S}}_1$. The following estimate is essential. The density function can be identified with the normal derivative of a related Green's function, which is bounded from above by the boundary gradient estimate and bounded below by using the Hopf lemma [13].

Lemma 10.8 *Given $\varepsilon_0 > 0$, there are $0 < c_1 < c_2 < \infty$ such that $c_1 \le \omega^\varepsilon(w|x) \le c_2$ for all $x \in \bar{\mathcal{S}}_1$, $w \in \bar{\mathcal{S}}_2$ and $\varepsilon \in (0, \varepsilon_0)$.*

Next let $p^\varepsilon(u|w)$ denote the density of the return location for Y^ε on $\bar{\mathcal{S}}_2$, conditioned on visiting $\bar{\mathcal{S}}_1$ before $\partial B_2/\sqrt{\varepsilon}$, and starting at $w \in \bar{\mathcal{S}}_2$. The last lemma then directly gives the following.

Lemma 10.9 *For $\varepsilon_0 > 0$ and c_1, c_2 as in the last lemma $c_1 \le p^\varepsilon(u|w) \le c_2$ for all $u, w \in \bar{\mathcal{S}}_2$ and $\varepsilon \in (0, \varepsilon_0)$.*

Let $r^\varepsilon(w)$ denote the unique stationary distribution of $p^\varepsilon(u|w)$, and let $p^{\varepsilon,n}(u|w)$ denote the n-step transition density. The preceding lemma, [14, Theorem 10.1 Chapter 3], and the existence of a uniform strictly positive lower bound on $r^\varepsilon(u)$ for all sufficiently small $\varepsilon > 0$ imply the following.

Lemma 10.10 *There is $K < \infty$ and $\alpha \in (0, 1)$ such that for all $\varepsilon \in (0, \varepsilon_0)$*

$$\sup_{w \in \bar{\mathcal{S}}_2} \left| p^{\varepsilon,n}(u|w) - r^\varepsilon(u) \right| / r^\varepsilon(u) \le K \alpha^n.$$

Let $\eta^\varepsilon(dx|w)$ denote the distribution of X^ε upon first hitting ∂B_2 given that X^ε reaches $\partial \mathcal{S}_1(\varepsilon)$ before ∂B_2 and starts at $w \in \partial \mathcal{S}_2(\varepsilon)$.

Lemma 10.11 *There is $\kappa > 0$ and $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$*

$$\sup_{x \in \partial B_1} P_x \left\{ X^\varepsilon \text{ reaches } \partial B_2 \text{ before } \mathcal{S}_2(\varepsilon) \right\} \le e^{-\kappa/\varepsilon}.$$

Lemma 10.12 *There are $\bar{\eta}^{\varepsilon}(dz) \in \mathcal{P}(\partial B_2)$, s^{ε} that tends to 0 as $\varepsilon \to 0$ and $\varepsilon_0 > 0$, such that for all $A \in \mathcal{B}(\partial B_2)$, $w \in \partial \mathcal{S}_2(\varepsilon)$ and $\varepsilon \in (0, \varepsilon_0)$*

$$\bar{\eta}^{\varepsilon}(A)[1 - s^{\varepsilon}K/(1-\alpha)] \leq \eta^{\varepsilon}(A|w) \leq \bar{\eta}^{\varepsilon}(A)[1 + s^{\varepsilon}K/(1-\alpha)],$$

where K and α are from Lemma 10.10.

Proof of Lemma 10.11 Recall that $a_1^2 = 2\sup_{x \in B_2} \text{tr}[\sigma(x)\sigma(x)^T M]$. We then use that $AM + MA^T = -I$ to get that with $U(x) \doteq \langle x, Mx \rangle$,

$$\langle DU(x), b(x) \rangle \leq -\varepsilon a_1^2 \tag{10.10}$$

for $x \in B_2 \setminus \mathcal{S}_2(\varepsilon)$, and

$$\langle DU(x), b(x) \rangle \leq -\frac{1}{8}b_0^2 \tag{10.11}$$

for $B_2 \setminus (B_0/2)$. By Itô's formula

$$dU(X_t^{\varepsilon}) = \langle DU(X_t^{\varepsilon}), b(X_t^{\varepsilon}) \rangle dt + \frac{\varepsilon}{2}\text{tr}[\sigma(X_t^{\varepsilon})\sigma(X_t^{\varepsilon})^T M]dt$$
$$+ \sqrt{\varepsilon} \langle DU(X_t^{\varepsilon}), \sigma(X_t^{\varepsilon})dW_t \rangle. \tag{10.12}$$

Starting at $x \in \partial B_1$, we are concerned with the probability

$$P_x \left\{ U(X_t^{\varepsilon}) \text{ reaches } b_2^2 \text{ before } a_2^2\varepsilon \right\},$$

where $U(x) = b_1^2$. However, according to (10.12) and (10.11), reaching b_2^2 before $b_0^2/4$ is a rare event, and its probability decays exponentially in the form $e^{-\kappa/\varepsilon}$ for some $\kappa > 0$ and uniformly in $x \in \partial B_1$. Once the process reaches $B_0/2$, (10.12) and (10.10) imply $U(X_t^{\varepsilon})$ is a supermartingale as long as it is in the interval $[0, b_0^2]$, and therefore after X_t^{ε} reaches $B_0/2$, the probability that $U(X_t^{\varepsilon})$ reaches $a_2^2\varepsilon$ before b_0^2 is greater than $1/2$. □

Proof of Lemma 10.12 Consider a starting position $w \in \partial \mathcal{S}_2(\varepsilon)$, and recall that $\eta^{\varepsilon}(dz|w)$ denotes the hitting distribution on ∂B_2 after starting at w. Let θ_k^{ε} denote the return times to $\partial \mathcal{S}_2(\varepsilon)$ after visiting $\partial \mathcal{S}_1(\varepsilon)$, and let $q_n^{\varepsilon}(w)$ denote the probability that the first k for which X^{ε} visits ∂B_2 before visiting $\partial \mathcal{S}_1(\varepsilon)$ during $[\theta_k^{\varepsilon}, \theta_{k+1}^{\varepsilon}]$ is n. Then by the strong Markov property and using the rescaled process

$$\int_{\partial B_2} g(z)\eta^{\varepsilon}(dz|w) = \sum_{n=0}^{\infty} \int_{\partial B_2} g(z)q_n^{\varepsilon}(w) \int_{\partial \mathcal{S}_2(\varepsilon)} \eta^{\varepsilon}(dz|u)J^{\varepsilon}(u)p^{\varepsilon,n}(\sqrt{\varepsilon}u|\sqrt{\varepsilon}w)du,$$

where $J^{\varepsilon}(u)$ is the Jacobian that accounts for the mapping between $\partial \mathcal{S}_2(\varepsilon)$ and $\partial \mathcal{S}_2(1)$ and is given by $u/\sqrt{\varepsilon}$. We next use that uniformly in $w \in \partial \mathcal{S}_2(\varepsilon)$

$$p^{\varepsilon,n}(\sqrt{\varepsilon}u|\sqrt{\varepsilon}w) \leq r^{\varepsilon}(\sqrt{\varepsilon}u)[1 + K\alpha^n]$$

to get

$$\sum_{n=0}^{\infty} \int_{\partial B_2} g(z) q_n^{\varepsilon}(w) \int_{\partial \mathcal{S}_2(\varepsilon)} \eta^{\varepsilon}(dz|u) J^{\varepsilon}(u) p^{\varepsilon, n}(\sqrt{\varepsilon} u | \sqrt{\varepsilon} w) du$$

$$\leq \left(\sum_{n=0}^{\infty} \int_{\partial B_2} g(z) q_n^{\varepsilon}(w) \int_{\partial \mathcal{S}_2(\varepsilon)} \eta^{\varepsilon}(dz|u) J^{\varepsilon}(u) r^{\varepsilon}(\sqrt{\varepsilon} u) du \right) [1 + K\alpha^n]$$

$$= \int_{\partial B_2} g(z) \int_{\partial \mathcal{S}_2(\varepsilon)} \eta^{\varepsilon}(dz|u) J^{\varepsilon}(u) r^{\varepsilon}(\sqrt{\varepsilon} u) du \left[1 + K \sum_{n=0}^{\infty} q_n^{\varepsilon}(w) \alpha^n \right].$$

Now, use that $K \sum_{n=0}^{\infty} \alpha^n = K/(1 - \alpha) < \infty$ and $\sup_{w \in \partial \mathcal{S}_2(\varepsilon)} \sup_{n \in \mathbb{N}_0} q_n^{\varepsilon}(w) \to 0$ as $\varepsilon \to 0$ to get the upper bound with

$$\bar{\eta}^{\varepsilon}(dz) \doteq \int_{\partial \mathcal{S}_2(\varepsilon)} \eta^{\varepsilon}(dz|u) J^{\varepsilon}(u) r^{\varepsilon}(\sqrt{\varepsilon} u) du.$$

When combined with the lower bound which has an analogous proof, Lemma 10.12 follows. □

10.5 Times to Reach Another Equilibrium Point after Starting with General Initial Distribution

Lemma 10.13 *For each $j \in L_{\mathrm{s}}$, there exist $\tilde{c} > 0$ and $\varepsilon_0 \in (0, 1)$ such that for any distribution λ^{ε} on $\partial B_{\delta}(O_j)$,*

$$P_{\lambda^{\varepsilon}}(\upsilon_j^{\varepsilon}/E_{\lambda^{\varepsilon}} \upsilon_j^{\varepsilon} > t) \leq e^{-\tilde{c}t}$$

for all $t > 0$ and $\varepsilon \in (0, \varepsilon_0)$.

Proof We give the proof for the case $j = 1$. We first show for any $r \in (0, 1)$ there is $\varepsilon_0 > 0$ such that for any $\varepsilon \in (0, \varepsilon_0)$ and $\lambda^{\varepsilon}, \theta^{\varepsilon} \in \mathcal{P}(\partial B_{\delta}(O_1))$

$$E_{\lambda^{\varepsilon}} \upsilon_1^{\varepsilon}/E_{\theta^{\varepsilon}} \upsilon_1^{\varepsilon} \geq r. \tag{10.13}$$

We use that υ_1^{ε} can be decomposed into $\bar{\upsilon}_1^{\varepsilon} + \hat{\upsilon}_1^{\varepsilon}$, where $\bar{\upsilon}_1^{\varepsilon}$ is the first hitting time to $\partial B_{2\delta}(O_1)$. Since by standard large deviation theory the exponential growth rate of the expected value of υ_1^{ε} is strictly greater than that of $\bar{\upsilon}_1^{\varepsilon}$ (uniformly in the initial distribution) $E_{\lambda^{\varepsilon}} \bar{\upsilon}_1^{\varepsilon}$ (respectively, $E_{\theta^{\varepsilon}} \bar{\upsilon}_1^{\varepsilon}$) is negligible compared to $E_{\lambda^{\varepsilon}} \upsilon_1^{\varepsilon}$ (respectively, $E_{\theta^{\varepsilon}} \upsilon_1^{\varepsilon}$), and so it is enough to show $E_{\lambda^{\varepsilon}} \hat{\upsilon}_1^{\varepsilon}/E_{\theta^{\varepsilon}} \hat{\upsilon}_1^{\varepsilon} \geq r$. Owing to Lemma 10.11 (and in particular because $\kappa > 0$), the contribution to either $E_{\lambda^{\varepsilon}} \hat{\upsilon}_1^{\varepsilon}$ or $E_{\theta^{\varepsilon}} \hat{\upsilon}_1^{\varepsilon}$ from trajectories that reach $\partial B_{2\delta}(O_1)$ before $\partial \mathcal{S}_2(\varepsilon)$ can be neglected. Using Lemma 10.12 and the strong Markov property gives

$$\inf_{w_1, w_2 \in \partial \mathcal{S}_2(\varepsilon)} \frac{E_{w_1} \hat{\upsilon}_1^{\varepsilon}}{E_{w_2} \hat{\upsilon}_1^{\varepsilon}} \geq \frac{[1 - s^{\varepsilon} K/(1 - \alpha)]}{[1 + s^{\varepsilon} K/(1 - \alpha)]},$$

and the lower bound follows since $s^\varepsilon \to 0$.

We next claim that a suitable bound can be found for $P_{\lambda^\varepsilon}(\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t)$. Recall that $u^\varepsilon \in \mathcal{P}(\partial B_{2\delta}(O_1))$ is the stationary probability for ψ^ε defined in (10.5). Let β^ε be the probability measure on $\partial B_\delta(O_1)$ obtained by integrating the transition kernel ψ_1^ε with respect to u^ε, and note that integrating ψ_2^ε with respect to β^ε returns u^ε. Since the diffusion matrix is uniformly nondegenerate, by using well-known "Gaussian type" bounds on the transition density for the process [2], there are $K \in (0, \infty)$ and $p \in (0, \infty)$ such that

$$ P_x \left\{ X_\theta^\varepsilon \in A | X_\theta^\varepsilon \in \partial B_{2\delta}(O_1) \right\} \leq K m(A)/\varepsilon^p $$

for all $x \in \partial B_\delta(O_1)$, where m is the uniform measure on $\partial B_{2\delta}(O_1)$ and $\theta = \inf\{t > 0 : X_t^\varepsilon \in \partial B_{2\delta}(O_1) \cup \mathcal{S}_2(\varepsilon)\}$. Together with Lemmas 10.11 and 10.12, this implies that for all sufficiently small $\varepsilon > 0$ and any bounded measurable function $h : \partial B_{2\delta}(O_1) \to \mathbb{R}$,

$$ \int_{\partial B_{2\delta}(O_1)} \int_{\partial B_\delta(O_1)} h(y)\psi_2^\varepsilon(dy|x)\lambda^\varepsilon(dx) \leq 2 \int_{\partial B_{2\delta}(O_1)} \int_{\partial B_\delta(O_1)} h(y)\psi_2^\varepsilon(dy|x)\beta^\varepsilon(dx) $$

$$ \leq 2 \int_{\partial B_{2\delta}(O_1)} h(y)u^\varepsilon(dy). $$

Using the last display for the first inequality, (10.13) for the second, that $\bar{\upsilon}_1^\varepsilon$ is small compared with $\hat{\upsilon}_1^\varepsilon$ for the third and Lemma 10.6 for the last, there is $\varepsilon_1 > 0$ such that

$$ P_{\lambda^\varepsilon}(\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t) = E_{\lambda^\varepsilon}(P_{X_{\bar{\upsilon}_1^\varepsilon}^\varepsilon}(\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t)) \leq 2 P_{u^\varepsilon}(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t) $$

$$ \leq 2 P_{u^\varepsilon}(\upsilon_1^\varepsilon / E_{\beta^\varepsilon} \upsilon_1^\varepsilon > t/2) \leq 2 P_{u^\varepsilon}(\upsilon_1^\varepsilon / E_{u^\varepsilon} \upsilon_1^\varepsilon > t/4) \leq 2 e^{-\tilde{c} t/4} $$

for all $\varepsilon \in (0, \varepsilon_1)$ and $t \geq 0$.

Since as noted previously $E_{\lambda^\varepsilon} \upsilon_1^\varepsilon \geq E_{\lambda^\varepsilon} \bar{\upsilon}_1^\varepsilon$ and since by [6, Theorem 4 and Corollary 1], there exists $\varepsilon_2 \in (0, 1)$ such that $P_{\lambda^\varepsilon}(\bar{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \bar{\upsilon}_1^\varepsilon > t) \leq 2 e^{-t/2}$ for any $t > 0$ and $\varepsilon \in (0, \varepsilon_2)$, we conclude that for any $t > 0$ $P_{\lambda^\varepsilon}(\bar{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t/2) \leq P_{\lambda^\varepsilon}(\bar{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \bar{\upsilon}_1^\varepsilon > t/2) \leq 2 e^{-t/2}$. The conclusion of the lemma follows from these two bounds and

$$ P_{\lambda^\varepsilon}(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t) \leq P_{\lambda^\varepsilon}(\bar{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t/2) + P_{\lambda^\varepsilon}(\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon > t/2). $$

$\square$

Lemma 10.14 *For any* $j \in L_s$ *and any distribution* λ^ε *on* $\partial B_\delta(O_j)$, $\upsilon_j^\varepsilon / E_{\lambda^\varepsilon} \upsilon_j^\varepsilon$ *converges in distribution to an* Exp(1) *random variable under* P_{λ^ε}. *Moreover,* $E_{\lambda^\varepsilon} e^{it \upsilon_j^\varepsilon / E_{\lambda^\varepsilon} \upsilon_j^\varepsilon} \to 1/(1 - it)$ *uniformly on any compact set in* $\mathbb{R}$.

Proof We give the proof for the case $j = 1$. Recall that $E_{u^\varepsilon} e^{it \upsilon_1^\varepsilon / E_{u^\varepsilon} \upsilon_1^\varepsilon} \to 1/(1 - it)$ uniformly on any compact set in $\mathbb{R}$ as $\varepsilon \to 0$ from Remark 10.5. We would like to show that $E_{\lambda^\varepsilon} e^{it \upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \to 1/(1-it)$ uniformly on any compact set in $\mathbb{R}$. Since $\upsilon_1^\varepsilon = \bar{\upsilon}_1^\varepsilon + \hat{\upsilon}_1^\varepsilon$ with $\bar{\upsilon}_1^\varepsilon$ the first hitting time to $\partial B_{2\delta}(O_1)$, we know that $E_{\lambda^\varepsilon} \bar{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon \to 0$

and thus $E_{\lambda^\varepsilon} \hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon \to 1$. Observe that

$$E_{\lambda^\varepsilon} e^{it\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} = E_{\lambda^\varepsilon} \left[e^{it\bar{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \cdot E_{X^\varepsilon(\bar{\upsilon}_1^\varepsilon)} \left(e^{it\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \right) \right],$$

$$E_{\lambda^\varepsilon} \left[E_{X^\varepsilon(\bar{\upsilon}_1^\varepsilon)} \left(e^{it\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \right) \right] \le \frac{[1 + s^\varepsilon K/(1-\alpha)]}{[1 - s^\varepsilon K/(1-\alpha)]} E_{u^\varepsilon} e^{it\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \to 1/(1-it)$$

and

$$E_{\lambda^\varepsilon} \left[E_{X^\varepsilon(\bar{\upsilon}_1^\varepsilon)} \left(e^{it\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \right) \right] \ge \frac{[1 - s^\varepsilon K/(1-\alpha)]}{[1 + s^\varepsilon K/(1-\alpha)]} E_{u^\varepsilon} e^{it\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \to 1/(1-it).$$

Since $E_{\lambda^\varepsilon} \bar{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon \to 0$ and e^{ix} is a bounded and continuous function, a conditioning argument gives

$$\left| E_{\lambda^\varepsilon} e^{it\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} - E_{\lambda^\varepsilon} \left[E_{X^\varepsilon(\bar{\upsilon}_1^\varepsilon)} \left(e^{it\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \right) \right] \right| \le E_{\lambda^\varepsilon} \left| e^{it\bar{\upsilon}^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} - 1 \right| \to 0.$$

We conclude that $E_{\lambda^\varepsilon} e^{it\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon} \to 1/(1-it)$ uniformly on any compact set in $\mathbb{R}$. □

10.6 Return Times (Single Cycles)

In this subsection, we will extend all the three results to return times for the single cycle case (i.e., when $h_1 > w$).

Lemma 10.15 *There exists $\delta_0 \in (0,1)$ such that for any $\delta \in (0, \delta_0)$ and any distribution λ^ε on $\partial B_\delta(O_1)$,*

$$\lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \tau_1^\varepsilon = \min_{y \in \cup_{k \in L \setminus \{1\}} \partial B_\delta(O_k)} V(O_1, y).$$

Proof We have $E_{\lambda^\varepsilon} \tau_1^\varepsilon = E_{\lambda^\varepsilon} \upsilon_1^\varepsilon + E_{\lambda^\varepsilon}(\tau_1^\varepsilon - \upsilon_1^\varepsilon)$, and by Lemma 10.2 we know that

$$\lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \upsilon_1^\varepsilon = \min_{y \in \cup_{k \in L \setminus \{1\}} \partial B_\delta(O_k)} V(O_1, y).$$

Moreover, observe that $W(O_j) > W(O_1)$ for any $j \in L \setminus \{1\}$ due to Remark 3.14. Note that υ_1^ε as defined in (10.1) coincides with σ_0^ε defined in (3.4). We can therefore apply Remark 7.22 with $f = 0$, $A = M$ and $\eta = [\min_{j \in L \setminus \{1\}} W(O_j) - W(O_1)]/3$,

we find that there exists $\delta_1 \in (0, 1)$ such that for any $\delta \in (0, \delta_1)$

$$
\begin{aligned}
\liminf_{\varepsilon \to 0} -\varepsilon \log &\left(\sup_{z \in \partial B_\delta(O_1)} E_z \left(\tau_1^\varepsilon - \upsilon_1^\varepsilon \right) \right) \\
&\geq \min_{j \in L \setminus \{1\}} W(O_j) - W(O_1) - \min_{j \in L \setminus \{1\}} V(O_1, O_j) - \eta \\
&= - \min_{j \in L \setminus \{1\}} V(O_1, O_j) + 2\eta.
\end{aligned}
$$

On the other hand, by continuity of $V(O_1, \cdot)$, for this given η, there exists $\delta_2 \in (0, 1)$ such that for any $\delta \in (0, \delta_2)$

$$
\min_{y \in \cup_{k \in L \setminus \{1\}} \partial B_\delta(O_k)} V(O_1, y) \geq \min_{j \in L \setminus \{1\}} V(O_1, O_j) - \eta.
$$

Thus, for any $\delta \in (0, \delta_0)$ with $\delta_0 \doteq \delta_1 \wedge \delta_2$

$$
\begin{aligned}
\limsup_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} (\tau_1^\varepsilon - \upsilon_1^\varepsilon) &\leq \limsup_{\varepsilon \to 0} \varepsilon \log \left(\sup_{z \in \partial B_\delta(O_1)} E_z(\tau_1^\varepsilon - \upsilon_1^\varepsilon) \right) \\
&\leq \min_{j \in L \setminus \{1\}} V(O_1, O_j) - 2\eta \leq \min_{y \in \cup_{k \in L \setminus \{1\}} \partial B_\delta(O_k)} V(O_1, y) - \eta \\
&= \lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \upsilon_1^\varepsilon - \eta
\end{aligned}
$$

and

$$
\lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \tau_1^\varepsilon = \lim_{\varepsilon \to 0} \varepsilon \log E_{\lambda^\varepsilon} \upsilon_1^\varepsilon = \min_{y \in \cup_{k \in L \setminus \{1\}} \partial B_\delta(O_k)} V(O_1, y).
$$

$\square$

Lemma 10.16 *Given $\delta > 0$ sufficiently small, and for any distribution λ^ε on $\partial B_\delta(O_1)$, there exist $\tilde{c} > 0$ and $\varepsilon_0 \in (0, 1)$ such that*

$$
P_{\lambda^\varepsilon} (\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon > t) \leq e^{-\tilde{c}t}
$$

for all $t \geq 1$ and $\varepsilon \in (0, \varepsilon_0)$.

Proof For any $t > 0$, $P_{\lambda^\varepsilon}(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon > t) \leq P_{\lambda^\varepsilon}(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon > t/2) + P_{\lambda^\varepsilon}((\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon > t/2)$. It is easy to see that the first term has this sort of bound due to Lemma 10.13 and $E_{\lambda^\varepsilon} \tau_1^\varepsilon \geq E_{\lambda^\varepsilon} \upsilon_1^\varepsilon$.

It suffices to show that this sort of bound holds for the second term, namely, there exists a constant $\tilde{c} > 0$ such that

$$
P_{\lambda^\varepsilon} \left((\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon > t \right) \leq e^{-\tilde{c}t}
$$

for all $t \in [0, \infty)$ and ε sufficiently small. By Chebyshev's inequality,

$$
P_{\lambda^\varepsilon} \left((\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon > t \right) = P_{\lambda^\varepsilon} (e^{(\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon} > e^t) \leq e^{-t} E_{\lambda^\varepsilon} e^{(\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon},
$$

and it therefore suffices to prove that $E_{\lambda^\varepsilon}e^{(\tau_1^\varepsilon-\upsilon_1^\varepsilon)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}$ is less than a constant for all ε sufficiently small. Observe that

$$\tau_1^\varepsilon-\upsilon_1^\varepsilon=\sum_{j\in L\setminus\{1\}}\sum_{k=1}^{N_j}\upsilon_j^\varepsilon(k),$$

where N_j is the number of visits of $\partial B_\delta(O_j)$, and $\upsilon_j^\varepsilon(k)$ is the k-th copy of the first hitting time to $\cup_{k\in L\setminus\{j\}}\partial B_\delta(O_k)$ after starting from $\partial B_\delta(O_j)$.

If we consider $\partial B_\delta(O_j)$ as the starting location of a regenerative cycle, as was done previously in the paper for $\partial B_\delta(O_1)$, then there will be a unique stationary distribution, and if the process starts with that as the initial distribution, then the times $\upsilon_j^\varepsilon(k)$ are independent from each other and from the number of returns to $\partial B_\delta(O_j)$ before first visiting $\partial B_\delta(O_1)$. While these random times as used here do not arise from starting with such a distribution, we can use Lemma 10.12 to bound the error in terms of a multiplicative factor that is independent of ε for small $\varepsilon>0$, and thereby justify treating N_j as though it is independent of the $\upsilon_j^\varepsilon(k)$.

Recalling that $l\doteq|L|$,

$$E_{\lambda^\varepsilon}e^{(\tau_1^\varepsilon-\upsilon_1^\varepsilon)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}=E_{\lambda^\varepsilon}\prod_{j\in L\setminus\{1\}}e^{\left(\sum_{k=1}^{N_j}\upsilon_j^\varepsilon(k)\right)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}$$

$$\leq\prod_{j\in L\setminus\{1\}}\left(E_{\lambda^\varepsilon}\left[e^{\left(\sum_{k=1}^{N_j}\upsilon_j^\varepsilon(k)\right)(l-1)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}\right]\right)^{1/(l-1)},$$

where we use the generalized Hölder's inequality for the last line. Thus, if we can show for each $j\in L\setminus\{1\}$ that $E_{\lambda^\varepsilon}\exp[(\sum_{k=1}^{N_j}\upsilon_j^\varepsilon(k))(l-1)/E_{\lambda^\varepsilon}\tau_1^\varepsilon]$ is less than a constant for all ε sufficiently small, then we are done.

Such an estimate is straightforward for the case of an unstable equilibrium, i.e., for $j\in L\setminus L_s$, and so we focus on the case $j\in L_s\setminus\{1\}$. For this case, we apply Lemma 10.13 to find that there exists $\tilde{c}>0$ and $\varepsilon_0\in(0,1)$ such that for any $j\in L$ and any distribution $\tilde{\lambda}^\varepsilon$ on $\partial B(O_j)$,

$$P_{\tilde{\lambda}^\varepsilon}(\upsilon_j^\varepsilon/E_{\tilde{\lambda}^\varepsilon}\upsilon_j^\varepsilon>t)\leq e^{-\tilde{c}t}\qquad(10.14)$$

for any $t>0$ and $\varepsilon\in(0,\varepsilon_0)$. Hence, given any $\eta>0$, there is $\bar{\varepsilon}_0\in(0,\varepsilon_0)$ such that for all $\varepsilon\in(0,\bar{\varepsilon}_0)$ and any $j\in L\setminus\{1\}$

$$E_{\lambda^\varepsilon}\left[e^{\upsilon_j^\varepsilon(l-1)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}\right]\leq1+\int_1^\infty P_{\lambda^\varepsilon}(e^{(l-1)\upsilon_j^\varepsilon/E_{\lambda^\varepsilon}\tau_1^\varepsilon}>t)dt$$

$$\leq1+\int_1^\infty P_{\lambda^\varepsilon}\left(\upsilon_j^\varepsilon/E_{\lambda^\varepsilon}\upsilon_j^\varepsilon>\log t\cdot E_{\lambda^\varepsilon}\tau_1^\varepsilon/((l-1)E_{\lambda^\varepsilon}\upsilon_j^\varepsilon)\right)dt$$

$$\leq1+\int_1^\infty t^{-\tilde{c}E_{\lambda^\varepsilon}\tau_1^\varepsilon/((l-1)E_{\lambda^\varepsilon}\upsilon_j^\varepsilon)}dt$$

$$=1+\left(\tilde{c}E_{\lambda^\varepsilon}\tau_1^\varepsilon/((l-1)E_{\lambda^\varepsilon}\upsilon_j^\varepsilon)-1\right)^{-1}\leq1+e^{-\frac{1}{\varepsilon}(h_1-h_j-\eta)},$$

where the last inequality comes from Lemmas 10.2 and 10.15, and by picking the range of ε small if it needs to be.

By using induction and a conditioning argument, it follows that for any $\eta > 0$, for any $j \in L \setminus \{1\}$ and for any $n \in \mathbb{N}$,

$$E_{\lambda^\varepsilon}\left[e^{\left(\sum_{k=1}^n v_j^\varepsilon(k)\right)(l-1)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}\right] \leq \left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)^n.$$

This implies that

$$E_{\lambda^\varepsilon}\left[e^{\left(\sum_{k=1}^{N_j} v_j^\varepsilon(k)\right)(l-1)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}\right] \leq E_{\lambda^\varepsilon}\left[\left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)^{N_j}\right].$$

The next thing we need to know is the distribution of N_j, i.e., $P_{\lambda^\varepsilon}(N_j = n)$ for $n \in \mathbb{N}$. Following a similar argument as in the proof of Lemma 7.3 and the proof of Lemma 7.6, for sufficiently small $\varepsilon > 0$ we find

$$P_{\lambda^\varepsilon}(N_j = n) \leq \left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)}\right)(1 - q_j)^{n-1} q_j,$$

where

$$q_j \doteq \frac{\inf_{x \in \partial B_\delta(O_j)} P_x(\tilde{T}_1 < \tilde{T}_j^+)}{1 - \sup_{y \in \partial B_\delta(O_j)} p(y, \partial B_\delta(O_j))} \geq e^{-\frac{1}{\varepsilon}(W(O_1) - W(O_1 \cup O_j) - h_j + \eta)}. \quad (10.15)$$

Therefore,

$$E_{\lambda^\varepsilon}\left[e^{\left(\sum_{k=1}^{N_j} v_j^\varepsilon(k)\right)(l-1)/E_{\lambda^\varepsilon}\tau_1^\varepsilon}\right]$$

$$\leq E_{\lambda^\varepsilon}\left[\left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)^{N_j}\right] = \sum_{n=1}^\infty \left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)^n P_{\lambda^\varepsilon}(N_j = n)$$

$$\leq \sum_{n=1}^\infty \left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)}\right)\left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)^n (1 - q_j)^{n-1} q_j$$

$$= \frac{\left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)}\right) q_j \left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)}{1 - \left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)(1 - q_j)}$$

$$\leq \frac{\left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)}\right)\left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)}{-e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}/q_j + 1}$$

$$\leq \frac{\left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)}\right)\left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}\right)}{-e^{-\frac{1}{\varepsilon}(h_1 + W(O_1 \cup O_j) - W(O_1) - 2\eta)} + 1}.$$

The second equality holds since $h_1 > w \geq h_j$ and (10.15) imply $(1 - q_j)(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)}) < 1$ for all ε sufficiently small; the last inequality is from (10.15).

Then, we use the fact that for $x \in (0, 1/2)$, $1/(1 - x) \leq 1 + 2x$ to find that

$$
E_{\lambda^\varepsilon} \left[e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)(l-1)/E_{\lambda^\varepsilon} \tau_1^\varepsilon} \right]
$$

$$
\leq \left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)} \right) \left(1 + e^{-\frac{1}{\varepsilon}(h_1 - h_j - \eta)} \right)
$$

$$
\times \left(1 + 2e^{-\frac{1}{\varepsilon}(h_1 + W(O_1 \cup O_j) - W(O_1) - 2\eta)} \right)
$$

$$
\leq \left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)} \right) \left(1 + 5e^{-\frac{1}{\varepsilon}(h_1 + W(O_1 \cup O_j) - W(O_1) - 2\eta)} \right)
$$

$$
\leq 1 \cdot 6 = 6. \tag{10.16}
$$

The third inequality holds due to the fact that $W(O_1) \geq W(O_1 \cup O_j) + h_j$ and the last inequality comes from the assumption that $h_1 > w$ and by picking η to be smaller than $(h_1 - w)/2$. This completes the proof. $\qquad\square$

Lemma 10.17 *Given $\delta > 0$ sufficiently small, and for any distribution λ^ε on $\partial B_\delta(O_1)$, $\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon$ converges in distribution to an $\mathrm{Exp}(1)$ random variable under P_{λ^ε}. Moreover, $E_{\lambda^\varepsilon}(e^{it(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)}) \to 1/(1 - it)$ uniformly on any compact set in $\mathbb{R}$.*

Proof Note that

$$
E_{\lambda^\varepsilon} \left(e^{it(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)} \right) = E_{\lambda^\varepsilon} \left(e^{it(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)} E_{X^\varepsilon(\upsilon_1^\varepsilon)} \left(e^{it((\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon)} \right) \right).
$$

Since

$$
E_{\lambda^\varepsilon} \left(e^{it(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)} \right) = E_{\lambda^\varepsilon} \left(e^{it(E_{\lambda^\varepsilon} \upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \upsilon_1^\varepsilon)} \right)
$$

and we know that $E_{\lambda^\varepsilon} \upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon \to 1$ from the proof of Lemma 10.15, by applying Lemma 10.14 we have $E_{\lambda^\varepsilon}(e^{it(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)}) \to 1/(1 - it)$ uniformly on any compact set in $\mathbb{R}$. Also

$$
\left| E_{\lambda^\varepsilon} \left(e^{it(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)} \right) - E_{\lambda^\varepsilon} \left(e^{it(\upsilon_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)} \right) \right| \leq E_{\lambda^\varepsilon} \left| E_{X^\varepsilon(\upsilon_1^\varepsilon)} \left(e^{it((\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon)} \right) - 1 \right|,
$$

where the right hand side converges to 0 using $E_{\lambda^\varepsilon}(\tau_1^\varepsilon - \upsilon_1^\varepsilon)/E_{\lambda^\varepsilon} \tau_1^\varepsilon \to 0$ and the dominated convergence theorem. The convergence of $\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon$ to an $\mathrm{Exp}(1)$ random variable under P_{λ^ε} and uniform convergence of $E_{\lambda^\varepsilon}(e^{it(\tau_1^\varepsilon / E_{\lambda^\varepsilon} \tau_1^\varepsilon)})$ to $1/(1 - it)$ on compact set in $\mathbb{R}$ follows. $\qquad\square$

10.7 Return Times (Multicycles)

In this subsection, we will extend all the three results to multi-regenerative cycles (when $w \geq h_1$). Recall that the multicycle times $\hat{\tau}_i^\varepsilon$ are defined according to (6.4)

where $\{\mathbf{M}_i^\varepsilon\}_{i\in\mathbb{N}}$ is a sequence of independent and geometrically distributed random variables with parameter $e^{-m/\varepsilon}$ for some $m > 0$ such that $m + h_1 > w$. In addition, $\{\mathbf{M}_i^\varepsilon\}$ is independent of $\{\tau_n^\varepsilon\}$.

Lemma 10.18 *There exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$ and any distribution λ^ε on $\partial B_\delta(O_1)$,*

$$\lim_{\varepsilon\to 0} \varepsilon \log E_{\lambda^\varepsilon} \hat{\tau}_1^\varepsilon = m + \min_{y\in\cup_{k\in L\setminus\{1\}}\partial B_\delta(O_k)} V(O_1, y).$$

Proof Since $\{\mathbf{M}_i^\varepsilon\}$ is independent of $\{\tau_n^\varepsilon\}$ and $E_{\lambda^\varepsilon}\mathbf{M}_i^\varepsilon = e^{m/\varepsilon}$, we apply Lemma 10.15 to complete the proof. $\qquad\qquad\square$

Lemma 10.19 *Given $\delta > 0$, for any distribution λ^ε on $\partial B_\delta(O_1)$, there exist $\tilde{c} > 0$ and $\varepsilon_0 \in (0, 1)$ such that*

$$P_{\lambda^\varepsilon}(\hat{\tau}_1^\varepsilon / E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon > t) \le e^{-\tilde{c}t}$$

for all $t \ge 1$ and $\varepsilon \in (0, \varepsilon_0)$.

Proof We divide the multicycle into a sum of two terms. The first term is the sum of all the hitting times to $\cup_{j\in L\setminus\{1\}}\partial B_\delta(O_j)$, and the second term is the sum of all residual times. That is, $\hat{\tau}_1^\varepsilon = \hat{\upsilon}_1^\varepsilon + (\hat{\tau}_1^\varepsilon - \hat{\upsilon}_1^\varepsilon)$, where

$$\hat{\upsilon}_1^\varepsilon = \sum_{i=1}^{\mathbf{M}_1^\varepsilon} \upsilon_1^\varepsilon(i) \text{ and } \hat{\tau}_1^\varepsilon - \hat{\upsilon}_1^\varepsilon = \sum_{i=1}^{\mathbf{M}_1^\varepsilon}\left(\sum_{j\in L\setminus\{1\}}\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(i,k)\right).$$

As discussed many times, it suffices to show that there exist $\tilde{c} > 0$ and $\varepsilon_0 \in (0, 1)$ such that

$$P_{\lambda^\varepsilon}\left(\hat{\upsilon}_1^\varepsilon / E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon > t\right) \le e^{-\tilde{c}t} \text{ and } P_{\lambda^\varepsilon}\left((\hat{\tau}_1^\varepsilon - \hat{\upsilon}_1^\varepsilon)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon > t\right) \le e^{-\tilde{c}t}$$

for all $t \ge 1$ and $\varepsilon \in (0, \varepsilon_0)$.

The first bound is relatively easy since $\upsilon_1^\varepsilon(i)$ is a sum of approximate exponentials with a tail bound of the given sort, and since the sum of geometrically many independent and identically distributed exponentials is again an exponential distribution.

For the second bound, we use Chebyshev's inequality again as in the proof of Lemma 10.16 to find that it suffices to prove that $E_{\lambda^\varepsilon}e^{(\hat{\tau}_1^\varepsilon - \hat{\upsilon}_1^\varepsilon)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon}$ is less than a constant for all ε sufficiently small. Now due to the independence of $\mathbf{M}_1^\varepsilon$ and $\{\upsilon_j^\varepsilon(i,k)\}$,

we have

$$
E_{\lambda^\varepsilon} e^{(\hat{\tau}_1^\varepsilon - \hat{\upsilon}_1^\varepsilon)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon}
$$

$$
= \sum_{i=1}^{\infty} \left(E_{\lambda^\varepsilon} \left[\prod_{j \in L \setminus \{1\}} e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right] \right)^i \cdot P_{\lambda^\varepsilon}(\mathbf{M}_1^\varepsilon = i)
$$

$$
= e^{-m/\varepsilon} \cdot E_{\lambda^\varepsilon} \left[\prod_{j \in L \setminus \{1\}} e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right]
$$

$$
\cdot \sum_{i=1}^{\infty} \left(E_{\lambda^\varepsilon} \left[\prod_{j \in L \setminus \{1\}} e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right] (1 - e^{-m/\varepsilon}) \right)^{i-1}. \quad (10.17)
$$

To do a further computation, we have to at least make sure that

$$
E_{\lambda^\varepsilon} \left[\prod_{j \in L \setminus \{1\}} e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right] (1 - e^{-m/\varepsilon}) < 1. \quad (10.18)
$$

To see this, we first use the generalized Hölder's inequality to find

$$
E_{\lambda^\varepsilon} \left[\prod_{j \in L \setminus \{1\}} e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right]
$$

$$
\leq \prod_{j \in L \setminus \{1\}} \left(E_{\lambda^\varepsilon} \left[e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)(l-1)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right] \right)^{1/(l-1)}.
$$

Moreover, since $m + h_1 > w$ and $E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon = E_{\lambda^\varepsilon}\tau_1^\varepsilon \cdot E_{\lambda^\varepsilon}\mathbf{M}_1^\varepsilon = e^{m/\varepsilon} E_{\lambda^\varepsilon}\tau_1^\varepsilon$, by the same argument that gives (10.16), for any $\eta > 0$ and $j \in L \setminus \{1\}$

$$
E_{\lambda^\varepsilon} \left[e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)(l-1)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right]
$$

$$
\leq \left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j) - W(O_1 \cup O_j) - h_1 - \eta)} \right) \left(1 + 5e^{-\frac{1}{\varepsilon}(m + h_1 + W(O_1 \cup O_j) - W(O_1) - 2\eta)} \right).
$$

Therefore,

$$
E_{\lambda^\varepsilon} \left[\prod_{j \in L \setminus \{1\}} e^{\left(\sum_{k=1}^{N_j} \upsilon_j^\varepsilon(k) \right)/E_{\lambda^\varepsilon}\hat{\tau}_1^\varepsilon} \right] (1 - e^{-m/\varepsilon}) \leq \prod_{j \in L \setminus \{1\}} s_j^{1/(l-1)},
$$

with

$$s_j \doteq \left(1 \wedge e^{-\frac{1}{\varepsilon}(W(O_j)-W(O_1 \cup O_j)-h_1-\eta)}\right)$$
$$\cdot \left(1 + 5e^{-\frac{1}{\varepsilon}(m+h_1+W(O_1 \cup O_j)-W(O_1)-2\eta)}\right)(1-e^{-m/\varepsilon}).$$

Using $(a \wedge b)(c+d) \leq ac + bd$ for positive numbers a, b, c, d,

$$s_j \leq \left(1 + 5e^{-\frac{1}{\varepsilon}(m+W(O_j)-W(O_1)-3\eta)}\right)(1-e^{-m/\varepsilon}) \leq 1 - e^{-m/\varepsilon}/2,$$

where we use $W(O_j) > W(O_1)$ for the second inequality and pick the range of ε small if it needs to be. Thus, (10.18) holds, and by (10.17)

$$E_{\lambda^\varepsilon} e^{(\hat{\tau}_1^\varepsilon - \hat{\upsilon}_1^\varepsilon)/E_{\lambda^\varepsilon} \hat{\tau}_1^\varepsilon} \leq e^{-m/\varepsilon} \cdot 2 \sum_{i=1}^\infty \left(1-e^{-m/\varepsilon}/2\right)^{i-1} = \frac{2e^{-m/\varepsilon}}{1-\left(1-e^{-m/\varepsilon}/2\right)} = 4.$$

We complete the proof. $\square$

Lemma 10.20 *Given $\delta > 0$, for any distribution λ^ε on $\partial B_\delta(O_1)$, $\hat{\tau}_1^\varepsilon / E_{\lambda^\varepsilon} \hat{\tau}_1^\varepsilon$ converges in distribution to an* Exp(1) *random variable under P_{λ^ε}.*

Proof Let $\mathbf{M}$ be a geometrically distributed random variables with parameter $p \in (0, 1)$ and assume that it is independent of $\{\tau_n^\varepsilon\}$. Then, $E_{\lambda^\varepsilon}\left(\sum_{n=1}^{\mathbf{M}} \tau_n^\varepsilon\right) = E_{\lambda^\varepsilon} \tau_1^\varepsilon / p$ and

$$E_{\lambda^\varepsilon} e^{it\left(\left(p\sum_{n=1}^{\mathbf{M}} \tau_n^\varepsilon\right)/E_{\lambda^\varepsilon} \tau_1^\varepsilon\right)}$$
$$= \sum_{k=1}^\infty \left(E_{\lambda^\varepsilon} e^{it(p\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)}\right)^k (1-p)^{k-1} p = \frac{p E_{\lambda^\varepsilon} e^{it(p\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)}}{1-(1-p)E_{\lambda^\varepsilon} e^{it(p\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)}}.$$
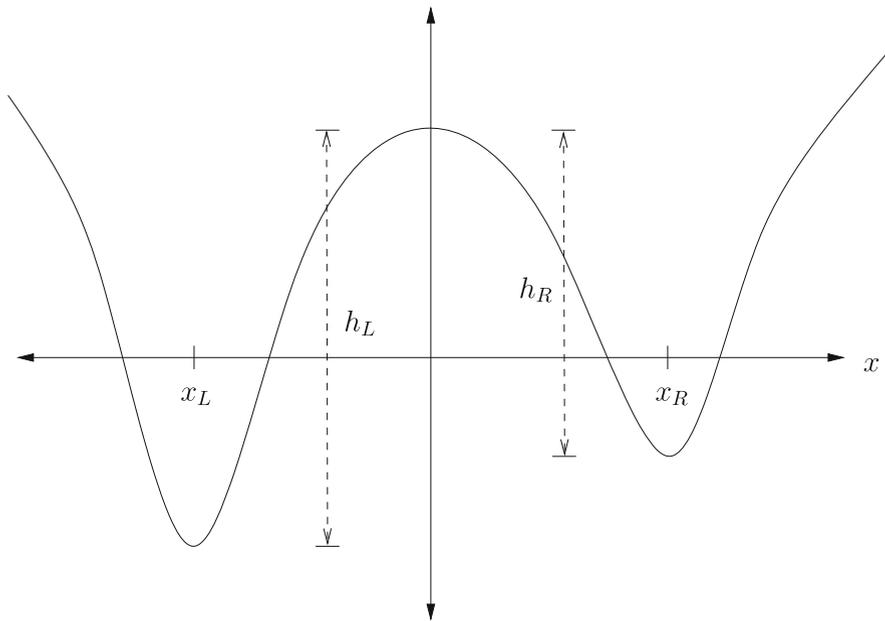
Given any fixed $t \in \mathbb{R}$, consider

$$f_\varepsilon(p) = \frac{p E_{\lambda^\varepsilon} e^{it(p\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)}}{1-(1-p)E_{\lambda^\varepsilon} e^{it(p\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)}} \text{ and } f(p) = \frac{1}{1-it}.$$

According to Lemma 10.17, $E_{\lambda^\varepsilon} e^{it(\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)} \to 1/(1-it)$ uniformly on any compact set in $\mathbb{R}$. This implies that

$$f_\varepsilon(p) = \frac{p E_{\lambda^\varepsilon} e^{it(p\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)}}{1-(1-p)E_{\lambda^\varepsilon} e^{it(p\tau_1^\varepsilon/E_{\lambda^\varepsilon} \tau_1^\varepsilon)}} \to \frac{p/(1-itp)}{1-(1-p)/(1-itp)} = \frac{1}{1-it} = f(p)$$

uniformly on $p \in (0, 1)$. Therefore, if we consider $p^\varepsilon \doteq e^{-m/\varepsilon} \to 0$, it follows from the uniform (in p) convergence that

$$E_{\lambda^\varepsilon} e^{it(\hat{\tau}_1^\varepsilon/E_{\lambda^\varepsilon} \hat{\tau}_1^\varepsilon)} = f_\varepsilon(p^\varepsilon) \to f(0) = \frac{1}{1-it}.$$

Fig. 3 Asymmetric well for lower bound

We complete the proof.   □

11 Sketch of the Proof of Conjecture 4.11 for a Special Case

In this section, we outline the proof of the upper bound on the decay rate (giving a
lower bound on the variance per unit time) that complements Theorem 4.5 for a special
case. Consider $U : \mathbb{R} \to \mathbb{R}$ shown as in Fig. 3.

In particular, assume U is a bounded C^2 function satisfying the following condi-
tions:

Condition 11.1   • *U is defined on a compact interval $D \doteq [\bar{x}_L, \bar{x}_R] \subset \mathbb{R}$ and extends
periodically as a C^2 function.*
  • *U has two local minima at x_L and x_R with values $U(x_L) < U(x_R)$ and $[x_L -
\delta, x_R + \delta] \subset D$ for some $\delta > 0$.*
  • *U has one local maximum at $0 \in (x_L, x_R)$.*
  • *$U(x_L) = 0$, $U(0) = h_L$ and $U(x_R) = h_L - h_R > 0$.*
  • *$\inf_{x \in \partial D} U(x) > h_L$.*

Consider the diffusion process $\{X_t^\varepsilon\}_{t \geq 0}$ satisfying the stochastic differential equa-
tion

$$dX_t^\varepsilon = -\nabla U\left(X_t^\varepsilon\right) dt + \sqrt{2\varepsilon} dW_t, \tag{11.1}$$

where W is a one-dimensional standard Wiener process. Then, there are just two stable equilibrium points $O_1 = x_L$ and $O_2 = x_R$, and one unstable equilibrium point $O_3 = 0$. Moreover, one easily finds that $V(O_1, O_2) = h_L$ and $V(O_2, O_1) = h_R$, and these give that $W(O_1) = V(O_2, O_1)$, $W(O_2) = V(O_1, O_2)$ and $W(O_1 \cup O_2) = 0$ (since $L_s = \{1, 2\}$, this implies that $G_s(1) = \{(2 \to 1)\}$ and $G_s(2) = \{(1 \to 2)\}$). Another observation is that $h_1 \doteq \min_{\ell \in \mathcal{M} \setminus \{1\}} V(O_1, O_\ell) = V(O_1, O_3) = h_L$ in this model.

If $f \equiv 0$, then one obtains

$$R_1^{(1)} \doteq \inf_{y \in A} V(O_1, y) + W(O_1) - W(O_1) = \inf_{y \in A} V(O_1, y);$$

$$R_1^{(2)} \doteq 2 \inf_{y \in A} V(O_1, y) - h_1 = 2 \inf_{y \in A} V(O_1, y) - h_L;$$

$$R_2^{(1)} \doteq \inf_{y \in A} V(O_2, y) + W(O_2) - W(O_1) = \inf_{y \in A} V(O_2, y) + h_L - h_R;$$

$$R_2^{(2)} \doteq 2 \inf_{y \in A} V(O_2, y) + W(O_2) - 2W(O_1) + 0 - W(O_1 \cup O_2)$$

$$= 2 \inf_{y \in A} V(O_2, y) + h_L - 2h_R.$$

Let $A \subset [0, \bar{x}_R]$ and assume that it contains a nonempty open interval, so that we are computing approximations to probabilities that are small under the stationary distribution (the case of bounded and continuous f can be dealt with by approximation, as in the case of the upper bound on the decay rate). We first compute the bounds one would obtain from Theorem 4.5.

Case I. If $x_R \in A$, then $\inf_{y \in A} V(O_1, y) = h_L$ and $\inf_{y \in A} V(O_2, y) = 0$. Thus, the decay rate of variance per unit time is bounded below by

$$\min_{j=1,2} \left[R_j^{(1)} \wedge R_j^{(2)} \right] = \min \{h_L, h_L - 2h_R\} = h_L - 2h_R.$$

Case II. If $A \subset [0, x_R - \delta]$ for some $\delta > 0$ and $\delta < x_R$, then $\inf_{y \in A} V(O_1, y) = h_L$ and $\inf_{y \in A} V(O_2, y) > 0$ (we denote it by $b \in (0, h_R]$). Thus, the decay rate of variance per unit time is bounded below by

$$\min_{j=1,2} \left[R_j^{(1)} \wedge R_j^{(2)} \right] = \min \{h_L, h_L + 2(b - h_R)\} = h_L + 2(b - h_R).$$

Case III. If $A \subset [x_R + \delta, x^*]$ with $U(x^*) = h_L$ for some $\delta > 0$ and $\delta < x^* - x_R$, then $\inf_{y \in A} V(O_1, y) = h_L + \inf_{y \in A} V(O_2, y)$ and $\inf_{y \in A} V(O_2, y) > 0$ (we denote it by $b \in (0, h_R]$). Thus, the decay rate of variance per unit time is bounded below by

$$\min_{j=1,2} \left[R_j^{(1)} \wedge R_j^{(2)} \right] = \min \{h_L + b, h_L + 2(b - h_R)\} = h_L + 2(b - h_R).$$

Case IV. If $A \subset [x^* + \delta, \bar{x}_R]$ with $U(x^*) = h_L$ for some $\delta > 0$ and $x^* > x_R$, then $\inf_{y \in A} V(O_1, y) = h_L + \inf_{y \in A} V(O_2, y)$ and $\inf_{y \in A} V(O_2, y) > 0$ (we denote it

by $\bar{b} > h_R$). Thus, the decay rate of variance per unit time is bounded below by

$$\min_{j=1,2} \left[R_j^{(1)} \wedge R_j^{(2)} \right] = \min \left\{ h_L + \bar{b}, h_L + (\bar{b} - h_R) \right\} = h_L + (\bar{b} - h_R).$$

To find an upper bound for the decay rate of variance per unit time, we recall that

$$\frac{1}{T^\varepsilon} \sum_{j=1}^{N^\varepsilon(T^\varepsilon)-1} \int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \le \frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \le \frac{1}{T^\varepsilon} \sum_{j=1}^{N^\varepsilon(T^\varepsilon)} \int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt$$

with τ_j^ε being the j-th regenerative cycle. In Case I, one might guess that

$$\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \tag{11.2}$$

has approximately the same distribution as the exit time from the shallow well, which has been shown to asymptotically have an exponential distribution with parameter $\exp(-h_R/\varepsilon)$. Additionally, since the exit time from the shallower well is exponentially smaller than τ_j^ε, it suggests that the random variables (11.2) can be taken as independent of $N^\varepsilon(T^\varepsilon)$ when ε is small. We also know that

$$EN^\varepsilon \left(T^\varepsilon \right) / T^\varepsilon \approx 1/E\tau_1^\varepsilon \approx \exp \left(-h_L(\delta)/\varepsilon \right),$$

where $h_L(\delta) \uparrow h_L$ as $\delta \downarrow 0$ and $\approx$ means that quantities on either side have the same exponential decay rate. Using Jensen's inequality to find that $E[N^\varepsilon(T^\varepsilon)]^2 \ge [EN^\varepsilon(T^\varepsilon)]^2$ and then applying Wald's identity, we obtain

$$T^\varepsilon \mathrm{Var} \left(\frac{1}{T^\varepsilon} \int_0^{T^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right)$$

$$\approx \frac{1}{T^\varepsilon} E \left[\sum_{j=1}^{N^\varepsilon(T^\varepsilon)} \int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt - EN^\varepsilon \left(T^\varepsilon \right) E \left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right) \right]^2$$

$$= \frac{1}{T^\varepsilon} E \left(\sum_{j=1}^{N^\varepsilon(T^\varepsilon)} \int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right)^2 - \frac{1}{T^\varepsilon} (E(N^\varepsilon(T^\varepsilon)))^2 \left(E \left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right) \right)^2$$

$$= \frac{1}{T^\varepsilon} EN^\varepsilon \left(T^\varepsilon \right) E \left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right)^2 - \frac{1}{T^\varepsilon} \left[EN^\varepsilon \left(T^\varepsilon \right) \right]^2 \left(E \left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right) \right)^2$$

$$+ \frac{1}{T^\varepsilon} \left(E \left[N^\varepsilon \left(T^\varepsilon \right) \right]^2 - EN^\varepsilon \left(T^\varepsilon \right) \right) \left(E \left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right) \right)^2$$

$$\ge \frac{EN^\varepsilon \left(T^\varepsilon \right)}{T^\varepsilon} \mathrm{Var} \left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A \left(X_t^\varepsilon \right) dt \right)$$

$$\approx \exp \left(-h_L(\delta)/\varepsilon \right) \cdot \exp(2h_R/\varepsilon) = \exp((2h_R - h_L(\delta)) /\varepsilon). \tag{11.3}$$

Letting $\delta \to 0$, we see that the decay rate of variance per unit time is bounded above by $h_L - 2h_R$, which is the same as lower bound found for Case I.

For the other three Cases II, III and IV, the process spends only a very small fraction of the time while in the shallower well in the set A. In fact, using the stopping time arguments of the sort that appear in [12, Chapter 4], the event that the process enters A during an excursion away from the neighborhood of x_R can be accurately approximated (as far as large deviation behavior goes) using independent Bernoulli random variables $\{B_i^\varepsilon\}$ with success parameter $e^{-b/\varepsilon}$, and when this occurs the process spends an order one amount of time in A before returning to the neighborhood of x_R. There is, however, another sequence of independent Bernoulli random variables with success parameter $e^{-h_R/\varepsilon}$, and the process accumulates time in A only up till the time of first success of this sequence.

Then, $\mathrm{Var}(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A(X_t^\varepsilon)\,dt)$ has the same logarithmic asymptotics as $\mathrm{Var}(\sum_{i=1}^{R^\varepsilon} 1_{\{B_i^\varepsilon=1\}})$, where R^ε is geometric with success parameter $e^{-h_R/\varepsilon}$ and independent of the $\{B_i^\varepsilon\}$. Straightforward calculation using Wald's identity then gives the exponential rate of decay $2h_R - 2b$ for Cases II, III and $h_R - \bar{b}$ for Case IV, so according to (11.3), we obtain

$$T^\varepsilon \mathrm{Var}\left(\frac{1}{T^\varepsilon}\int_0^{T^\varepsilon} 1_A(X_t^\varepsilon)\,dt\right) \geq \frac{EN^\varepsilon(T^\varepsilon)}{T^\varepsilon}\mathrm{Var}\left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A(X_t^\varepsilon)\,dt\right) \approx e^{[(2(h_R-b)-h_L(\delta))/\varepsilon]}$$

for Cases II and III and

$$T^\varepsilon \mathrm{Var}\left(\frac{1}{T^\varepsilon}\int_0^{T^\varepsilon} 1_A(X_t^\varepsilon)\,dt\right) \geq \frac{EN^\varepsilon(T^\varepsilon)}{T^\varepsilon}\mathrm{Var}\left(\int_{\tau_{j-1}^\varepsilon}^{\tau_j^\varepsilon} 1_A(X_t^\varepsilon)\,dt\right) \approx e^{[((h_R-\bar{b})-h_L(\delta))/\varepsilon]}$$

for Case IV.

Letting $\delta \to 0$, this means that the decay rate of variance per unit time is bounded above by $h_L + 2(b - h_R)$ for Case II and III, and by $h_L + (\bar{b} - h_R)$ for Case IV which is again the same as the corresponding lower bound.

12 Appendix

Proof of Lemma 7.8 Given a function g, we define the notation

$$I(t_1, t_2; g) \doteq \int_{t_1}^{t_2} g(X_s^\varepsilon) ds,$$

for any $0 \le t_1 \le t_2$. By definition, $\tau_1^\varepsilon = \tau_N$ and observe that

$$I(0, \tau_N; g) = \sum_{\ell=1}^{N} I(\tau_{\ell-1}, \tau_\ell; g) = \sum_{\ell=1}^{\infty} I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\ell \le N\}}$$

$$= \sum_{\ell=1}^{\infty} I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\ell \le \hat{N}\}} + \sum_{\ell=1}^{\infty} I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\hat{N}+1 \le \ell \le N\}}$$

$$+ \sum_{j \in L \setminus \{1\}} \sum_{\ell=1}^{\infty} \left(I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\hat{N}+1 \le \ell \le N, Z_{\ell-1} \in \partial B_\delta(O_j)\}} \right).$$

Since $\hat{N}$ and N are stopping times with respect to the filtration $\{\mathcal{G}_n\}_n$, it implies that $\{\ell \le \hat{N}\} = \{\hat{N} \le \ell - 1\}^c \in \mathcal{G}_{\ell-1}$ and $\{\hat{N} + 1 \le \ell \le N, Z_{\ell-1} \in \partial B_\delta(O_j)\} \in \mathcal{G}_{\ell-1}$. Let

$$\mathfrak{S}_1 = \sum_{\ell=1}^{\infty} I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\ell \le \hat{N}\}} \text{ and}$$

$$\mathfrak{S}_j = \sum_{\ell=1}^{\infty} \left(I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\hat{N}+1 \le \ell \le N, Z_{\ell-1} \in \partial B_\delta(O_j)\}} \right)$$

for all $j \in L \setminus \{1\}$. We find

$$E_x(\mathfrak{S}_1) = \sum_{\ell=1}^{\infty} E_x \left(E_x \left[I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\ell \le \hat{N}\}} \middle| \mathcal{G}_{\ell-1} \right] \right)$$

$$= \sum_{\ell=1}^{\infty} E_x \left(1_{\{\ell \le \hat{N}\}} E_{Z_{\ell-1}} [I(0, \tau_1; g)] \right)$$

$$\le \sup_{y \in \partial B_\delta(O_1)} E_y [I(0, \tau_1; g)] \cdot \left(\sum_{\ell=1}^{\infty} P_x(\hat{N} \ge \ell) \right).$$

In addition, for $j \in L \setminus \{1\}$,

$$E_x(\mathfrak{S}_j) = \sum_{\ell=1}^{\infty} E_x \left(I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\hat{N}+1 \le \ell \le N, Z_{\ell-1} \in \partial B_\delta(O_j)\}} \right)$$

$$= \sum_{\ell=1}^{\infty} E_x \left(E_x \left[I(\tau_{\ell-1}, \tau_\ell; g) \cdot 1_{\{\hat{N}+1 \le \ell \le N, Z_{\ell-1} \in \partial B_\delta(O_j)\}} \middle| \mathcal{G}_{\ell-1} \right] \right)$$

$$= \sum_{\ell=1}^{\infty} E_x \left(1_{\{\hat{N}+1 \le \ell \le N, Z_{\ell-1} \in \partial B_\delta(O_j)\}} E_{Z_{\ell-1}} [I(0, \tau_1; g)] \right)$$

$$\le \sup_{y \in \partial B_\delta(O_j)} E_y [I(0, \tau_1; g)] \cdot \left(\sum_{\ell=1}^{\infty} E_x \left(1_{\{\hat{N}+1 \le \ell \le N, Z_{\ell-1} \in \partial B_\delta(O_j)\}} \right) \right).$$

It is straightforward to see that $\hat{N} = N_1$. This implies that $\sum_{\ell=1}^{\infty} P_x(\hat{N} \geq \ell) = E_x \hat{N} = E_x N_1$. Moreover, observe that for any $j \in L \setminus \{1\}$ $\sum_{\ell=1}^{\infty} 1_{\{\hat{N}+1 \leq \ell \leq N, Z_{\ell-1} \in \partial B_\delta(O_j)\}} = N_j$, which gives that $\sum_{\ell=1}^{\infty} E_x(1_{\{\hat{N}+1 \leq \ell \leq N, Z_{\ell-1} \in \partial B_\delta(O_j)\}}) = E_x N_j$. Hence,

$$E_x(I(0, \tau_N; g)) = \sum_{j \in L} E_x(\mathfrak{S}_j) \leq \sum_{j \in L} \left[\sup_{y \in \partial B_\delta(O_j)} E_y(I(0, \tau_1; g)) \right] \cdot E_x N_j.$$

$\square$

Proof of Lemma 7.9 Let $l = |L|$. For any $j \in L$ and $n \in \mathbb{N}$, $\xi_1^{(j)} = \inf\{k \in \mathbb{N}_0 : Z_k \in \partial B_\delta(O_j)\}$, $\xi_n^{(j)} = \inf\{k \in \mathbb{N} : k > \xi_{n-1}^{(j)} \text{ and } Z_k \in \partial B_\delta(O_j)\}$, i.e., $\xi_n^{(j)}$ is the n-th time of hitting $\partial B_\delta(O_j)$. Moreover, we define $N^{(j)} = \inf\{n \in \mathbb{N} : \xi_n^{(j)} \geq N\}$, recalling that $N \doteq \inf\{n \geq \hat{N} : Z_n \in \partial B_\delta(O_1)\}$ and $\hat{N} \doteq \inf\{n \in \mathbb{N} : Z_n \in \cup_{j \in L \setminus \{1\}} \partial B_\delta(O_j)\}$. Since $\xi_n^{(j)}$ is a stopping time with respect to $\{\mathcal{G}_n\}_n$, we can define the filtration $\{\mathcal{G}_{\xi_n^{(j)}}\}$, and one can verify that $N^{(j)}$ is a stopping time with respect to $\{\mathcal{G}_{\xi_n^{(j)}}\}_n$. As in the proof just given, for any function g and for any $0 \leq t_1 \leq t_2$ we define $I(t_1, t_2; g) \doteq \int_{t_1}^{t_2} g(X_s^\varepsilon) ds$. With this notation and since by definition $\tau_1^\varepsilon = \tau_N$, we can write

$$I(0, \tau_N; g) = \sum_{j \in L} \sum_{\ell=1}^{\infty} I(\tau_{\xi_\ell^{(j)}}, \tau_{\xi_\ell^{(j)}+1}; g) \cdot 1_{\{\ell \leq N^{(j)}-1\}}.$$

Since $(x_1 + \cdots + x_l)^2 \leq l(x_1^2 + \cdots + x_l^2)$ for any $(x_1, \ldots, x_l) \in \mathbb{R}^l$ and $l \in \mathbb{N}$,

$$I(0, \tau_N; g)^2 \leq l \sum_{j \in L} \left(\sum_{\ell=1}^{\infty} I(\tau_{\xi_\ell^{(j)}}, \tau_{\xi_\ell^{(j)}+1}; g) \cdot 1_{\{\ell \leq N^{(j)}-1\}} \right)^2.$$

Now for any $j \in L$, each square term from the right can be written an addition of two sums, where the first sum is summation of $I(\tau_{\xi_\ell^{(j)}}, \tau_{\xi_\ell^{(j)}+1}; g)^2 \cdot 1_{\{\ell \leq N^{(j)}-1\}}$ over all ℓ, and the second sum is twice of summation of $I(\tau_{\xi_\ell^{(j)}}, \tau_{\xi_\ell^{(j)}+1}; g) \cdot 1_{\{\ell \leq N^{(j)}-1\}} I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{\ell \leq N^{(j)}-1\}}$ over k, ℓ with $k < \ell$. For the expected value of the first sum, note that $\{\ell \leq N^{(j)} - 1\} = \{N^{(j)} \leq \ell\}^c \in \mathcal{G}_{\xi_\ell^{(j)}}$, we have

$$\sum_{\ell=1}^{\infty} E_x \left[I(\tau_{\xi_\ell^{(j)}}, \tau_{\xi_\ell^{(j)}+1}; g)^2 1_{\{\ell \leq N^{(j)}-1\}} \right]$$

$$= \sum_{\ell=1}^{\infty} E_x \left[1_{\{\ell \leq N^{(j)}-1\}} E_x \left[I(\tau_{\xi_\ell^{(j)}}, \tau_{\xi_\ell^{(j)}+1}; g)^2 \middle| \mathcal{G}_{\xi_\ell^{(j)}} \right] \right]$$

$$\leq \sup_{y \in \partial B_\delta(O_j)} E_y I(0, \tau_1; g)^2 \sum_{\ell=1}^{\infty} P_x(N^{(j)} - 1 \geq \ell)$$

$$= \sup_{y \in \partial B_\delta(O_j)} E_y I(0, \tau_1; g)^2 E_x(N_j).$$

The last equality holds since $N^{(j)} - 1 = N_j$ (recall that N_j is the number of visits of $\{Z_n\}_{n \in \mathbb{N}_0}$ to $\partial B_\delta(O_j)$ before N including the initial position) this implies that $\sum_{\ell=1}^{\infty} P_x(N^{(j)} - 1 \geq \ell) = \sum_{\ell=1}^{\infty} P_x(N_j \geq \ell) = E_x(N_j)$.

Turning to the expected value of the second sum, by conditioning on $\mathcal{G}_{\xi_\ell^{(j)}}$ gives

$$\sum_{\ell=2}^{\infty} \sum_{k=1}^{\ell-1} E_x \left[I(\tau_{\xi_\ell^{(j)}}, \tau_{\xi_\ell^{(j)}+1}; g) \cdot 1_{\{\ell \leq N^{(j)}-1\}} I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{k \leq N^{(j)}-1\}} \right]$$

$$\leq \sup_{y \in \partial B_\delta(O_j)} E_y(I(0, \tau_1; g)) \sum_{\ell=2}^{\infty} \sum_{k=1}^{\ell-1} E_x \left[1_{\{\ell \leq N^{(j)}-1\}} I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{k \leq N^{(j)}-1\}} \right].$$

Now, since for any $k \leq \ell - 1$, i.e., $k + 1 \leq \ell$, $I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \in \mathcal{G}_{\xi_k^{(j)}+1}$ and $1_{\{\ell \leq N^{(j)}-1\}} \in \mathcal{G}_{\xi_\ell^{(j)}}$, we have

$$E_x \left[1_{\{\ell \leq N^{(j)}-1\}} I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{k \leq N^{(j)}-1\}} \right]$$

$$= E_x \left[I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{\tau_{\xi_1^{(j)}} < N, \dots, \tau_{\xi_\ell^{(j)}} < N\}} \right]$$

$$= E_x \left[E_{Z_{\xi_{k+1}^{(j)}}} \left[1_{\{\tau_{\xi_1^{(j)}} < N, \dots, \tau_{\xi_{\ell-k}^{(j)}} < N\}} \right] 1_{\{\tau_{\xi_{k+1}^{(j)}} < N\}} I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{\tau_{\xi_1^{(j)}} < N, \dots, \tau_{\xi_k^{(j)}} < N\}} \right]$$

$$= E_x \left[E_{Z_{\xi_{k+1}^{(j)}}} \left[1_{\{\ell-k \leq N^{(j)}-1\}} \right] 1_{\{\tau_{\xi_{k+1}^{(j)}} < N\}} I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{k \leq N^{(j)}-1\}} \right]$$

$$\leq \sup_{y \in \partial B_\delta(O_j)} P_y(\ell - k \leq N^{(j)} - 1) E_x \left[I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \cdot 1_{\{k \leq N^{(j)}-1\}} \right]$$

$$= \sup_{y \in \partial B_\delta(O_j)} P_y(\ell - k \leq N_j) E_x \left[E_x \left[I(\tau_{\xi_k^{(j)}}, \tau_{\xi_k^{(j)}+1}; g) \middle| \mathcal{G}_{\xi_k^{(j)}} \right] \cdot 1_{\{k \leq N^{(j)}-1\}} \right]$$

$$\leq \sup_{y \in \partial B_\delta(O_j)} E_y(I(0, \tau_1; g)) \cdot \sup_{y \in \partial B_\delta(O_j)} P_y(\ell - k \leq N_j) \cdot P_x(k \leq N_j).$$

This gives that the expected value of the second sum is less than or equal to

$$\left(\sup_{y \in \partial B_\delta(O_j)} E_y I(0, \tau_1; g) \right)^2 \sum_{\ell=2}^{\infty} \sum_{k=1}^{\ell-1} \sup_{y \in \partial B_\delta(O_j)} P_y(\ell - k \leq N_j) \cdot P_x(k \leq N_j)$$

$$= \left(\sup_{y \in \partial B_\delta(O_j)} E_y I(0, \tau_1; g) \right)^2 \sum_{k=1}^{\infty} \sup_{y \in \partial B_\delta(O_j)} P_y(k \leq N_j) \cdot E_x N_j.$$

Therefore, putting the estimates together gives

$$E_x I(0, \tau_1^\varepsilon; g)^2 \leq 2l \sum_{j \in L} \left[\sup_{y \in \partial B_\delta(O_j)} E_y I(0, \tau_1; g) \right]^2 \cdot E_x N_j \cdot \sum_{\ell=1}^{\infty} \sup_{y \in \partial B_\delta(O_j)} P_y(\ell \leq N_j)$$

$$+ l \sum_{j \in L} \left[\sup_{y \in \partial B_\delta(O_j)} E_y I(0, \tau_1; g)^2 \right] \cdot E_x N_j.$$

$\square$

Proof of Lemma 9.2 The main idea of the proof comes from [18, Theorem 3.16].

Given any $\varepsilon > 0$, we define $g^\varepsilon(t) \doteq E_{\lambda^\varepsilon} S^\varepsilon_{N^\varepsilon(t)}$ for any $t \geq 0$. Conditioning on τ^ε_1 yields

$$g^\varepsilon(t) = \int_0^\infty E_{\lambda^\varepsilon}[S^\varepsilon_{N^\varepsilon(t)}|\tau^\varepsilon_1 = x]dF^\varepsilon(x),$$

where $F^\varepsilon(\cdot)$ is the distribution function of τ^ε_1. Note that

$$E_{\lambda^\varepsilon}\left[S^\varepsilon_{N^\varepsilon(t)}|\tau^\varepsilon_1 = x\right] = \begin{cases} g^\varepsilon(t-x) & \text{if } x \leq t \\ E_{\lambda^\varepsilon}\left[S^\varepsilon_1|\tau^\varepsilon_1 = x\right] & \text{if } x > t \end{cases},$$

which implies

$$g^\varepsilon(t) = \int_0^t g^\varepsilon(t-x)\,dF^\varepsilon(x) + h^\varepsilon(t),$$

with

$$h^\varepsilon(t) = \int_t^\infty E_{\lambda^\varepsilon}\left[S^\varepsilon_1|\tau^\varepsilon_1 = x\right]dF^\varepsilon(x).$$

Since $E_{\lambda^\varepsilon}S^\varepsilon_1 = \int_0^\infty E_{\lambda^\varepsilon}\left[S^\varepsilon_1|\tau^\varepsilon_1 = x\right]dF^\varepsilon(x) < \infty$, we have $h^\varepsilon(t) \leq E_{\lambda^\varepsilon}S^\varepsilon_1$ for all $t \geq 0$. Moreover, if we apply Hölder's inequality first and then the conditional Jensen's inequality, we find that for all $t \geq 0$,

$$h^\varepsilon(t) \leq \left(\int_t^\infty \left(E_{\lambda^\varepsilon}\left[S^\varepsilon_1|\tau^\varepsilon_1 = x\right]\right)^2 dF^\varepsilon(x)\right)^{\frac{1}{2}} \left(\int_t^\infty 1^2 dF^\varepsilon(x)\right)^{\frac{1}{2}}$$

$$\leq \left(1 - F^\varepsilon(t)\right)^{\frac{1}{2}} \left(\int_t^\infty E_{\lambda^\varepsilon}[(S^\varepsilon_1)^2|\tau^\varepsilon_1 = x]dF^\varepsilon(x)\right)^{\frac{1}{2}} \leq \left(1 - F^\varepsilon(t)\right)^{\frac{1}{2}} (E_{\lambda^\varepsilon}(S^\varepsilon_1)^2)^{\frac{1}{2}}.$$

Given $\ell \in (0, c - h_1)$ let $U^\varepsilon \doteq e^{\ell/\varepsilon} E_{\lambda^\varepsilon}\tau^\varepsilon_1$. According to Theorem 8.5, there exists $\varepsilon_0 \in (0, 1)$ and a constant $\tilde{c} > 0$ such that

$$1 - F^\varepsilon(U^\varepsilon) = P_{\lambda^\varepsilon}(\tau^\varepsilon_1/E_{\lambda^\varepsilon}\tau^\varepsilon_1 > e^{\ell/\varepsilon}) \leq e^{-\tilde{c}e^{\ell/\varepsilon}}$$

for any $\varepsilon \in (0, \varepsilon_0)$. Also by Theorem 8.5, $U^\varepsilon < T^\varepsilon$ for all ε small enough. Hence, for any $t \geq U^\varepsilon$,

$$1 - F^\varepsilon(t) \leq 1 - F^\varepsilon(U^\varepsilon) \leq e^{-\tilde{c}e^{\ell/\varepsilon}} \text{ and } h^\varepsilon(t) \leq e^{-\tilde{c}e^{\ell/\varepsilon}/2}(E_{\lambda^\varepsilon}(S^\varepsilon_1)^2)^{\frac{1}{2}}.$$

By Proposition 3.4 in [18], we know that for any $\varepsilon > 0$, for $t \in [0, \infty)$

$$g^\varepsilon(t) = h^\varepsilon(t) + \int_0^t h^\varepsilon(t-x)\,da^\varepsilon(x),$$

where

$$a^\varepsilon(t) \doteq \int_0^\infty E_{\lambda^\varepsilon}\left[N^\varepsilon(t) \mid \tau_1^\varepsilon = x\right] dF^\varepsilon(x) = E_{\lambda^\varepsilon}\left(N^\varepsilon(t)\right).$$

This implies

$$\frac{E_{\lambda^\varepsilon} S_{N^\varepsilon(T^\varepsilon)}^\varepsilon}{T^\varepsilon} = \frac{h^\varepsilon(T^\varepsilon)}{T^\varepsilon} + \frac{1}{T^\varepsilon}\int_0^{T^\varepsilon - U^\varepsilon} h^\varepsilon(T^\varepsilon - x) da^\varepsilon(x) + \frac{1}{T^\varepsilon}\int_{T^\varepsilon - U^\varepsilon}^{T^\varepsilon} h^\varepsilon(T^\varepsilon - x) da^\varepsilon(x),$$

$$\leq \frac{E_{\lambda^\varepsilon} S_1^\varepsilon}{T^\varepsilon} + e^{-\tilde{c}e^{\ell/\varepsilon}/2}(E_{\lambda^\varepsilon}(S_1^\varepsilon)^2)^{\frac{1}{2}}\frac{a^\varepsilon(T^\varepsilon - U^\varepsilon)}{T^\varepsilon} + E_{\lambda^\varepsilon} S_1^\varepsilon \frac{a^\varepsilon(T^\varepsilon) - a^\varepsilon(T^\varepsilon - U^\varepsilon)}{T^\varepsilon},$$

where we use $h^\varepsilon(t) \leq E_{\lambda^\varepsilon} S_1^\varepsilon$ to bound the first term and the third term, and $h^\varepsilon(t) \leq e^{-\tilde{c}e^{\ell/\varepsilon}/2}(E_{\lambda^\varepsilon}(S_1^\varepsilon)^2)^{1/2}$ for any $t \geq U^\varepsilon$ for the second term.

To calculate the decay rate of the first term, we apply Lemma 7.21 to find that for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon} S_1^\varepsilon}{T^\varepsilon} \geq \inf_{x \in A}[f(x) + W(x)] - W(O_1) + c - h_1 - \eta. \quad (12.1)$$

For the decay rate of the second term, given any $\delta > 0$

$$\liminf_{\varepsilon \to 0} -\varepsilon \log\left(e^{-\tilde{c}e^{\ell/\varepsilon}/4}(E_{\lambda^\varepsilon}(S_1^\varepsilon)^2)^{\frac{1}{2}}\frac{a^\varepsilon(T^\varepsilon - U^\varepsilon)}{T^\varepsilon}\right)$$

$$= \frac{\tilde{c}}{4}\liminf_{\varepsilon \to 0} \varepsilon e^{\ell/\varepsilon} + \liminf_{\varepsilon \to 0} -\varepsilon \log\left((E_{\lambda^\varepsilon}(S_1^\varepsilon)^2)^{\frac{1}{2}}\frac{a^\varepsilon(T^\varepsilon - U^\varepsilon)}{T^\varepsilon}\right) = \infty, \quad (12.2)$$

where the last equality holds since $\ell > 0$ implies $\liminf_{\varepsilon \to 0} \varepsilon e^{\ell/\varepsilon} = \infty$ and also because Lemma 7.23 and Corollary 8.3 ensure that

$\liminf_{\varepsilon \to 0} -\varepsilon \log((E_{\lambda^\varepsilon}(S_1^\varepsilon)^2)^{1/2} a^\varepsilon(T^\varepsilon - U^\varepsilon)/T^\varepsilon)$ is bounded below by a constant.

For the last term, note that for any ε fixed, the renewal function $a^\varepsilon(t)$ is subadditive in t (see for example Lemma 1.2 in [17]), so we have $a^\varepsilon(T^\varepsilon) - a^\varepsilon(T^\varepsilon - U^\varepsilon) \leq a^\varepsilon(U^\varepsilon)$. Thus, we apply by Lemma 7.21, Corollary 8.3 and Theorem 8.5 to find that for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log\left(E_{\lambda^\varepsilon} S_1^\varepsilon \frac{a^\varepsilon(T^\varepsilon) - a^\varepsilon(T^\varepsilon - U^\varepsilon)}{T^\varepsilon}\right)$$

$$\geq \liminf_{\varepsilon \to 0} -\varepsilon \log E_{\lambda^\varepsilon} S_1^\varepsilon + \liminf_{\varepsilon \to 0} -\varepsilon \log \frac{a^\varepsilon(U^\varepsilon)}{U^\varepsilon} + \liminf_{\varepsilon \to 0} -\varepsilon \log \frac{U^\varepsilon}{T^\varepsilon}$$

$$\geq \inf_{x \in A}[f(x) + W(x)] - W(O_1) + (c - h_1 - \ell) - \eta. \quad (12.3)$$

Since (12.3) holds for all $\ell > 0$, by sending ℓ to 0, we know that (12.3) holds with $\ell = 0$.

Putting the bounds (12.1), (12.2) and (12.3) with $\ell = 0$ together gives that for any $\eta > 0$, there exists $\delta_0 \in (0, 1)$ such that for any $\delta \in (0, \delta_0)$,

$$\liminf_{\varepsilon \to 0} -\varepsilon \log \frac{E_{\lambda^\varepsilon} S^\varepsilon_{N^\varepsilon(T^\varepsilon)}}{T^\varepsilon} \geq \inf_{x \in A} [f(x) + W(x)] - W(O_1) + c - h_1 - \eta.$$

$\square$

Proof of Lemma 9.5 By the definition of $W(x)$,

$$2 \inf_{x \in A} [f(x) + W(x)] - 2W(O_1) - h_1$$

$$= 2 \inf_{x \in A} [f(x) + \min_{j \in L} (V(O_j, x) + W(O_j))] - 2W(O_1) - h_1$$

$$= \min_{j \in L} \{2 \inf_{x \in A} [f(x) + V(O_j, x)] + 2W(O_j) - 2W(O_1) - h_1\}.$$

Define $Q_j \doteq 2 \inf_{x \in A} [f(x) + V(O_j, x)] + 2W(O_j) - 2W(O_1) - h_1$. Then, it suffices to show that $Q_j \geq R_j^{(2)}$ for all $j \in L$.

For $j = 1$, $Q_1 = 2 \inf_{x \in A} [f(x) + V(O_1, x)] - h_1 = R_1^{(2)}$. For $j \in L \setminus \{1\}$, $Q_j \geq R_j^{(2)}$ if and only if $W(O_j) - h_1 \geq W(O_1 \cup O_j)$. Recall that

$$W(O_j) = \min_{g \in G(j)} \left[\sum_{(m \to n) \in g} V(O_m, O_n) \right] \text{ and}$$

$$W(O_1 \cup O_j) = \min_{g \in G(1, j)} \left[\sum_{(m \to n) \in g} V(O_m, O_n) \right].$$

Therefore, for any $\tilde{g} \in G(j)$ such that $W(O_j) = \sum_{(m \to n) \in \tilde{g}} V(O_m, O_n)$, if we remove the arrow starting from 1, and assume that it goes to i, then it is easy to see that $\hat{g} \doteq \tilde{g} \setminus \{(1, i)\} \in G(1, j)$. Since $V(O_1, O_j) \geq h_1$, we find that

$$W(O_j) - h_1 = \sum_{(m \to n) \in \tilde{g}} V(O_m, O_n) - h_1$$

$$= \sum_{(m \to n) \in \hat{g}} V(O_m, O_n) + V(O_1, O_j) - h_1$$

$$\geq \min_{g \in G(1, j)} \left[\sum_{(m \to n) \in g} V(O_m, O_n) \right] = W(O_1 \cup O_j).$$

$\square$

References

1. Aldous, D., Fill, J.: Reversible markov chains and random walks on graphs (monograph), 2002. Available at https://www.stat.berkeley.edu/users/aldous/RWG/book.pdf
2. Aronson, D.G.: Bounds for the fundamental solution of a parabolic equation. Bull. Am. Math. Soc. **73**(6), 890–896 (1967)
3. Asmussen, S., Glynn, P.W.: Stoch. Simul. Algorithms Anal. Appl. Math. Springer Science+Business Media, LLC (2007)

4. Budhiraja, A., Dupuis, P.: Analysis and Approximation of Rare Events. Representations and Weak Convergence Methods: Number 94 in Probability Theory and Stochastic Modelling. Springer, New York (2019)
5. Collet, P., Martínez, S., San Martín, J.: Quasi-Stationary Distributions. Springer, Berlin (2013)
6. Day, M.V.: On the exponential exit law in the small parameter exit problem. Stochastics **8**(4), 297–323 (1983)
7. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time I. Comm. Pure Appl. Math. **28**, 1–47 (1975)
8. Donsker, M.D., Varadhan, S.R.S.: Asymptotic evaluation of certain Markov process expectations for large time III. Comm. Pure Appl. Math. **29**, 389–461 (1976)
9. Dupuis, P., Liu, Y., Plattner, N., Doll, J.D.: On the infinite swapping limit for parallel tempering. SIAM J. Multiscale Model. Simul. **10**, 986–1022 (2012)
10. Dupuis, P.,Wu, G.-J.: Analysis and optimization of certain parallel Monte Carlo methods in the low temperature limit. Working paper (2020)
11. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. 2. Wiley, New York (1971)
12. Freidlin, M.I., Wentzell, A.D.: Random Perturbations of Dynamical Systems, 3rd edn. Springer, New York (2012)
13. Gilbarg, D., Trudinger, N.S.: Elliptic Partial Differential Equations of Second Order, 2nd edn. Springer, Berlin (1983)
14. Harris, T.E.: The Theory of Branching Processes Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen. Springer, Newyork (1963)
15. Dupuis, P., Doll, J., Nyquist, P.: A large deviations analysis of certain qualitative properties of parallel tempering and infinite swapping algorithms. Appl. Math. and Opt. **78**, 103–144 (2018)
16. Khasminskii, R.: Stochastic Stability of Differential Equations, 2nd edn. Springer, Berlin Heidelberg (2012)
17. Limnios, N., Oprisan, G.: Semi-Markov Processes and Reliability. Statistics for Industry and Technology. Birkhäuser, Boston (2001)
18. Ross, S.: Applied Probability Models with Optimization Applications. Dover Publications, New York (1992)
19. Shwartz, A., Weiss, A.: Large Deviations for Performance Analysis: Queues. Communication and Computing. Chapman and Hall, New York (1995)