Stochastic Processes and their Applications 190 (2025) 104763

Contents lists available at ScienceDirect

Stochastic Processes and their Applications

journal homepage: www.elsevier.com/locate/spa

Essential barrier height and a probabilistic approach in characterizing potential landscape☆

Yao Li [a], Molei Tao [b], Shirou Wang [c,d,*]

[a] Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA, 01002, USA
[b] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
[c] School of Mathematics, Jilin University, Changchun 130012, China
[d] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2G1, Canada

ARTICLE INFO

MSC:
primary 37H10
secondary 60H10
60J22
60J60

Keywords:
Potential landscape
Coupling method
Overdamped Langevin dynamics
Essential barrier height
Non-convexity

ABSTRACT

This paper proposes a probabilistic approach to investigate the shape of landscapes of multi-dimensional potential functions. Under a suitable coupling scheme, two copies of the over-damped Langevin dynamics associated with the potential function are coupled, and the coupling times are collected. Assuming a set of intuitive yet technically challenging conditions on the coupling scheme, it is shown that the tail distributions of the coupling times exhibit qualitatively different dependencies on the noise magnitude for single-well versus multi-well potential functions. More specifically, for convex single-well potentials, the negative tail exponent of the coupling time distribution is uniformly bounded away from zero by the convexity parameter and is independent of the noise magnitude. In contrast, for multi-well potentials, the negative tail exponent decreases exponentially as the noise vanishes, with the decay rate governed by the *essential barrier height*, a quantity introduced in this paper to characterize the non-convex nature of the potential function. Numerical investigations are conducted for a variety of examples, including the Rosenbrock function, interacting particle systems, and loss functions arising in artificial neural networks. These examples not only illustrate the theoretical results in various contexts but also provide crucial numerical validation of the conjectured assumptions, which are essential to the theoretical analysis yet lie beyond the reach of standard technical tools.

1. Introduction

The concept of potential functions is fundamental in both continuous and discrete time dynamics. In continuous-time dynamics, it arises in both conservative systems (e.g., Hamiltonian dynamics) and dissipative systems (e.g., gradient flow and damped mechanical systems). In discrete-time dynamics, it often corresponds to the objective function of an optimization algorithm or, more generally, to a variational inequality. In all these contexts, characterizing the landscape of the potential function, particularly in high dimensions, is often crucial. For example, understanding the existence, locations, and connections of local minima, saddle points, and global minima of neural network training objectives is essential for comprehending both the training dynamics and the generalization capabilities of machine learning models (e.g., [1–6]).

☆ Y. Li was partially supported by NSF DMS-2108628. M. Tao is grateful for partial support from NSF DMS-1847802, NSF DMS-2513699, DOE Grant DE-NA0004261, Cullen-Peck Scholarship, and Emory-GT AI.Humanity Award. S. Wang was partially supported by NSFC (12201244, 12271204), and a faculty development grant from Jilin University.

* Corresponding author at: School of Mathematics, Jilin University, Changchun 130012, China.
E-mail addresses: yaoli@math.umass.edu (Y. Li), mtao@gatech.edu (M. Tao), shirou@jlu.edu.cn (S. Wang).

https://doi.org/10.1016/j.spa.2025.104763
Received 20 April 2024; Received in revised form 8 May 2025; Accepted 3 August 2025
Available online 22 August 2025

This paper proposes a probabilistic approach to understanding how local minima are globally connected in a potential landscape. Let U be a smooth function defined on a regular domain $D \subseteq \mathbb{R}^k (k \geq 1)$ with finitely many local minima $x_1, \dots, x_L$. Generically, denote by $(\varphi^t)_{t\geq 0}$ the negative gradient flow of U. Then each x_i is a stable equilibrium of $(\varphi^t)_{t\geq 0}$ with the basins (of attraction) given by

$$B_i = \{x \in D : \varphi^t(x) \to x_i \text{ as } t \to \infty\}.$$

Call U a single-well potential if it has only one local minimum x_1 (i.e., $L = 1$) such that $D = B_1$, and call U a multi-well potential if $L \geq 2$ and $D = \cup_{1\leq i\leq L} B_i$ up to a Lebesgue null set. A multi-well potential is, in particular, called a double-well potential if $L = 2$. Throughout the paper, the following is always assumed for U:

(U1) The potential function $U \in C^3(D)$, where D is open, convex and connected, such that $\lim_{x\to\partial D} U(x) = \infty$, and if D is unbounded, it further holds that

$$\lim_{x\to\partial D} |\nabla U| = \infty, \quad \lim_{x\to\partial D} |\nabla U(x)| - 2\Delta U(x) = \infty,$$

where $|\cdot|$ denotes the Euclidean norm.

Remark. In the single-well setting, **(U1)** ensures the existence of a global strong solution of (2). In the multi-well setting, further assumptions on the finiteness and non-degeneracy of the saddle points and local minima, as stated in **(U2)** or **(U3)**(iii), guarantee this [7,8].

Our approach makes strong use of the coupling idea in probability. Given two stochastic processes $\boldsymbol{X} = \{\mathcal{X}_t; t \geq 0\}, \boldsymbol{Y} = \{\mathcal{Y}_t; t \geq 0\}$ on $\mathbb{R}^k$, a *coupling* of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is a stochastic process $\{(X_t, Y_t); t \geq 0\}$ on $\mathbb{R}^{2k}$ satisfying the following:

(i) For any $t > 0$, X_t (*resp.* Y_t) has the same law as $\mathcal{X}_t$ (*resp.* $\mathcal{Y}_t$);

(ii) If $X_s = Y_s$ for certain $s > 0$, then $X_t = Y_t$ for all $t \geq s$. The coupling time τ_c is defined to be the first meeting time between X_t and Y_t, i.e.,

$$\tau_c = \inf\{t \geq 0 : X_t = Y_t\}. \tag{1}$$

A coupling is said to be successful if $\tau_c < \infty$ almost surely. Henceforth, a coupling is denoted by (X_t, Y_t) for simplicity and clarity.

Coupling is a classical tool for comparing two probability measures and, in the context of stochastic processes, provides a probabilistic approach to investigate the distributional convergence of the process [9–13]. In this paper, the coupling method is utilized to characterize the landscape of a potential function U. The two stochastic processes being coupled are the overdamped Langevin dynamics, which satisfy the stochastic differential equation (SDE)

$$dZ_t = -\nabla U(Z_t)dt + \varepsilon dB_t, \tag{2}$$

where $\{B_t; t \geq 0\}$ is a k-dimensional Brownian motion and $\varepsilon > 0$ is the noise magnitude. Under effective coupling methods, the coupling time distribution for Langevin dynamics usually exhibits exponential tails (e.g., [12,14]), indicating intuitive connections with the characteristics of the potential function U.

We will focus on how the exponential tails of the coupling time distributions depend on the noise magnitude ε. The main message is that, under a *suitable* coupling scheme, this dependence exhibits both quantitatively and qualitatively different behaviors between potential functions with only a single well and those with multiple wells. More specifically, if denote $r(\varepsilon) = -\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}[\tau_c > t]$, then for a single-well potential U, $r(\varepsilon)$ is uniformly bounded away from zero, independent of ε (see Theorem 1.1); whereas for a multi-well potential U, $r(\varepsilon)$ decreases exponentially with respect to ε, leading to the emergence of a quantity called *essential barrier height*, which quantifies the level of non-convexity of the potential U in a certain sense (see Theorems 1.2 – 1.3).

Various coupling methods have been developed in different contexts since the pioneer work of Doeblin [15]. In this paper, for the purpose of coupling efficiency, we use a *mixture* of two particular coupling methods: reflection coupling and maximal coupling (see Section 2 for details on these two methods). Specifically, for a certain threshold distance $d > 0$, the coupling method between X_t and Y_t is switched between the reflection and maximal coupling in such a way that (X_t, Y_t) evolves according to the reflection (*resp.* maximal) coupling whenever $|X_t - Y_t| > d$ (*resp.* $|X_t - Y_t| \leq d$) until a successful coupling is attained (i.e., $X_t = Y_t$ for some t). This coupling scheme is referred to as the *reflection-maximal coupling*. It was developed in [14] to compute the geometric convergence rate of stochastic dynamics, and a similar scheme is utilized to compute the convergence rate for Markov processes [16].

How should the threshold d be chosen? We note that the maximal coupling is defined in the discrete-time setting, specifically for the time-h sampled chain of the SDE. The choice of d should be chosen so that, if $|X^h_{n-1} - Y^h_{n-1}| < d$, then the distributions of the time-h sampled chains X^h_n and Y^h_n have sufficient overlap to ensure that the probability of successful coupling, $\mathbb{P}[X^h_n = Y^h_n]$, is of order $\mathcal{O}(1)$. Lemma 2.3 shows that by taking $d = \mathcal{O}(\varepsilon\sqrt{h})$, both the coupling probability and the expected distance between X^h_n and Y^h_n can be controlled suitably. Hereafter, we refer to the scheme as the "h-reflection-maximal coupling" when emphasizing the time step size h; otherwise, we simply refer to it as the reflection-maximal coupling, typically assuming a small h without specifying its exact value.

Although the theoretical results established in this paper do not depend on the choice of discretization scheme, in numerical simulations, the Euler–Maruyama scheme is adopted for all numerical examples. This is because its probability density function at any given point can be explicitly computed, which is required for the implementation of the maximal coupling. With additional effort to evaluate the relevant densities, the reflection-maximal coupling can also be adapted to other numerical schemes, such as

the Milstein scheme. Since the primary goal of this paper is to demonstrate the effectiveness of the reflection-maximal coupling method in characterizing the potential landscape, the Euler–Maruyama scheme is used throughout the numerical examples.

The first main result of this paper concerns the single-well potential. Let U be a single-well potential on a convex domain D. The function U is said to be *strongly convex* (with constant $m_0 > 0$) if

$$\langle \nabla U(x) - \nabla U(y), x - y \rangle \geq m_0 |x - y|^2, \quad \forall x, y \in D, \tag{3}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^k$. The supremum of all positive values of m_0 satisfying (3) is called the convexity parameter of U. Henceforth, m_0 always denotes the convexity parameter.

Theorem 1.1. *Let U be a single-well potential satisfying* **(U1)** *and strongly convex with constant $m_0 > 0$. Given any $\delta > 0$, there exists $h_0 > 0$ such that for any $h \in (0, h_0)$, if (X_t, Y_t) is an h-reflection-maximal coupling of two solutions of (2) satisfying $\mathbb{E}[|X_0 - Y_0|] < \infty$, then for any $\varepsilon > 0$, it holds that*

$$\limsup_{t \to \infty} \frac{1}{t} \log \mathbb{P}[\tau_c > t] \leq -m_0 + \delta.$$

Remark. Theorem 1.1 provides only an upper bound for the coupling time in the single-well case, in contrast to the asymptotic characterizations established for the multi-well case in Theorems 1.2 and 1.3 below. Deriving a lower bound would require identifying a mechanism by which two coupled trajectories fail to meet within a sufficiently long time. In the absence of energy barriers, as in the single-well case, such a mechanism is not straightforward. Even under the simplifying assumption that the potential is quadratic, estimating the probability of near-coupling without success involves estimates on the first hitting times of the Ornstein–Uhlenbeck process, for which explicit formulas are generally not available [17]. In practice, the exponential tail of the coupling time distribution for the single-well case is expected to be governed by the smallest eigenvalue of the Hessian at the global minimum; see Section 5.2.

When U has multiple wells, a crucial quantity is the least barrier height of any continuous path connecting two local minima of U. More specifically, given two subsets $A, B \subseteq D$, the communication height between A and B is defined as

$$\Phi(A, B) = \inf_{\substack{\phi \in C([0,1], D), \\ \phi(0) \in A,\ \phi(1) \in B}} \sup_{t \in [0,1]} U(\phi(t)), \tag{4}$$

where the infimum is taken over all continuous paths in D. It is straightforward to observe that $\Phi(A, B) = \Phi(B, A)$.

For a double-well potential U with two local minima x_1, x_2, define the *essential barrier height* as

$$H_U = \min\{\Phi(x_1, x_2) - U(x_1), \Phi(x_1, x_2) - U(x_2)\}, \tag{5}$$

which represents the lower of the two barrier heights that must be crossed when transitioning from one local minimum to the other. In the double-well setting, the potential function is assumed to satisfy the following generic conditions.

(U2) Let $U : D \to \mathbb{R}$ be a double-well potential function satisfying **(U1)** with two local minima x_1 and x_2. The following hold:

(i) The communication height between x_1 and x_2 is attained at a unique saddle point $z^*(x_1, x_2)$, i.e.,

$$U(z^*(x_1, x_2)) = \Phi(x_1, x_2);$$

(ii) U is non-degenerate (i.e., the Hessian of U has only non-zero eigenvalues) at the two local minima x_1, x_2, and at the saddle point $z^*(x_1, x_2)$.

In the multi-well setting, in addition to assumptions on the potential function, several key properties of the coupling scheme are also required; see **(H1)–(H3)** in Section 4. These property assumptions, while technical in form, are supported by intuitive reasoning and are numerically validated in Section 5.

When multiple wells are present, the coupling process is assumed to be initially related to all basins, ensuring that all typical scenarios are considered. More specifically, a probability measure μ on $D \times D$ is said to be fully supported (with respect to all local minima) if for any $\delta > 0$,

$$\mu(B_\delta(x_i) \times B_\delta(x_j)) > 0, \quad i, j = 1, \ldots, L,$$

where $B_\delta(x)$ denotes the ball centered at x with radius δ. A coupling (X, Y) is said to be fully supported if its distribution is fully supported. Analogously, a probability measure μ on D is said to be fully supported if for any $\delta > 0$,

$$\mu(B_\delta(x_i)) > 0, \quad i = 1, \ldots, L.$$

A random variable X is said to be fully supported if its distribution is fully supported. Note that any probability measure equivalent to the Lebesgue measure is fully supported.

Throughout this paper, the notation $x \lesssim y$ (*resp.* $x \gtrsim y$) indicates that x is bounded from above (*resp.* below) by a constant, which is independent of t and ε, multiplied by y. The notation $x \simeq y$ means that both $x \lesssim y$ and $y \gtrsim x$ hold.

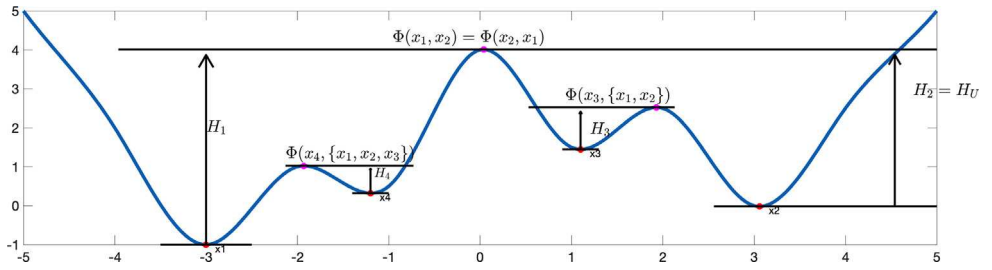


Fig. 1. Example of 1D potential function, local minima, communication height, and essential barrier height H_U. Four relative depths $H_1 = \Phi(x_1,x_2)-U(x_1)$, $H_2 = \Phi(x_2,x_1)-U(x_2)$, $H_3 = \Phi(x_3,\{x_1,x_2\})-U(x_3)$, and $H_4 = \Phi(x_4,\{x_1,x_2,x_3\})-U(x_4)$ are demonstrated. Note that $H_2 > H_3 > H_4$. In this example $H_U = H_2 < H_1$.

Theorem 1.2. *Let U be a double-well potential satisfying* **(U2)**, *and (X_t,Y_t) be a coupling of two solutions of (2) such that (X_0,Y_0) is fully supported. Then, if the coupling (X_t,Y_t) satisfies* **(H1)–(H2)**, *for any $\varepsilon>0$ sufficiently small, it holds that*

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}[\tau_c>t]\simeq -C_\varepsilon e^{-2H_U/\varepsilon^2},$$

where H_U is defined in (5), and $C_\varepsilon>0$ is a constant such that the limit $\lim_{\varepsilon\to0}C_\varepsilon$ exists and depends only on U.

In the general setting of multi-well potentials, in addition to the degeneracy of the critical points and the uniqueness of the saddle, as specified in **(U2)**, the potential function U is also assumed to exhibit distinct potential values and depths corresponding to the different local minima.

(U3) Let $U:D\to\mathbb{R}$ be a multi-well potential function satisfying **(U1)** with local minima $x_1,\dots,x_L$. The following hold:

(i) U has different potential values at the different local minima. In particular, U admits a unique global minimum, denoted by x_1;

(ii) The different basins of potential U admit different depths. More precisely, there exists some $\delta>0$ such that the L local minima of U can be labeled in such a way that

$$\Phi(x_i,\mathcal{M}_{i-1})-U(x_i)\le\min_{\ell<i}\{\Phi(x_\ell,\mathcal{M}_i\backslash x_\ell)-U(x_\ell)\}-\delta,\quad i=1,\dots,L, \tag{6}$$

where $\mathcal{M}_0=D^c$, $\mathcal{M}_i=\{x_1,\dots,x_i\}$, $i=1,\dots,L$;

(iii) Let $\mathcal{M}_i$ be as in (ii). Then for each $i\in\{1,\dots,L\}$, the communication height between x_i and $\mathcal{M}_{i-1}$ is reached at the unique saddle point $z^*(x_i,\mathcal{M}_{i-1})$, i.e.,

$$U(z^*(x_i,\mathcal{M}_{i-1}))=\Phi(x_i,\mathcal{M}_{i-1}).$$

Moreover, U is non-degenerate at all the local minima $x_1,\dots,x_L$, and at the associated saddle points $z^*(x_i,\mathcal{M}_{i-1})$, $1\le i\le L$.

Note that **(U3)**(iii) reduces to **(U2)** when $L=2$. We refer to Fig. 1 for an example of the potential function $U(x)$ in one dimension, which illustrates the local minima, the communication heights, the essential barrier height, the relative depths $\Phi(x_i,\mathcal{M}_i\backslash x_i)-U(x_i)$, and their relationships. It should be noted that the essential barrier height H_U is *not* the highest communication height among the local minima, as indicated by H_1.

Condition **(U3)** comes from a nice work on metastability [18,19], in which a sharp estimate of the first hitting time from a local minimum to an *appropriate set* is rigorously proved. We will extensively apply this result to derive an estimate of the first hitting time to the basin of the global minimum (see Lemma 2.6), naturally introducing the notion of essential barrier height defined in (7) below. This ultimately yields the coupling time estimate for the multi-well case.

We now define the *essential barrier height* in the general context. Let U be a multi-well potential satisfying **(U3)**, with x_1 denoting the (unique) global minimum. The essential barrier height of U is defined as

$$H_U=\max_{2\le i\le L}\big\{\Phi(x_i,x_1)-U(x_i)\big\}. \tag{7}$$

Note that when $L=2$, (7) reduces to (5), so the definitions of essential barrier height for double- and multi-well potentials coincide.

We note that the essential barrier height defined in (7) differs from the usual notion of barrier height in the literature. The latter is a local characterization of the potential landscape by focusing only on the relevant barriers that must be crossed when transitioning from one local minimum to another. In contrast, the essential barrier height considered in this paper is a *global* characterization, as it captures the greatest height of the barriers that must be passed by any continuous path going towards the global minimum from any of the local minima. An equivalent characterization of H_U will be given in Section 2.3.

Theorem 1.3. *Let U be a multi-well potential satisfying* **(U3)**, *and let (X_t,Y_t) be a coupling of two solutions of (2) such that (X_0,Y_0) is fully supported. Then, if the coupling (X_t,Y_t) satisfies* **(H1)–(H3)**, *for any $\varepsilon>0$ sufficiently small, it holds that*

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}[\tau_c>t]\simeq C_\varepsilon e^{-2H_U/\varepsilon^2},$$

where H_U is given in (7), and $C_\varepsilon > 0$ is a constant such that the limit $\lim_{\varepsilon\to 0} C_\varepsilon$ exists and depends only on U.

The intuitive ideas underlying Theorems 1.2 and 1.3, which relate the coupling times to the essential barrier height H_U, are as follows. In the double-well case, the typical scenario is that both processes enter the basin associated with the global minimum and be coupled within that basin, as they overcome a lower barrier when transitioning from the local minimum to the global one than in the reverse direction. This intuition analogously extends to multi-well cases: when the two coupled processes start in different basins, the minimal height of the barriers they must overcome to reach the same basin is always no greater than H_U. Specifically, it is no greater than the lower barrier when transitioning to the basin of the global minimum. Such height can be attained when the initial basins of the two processes are sufficiently "distant" from each other (see Section 2.4).

The essential barrier height, in a certain sense, quantifies the "global non-convexity" of multi-well potentials, which is of crucial importance in non-convex optimization problems arising in various fields. In Section 5, we propose a numerical algorithm to compute the essential barrier height, based on the linear extrapolation of the exponential tails of coupling time distributions. The computed values are validated for both a one-dimensional double-well potential and a multi-dimensional interacting particle system, with numerical results shown to closely match the theoretical values. We further apply this algorithm to detect the loss landscapes of artificial neural networks. In a two-layer neural network model, it is shown that the loss functions of large artificial neural networks (over-parameterized) have lower essential barrier heights than that of small ones (under-parameterized). This is largely consistent with observations in the machine learning community, suggesting a promising criterion for training artificial neural networks based on the essential barrier height of the training loss function.

This paper is organized as follows. Section 2 presents basic facts and results that will be used in the subsequent sections, including estimates for reflection and maximal couplings, first hitting times of Langevin dynamics under multi-well potentials, as well as probability generating functions. Section 3 studies the case of the single-well potential and proves Theorem 1.1. Section 4 investigates both double-well and multi-well potentials, and proves Theorems 1.2 and 1.3. Section 5 explores various examples of single- and multi-well potentials, in which both the theoretical findings and the assumptions on the coupling scheme are numerically verified.

2. Preliminary

This section prepares key preliminary results that will be used in the rest of the paper.

2.1. Reflection coupling and single-well potential

Consider two stochastic processes $\boldsymbol{X}, \boldsymbol{Y}$ satisfying the following stochastic differential equation

$$d Z_t = g(Z_t)dt + \varepsilon d B_t, \quad Z_t \in \mathbb{R}^k \tag{8}$$

with initial conditions μ and ν respectively. Assume that $g : \mathbb{R}^k \to \mathbb{R}^k$ is Lipschitz continuous and satisfies additional conditions, ensuring the unique existence of non-explosive strong solutions of (8) from any initial condition.

A *reflection coupling* of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is a stochastic process $\{(X_t, Y_t); t \geq 0\}$ taking values in $\mathbb{R}^{2k}$ such that $X_0 \sim \mu, Y_0 \sim \nu$, and

$$\begin{aligned} &d X_t = g(X_t)dt + \varepsilon d B_t, \\ &d Y_t = g(Y_t)dt + \varepsilon P_t d B_t, \quad 0 < t < \tau_c; \quad Y_t = X_t, \quad t \geq \tau_c, \end{aligned} \tag{9}$$

where $P_t = I_k - 2e_t e_t^\top$ is the orthogonal matrix in which $e_t = (X_t - Y_t)/|X_t - Y_t|$, and τ_c is the coupling time defined in (1).

The reflection coupling, as its name suggests, is to make the noise terms in X_t and Y_t the mirror reflection of each other with respect to the middle hyperplane between X_t and Y_t [9]. It is a particularly efficient coupling method in high-dimension, achieved by only keeping the noise along the vertical direction (which is one-dimensional) of the hyperplane with noise in other directions being canceled out.

The following proposition states that under the method of reflection coupling, the distributions of coupling time of the overdamped Langevin dynamics along a strongly convex single-well potential have exponential tails, bounding away from zero by the convexity parameter.

Proposition 2.1. *Let U be a single-well potential satisfying* **(U1)**. *Assume that U is strongly convex with constant $m_0 > 0$. Then given any $t_0 > 0$, there exists $c_0 > 0$ such that, if (X_t, Y_t) is a reflection coupling of two solutions of (2) with initial conditions $X_0 = x_0, Y_0 = y_0$, for any $\varepsilon > 0$, it holds that*

$$\mathbb{P}[\tau_c > t] \leq c_0\big(|x_0 - y_0|/2\varepsilon\big)e^{-m_0 t}, \quad \forall t \geq t_0.$$

Proof. Denote $R_t = |X_t - Y_t|/2\varepsilon$. It is not hard to see that $\{R_t; t \geq 0\}$ is a one-dimensional stochastic process satisfying

$$d R_t = -R_t^{-1}\langle \nabla U(X_t) - \nabla U(Y_t), X_t - Y_t\rangle dt + 2\varepsilon d\bar{B}_t, \ 0 \leq t < \tau_c, \tag{10}$$

where $\{\bar{B}_t; t \geq 0\}$ is a one-dimensional Brownian motion.

By the strong convexity of U, the drift term in (10) is upper bounded by $-m_0 R_t$. Thus, for $t \in [0, \tau_c)$, R_t is always bounded by the following one-dimensional Ornstein–Uhlenbeck process $\{S_t; t \geq 0\}$

$$d S_t = -m_0 S_t dt + d\bar{B}_t, \quad S_0 = |x_0 - y_0|/2\varepsilon. \tag{11}$$

Let $\tau_0 = \inf\{t \geq 0 : S_t = 0\}$. It is now sufficient to estimate $\mathbb{P}[\tau_0 > t]$.

By Proposition 1 in [20] (see also [17]), the probability density function of τ_0 has an analytic expression as follows

$$p(t) = \frac{S_0}{\sqrt{2\pi}}\Big(\frac{m_0}{\sinh(t)}\Big)^{3/2} \exp\Big\{\frac{m_0(t-S_0^2)}{2} - \frac{m_0 S_0^2}{2}\coth(m_0 t)\Big\}, \quad t \geq 0. \tag{12}$$

Note that

$$p(t) \leq \frac{S_0}{\sqrt{2\pi}}\Big(\frac{m_0}{\sinh(t)}\Big)^{3/2} \exp\Big\{\frac{m_0(t-S_0^2)}{2}\Big\} = c_0 S_0 e^{\frac{m_0 t}{2}} \big/ (e^{m_0 t} - e^{-m_0 t})^{3/2}$$

where c_0 is a constant independent of t, x_0, y_0 and ε. Thus, we have

$$\mathbb{P}[\tau_0 \geq t] = \int_t^\infty p(s)ds \leq \frac{c_0|x_0 - y_0|}{2\varepsilon}\int_t^\infty \frac{e^{2m_0 s}}{(e^{2m_0 s}-1)^{3/2}}ds.$$

Note that for any $p \in (0,1)$,

$$e^{2m_0 s} - 1 \geq p e^{2m_0 s}, \quad \forall s \geq |\ln(1-p)|/2m_0.$$

Thus, for any given $t_0 > 0$, by letting $p \in (0,1)$ be such that $p \geq 1 - e^{-2m_0 t_0}$ and suitably enlarging the constant c_0, one obtains

$$\mathbb{P}[\tau_0 \geq t] \leq \frac{c_0|x_0 - y_0|}{2\varepsilon}\int_t^\infty \frac{e^{2m_0 s}}{(e^{2m_0 s})^{3/2}}ds \leq c_0\big(|x_0 - y_0|/2\varepsilon\big)e^{-m_0 t}, \quad \forall t \geq t_0,$$

where c_0 is independent of t, x_0, y_0 and ε. □

2.2. Maximal coupling and estimations

Let μ_1 and μ_2 be two probability distributions on $\mathbb{R}^k$. Call (X,Y) a coupling of μ_1 and μ_2 if $X \sim \mu_1, Y \sim \mu_2$. By the well-known coupling inequality (see, for instance, Lemma 3.6 in [21]),

$$\mathrm{TV}(\mu_1,\mu_2) \leq 2\mathbb{P}[X \neq Y], \tag{13}$$

where $\mathrm{TV}(\mu_1,\mu_2) := 2\sup_{A\subseteq\mathbb{R}^k}|\mu_1(A) - \mu_2(A)|$ denotes the total variation distance between probability measures on $\mathbb{R}^k$. A coupling (X,Y) is said to be a *maximal coupling* if the equality in (13) is attained, i.e., the probability $\mathbb{P}[X = Y]$ is maximized.

A particular way to obtain maximal coupling is as follows: Denote the "minimum" distribution of μ_1 and μ_2 by $\nu(\cdot) = \alpha^{-1}\min\{\mu_1(\cdot),\mu_2(\cdot)\}$, where α is the normalizer satisfying $\alpha = \mathbb{P}[X = Y]$. With probability α, let $X = Y \sim \nu$, and with probability $(1-\alpha)$, let X and Y be independently sampled such that

$$X \sim (1-\alpha)^{-1}(\mu_1 - \alpha\nu), \quad Y \sim (1-\alpha)^{-1}(\mu_2 - \alpha\nu). \tag{14}$$

It is not hard to verify that $\mathbb{P}[X \neq Y] = \mathrm{TV}(\mu_1,\mu_2)/2$ (see [22], Theorem 1).

In the context of stochastic processes, the maximal coupling is defined in terms of conditional distributions of the associated discrete-time chains. Let $\{(X_n^h, Y_n^h); n \geq 0\}$ be the time-h sampled chain of a coupling of two solutions of (8). Assume that at step $n-1$, (X_{n-1}^h, Y_{n-1}^h) takes the value $(x,y) \in \mathbb{R}^k \times \mathbb{R}^k$, Then (X_n^h, Y_n^h) is a maximal coupling at step n if

$$\mathrm{TV}(\mu_x,\mu_y) = 2\mathbb{P}[X_n^h \neq Y_n^h | X_{n-1}^h = x, Y_{n-1}^h = y]. \tag{15}$$

where μ_x and μ_y denote the probability distribution of X_n^h and Y_n^h conditioning on $X_{n-1}^h = x$ and $Y_{n-1}^h = y$, respectively.

In the proof of Theorem 1.1, a key step is to bound the expected distance between two coupled processes under the maximal coupling. This estimate can be derived using the independent coupling, in which the two coupled random variables are independent.

Proposition 2.2. *Let (X,Y) be a coupling of two random variables such that $X \sim \mu_1, Y \sim \mu_2$. Assume that $\alpha := \mathbb{P}[X = Y] < 1$. Then*

$$\mathbb{E}_{\max}[|X-Y|^2] \leq \frac{2}{1-\alpha}\mathbb{E}_{\mathrm{ind}}[|X-Y|^2]$$

where $\mathbb{E}_{\max}$ and $\mathbb{E}_{\mathrm{ind}}$ denote expectations with respect to the maximal coupling and independent coupling, respectively.

Proof. By the construction of the maximal coupling, we have

$$\begin{aligned}\mathbb{E}_{\max}[|X-Y|^2] &= (1-\alpha)\int_{\mathbb{R}^{2k}}|x-y|^2\frac{(\mu_1-\alpha\nu)(dx)}{1-\alpha}\cdot\frac{(\mu_2-\alpha\nu)(dy)}{1-\alpha}\\ &\leq \frac{1}{1-\alpha}\int_{\mathbb{R}^{2k}}|x-y|^2\mu_1(dx)\mu_2(dy) + \frac{1}{1-\alpha}\int_{\mathbb{R}^{2k}}|x-y|^2(\alpha\nu)(dx)(\alpha\nu)(dy)\end{aligned}$$

where $\nu = \alpha^{-1}\min\{\mu_1,\mu_2\}$. Hence,

$$\mathbb{E}_{\max}[|X-Y|^2] \leq \frac{2}{1-\alpha}\int_{\mathbb{R}^{2k}}|x-y|^2\mu_1(dx)\mu_2(dy) = \frac{2}{1-\alpha}\mathbb{E}_{\mathrm{ind}}[|X-Y|^2].$$ □

As shown in Proposition 2.2, in the context stochastic processes, obtaining an upper bound of $\mathbb{E}_{\max}[|X_n^h - Y_n^h|^2]$ requires that $\alpha_n := \mathbb{P}[X_n^h = Y_n^h]$ remains uniformly bounded away from 1, independent of n. Although in practical simulations, it is rarely observed that $\mathbb{P}[X_n^h = Y_n^h]$ exceeds 0.8, a rigorous theoretical verification of this uniform bound remains challenging. To address this issue, we introduce a modified construction of maximal coupling. For any given $\alpha_0 \in (0, 1]$, define $\tilde{\alpha}_n = \min\{\alpha_n, \alpha_0\}$ where $\alpha_n = \mathbb{P}[X_n^h = Y_n^h]$. Let $\mu_{1,n}$ and $\mu_{2,n}$ denote the distributions of X_n^h and Y_n^h, respectively, and define $\nu_n = \alpha_n^{-1} \min\{\mu_{1,n}(\cdot), \mu_{2,n}(\cdot)\}$. Then with probability $\tilde{\alpha}_n$, let $X_n^h = Y_n^h \sim \nu_n$, and with probability $(1 - \tilde{\alpha}_n)$, let X_n^h and Y_n^h be independently sampled according to

$$X_n^h \sim (1 - \tilde{\alpha}_n)^{-1}(\mu_{1,n} - \tilde{\alpha}_n \nu_n), \quad Y_n^h \sim (1 - \tilde{\alpha}_n)^{-1}(\mu_{2,n} - \tilde{\alpha}_n \nu_n) \tag{16}$$

such that $X_n^h \neq Y_n^h$. This modification ensures $\mathbb{P}[X_n^h = Y_n^h] = \tilde{\alpha}_n \leq \alpha_0$, so the coupling probability is uniformly bounded by α_0. Note that (16) reduces to the standard maximal coupling (14) when $\alpha_0 \geq \alpha_n$.

The modified construction of the maximal coupling in (16) is referred to as the α_0-maximal coupling for $\alpha_0 \in (0, 1]$. Henceforth, the term "maximal coupling" refers to the α_0-maximal coupling with α_0 fixed at 0.8.

Under the reflection-maximal coupling scheme, a maximal coupling is implemented whenever triggered in the previous step. More specifically, (X_n^h, Y_n^h) is a maximal coupling, if at the previous step $n-1$, the distance between X_{n-1}^h and Y_{n-1}^h does not exceed a threshold d. In the numerical implementation, the distance between X_n^h and Y_n^h is evaluated at each step n to determine whether maximal coupling should be triggered for the next step $n+1$. If the condition is not met, reflection coupling is applied instead at step $n+1$.

The triggering of maximal coupling is a crucial mechanism, especially in the numerical schemes, for ensuring a successful coupling. It guarantees a positive success rate of coupling in the following step when the two processes are sufficiently close. In the absence of maximal coupling, numerical errors may cause two processes to "miss" each other, even if they should theoretically be coupled successfully. Moreover, maximal coupling exhibits robustness to small perturbations, making it a reliable method in numerical simulations.

The following lemma shows that in the single-well setting, choosing the threshold $d = \mathcal{O}(\varepsilon\sqrt{h})$ ensures both an $\mathcal{O}(1)$ coupling probability and a uniform bound on the expected one-step distance between the two processes. It provides crucial estimates for the proof of Lemma 3.3 in Section 3.

Lemma 2.3. *Let U be a single-well potential satisfying* **(U1)** *which is strongly convex, and let (X_n^h, Y_n^h) be a coupling of the time-h sampled chains of two solutions of (2). Assume for $n \geq 1$, (X_n^h, Y_n^h) is a maximal coupling conditional on $X_{n-1}^h = x_0, Y_{n-1}^h = y_0$, where $x_0, y_0 \in \mathbb{R}^k$ satisfy $|x_0 - y_0| \leq d = 2\varepsilon\sqrt{h}$. Then the following hold:*

(i) There exists a constant $\gamma \in (0, 1)$ such that for any $n \geq 1$ and $h > 0$,

$$\mathbb{P}[|X_n^h - Y_n^h| > 0 | X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \leq \gamma.$$

(ii) For any $n \geq 1$ and any $h > 0$ sufficiently small,

$$\mathbb{E}[|X_n^h - Y_n^h| \Big| X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \leq c_1 \varepsilon \sqrt{h}.$$

where the constant $c_1 > 0$ is independent of h, ε, and n.

Proof. (i) By definition, the one-step conditional probability

$$\mathbb{P}[X_n^h = Y_n^h | X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \tag{17}$$

is maximized under the standard maximal coupling (i.e., the 1-maximal coupling). In particular, for any alternative coupling method, such as the reflection coupling, the probability in (17) is no greater than that achieved under the maximal coupling. Hence, it suffices to establish that under the reflection coupling, with initial condition $X_0 = x_0, Y_0 = y_0$ and $|x_0 - y_0| \leq d = 2\varepsilon\sqrt{h}$, the probability $\mathbb{P}[X_h = Y_h]$ remains uniformly away from 0 for all sufficiently small $h > 0$.

Denote by m_0 the convexity parameter of U. From the proof of Proposition 2.1, in the single-well case, the coupling time of the reflection coupling is bounded by that of one-dimensional Ornstein–Uhlenbeck process $\{S_t\}$ governed by (11), whose probability density function $p(t)$ is given by (12). Without loss of generality, assume $m_0 = 1$. Then for any sufficiently small $h > 0$ and $0 < t < h$,

$$\begin{aligned} p(t) &\geq a_0 S_0 \exp\left\{ \frac{t}{2} - \frac{S_0^2}{2} \cdot \frac{e^{2t} + 1}{e^{2t} - 1} \right\} \Big/ \left(e^t - e^{-t}\right)^{3/2} \\ &\geq a_0 S_0 \exp\left\{ -\frac{2S_0^2}{e^{2t} - 1} \right\} \Big/ \left(e^{\frac{2t}{3}} - e^{-\frac{4t}{3}}\right)^{3/2} \\ &\geq a_0 S_0 e^{-\frac{S_0^2}{t}} \Big/ t^{\frac{3}{2}}, \end{aligned}$$

for some constant $a_0 > 0$ independent of t and h. Integrating over $[0, h]$ yields

$$\mathbb{P}[X_h = Y_h] = \int_0^h p(t)dt \geq a_0 S_0 \int_0^h e^{-\frac{S_0^2}{t}} t^{-\frac{3}{2}} dt.$$

Applying the change of variable $u = S_0^2/t$ and using the assumption $S_0 = |x_0 - y_0|/2\varepsilon \leq \sqrt{h}$ yields

$$\mathbb{P}[X_h = Y_h] \geq a_0 \int_{S_0^2/h}^{\infty} e^{-u} u^{-\frac{1}{2}} du \geq a_0 \int_1^{\infty} e^{-u} u^{-\frac{1}{2}} du := a_1 > 0. \tag{18}$$

Now, (18) implies that under the 1-maximal coupling,

$$\mathbb{P}[X_n^h = Y_n^h | X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \geq a_1 > 0. \tag{19}$$

Thus, for the α_0-maximal coupling, by setting $\tilde{a}_1 := \min\{a_1, \alpha_0\} < 1$, (19) yields

$$\mathbb{P}[X_n^h = Y_n^h | X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \geq \tilde{a}_1 > 0,$$

and hence

$$\mathbb{P}[|X_n^h - Y_n^h| > 0 | X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \leq 1 - \tilde{a}_1 := \gamma.$$

In particular, $\gamma \in (0,1)$ is independent of h and n.

(ii) Let X_t and Y_t be solutions of (2) with the initial condition $X_0 = x_0$ and $Y_0 = y_0$. Under the independent coupling, where the noise terms driving X_t and Y_t are independent, the following holds

$$\mathbb{E}_{\text{ind}}[|X_h - Y_h|^2] \leq 6\varepsilon^2 h. \tag{20}$$

To verify (20), apply Dynkin's formula to obtain

$$\mathbb{E}_{\text{ind}}[|X_h - Y_h|^2] = |x_0 - y_0|^2 + \mathbb{E}_{\text{ind}}\Big[\int_0^h \big(-2\langle \nabla U(X_s) - \nabla U(Y_s), X_s - Y_s\rangle + 2\varepsilon^2\big) ds\Big],$$

which, together with the strong convexity of U in (3), leads to

$$\mathbb{E}_{\text{ind}}[|X_h - Y_h|^2] \leq |x_0 - y_0|^2 + 2\varepsilon^2 h.$$

Since $|x_0 - y_0| \leq 2\varepsilon\sqrt{h}$, it follows that $|x_0 - y_0|^2 \leq 4\varepsilon^2 h$, and thus (20) follows.

Now, consider $\mathbb{E}[|X_n^h - Y_n^h| \big| X_{n-1}^h = x_0, Y_{n-1}^h = y_0]$ under the (α_0-)maximal coupling. Applying Proposition 2.2 with $\alpha = \alpha_0$, together with the bound in (20), it follows that

$$\mathbb{E}[|X_n^h - Y_n^h|^2 \big| X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \leq \frac{2}{1-\alpha_0}\mathbb{E}_{\text{ind}}[|X_h - Y_h|^2] \leq \frac{12}{1-\alpha_0}\varepsilon^2 h.$$

Applying Hölder's inequality, one obtains

$$\mathbb{E}[|X_n^h - Y_n^h| \big| X_{n-1}^h = x_0, Y_{n-1}^h = y_0] \leq c_1 \varepsilon \sqrt{h}$$

where $c_1 > 0$ is a constant independent of h, n and ε. This completes the proof of (ii). □

In concluding this subsection, we remark that the maximal coupling, as employed for numerical efficiency, is formulated for discrete-time processes. However, the theoretical results in this paper are presented in the continuous-time setting. To ensure consistency between the discrete-time numerical scheme and its continuous-time theoretical counterpart, we assume that when the maximal coupling is applied, the intermediate values of the processes between the discrete steps are disregarded. That is, only the values at times $t = nh$ are relevant, and the behavior of the coupling process at times between the discrete steps has no influence on the analysis.

2.3. An equivalent characterization of essential barrier height

Let $U : D \to \mathbb{R}$ be a multi-well potential satisfying **(U3)**. Throughout the paper, let the L local minima of U be labeled according to (6), with x_1 being the unique global minimum.

The following proposition provides an equivalent characterization of the essential barrier height H_U defined in (7).

Proposition 2.4. *Let U be a multi-well potential on D with L local minima $x_i, i = 1, \dots, L$. Then*

$$H_U = \max_{2\leq i\leq L}\big\{\Phi(x_i, \mathcal{M}_{i-1}) - U(x_i)\big\}, \tag{21}$$

where $\mathcal{M}_i$ is defined as in (6). In particular,

$$H_U = \Phi(x_2, x_1) - U(x_2). \tag{22}$$

Proof. Since $x_1 \in \mathcal{M}_{i-1}$ for all $i \in \{2, \dots, L\}$, it follows that

$$H_U \geq \Phi(x_i, x_1) - U(x_i) \geq \Phi(x_i, \mathcal{M}_{i-1}) - U(x_i), \tag{23}$$

which yields

$$H_U \geq \max_{2\leq i\leq L}\big\{\Phi(x_i, \mathcal{M}_{i-1}) - U(x_i)\big\}. \tag{24}$$

It remains to prove that the inequality in (24) is in fact an equality. Suppose, by contradiction, that the inequality is strict; that is,

$$\Phi(x_i, \mathcal{M}_{i-1}) - U(x_i) < H_U, \quad \forall i \in \{2, \dots, L\}. \tag{25}$$

Under this assumption, we *claim* that for each $i \in \{2, \dots, L\}$, one has

$$\Phi(x_i, x_1) - U(x_i) < H_U, \tag{26}$$

which further implies

$$H_U = \max_{2\le i\le L} \{\Phi(x_i, x_1) - U(x_i)\} < H_U,$$

yielding a contradiction.

Now, it only needs to prove (26). Fix $i_0 \in \{2, \dots, L\}$. By (25), there exists $x_{i_1} \in \mathcal{M}_{i_0-1}$ such that

$$\Phi(x_{i_0}, x_{i_1}) - U(x_{i_0}) = \Phi(x_{i_0}, \mathcal{M}_{i_0-1}) - U(x_{i_0}) < H_U. \tag{27}$$

Since $x_{i_1} \in \mathcal{M}_{i_0-1}$, it follows that $i_1 < i_0$. Moreover, the ordering in (6) yields

$$\Phi(x_{i_0}, \mathcal{M}_{i_0-1}) - U(x_{i_0}) < \Phi(x_{i_1}, \mathcal{M}_{i_0}\backslash x_{i_1}) - U(x_{i_1}).$$

Using the fact that $\Phi(x_{i_0}, \mathcal{M}_{i_0-1}) = \Phi(x_{i_0}, x_{i_1})$ and $\Phi(x_{i_1}, \mathcal{M}_{i_0}\backslash x_{i_1}) \le \Phi(x_{i_1}, x_{i_0})$, we obtain

$$\Phi(x_{i_0}, x_{i_1}) - U(x_{i_0}) < \Phi(x_{i_1}, x_{i_0}) - U(x_{i_1}).$$

Hence, $U(x_{i_1}) < U(x_{i_0})$.

If $i_1 = 1$, then (26) follows directly from (27). Otherwise, the same argument can be applied recursively: for $i_1 \in \{2, \dots, L\}$, there exists $i_2 \in \mathcal{M}_{i_1-1}$ such that $\Phi(x_{i_1}, x_{i_2}) - U(x_{i_1}) < H_U$, with $i_2 < i_1$ and $U(x_{i_2}) < U(x_{i_1})$. Continuing inductively, a finite sequence of indices $i_0 > i_1 > \cdots > i_k = 1$, with finite $k \le L$, is obtained such that

$$U(x_{i_k}) < \cdots < U(x_{i_0}). \tag{28}$$

Hence,

$$\begin{aligned}\Phi(x_{i_0}, x_1) - U(x_{i_0}) &\le \max_{0\le j<k} \Phi(x_{i_j}, x_{i_{j+1}}) - U(x_{i_0}) \\ &\le \max_{0\le j<k} \left\{\Phi(x_{i_j}, x_{i_{j+1}}) - U(x_{i_j})\right\} < H_U,\end{aligned}$$

where the final inequality follows from (27) and (28). This obtains (26).

Since the local minima x_i are labeled according to (6), identity (22) follows directly. □

Remark 2.5. In fact, x_2 is the unique local minimum such that (22) is satisfied. In other words,

$$\Phi(x_i, x_1) - U(x_i) < H_U, \quad \forall i > 2. \tag{29}$$

To see this, suppose for the sake of contradiction that there exists $i_0 > 2$ such that (29) does not hold. Then it follows that $\Phi(x_{i_0}, x_1) - U(x_{i_0}) = H_U$, as it always holds that $\Phi(x_{i_0}, x_1) - U(x_{i_0}) \le H_U$. Since, by (6), $\Phi(x_{i_0}, \mathcal{M}_{i_0-1}) - U(x_{i_0}) < H_U$, there exists an index $1 < i_1 < i_0$ such that

$$\Phi(x_{i_0}, x_{i_1}) - U(x_{i_0}) = \Phi(x_{i_0}, \mathcal{M}_{i_0-1}) - U(x_{i_0}) < H_U.$$

This leads to a contradiction, as this argument can be applied repeatedly until eventually arriving at $i_k = 1$ for some finite k.

2.4. Multi-well potential and first hitting time

Given a multi-well potential $U : D \to \mathbb{R}$ satisfying **(U3)**, let $Z = \{Z_t; t \ge 0\}$ be a solution of (2) and $A \subseteq D$ be a subset. Denote the *first hitting time* of Z_t to A as

$$\kappa_Z(A) = \inf\{t > 0 : Z_t \in A\}. \tag{30}$$

It is well known, from large deviation theory, that the first hitting time from a local minimum x_i to an appropriate subset is asymptotically exponentially distributed, with the exponent determined by the associated (i.e., *local*) barrier heights [18,19,23,24]. The essential barrier height H_U plays a similar role in a *global* sense, characterizing the first hitting time to the basin of the global minimum from any local minimum.

Lemma 2.6. *Let $Z = \{Z_t; t \ge 0\}$ be a solution of (2) with initial condition $Z_0 = z$. Then for any $t > 0$ and any $\varepsilon > 0$ sufficiently small,*

$$\mathbb{P}_z[\kappa_Z(B_1) > t] \le A_{z,\varepsilon} \exp\left\{-C_\varepsilon e^{-2H_U/\varepsilon^2} t\right\}, \tag{31}$$

where $C_\varepsilon > 0$ and $A_{z,\varepsilon} > 0$ are constants such that the limit $\lim_{\varepsilon\to 0} C_\varepsilon$ exists, and $A_{z,\varepsilon}$ depends on both the initial value z and the noise strength ε, but is independent of t. Moreover, if the initial condition Z_0 is fully supported with distribution μ, then

$$\mathbb{P}_z[\kappa_Z(B_1) > t] \simeq A_{\mu,\varepsilon} \exp\left\{-C_\varepsilon e^{-2H_U/\varepsilon^2} t\right\}, \tag{32}$$

where constant $A_{\mu,\varepsilon} > 0$ depends on both μ and ε.

The proof of Lemma 2.6 follows closely the approach in [18,19], relying on estimates for the eigenvalues and eigenfunctions of the generator. As it is not directly relevant to the main focus of the paper, the proof is deferred to Appendix A.

Define

$$\mathcal{I} = \{1 \le i \le L : \Phi(x_2, x_i) = \Phi(x_2, x_1)\} \tag{33}$$

to be the set of indices corresponding to local minima whose communication height with x_2 equals that between x_2 and x_1. Let

$$\boldsymbol{B}_1 = \bigcup_{i \in \mathcal{I}} B_i. \tag{34}$$

Clearly, $B_1 \subseteq \boldsymbol{B}_1$ since $1 \in \mathcal{I}$. As argued in Remark A.4, the conclusion in Lemma 2.6 remains valid if the set B_1 is replaced by the larger set $\boldsymbol{B}_1$, that is,

$$\mathbb{P}_z[\kappa_{\boldsymbol{Z}}(\boldsymbol{B}_1) > t] \simeq A_{\mu,\varepsilon} \exp\left\{-C_\varepsilon e^{-2H_U/\varepsilon^2} t\right\}, \tag{35}$$

for process $\boldsymbol{Z}$ with initial distribution μ.

The following proposition establishes a property for indices not belonging to $\mathcal{I}$.

Proposition 2.7. *For any $j \in \{1, \dots, L\} \backslash \mathcal{I}$, $\Phi(x_1, x_j) - U(x_1) > H_U$.*

Proof. By the definition of $\mathcal{I}$, it follows directly from (22) that

$$\Phi(x_2, x_i) - U(x_2) = H_U, \quad \forall i \in \mathcal{I}.$$

Thus, for $j \notin \mathcal{I}$, it follows that either (i) $\Phi(x_2, x_j) - U(x_2) < H_U$, or (ii) $\Phi(x_2, x_j) - U(x_2) > H_U$.

First, consider case (i). We claim that

$$\Phi(x_j, x_1) > \Phi(x_j, x_2). \tag{36}$$

Indeed, if (36) fails, then

$$\Phi(x_2, x_1) \le \max\{\Phi(x_2, x_j), \Phi(x_j, x_1)\} \le \Phi(x_2, x_j),$$

which implies

$$H_U = \Phi(x_2, x_1) - U(x_2) \le \Phi(x_2, x_j) - U(x_2) < H_U,$$

a contradiction. Thus, (36) holds. It then follows that

$$\Phi(x_1, x_2) - U(x_1) \le \max\{\Phi(x_1, x_j), \Phi(x_j, x_2)\} - U(x_1) = \Phi(x_1, x_j) - U(x_1).$$

Since $\Phi(x_1, x_2) - U(x_1) > \Phi(x_2, x_1) - U(x_2) = H_U$, we conclude that

$$\Phi(x_1, x_j) - U(x_1) > H_U.$$

Next, consider case (ii). Assume that $\Phi(x_2, x_j) - U(x_2) > H_U$. Suppose, by contradiction, that $\Phi(x_1, x_j) - U(x_1) \le H_U$. Then

$$\begin{aligned} \Phi(x_2, x_j) - U(x_2) &\le \max\{\Phi(x_2, x_1), \Phi(x_1, x_j)\} - U(x_2) \\ &\le \max\{\Phi(x_2, x_1) - U(x_2), \Phi(x_1, x_j) - U(x_1)\} \le H_U, \end{aligned}$$

contradicting the assumption that $\Phi(x_2, x_j) - U(x_2) > H_U$. Hence, $\Phi(x_1, x_j) - U(x_1) > H_U$. □

From the proof of Proposition 2.7, we see that $i \in \mathcal{I}$ if and only if

$$\Phi(x_i, x_1) < \Phi(x_i, x_2) \quad \text{and} \quad \Phi(x_2, x_i) - U(x_2) \le H_U. \tag{37}$$

That is, among the local minima that can be reached from x_2 via a barrier not exceeding H_U, the set $\mathcal{I}$ consists precisely of the indices for which the corresponding minima are more accessible to the global minimum x_1 than to the local minimum x_2. Accordingly, each basin in the collection $\boldsymbol{B}_1$ is referred to as a *nearby* basin (relative to x_1), whereas B_2 is referred to as the *distant* basin.

2.5. An upper bound in the form of probability generating function

Given $C_0 > 0, \lambda_0 > 1$, and $m \in \mathbb{N}$, define

$$g(\lambda; C_0, \lambda_0, m) = \lambda^m + C_0 \sum_{n=m}^{\infty} (\lambda^{n+1} - \lambda^n) \lambda_0^{-n}, \quad \lambda \in \mathbb{R}. \tag{38}$$

It is not hard to see that $g(\lambda; C_0, \lambda_0, m) < \infty$, $\forall \lambda \in (1, \lambda_0)$.

The right-hand side of (38) is motivated by the probability generating function. The following proposition states that if a random variable T exhibits exponential decay, then $\mathbb{E}[\lambda^T]$ is bounded above by a certain function g.

Proposition 2.8. *Let T be a random variable taking positive real values. Assume that for some constant $t_0 > 0$,*

$$\mathbb{P}[T > t] \le C_0 \lambda_0^{-t}, \quad \forall t \ge t_0. \tag{39}$$

Then for any $\lambda \in (1, \lambda_0)$, it holds that

$$\mathbb{E}[\lambda^T] \le g(\lambda; C_0, \lambda_0, n_0) < \infty$$

where $n_0 = \lfloor t_0 \rfloor + 1$. In particular, if $t_0 \in (0, 1)$, then

$$\mathbb{E}[\lambda^T] \le g(\lambda; C_0, \lambda_0, 1) < \infty.$$

Proof. Note that

$$\mathbb{E}[\lambda^T] \le \sum_{n=0}^{\infty} \lambda^{n+1} \mathbb{P}[n < T \le n+1] \le \lambda + \sum_{n=1}^{\infty} (\lambda^{n+1} - \lambda^n) \mathbb{P}[T > n],$$

where, for $n_0 > 1$, the right-hand side can be rewritten as

$$\lambda + \sum_{n=1}^{n_0-1} (\lambda^{n+1} - \lambda^n) \mathbb{P}[T > n] + \sum_{n=n_0}^{\infty} (\lambda^{n+1} - \lambda^n) \mathbb{P}[T > n].$$

By (39), for any $n \ge n_0$, $\mathbb{P}[T > n] \le C_0 \lambda_0^{-n}$. It then follows that

$$\mathbb{E}[\lambda^T] \le \lambda^{n_0} + C_0 \sum_{n=n_0}^{\infty} (\lambda^{n+1} - \lambda^n) \lambda_0^{-n} = g(\lambda; C_0, \lambda_0, n_0) < \infty. \quad \square$$

Note that for $m = 1$,

$$g(\lambda; C_0, \lambda_0, 1) \to 1, \quad \text{as } \lambda \to 1. \tag{40}$$

Given $\rho > 1$, $\lambda_0 > 1$, and $C_0 > 0$, define

$$\beta(\rho; C_0, \lambda_0) = \min\{1, \beta^*\}, \tag{41}$$

where

$$\beta^* = \begin{cases} \dfrac{-(\lambda_0 + \rho + C_0 - 2) + \sqrt{(\lambda_0 + \rho + C_0 - 2)^2 + 4(C_0 - 1)(\lambda_0 - 1)(\rho - 1)}}{2(C_0 - 1)(\lambda_0 - 1)}, & \text{if } C_0 \ne 1, \\ \dfrac{\rho - 1}{\lambda_0 + \rho - 1}, & \text{if } C_0 = 1. \end{cases} \tag{42}$$

The following Proposition 2.9 provides a quantitative characterization of the approximation in (40).

Proposition 2.9. *Given $\rho > 1$, $\lambda_0 > 1$ and $C_0 > 0$, let $\beta = \beta(\rho; C_0, \lambda_0)$ be as in (41). Then for any $\lambda \in \big(1, 1 + \beta(\lambda_0 - 1)\big)$, it holds that*

$$g(\lambda; C_0, \lambda_0, 1) < \rho.$$

Proof. Write $\lambda = 1 + \beta(\lambda_0 - 1)$ with $\beta \in (0, 1)$, we have

$$\begin{aligned} g(\lambda; C_0, \lambda_0, 1) &= \lambda + C_0(\lambda - 1) \sum_{n=1}^{\infty} (\lambda/\lambda_0)^n \\ &= (1 + \beta(\lambda_0 - 1)) \Big(1 + C_0 \frac{\beta}{1 - \beta}\Big) \end{aligned}$$

To have $g(\lambda; C_0, \lambda_0, 1) < \rho$, it suffices

$$(C_0 - 1)(\lambda_0 - 1)\beta^2 + (\lambda_0 + C_0 + \rho - 2)\beta + 1 - \rho < 0.$$

This specifies the definition of β in (41). Proposition 2.9 is proved. $\square$

3. Single-well potential and proof of Theorem 1.1

Throughout this section, U is a strongly convex single-well potential satisfying **(U1)**. Let (X_t, Y_t) denote an h-reflection-maximal coupling of two solutions of (2). The threshold $d = \mathcal{O}(\varepsilon\sqrt{h})$, at which the coupling (X_t, Y_t) switches between the reflection and maximal couplings, is set to $2\varepsilon\sqrt{h}$.

Define

$$\tau_h^{(1)} = \inf\Big\{ t \ge 0 : |X_t - Y_t| \in (0, 2\varepsilon\sqrt{h}], \text{ and for some } s \in (0, t),\ |X_s - Y_s| > 2\varepsilon\sqrt{h} \Big\},$$

with the convention that $\tau_h^{(1)} = \infty$ if the set is empty. Note that $\tau_h^{(1)}$ is the infimum time at which the distance between X_t and Y_t attains the threshold $d = 2\varepsilon\sqrt{h}$ from a distance greater than this value.

Since it is possible for $|X_t - Y_t|$ to never exceed the threshold d before a successful coupling occurs, in which case $\tau_h^{(1)} = \infty$, define

$$\tau_h = \tau_h^{(1)} \wedge \tau_c.$$

Note that $\tau_h < \infty$ holds almost surely. It will be shown later that the coupling time τ_c is almost surely a finite iteration of τ_h.

3.1. Estimation of τ_h.

In this subsection, estimates of τ_h are provided under the two initial conditions $|X_0 - Y_0| > 2\varepsilon\sqrt{h}$ and $|X_0 - Y_0| \le 2\varepsilon\sqrt{h}$, respectively.

If $|X_0 - Y_0| > 2\varepsilon\sqrt{h}$, then (X_t, Y_t) remains a reflection coupling until $t = \tau_h^{(1)}$, when (X_t, Y_t) switches to the maximal coupling. Proposition 2.1 immediately yields the following.

Lemma 3.1. *Assume $|X_0 - Y_0|/2\varepsilon = r_0 > \sqrt{h}$. Then $\tau_h = \tau_h^{(1)}$ holds $\mathbb{P}$-a.s., and for any $t_0 > 0$, there exists a constant $c_0 > 0$ such that*

$$\mathbb{P}[\tau_h > t] \le c_0 r_0 e^{-m_0 t}, \quad \forall t \ge t_0. \tag{43}$$

In particular, by letting $0 < t_0 < 1$ and applying Proposition 2.8, for any $\lambda \in (1, e^{m_0})$,

$$\mathbb{E}[\lambda^{\tau_h}] \le g(\lambda; c_0 r_0, e^{m_0}, 1). \tag{44}$$

Remark 3.2. The estimation in (43) is for the continuous-time process instead of its time-h sampled chain, which the numerical scheme truly approximates. Let τ_h^0 (*resp.* τ_h^h) be the first passage time of the coupling process (X_t, Y_t) (*resp.* its time-h sampled chain (X_n^h, Y_n^h)) to the set $\{(x, y) \in \mathbb{R}^k \times \mathbb{R}^k : |x - y| \le 2\varepsilon\sqrt{h}\}$. It is obvious that $\tau_h^h \ge \tau_h^0$, and it is intuitive that their difference, which is generally difficult to theoretically estimate, should approach to zero as h tends to zero, i.e.,

$$\lim_{h\to 0}(\tau_h^h - \tau_h^0) = 0, \quad \mathbb{P}\text{-a.s.} \tag{45}$$

Throughout this section, (45) is always assumed and will be numerically verified in Section 5 for the example of symmetric quadratic potential functions. Therefore, the estimation (43) applies to the time-h sampled chain (X_n^h, Y_n^h) (with a possible slight enlargement of c_0 if necessary) whenever h is sufficiently small.

The analysis becomes more intricate for the initial condition $|X_0 - Y_0| \le 2\varepsilon\sqrt{h}$, as the coupling method between X_t and Y_t may switch during the time interval $(0, \tau_h)$. Specifically, there exists $n > 0$ such that the coupling between the time-h sampled chains X_{ih} and Y_{ih} remains a maximal coupling for $0 \le i < n$. At the step $i = n$, either $X_{nh} = Y_{nh}$, indicating a successful coupling, or $|X_{nh} - Y_{nh}| > 2\varepsilon\sqrt{h}$. In the former case, $\tau_h = \tau_c$; in the latter, $\tau_h = \tau_h^{(1)}$, and (X_t, Y_t) evolves under a reflection coupling until the condition $|X_t - Y_t| \le 2\varepsilon\sqrt{h}$ is satisfied again.

Lemma 3.3. *Given any $t_1 > 0$, there exist $h_0 > 0$ and $C_0 > 0$ such that for all $h \in (0, h_0)$, if $|X_0 - Y_0| \le 2\varepsilon\sqrt{h}$, then*

$$\mathbb{P}[\tau_h > t] \le C_0\sqrt{h}e^{-m_0 t}, \quad \forall t \ge t_1.$$

In particular, by choosing $0 < t_1 < 1$ and applying Proposition 2.8, for any $\lambda \in (1, e^{m_0})$ and $h > 0$ sufficiently small,

$$\mathbb{E}[\lambda^{\tau_h}] \le g(\lambda; C_0\sqrt{h}, e^{m_0}, 1) < \infty. \tag{46}$$

Proof. Recall from the proof of Proposition 2.1 that the process $R_t = |X_t - Y_t|/2\varepsilon$ is a one-dimensional stochastic process induced by the coupling (X_t, Y_t). Let $n = \lfloor t/h \rfloor \in \mathbb{N}$. Based on the coupling behaviors between X_t and Y_t before the stopping time τ_h, one has

$$\begin{aligned}
\mathbb{P}[\tau_h > t] &\le \mathbb{P}[\tau_h > nh] \\
&= \sum_{j=1}^{n} \mathbb{P}[R_{ih} \in (0, \sqrt{h}], 0 \le i \le j-1] \\
&\quad \cdot \left(\mathbb{P}[R_{jh} > \sqrt{h} | R_{(j-1)h} \in (0, \sqrt{h}]] \cdot \mathbb{P}[\tau_h^{(1)} \circ \theta^{jh} > t - jh | R_{jh} > \sqrt{h}]\right) \\
&\quad + \mathbb{P}[R_{ih} \in (0, \sqrt{h}], 0 \le i \le n],
\end{aligned}$$

where θ is the usual shift operator.

For any $i \ge 1$, since (X_{ih}, Y_{ih}) is a maximal coupling whenever $|X_{(i-1)h} - Y_{(i-1)h}| \le 2\varepsilon\sqrt{h}$, it follows from Lemma 2.3(i) that

$$\mathbb{P}\left[R_{ih} > 0 | R_{(i-1)h} \in (0, \sqrt{h}]\right] \le \gamma,$$

where $\gamma \in (0, 1)$ is independent of i and h. Therefor, by the Markov property, for any $2 \le j \le n$,

$$\mathbb{P}\left[R_{ih} \in (0, \sqrt{h}], 0 \le i \le j-1\right] = \prod_{i=1}^{j-1} \mathbb{P}\left[R_{ih} \in (0, \sqrt{h}] | R_{(i-1)h} \in (0, \sqrt{h}]\right] \le \gamma^{j-1},$$

and this also holds trivially for $j = 1$. Consequently,

$$\mathbb{P}[\tau_h > t] \le \sum_{j=1}^{n} \gamma^{j-1} \Big(\mathbb{P}[R_{jh} > \sqrt{h} | R_{(j-1)h} \in (0, \sqrt{h}]] \cdot \mathbb{P}[\tau_h \circ \theta^{jh} > t - jh | R_{jh} > \sqrt{h}] \Big) + \gamma^n. \tag{47}$$

Now, for $1 \le j \le n$, consider estimating

$$\mathbb{P}[R_{jh} > \sqrt{h} | R_{(j-1)h} \in (0, \sqrt{h}]] \cdot \mathbb{P}[\tau_h \circ \theta^{jh} > t - jh | R_{jh} > \sqrt{h}], \tag{48}$$

which equals

$$\int_{\sqrt{h}}^{\infty} \mathbb{P}\big[\tau_h \circ \theta^{jh} > t - jh | R_{jh} = r\big] \mathbb{P}\big[R_{jh} = dr | R_{(j-1)h} \in (0, \sqrt{h}]\big]. \tag{49}$$

Fix $t_0 \in (0, t_1)$. Then for any $1 \le j \le \lfloor \frac{t-t_0}{h} \rfloor$, it holds that $t - jh \ge t_0$. Since $R_0 = r > \sqrt{h}$, Lemma 3.1 and the Markov property implies that

$$\mathbb{P}\big[\tau_h \circ \theta^{jh} > t - jh | R_{jh} = r\big] = \mathbb{P}[\tau_h > t - jh | R_0 = r] \le c_0 r e^{-m_0(t-jh)},$$

where $c_0 > 0$ is the constant given in Lemma 3.1. Therefore, for any such j,

$$\begin{aligned} (49) &\le c_0 e^{-m_0(t-jh)} \int_{\sqrt{h}}^{\infty} r \mathbb{P}[R_{jh} = dr | R_{(j-1)h} \in (0, \sqrt{h}]] \\ &\le c_0 e^{-m_0(t-jh)} \cdot \mathbb{E}[R_{jh} | R_{(j-1)h} \in (0, \sqrt{h}]]. \end{aligned}$$

Since $R_{(j-1)h} \le \sqrt{h}$, and hence $|X_{(j-1)h} - Y_{(j-1)h}| \le 2\varepsilon\sqrt{h}$, Lemma 2.3 (ii) implies that for sufficiently small $h > 0$,

$$\begin{aligned} \mathbb{E}\big[R_{jh} \big| R_{(j-1)h} \in (0, \sqrt{h}]\big] &= \frac{1}{2\varepsilon} \mathbb{E}\big[|X_{jh} - Y_{jh}| \big| |X_{(j-1)h} - Y_{(j-1)h}| \le 2\varepsilon\sqrt{h}\big] \\ &\le \frac{1}{2} c_1 \sqrt{h}. \end{aligned}$$

Thus,

$$(48) \le C_0 \sqrt{h} e^{-m_0(t-jh)}, \quad \forall 1 \le j \le \lfloor \frac{t-t_0}{h} \rfloor, \tag{50}$$

for some constant $C_0 > 0$ independent of h, ε, and j. Moreover,

$$(48) \le 1, \quad \lfloor \frac{t-t_0}{h} \rfloor < j \le \lfloor \frac{t}{h} \rfloor := n. \tag{51}$$

Combining (47), (50), and (51), it follows that

$$\begin{aligned} \mathbb{P}[\tau_h > t] &\le C_0 \sqrt{h} \sum_{j=1}^{\lfloor \frac{t-t_0}{h} \rfloor} \gamma^{j-1} e^{-m_0(t-jh)} + \sum_{j=\lfloor \frac{t-t_0}{h} \rfloor}^{n} \gamma^j \\ &\le C_0 \sqrt{h} \frac{e^{m_0 h}}{1 - \gamma e^{m_0 h}} e^{-m_0 t} + \frac{\gamma^{\lfloor \frac{t-t_0}{h} \rfloor}}{1-\gamma} \end{aligned} \tag{52}$$

Let $0 < h_0 \le |\ln \gamma| / m_0$ be sufficiently small so that for any $h \in (0, h_0)$,

$$\gamma^{\lfloor \frac{t-t_0}{h} \rfloor} \le \sqrt{h} e^{-m_0 t}, \quad \forall t \ge t_1.$$

Since $e^{m_0 h}/(1 - \gamma e^{m_0 h}) \to 1/(1-\gamma)$ as $h \to 0$, by enlarging C_0 in (52) if necessary, it follows that

$$\mathbb{P}[\tau_h > t] \le C_0 \sqrt{h} e^{-m_0 t}, \quad \forall t \ge t_1. \quad \square$$

Combining Lemmas 3.1 and 3.3, the following holds.

Lemma 3.4. *Assume $\mathbb{E}[|X_0 - Y_0|] < \infty$. Then for any $h \in (0, h_0)$ where $h_0 > 0$ is as in Lemma 3.3, for any $\lambda \in (1, e^{m_0})$, it holds that*

$$\mathbb{E}[\lambda^{\tau_h}] < \infty.$$

Proof. Recall the one-dimensional stochastic process $R_t = |X_t - Y_t|/(2\varepsilon)$, $t \ge 0$. Let $\mathbb{E}_r[\cdot]$ denote the expectation with respect to the initial condition $R_0 = r$. Then

$$\mathbb{E}[\lambda^{\tau_h}] = \int_0^{\sqrt{h}} \mathbb{E}_r[\lambda^{\tau_h}] \mu(dr) + \int_{\sqrt{h}}^{\infty} \mathbb{E}_r[\lambda^{\tau_h}] \mu(dr) \tag{53}$$

where μ denotes the distribution of R_0.

By Lemma 3.3, for sufficiently small $h > 0$, the first term on the right-hand side of (53) satisfies

$$\int_0^{\sqrt{h}} \mathbb{E}_r[\lambda^{\tau_h}] \mu(dr) \le \int_0^{\sqrt{h}} g(\lambda; C_0\sqrt{h}, e^{m_0}, 1) \sqrt{h} \mu(dr) \le g(\lambda; C_0\sqrt{h}, e^{m_0}, 1)\sqrt{h}. \tag{54}$$

By (44), the second term on the right-hand side of (53) is bounded as

$$\int_{\sqrt{h}}^{\infty} \mathbb{E}_r[\lambda^{\tau_h}]\mu(dr) \le \int_{\sqrt{h}}^{\infty} g(\lambda; c_0 r, e^{m_0}, 1)\mu(dr) = g(\lambda; c_0 \int_{\sqrt{h}}^{\infty} r\mu(dr), e^{m_0}, 1) \le g(\lambda; c_0\mathbb{E}[R_0], e^{m_0}, 1). \tag{55}$$

Combining (54) and (55), we have

$$\mathbb{E}[\lambda^{\tau_h}] \le g(\lambda; C_0\sqrt{h}, e^{m_0}, 1)\sqrt{h} + g(\lambda; c_0\mathbb{E}[R_0], e^{m_0}, 1).$$

Since $\mathbb{E}[R_0] = \mathbb{E}[|X_0 - Y_0|]/2\varepsilon < \infty$, the lemma is proved. □

3.2. Iteration of τ_h and coupling times

The coupling time τ_c is in fact a finite iteration of τ_h. To see this, define

$$\tau_h^0 = 0, \quad \tau_h^k = \tau_h^{k-1} + \tau_h \circ \theta^{\tau_h^{k-1}}, \ k \ge 1$$

where θ is the usual shift operator, and let

$$\eta = \inf\{k \ge 1 : X_{\tau_h^k} = Y_{\tau_h^k}\}.$$

The following proposition immediately follows from the definition of τ_h.

Proposition 3.5. *Given any $h > 0$ and $k \ge 1$. The following hold:*

(i) $|X_{\tau_h^k} - Y_{\tau_h^k}| = 2\varepsilon\sqrt{h}$ or 0, where $X_{\tau_h^k} = Y_{\tau_h^k}$ if and only if $k \ge \eta$;

(ii) If $k > 1$, then

$$\mathbb{P}[|X_{\tau_h^k} - Y_{\tau_h^k}| > 0 | \mathcal{F}_{\tau_h^{k-1}}] < \gamma.$$

where γ is as in Lemma 2.3.

By Proposition 3.5(i),

$$\tau_c = \tau_h^{\eta}, \quad \mathbb{P}\text{-a.s.}$$

Hence, the estimation of τ_c is reduced to the estimation of τ_h^{η}.

Theorem 3.6. *Assume $\mathbb{E}[|X_0 - Y_0|] < \infty$. Then for any $\delta > 0$, there exists $h_0 > 0$ such that for any $h \in (0, h_0)$ and any $\lambda \in (1, e^{m_0 - \delta})$, it holds that*

$$\mathbb{E}[\lambda^{\tau_h^{\eta}}] < \infty.$$

Proof. The proof follows the approach of Lemma 2.9 in [25]. Note that

$$\mathbb{E}[\lambda^{\tau_h^{\eta}}] \le \sum_{k=1}^{\infty} \mathbb{E}[\lambda^{\tau_h^k}\mathbb{I}_{\eta \ge k}] = \mathbb{E}\left[\lambda^{\tau_h}\mathbb{I}_{\eta \ge 1}\right] + \sum_{k=2}^{\infty} \mathbb{E}\left[\mathbb{I}_{\eta \ge k}\lambda^{\tau_h^{k-1}}\mathbb{E}[\lambda^{\tau_h \circ \theta^{\tau_h^{k-1}}} | \mathcal{F}_{\tau_h^{k-1}}]\right], \tag{56}$$

where the last equality follows from the fact that $\lambda^{\tau_h^{k-1}} \in \mathcal{F}_{\tau_h^{k-1}}$ and $\{\eta \ge k\} \in \mathcal{F}_{\tau_h^{k-1}}$.

We retain the notation $R_t = |X_t - Y_t|/2\varepsilon$ for $t \ge 0$, and let $\mathbb{E}_r$ denote the expectation with respect to the initial condition $R_0 = r$. By Proposition 3.5(i), $R_{\tau_h^{k-1}} = \sqrt{h}$ for $2 \le k \le \eta$. By (46) and the strong Markov property,

$$\mathbb{E}\left[\lambda^{\tau_h \circ \theta^{\tau_h^{k-1}}} | \mathcal{F}_{\tau_h^{k-1}}\right] \le \mathbb{E}_{\sqrt{h}}[\lambda^{\tau_h}] \le g(\lambda; C_0\sqrt{h}, e^{m_0}, 1) < \infty, \quad \forall k \ge 1,$$

where $C_0 > 0$ is as in Lemma 3.3. Thus,

$$\mathbb{E}[\lambda^{\tau_h^{\eta}}] \le \mathbb{E}[\lambda^{\tau_h}\mathbb{I}_{\eta \ge 1}] + g(\lambda; C_0\sqrt{h}, e^{m_0}, 1) \sum_{k=2}^{\infty} \mathbb{E}[\mathbb{I}_{\eta \ge k}\lambda^{\tau_h^{k-1}}]. \tag{57}$$

Now, for $k \ge 2$, we estimate $\mathbb{E}[\mathbb{I}_{\eta \ge k}\lambda^{\tau_h^{k-1}}]$. Write $\mathbb{I}_{\eta \ge k} = \mathbb{I}_{\eta \ge k-1}\mathbb{I}_{R_{\tau_h^{k-1}} > 0}$. Then we have

$$\mathbb{E}\left[\mathbb{I}_{\eta \ge k}\lambda^{\tau_h^{k-1}}\right] = \mathbb{E}\left[\mathbb{I}_{\eta \ge k-1}\lambda^{\tau_h^{k-2}}\mathbb{E}\left[\mathbb{I}_{R_{\tau_h^{k-1}} > 0}\lambda^{\tau_h \circ \theta^{\tau_h^{k-2}}} | \mathcal{F}_{\tau_h^{k-2}}\right]\right] = \mathbb{E}\left[\mathbb{I}_{\eta \ge k-1}\lambda^{\tau_h^{k-2}}\mathbb{E}_{R_{\tau_h^{k-2}}}[\mathbb{I}_{R_{\tau_h} > 0}\lambda^{\tau_h}]\right]. \tag{58}$$

Note that for $k = 2$,

$$\mathbb{E}_{R_{\tau_h^{k-2}}}[\mathbb{I}_{R_{\tau_h}>0}\lambda^{\tau_h}] = \mathbb{E}_{R_0}[\mathbb{I}_{R_{\tau_h}>0}\lambda^{\tau_h}] \le \mathbb{E}[\lambda^{\tau_h}]. \tag{59}$$

For $k > 2$, since $R_{\tau_h^{k-2}} = \sqrt{h}$, the strong Markov property implies that

$$\mathbb{E}_{R_{\tau_h^{k-2}}}[\mathbb{I}_{R_{\tau_h}>0}\lambda^{\tau_h}] \le \mathbb{E}_{\sqrt{h}}[\mathbb{I}_{R_{\tau_h}>0}\lambda^{\tau_h}] \tag{60}$$

By the Hölder's inequality, for any $p \in (0,1)$,

$$\begin{aligned}\mathbb{E}_{\sqrt{h}}[\mathbb{I}_{R_{\tau_h}>0}\lambda^{\tau_h}] &\le (\mathbb{E}_{\sqrt{h}}[\mathbb{I}_{R_{\tau_h}>0}])^{1-p}\cdot(\mathbb{E}_{\sqrt{h}}[\lambda^{\tau_h/p}])^p\\ &= (\mathbb{P}_{\sqrt{h}}[R_{\tau_h}>0])^{1-p}\cdot(\mathbb{E}_{\sqrt{h}}[\lambda^{\tau_h/p}])^p.\end{aligned}$$

Then it follows from Proposition 3.5 (ii) and (46) that

$$\mathbb{E}_{\sqrt{h}}[\mathbb{I}_{R_{\tau_h}>0}\lambda^{\tau_h}] \le \gamma^{1-p} g(\lambda^{1/p}; C_0\sqrt{h}, e^{m_0}, 1)^p. \tag{61}$$

Substituting (59)–(61) into (58), we obtain

$$\mathbb{E}[\mathbb{I}_{\eta\ge k}\lambda^{\tau_h^{k-1}}] \le \begin{cases}\mathbb{E}[\lambda^{\tau_h}], & k=2\\ \gamma^{1-p} g(\lambda^{1/p}; C_0\sqrt{h}, e^{m_0}, 1)^p\cdot\mathbb{E}[\mathbb{I}_{\eta\ge k-1}\lambda^{\tau_h^{k-2}}], & k>2.\end{cases}$$

By induction, for $k \ge 2$,

$$\mathbb{E}[\mathbb{I}_{\eta\ge k}\lambda^{\tau_h^{k-1}}] \le \gamma^{(1-p)(k-2)} g(\lambda^{1/p}; C_0\sqrt{h}, e^{m_0}, 1)^{p(k-2)}\cdot\mathbb{E}[\lambda^{\tau_h}]$$

Therefore, (57) yields

$$\mathbb{E}[\lambda^{\tau_h^\eta}] \le \mathbb{E}[\lambda^{\tau_h}]\Big(1 + g(\lambda; C_0\sqrt{h}, e^{m_0}, 1)\sum_{k=2}^{\infty}\Big(\gamma^{1-p} g(\lambda^{1/p}; C_0\sqrt{h}, e^{m_0}, 1)^p\Big)^{k-2}\Big)$$

By Lemma 3.4, $\mathbb{E}[\lambda^{\tau_h}] < \infty$. Thus, to guarantee $\mathbb{E}[\lambda^{\tau_h^\eta}] < \infty$, it suffices

$$g(\lambda^{1/p}; C_0\sqrt{h}, e^{m_0}, 1) < \gamma^{-(1-p)/p} \tag{62}$$

By Proposition 2.9, (62) holds for any $\lambda > 1$ satisfying $\lambda^{1/p} \in \big(1, 1+\beta(e^{m_0}-1)\big)$, where

$$\beta = \min\{1, \beta^*\},$$

and β^* is given by (42) with $\lambda_0 = e^{m_0}$, $\rho = \gamma^{-(1-p)/p}$, and C_0 replaced by $C_0\sqrt{h}$.

Note that $\beta^* \to 1$, and hence $\beta \to 1$, as $h \to 0$. Since p can be arbitrarily close to 1, by choosing $h > 0$ sufficiently small, it follows that

$$\mathbb{E}[\lambda^{\tau_h^\eta}] < \infty, \quad \forall \lambda \in (1, e^{m_0-\delta}),$$

where $\delta > 0$ is arbitrarily small. □

The proof of Theorem 1.1 is now straightforward.

Proof of Theorem 1.1. Note that for any $\lambda \in (1,\infty)$ satisfying $\mathbb{E}[\lambda^{\tau_h^\eta}] < \infty$, we have

$$\mathbb{P}[\tau_h^\eta > t] \le \mathbb{E}[\lambda^{\tau_h^\eta}]\lambda^{-t}, \quad \forall t > 0.$$

By Theorem 3.6, for any $\lambda \in (1, e^{m_0-\delta})$ and any $\delta > 0$,

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}[\tau_h^\eta > t] \le -m_0 + m_0\delta.$$

Theorem 1.1 is proved by taking δ as $m_0\delta$. □

4. Multi-well potentials and proof of Theorems 1.2 – 1.3

Throughout this section, let U be a multi-well potential, and let (X_t, Y_t) be a coupling of two solutions of (2). Sections 4.1–4.3 focus primarily on the double-well potential U, while the general case of more than two wells is discussed in Section 4.4.

4.1. Key stopping times for double-well potential

Assume that U is a double-well potential satisfying **(U2)** with two basins B_1 and B_2. Let

$$\tau_\varepsilon^{(1)} = \inf\left\{t > h : (X_t, Y_t) \in B_1 \times B_1 \text{ or } B_2 \times B_2\right\}$$

denote the infimum time when X_t and Y_t lie in the *same* basin of U. Here, the subscript "ε" emphasizes the role of the noise magnitude ε in determining the stopping times in the multi-well setting. We note that $\tau_\varepsilon^{(1)}$ is finite $\mathbb{P}$-almost surely.

Remark 4.1. When X_t or Y_t is initiated near a basin boundary, repeated boundary crossings within an infinitesimal time interval may occur, making the analysis cumbersome. To circumvent these non-essential complications, $\tau_\varepsilon^{(1)}$ is defined after a small positive time. Specifically, this positive time is chosen as the numerical step size h to ensure compatibility with the numerical simulations. This convention is adopted for all stopping times defined in this section.

If initially, X_t and Y_t already belong to the same basin, then $\tau_\varepsilon^{(1)} = h$ with probability close to 1. Now, assume that X_0 and Y_0 belong to different basins. Without loss of generality, let $Y_0 \in B_1$. Then we have

$$\tau_\varepsilon^{(1)} = \kappa_{\boldsymbol{X}}(B_1) \wedge \kappa_{\boldsymbol{Y}}(B_2),$$

where recall that $\kappa_{\boldsymbol{X}}(B_1)$ (*resp.* $\kappa_{\boldsymbol{Y}}(B_2)$) denotes the first hitting time of the process X_t (*resp.* Y_t) to the basin B_1 (*resp.* B_2).

Throughout this section, let

$$\lambda_\varepsilon = \exp\left\{C_\varepsilon e^{-2H_U/\varepsilon^2}\right\}, \tag{63}$$

where, in the double-well case, H_U is the essential barrier height defined in (5), and $C_\varepsilon > 0$ is any constant that is not uniquely determined and satisfies $\lim_{\varepsilon\to 0} C_\varepsilon > 0$, with the limit depending only on U. By Lemma 2.6,

$$\mathbb{P}[\tau_\varepsilon^{(1)} > t] \le \mathbb{P}[\kappa_{\boldsymbol{X}}(B_1) > t] \lesssim \lambda_\varepsilon^{-t}, \quad \forall t > 0. \tag{64}$$

Before proceeding, recall from Section 2.4 that for the multi-well potential, an enlarged set $\boldsymbol{B}_1$, defined in (34), consists of all nearby basins, in particular, including B_1. By Proposition 2.7, for any local minimum x_j with the corresponding basin B_j not contained in $\boldsymbol{B}_1$, one has

$$\Phi(x_1, x_j) - U(x_1) > \Phi(x_2, x_i) - U(x_2) = H_U, \quad \forall i \in \mathcal{I}. \tag{65}$$

This suggests that a process starting from the global minimum x_1 must overcome a higher barrier to exit the enlarged set $\boldsymbol{B}_1$ than a process starting from x_2 has to overcome to enter it.

In light of (65), in the multi-well setting, which in particular includes the double-well case, we assume the following **(H1)** for the coupling scheme.

(H1) There exist constants $\delta_0 > 0$ and $\gamma_0 > 0$ such that if the coupling (X_t, Y_t) satisfies the initial condition $X_0 \in B_{\delta_0}(x_2), Y_0 \in B_{\delta_0}(x_1)$, then for any $\varepsilon > 0$ sufficiently small,

$$\mathbb{P}\left[Y_s \in \boldsymbol{B}_1 \text{ for all } s \in [0,t] \middle| \kappa_{\boldsymbol{X}}(\boldsymbol{B}_1) > t\right] \ge \gamma_0, \quad \forall t > 0. \tag{66}$$

The assumption **(H1)** states that when X_t and Y_t start from the bottom of the basins B_2 and B_1, respectively, the probability that Y_t remains in the enlarged set $\boldsymbol{B}_1$, given that X_t has not yet entered $\boldsymbol{B}_1$, is uniformly positive and independent of ε and t. In Section 5, **(H1)** is numerically verified. At present, although **(H1)**, along with the forthcoming assumptions **(H2)–(H3)**, cannot be rigorously verified, Section 5.7 provides numerical evidence supporting their validity in an interacting particle system with multiple local minima.

Remark 4.2. The assumption **(H1)** naturally arises in the context of reflection coupling. For simplicity, consider the one-dimensional double-well potential. Let $\varphi(t), t \in [0,T]$, be a C^1 function satisfying $\varphi(0) = 0$, and let $\delta > 0$ be a small constant such that

$$|(X_t - X_0) - \varphi(t)| < \delta, \quad \forall t \in [0,T].$$

Then the Wiener process B_t^x associated with X_t must stay in the neighborhood of $\varphi(t) + \int_0^t \nabla U(X_0 + \varphi(s))ds$. Due to the reflection, the Brownian motion terms in X_t and Y_t are symmetric. Consequently, the corresponding Wiener process B_t^y of Y_t must stay in the small neighborhood of $-\varphi(t) - \int_0^t \nabla U(X_0 + \varphi(s))ds$, which has the same action functional as $\varphi(t) + \int_0^t \nabla U(X_0 + \varphi(s))ds$. If X_t exits the shallower well B_2 at some $T > 0$, then with high probability, the final segment of X_t remains in a small neighborhood of the minimum energy path, denoted by $\phi(t)$ (see, for instance, Theorem 2.3 in Chapter 4 of [24]). This implies that the Brownian motion term B_t^x stays in the neighborhood of

$$\phi(t) + \int_0^t \nabla U(\phi(s))\mathrm{d}s,$$

whose action functional equals $2(\Phi(x_1, x_2) - U(x_2))$, which is strictly less than $2H_U$. On the other hand, for Y_t to exit B_1 from the neighborhood of x_1, its trajectory must have an action functional of at least $2H_U$. Therefore, when X_t exits from the shallower well, it is highly likely that Y_t remains in B_1. Unfortunately, to the best of our knowledge, this argument is difficult to establish rigorously, as the Freidlin–Wentzell large deviation theory applies only to a *fixed* time span $[0,T]$ as $\varepsilon \to 0$. However, the tail estimates required in this paper necessitate estimates that hold for arbitrarily large t.

In the double-well setting, $\boldsymbol{B}_1$ is simply B_1. Hence, (66) is reduced to

$$\mathbb{P}[\kappa_{\boldsymbol{Y}}(B_2) > t | \kappa_{\boldsymbol{X}}(B_1) > t] \geq \gamma_0. \tag{67}$$

Under **(H1)**, the reverse of (64) holds if $X_0 \in B_{\delta_0}(x_1)$ and $Y_0 \in B_{\delta_0}(x_2)$:

$$\begin{aligned}\mathbb{P}[\tau_\varepsilon^{(1)} > t] &\geq \mathbb{P}[\kappa_{\boldsymbol{X}}(B_1) > t, \kappa_{\boldsymbol{Y}}(B_2) > t] \\ &= \mathbb{P}[\kappa_{\boldsymbol{Y}}(B_2) > t | \kappa_{\boldsymbol{X}}(B_1) > t] \cdot \mathbb{P}[\kappa_{\boldsymbol{X}}(B_1) > t] \\ &\geq \gamma_0 \cdot \mathbb{P}[\kappa_{\boldsymbol{X}}(B_1) > t] \ \simeq \ \lambda_\varepsilon^{-t}\end{aligned} \tag{68}$$

where the last "$\simeq$" follows from Lemma 2.6, by choosing $\delta_0 > 0$ sufficiently small.

In contrast to $\tau_\varepsilon^{(1)}$, define another stopping time

$$\begin{aligned}\tau_\varepsilon^{(2)} = \inf \Big\{ & t > h : (X_t, Y_t) \in B_1 \times B_2 \text{ or } B_2 \times B_1, \text{ and for some } s \in (h, t), \\ & (X_s, Y_s) \in B_1 \times B_1 \text{ or } B_2 \times B_2 \Big\},\end{aligned}$$

and let $\tau_\varepsilon^{(2)} = \infty$ if the set is empty. Note that $\tau_\varepsilon^{(2)}$ captures the infimum time when X_t, Y_t are separated (again) by *different* basins, where "again" applies if X_t and Y_t already belong to different basins at the very beginning.

Let

$$\tau_\varepsilon = \tau_\varepsilon^{(2)} \wedge \tau_c.$$

We note that $\tau_\varepsilon = \tau_\varepsilon^{(2)} < \tau_c$ if X_t, Y_t are *not* coupled while staying in the same basin; otherwise, $\tau_\varepsilon = \tau_c$ and $\tau_\varepsilon^{(2)} = \infty$. As will be seen in Section 4.3, the coupling time τ_c is $\mathbb{P}$-a.s. a finite iteration of τ_ε.

4.2. Estimation of τ_ε

The following assumption **(H2)** is made in both double and multi-well settings, characterizing *local* coupling properties when X_t and Y_t lie in the same basin.

(H2) Let (X_t, Y_t) be a coupling of two solutions of (2) such that $(X_0, Y_0) \in \bigcup_{1 \leq i \leq L} \overline{B_i \times B_i}$. The following hold:

(i) There exists $\gamma_1 \in (0, 1)$ such that

$$\mathbb{P}[X_{\tau_\varepsilon} \neq Y_{\tau_\varepsilon}] < \gamma_1;$$

(ii) For any $\varepsilon > 0$ sufficiently small, there exists $r_0(\varepsilon) = \mathcal{O}(-1/\log \varepsilon) > 0$ such that

$$\mathbb{P}[\tau_\varepsilon > t] \lesssim e^{-r_0(\varepsilon)t}, \quad \forall t > 0.$$

Assumption **(H2)**(i) asserts that when X_t and Y_t belong to the same basin, there is a positive probability of successful coupling. Assumption **(H2)**(ii) states that as ε tends to zero, the exponential tail of τ_ε vanishes at the rate $\mathcal{O}(-1/\log \varepsilon)$. This is expected, since in the limiting case $\varepsilon = 0$, one process may be trapped at the saddle point on the boundary and cannot couple with the other one. The rate $\mathcal{O}(-1/\log \varepsilon$ can be derived by explicitly solving the linearized dynamics near the saddle point; see also [26,27] for rigorous results on the passage time of a small-noise perturbation of a deterministic dynamical system through a hyperbolic equilibrium.

The following proposition provides sufficient conditions for **(H2)**, which will be numerically verified in Section 5.

Proposition 4.3. *Assume that for each $i \in \{1, \dots, L\}$, the following conditions hold:*

(a) There exist constants $T_0 = \mathcal{O}(-\log \varepsilon)$, $\delta > 0$, and $\gamma_0 > 0$ such that

$$\mathbb{P}[(X_t, Y_t) \in B_\delta^i \times B_\delta^i, \text{for all } h \leq t \leq T_0 \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \geq \gamma_0$$

is uniform for all $(X_0, Y_0) \in \overline{B_i \times B_i}$, where $B_\delta^i = \{x \in B_i \mid d(x, \partial B_i) > \delta\}$ denotes the δ-interior of B_i;

(b) There exists a strongly convex neighborhood B_c^i of x_i such that $B_c^i \subseteq B_i$.

Then Assumption **(H2)** *holds.*

For the proof of Proposition 4.3, refer to Appendix B.

Remark 4.4. Assumption (b) in Proposition 4.3 holds if U has non-vanishing second-order derivatives at the minimum x_i. Assumption (a) asserts that if the two processes start from the same basin, the probability that both strictly remain in that basin for an extended period of time of order $\mathcal{O}(-\log \varepsilon)$ is positive. While this is intuitive, a rigorous proof is technically challenging and beyond the scope of this paper. It would require specifying a normal form for $-\nabla U$ near the boundary and analyzing the exit behavior from the separatrix of the reflection-coupled processes governed by (2). Therefore, we choose to verify this assumption numerically.

The analysis becomes more intricate when X_t and Y_t initially belong to different basins. The coupling process (X_t, Y_t) typically evolves in two stages during (h, τ_ε). In Stage 1, X_t and Y_t lie in different basins until one of them, either X_t or Y_t, jumps out of its initial basin and enters the other, making both of them stay in the same basin. Then Stage 2 begins, where X_t and Y_t are in the same basin for a period of time, until they are either successfully coupled or fail to couple with one of them jumping out of the basin again. Accordingly, write

$$\tau_\varepsilon = \tau_\varepsilon \circ \theta^{\tau_\varepsilon^{(1)}} + \tau_\varepsilon^{(1)}, \quad \mathbb{P}\text{- a.s.} \tag{69}$$

where $\tau_\varepsilon^{(1)}$ and $\tau_\varepsilon \circ \theta^{\tau_\varepsilon^{(1)}}$ correspond to the Stage 1 and Stage 2, respectively, and θ denotes the usual shift operator.

We note that Stage 1 and Stage 2 exhibit different time scales: Stage 1 corresponds to a slow time scale, typically persisting over an exponentially long period, with the tail exponent diminishing exponentially in terms of ε; Stage 2 is associated with the fast time scale, and as shown in **(H2)**, the exponent of the tail distribution remains uniformly away from zero, independent of ε.

Based on the above analysis, we obtain the following estimate of τ_ε when X_t and Y_t initially belong to different basins.

Lemma 4.5. *Let (X_t, Y_t) be a coupling of two solutions of (2) such that $(X_0, Y_0) \in \overline{B_1 \times B_2}$ or $\overline{B_2 \times B_1}$. Assume* **(H2)**. *Then there exists $C_1 > 0$ such that for any $t > 0$ and any $\varepsilon > 0$ sufficiently small,*

$$\mathbb{P}[\tau_\varepsilon > t] \le C_1 \lambda_\varepsilon^{-t}.$$

Consequently, by Proposition 2.8, for any $\lambda \in (1, \lambda_\varepsilon)$,

$$\mathbb{E}[\lambda^{\tau_\varepsilon}] \le g(\lambda; C_1, \lambda_\varepsilon, 1) < \infty. \tag{70}$$

Proof. According to (69),

$$\begin{aligned} \mathbb{P}[\tau_\varepsilon > t] &= \int_h^t \mathbb{P}[\tau_\varepsilon^{(1)} = s]\mathbb{P}[\tau_\varepsilon > t | \tau_\varepsilon^{(1)} = s] ds \\ &\le \int_h^t \mathbb{P}[\tau_\varepsilon^{(1)} > s - \delta]\mathbb{P}[\tau_\varepsilon \circ \theta^s > t - s] ds \end{aligned} \tag{71}$$

where $\delta \in (0, h)$ is sufficiently small. By (64), there exists a constant $C_2 > 0$ such that

$$\mathbb{P}[\tau_\varepsilon^{(1)} > s - \delta] \le C_2 \lambda_\varepsilon^{-(s-\delta)}.$$

Moreover, since X_t, Y_t belong to the same basin at $t = \tau_\varepsilon^{(1)}$, it follows from **(H2)**(ii) that there exists a constant $C_3 > 0$ such that

$$\mathbb{P}[\tau_\varepsilon \circ \theta^s > t - s] \le C_3 e^{-r_0(\varepsilon)(t-s)}.$$

Since δ is arbitrarily small, it follows that

$$(71) \le C_1 \lambda_\varepsilon^{-t} \int_0^t \left(\lambda_\varepsilon e^{-r_0(\varepsilon)}\right)^{t-s} ds,$$

where $C_1 > 0$ is a constant independent of t and ε.

Note that $\lambda_\varepsilon e^{-r_0(\varepsilon)} < 1$ for $\varepsilon > 0$ sufficiently small. Thus, by enlarging C_1 if necessary,

$$(71) \le C_1 \lambda_\varepsilon^{-t}.$$

The lemma is proved. □

Combining **(H2)**(ii) and Lemma 4.5, we immediately obtain the following result.

Proposition 4.6. *Let (X_t, Y_t) be a coupling of two solutions of (2) such that (X_0, Y_0) is fully supported. Assume* **(H2)**. *Then for any $\varepsilon > 0$ sufficiently small and any $\lambda \in (1, \lambda_\varepsilon)$,*

$$\mathbb{E}[\lambda^{\tau_\varepsilon}] < \infty.$$

4.3. Proof of Theorem 1.2

Let U be a double-well potential and (X_t, Y_t) be a coupling of two solutions of (2). In this section, the coupling (X_t, Y_t) is assumed to satisfy **(H1)–(H2)**. As in the proof of Theorem 1.1, a sequence of random times is defined inductively as

$$\tau_\varepsilon^0 = 0, \ \tau_\varepsilon^k = \tau_\varepsilon^{k-1} + \tau_\varepsilon \circ \theta^{\tau_\varepsilon^{k-1}}, \quad k \ge 1. \tag{72}$$

where θ is the usual shift operator. Note that by the definition of τ_ε, for each $k \ge 1$, either $X_{\tau_\varepsilon^k} = Y_{\tau_\varepsilon^k}$, or $X_{\tau_\varepsilon^k}$ and $Y_{\tau_\varepsilon^k}$ belong to different basins. Let

$$\eta = \inf\{k \ge 1 : X_{\tau_\varepsilon^k} = Y_{\tau_\varepsilon^k}\}. \tag{73}$$

The following Proposition 4.7 and Theorem 4.8 are analogues of Proposition 3.5 and Theorem 3.6, respectively, in the double-well setting.

Proposition 4.7. *For any $\varepsilon > 0$ sufficiently small, the following hold:*
(i) $X_{\tau_\varepsilon^k} = Y_{\tau_\varepsilon^k}$ if and only if $k \geq \eta$;
(ii) For any $\varepsilon > 0$ sufficiently small, it holds that

$$\mathbb{P}[X_{\tau_\varepsilon^k} \neq Y_{\tau_\varepsilon^k} | \mathcal{F}_{\tau_\varepsilon^{k-1}}] < \gamma_1, \quad \forall k \geq 1,$$

where $\gamma_1 \in (0,1)$ is as in **(H2)**.

Note that Proposition 4.7(i) yields

$$\tau_c = \tau_\varepsilon^\eta, \quad \mathbb{P}\text{-a.s.}$$

Proposition 4.7 (ii) directly follows from **(H2)**(i) by the strong Markov property.

Theorem 4.8. *Let (X_t, Y_t) be a coupling of two solutions of (2) such that (X_0, Y_0) is fully supported. Assume* **(H2)**. *Then for any $\varepsilon > 0$ sufficiently small and any $\lambda \in (1, \lambda_\varepsilon)$, it holds that*

$$\mathbb{E}[\lambda^{\tau_\varepsilon^\eta}] < \infty.$$

Proof. The proof follows the same approach as that of Theorem 3.6. By replacing $\mathbb{I}_{\{R_{\tau_\varepsilon^k} > 0\}}$ in the proof of Theorem 3.6 with $\mathbb{I}_{\{X_{\tau_\varepsilon^k} \neq Y_{\tau_\varepsilon^k}\}}$, we have

$$\mathbb{E}[\lambda^{\tau_\varepsilon^\eta}] \leq \mathbb{E}[\lambda^{\tau_\varepsilon}] \Big(1 + g(\lambda; C_1, \lambda_\varepsilon, 1) \sum_{k=0}^{\infty} \Big(\gamma_1^{1-p} g(\lambda^{1/p}; C_1, \lambda_\varepsilon, 1)^p \Big)^k \Big),$$

where $C_1 > 0$ is the constant given in Lemma 4.5, γ_1 is as in Proposition 4.7(ii), and $p \in (0,1)$ is an arbitrary number. Thus, $\mathbb{E}[\lambda^{\tau_\varepsilon^\eta}] < \infty$ holds if the inequality

$$g(\lambda^{1/p}; C_1, \lambda_\varepsilon, 1) < \gamma_1^{-(1-p)/p} \tag{74}$$

is satisfied.

By Proposition 2.9, (74) holds for any $\lambda > 1$ satisfying $\lambda^{1/p} \in (1, 1 + \beta(\lambda_\varepsilon - 1))$, where $\beta = \min\{1, \beta^*\}$ and β^* is given by (42). Note that

$$\beta^* \to \gamma_1^{-(1-p)/p} - 1, \quad \text{as } \varepsilon \to 0.$$

Since $\gamma_1^{-(1-p)/p} - 1 > 0$ and diverges to infinity as $p \to 0$, one can choose $p \in (0,1)$ such that $\gamma_1^{-(1-p)/p} - 1 > 1$. It then follows that $\beta^* > 1$, and hence $\beta = 1$, for any sufficiently small $\varepsilon > 0$. Hence, $\lambda^{1/p}$ can be arbitrarily close to λ_ε. By the definition of λ_ε in (63), it follows that

$$\ln \lambda \simeq C_\varepsilon e^{-2H_U/\varepsilon^2}. \quad \square$$

Proof of Theorem 1.2. Note that for any $\lambda > 1$ satisfying $\mathbb{E}[\lambda^{\tau_\varepsilon^\eta}] < \infty$, it holds that

$$\mathbb{P}[\tau_\varepsilon^\eta > t] \lambda^t \leq \mathbb{E}[\lambda^{\tau_\varepsilon^\eta}].$$

It then follows from Proposition 4.7(i) and Theorem 4.8 that for any $t > 0$ and $\varepsilon > 0$ sufficiently small,

$$\mathbb{P}[\tau_c > t] \lesssim \lambda^{-t},$$

where λ satisfies $\ln \lambda \simeq C_\varepsilon e^{-2H_U/\varepsilon^2}$. Hence,

$$\limsup_{t \to \infty} \frac{1}{t} \log \mathbb{P}[\tau_c > t] \lesssim -\ln \lambda \simeq -C_\varepsilon e^{-2H_U/\varepsilon^2}. \tag{75}$$

For the reverse inequality, since (X_0, Y_0) is fully supported, it follows that for any $\delta > 0$ sufficiently small,

$$\mathbb{P}[\tau_c > t] \geq \mathbb{P}[\tau_c > t, (X_0, Y_0) \in B_\delta(x_1) \times B_\delta(x_2) \text{ or } B_\delta(x_2) \times B_\delta(x_1)].$$

Note that when X_0 and Y_0 belong to different basins, it holds that $\tau_c \geq \tau_\varepsilon^{(1)}$. Thus, by (68)

$$\begin{aligned} &\mathbb{P}[\tau_c > t, (X_0, Y_0) \in B_\delta(x_1) \times B_\delta(x_2) \text{ or } B_\delta(x_2) \times B_\delta(x_1)] \\ \geq\ &\mathbb{P}[\tau_\varepsilon^{(1)} > t, (X_0, Y_0) \in B_\delta(x_1) \times B_\delta(x_2) \text{ or } B_\delta(x_2) \times B_\delta(x_1)] \ \gtrsim\ \lambda_\varepsilon^{-t}, \end{aligned}$$

and hence

$$\limsup_{t \to \infty} \frac{1}{t} \log \mathbb{P}[\tau_c > t] \gtrsim -C_\varepsilon e^{-2H_U/\varepsilon^2}. \tag{76}$$

Combining (75) and (76) completes the proof of Theorem 1.2. $\square$

4.4. Multi-well potential and proof of Theorem 1.3

In this section, we study the general case of multi-well potentials. Let U be a multi-well potential satisfying **(U3)**, with $L(L>2)$ local minima $x_1,\dots,x_L$ and the corresponding basins $B_1,\dots,B_L$. Let (X_t,Y_t) be a coupling of two solutions of (2).

Similar to the double-well case, several key stopping times need to be defined to estimate the coupling time. By a slight abuse of notation, we continue to use $\tau_\varepsilon^{(1)}$ to denote the infimum time at which X_t and Y_t lie in the same basin, i.e.,

$$\tau_\varepsilon^{(1)}=\inf\Big\{t>h:(X_t,Y_t)\in\cup_{1\le i\le L}B_i\times B_i\Big\}.$$

Define

$$\tau_\varepsilon^{(3)}=\inf\Big\{t>h:(X_t,Y_t)\in\bigcup_{1\le i,j\le L,i\neq j}B_i\times B_j,\text{and for some } s\in(h,t),\ (X_s,Y_s)\in\bigcup_{1\le i\le L}B_i\times B_i\Big\},$$

and let $\tau_\varepsilon^{(3)}=\infty$ if the set is empty. Note that $\tau_\varepsilon^{(3)}$ generalizes $\tau_\varepsilon^{(2)}$ to the case of multiple wells and coincides with $\tau_\varepsilon^{(2)}$ when $L=2$.

In the multi-well setting, a key stopping time of interest is when both X_t and Y_t lie in the vicinity of the (unique) global minimum x_1. Let ξ_1 denote the infimum time at which both X_t and Y_t lie in the basin B_1. Recall, as defined in (30), that $\kappa_{\boldsymbol{X}}(B_1)$ (*resp.* $\kappa_{\boldsymbol{Y}}(B_1)$) denotes the infimum time at which X_t (*resp.* Y_t) enters B_1. Then

$$\xi_1\ge\max\big\{\kappa_{\boldsymbol{X}}(B_1),\kappa_{\boldsymbol{Y}}(B_1)\big\}. \tag{77}$$

Note that X_t and Y_t may enter and exit the basin B_1 multiple times before ξ_1. However, as long as ε is sufficiently small, the typical scenario is that one of the two processes, say X_t, first enters B_1 and "waits" for Y_t to arrive. Although X_t may leave B_1 before Y_t enters, it is highly probable that X_t will stay in nearby basins and return to B_1 shortly after Y_t enters.

The following **(H3)** assumes that ξ_1 is no greater than $\kappa_{\boldsymbol{X}}(B_1)$ (or $\kappa_{\boldsymbol{Y}}(B_1)$) up to an infinitesimal of the same order as $\kappa_{\boldsymbol{X}}(B_1)$ (or $\kappa_{\boldsymbol{Y}}(B_1)$).

(H3) Let (X_t,Y_t) be a coupling of two solutions of (2) such that $(X_0,Y_0)\in\bigcup_{1\le i,j\le L,i\neq j}\overline{B_i\times B_j}$. Then for any $\varepsilon>0$ sufficiently small,

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{P}\big[\big(\xi_1-\max\{\kappa_{\boldsymbol{X}}(B_1),\kappa_{\boldsymbol{Y}}(B_1)\}\big)>t\big]\lesssim e^{-2H_U/\varepsilon^2}. \tag{78}$$

We note that in contrast to **(H2)**, which provides a local characterization of the coupling properties between X_t and Y_t, **(H3)** imposes a *global* condition on the coupling between X_t and Y_t as both processes evolve across the entire potential landscape. In Section 5.4, **(H3)** is numerically verified for the reflection-maximal coupling scheme.

Remark 4.9. A rigorous justification of **(H3)** is highly challenging, as it requires estimating the simultaneous hitting time, i.e., ξ_1, of the coupled process (X_t,Y_t). Although there are some results on the simultaneous hitting time of independent processes [11], to the best of knowledge of the authors, no such result exists for two reflection-coupled stochastic differential equations.

For the multi-well case with $L>2$, **(H1)–(H3)** are assumed. As in the double-well case, the quantity λ_ε is defined as

$$\lambda_\varepsilon=\exp\{C_\varepsilon e^{-2H_U/\varepsilon^2}\},$$

where H_U now represents the essential barrier height in the general form (7), applicable to multi-well potentials. Still, $C_\varepsilon>0$ is any constant, not uniquely determined, such that $\lim_{\varepsilon\to0}C_\varepsilon$ exists and depends only on U.

Under assumption **(H3)** and the initial condition $(X_0,Y_0)\in\bigcup_{1\le i,j\le L,i\neq j}\overline{B_i\times B_j}$, Lemma 2.6 implies

$$\mathbb{P}[\xi_1>t]\lesssim\lambda_\varepsilon^{-t},$$

which, since $\tau_\varepsilon^{(1)}\le\xi_1$, further yields

$$\mathbb{P}[\tau_\varepsilon^{(1)}>t]\lesssim\lambda_\varepsilon^{-t}. \tag{79}$$

Still, similar to the double-well case, we denote

$$\tau_\varepsilon=\tau_\varepsilon^{(3)}\wedge\tau_c.$$

The estimation of τ_ε follows the same reasoning as in the double-well case: If X_t and Y_t initially belong to the same basin, the result directly follows from **(H2)**(ii). If X_t and Y_t initially belong to different basins, the coupling process (X_t,Y_t) is typically decomposed into two stages over the time interval (h,τ_ε). In Stage 1, X_t and Y_t remain in different basins until $\tau_\varepsilon^{(1)}$, at which time they are in the same basin. Stage 2 then follows, during which X_t and Y_t are either successfully coupled within the same basin or remain uncoupled before being separated again by different basins. Hence, similar to (69) for the double-well case, it holds for the multi-well case as well that

$$\tau_\varepsilon=\tau_\varepsilon\circ\theta^{\tau_\varepsilon^{(1)}}+\tau_\varepsilon^{(1)},\quad\mathbb{P}\text{- a.s.}$$

where $\tau_\varepsilon^{(1)}$ and τ_ε are defined in the setting of multi-well potential.

The following is the "multi-well version" of Lemma 4.5 for the general case of $L\ge2$.

Lemma 4.10. *Let (X_t, Y_t) be a coupling of two solutions of (2) such that $(X_0, Y_0) \in \bigcup_{1\le i,j\le L, i\neq j} \overline{B_i \times B_j}$. Assume* **(H2)–(H3)**. *Then for any $t > 0$,*

$$\mathbb{P}[\tau_\varepsilon > t] \lesssim \lambda_\varepsilon^{-t}.$$

Proof. The proof follows similarly to that of Lemma 4.5. As in (71), for any $\delta \in (0, h)$,

$$\mathbb{P}[\tau_\varepsilon > t] \le \int_h^t \mathbb{P}[\tau_\varepsilon^{(1)} > s - \delta]\mathbb{P}[\tau_\varepsilon \circ \theta^s > t - s]ds. \tag{80}$$

Applying (79) and **(H2)**(ii) yields

$$\mathbb{P}[\tau_\varepsilon^{(1)} > s - \delta] \lesssim \lambda_\varepsilon^{-(s-\delta)}, \quad \mathbb{P}[\tau_\varepsilon \circ \theta^s > t - s] \lesssim e^{-r_0(\varepsilon)(t-s)}.$$

Since δ can be arbitrarily small, substituting into (80) yields

$$\mathbb{P}[\tau_\varepsilon > t] \lesssim \lambda_\varepsilon^{-t} \int_0^t \left(\lambda_\varepsilon e^{-r_0(\varepsilon)}\right)^{t-s} ds$$

Thus, for any $\varepsilon > 0$ sufficiently small such that $\lambda_\varepsilon e^{-r_0(\varepsilon)} < 1$, and therefore the integral remains bounded, we obtain

$$\mathbb{P}[\tau_\varepsilon > t] \lesssim \lambda_\varepsilon^{-t}. \quad \square$$

The proof of Theorem 1.3 is analogous to that of Theorem 1.2.

Proof of Theorem 1.3. Analogous to Proposition 4.6 in the double-well case, a combination of assumption **(H2)**(ii) and Lemma 4.10 yields the following result for the multi-well case: for any $\varepsilon > 0$ sufficiently small and any $\lambda \in (1, \lambda_\varepsilon)$,

$$\mathbb{E}[\lambda^{\tau_\epsilon}] < \infty.$$

As demonstrated in both the single and double-well cases, the coupling time τ_c can be written as a finite iteration of τ_ε, specifically $\tau_c = \tau_\varepsilon^\eta$, where η is defined in (73). In analogy with the double-well case in Theorem 4.8, it then follows that for $\varepsilon > 0$ sufficiently small and any $\lambda \in (1, \lambda_\varepsilon)$,

$$\mathbb{E}[\lambda^{\tau_\epsilon^\eta}] < \infty,$$

which implies the exponential tail estimate

$$\mathbb{P}[\tau_c > t] \lesssim \lambda_\varepsilon^{-t}.$$

To establish the corresponding lower bound, the assumption that the initial distribution of (X_0, Y_0) is fully supported is employed. Consider the case where X_t is initialized in the distant basin B_2, and Y_t starts in the basin B_1 associated with the global minimum. Under the assumption **(H1)**, the event that Y_t does not exit the region $\boldsymbol{B}_1$ prior to the entrance of X_t occurs with positive probability, uniformly in both ε and t. Hence,

$$\begin{aligned}\mathbb{P}[\tau_c > t] &\ge \mathbb{P}\left[Y_s \in \boldsymbol{B}_1 \text{ for all } s \in [0,t] \middle| \kappa_{\boldsymbol{X}}(\boldsymbol{B}_1) > t\right] \cdot \mathbb{P}[\kappa_{\boldsymbol{X}}(\boldsymbol{B}_1) > t] \\ &\ge \gamma_0 \cdot \mathbb{P}[\kappa_{\boldsymbol{X}}(\boldsymbol{B}_1) > t] \simeq \lambda_\varepsilon^{-t}\end{aligned} \tag{81}$$

where the last approximation follows from (35), the strengthened version of Lemma 2.6. This completes the proof of Theorem 1.3. $\square$

5. Numerical examples

This section presents numerical examples to verify the theoretical results and the assumptions **(H1)–(H3)** concerning the coupling scheme introduced in the preceding sections. An algorithm is first proposed in Section 5.1 to obtain accurate numerical estimates of the exponential tails of the coupling times. For further details on the coupling algorithm, the reader is referred to [14].

5.1. An algorithm for exponential tail estimation

Let τ_c denote the coupling time. While the rigorous results only establish bounds on the limit superior of $\frac{1}{t}\log\mathbb{P}[\tau_c > t]$, numerical simulations consistently indicate convergence of $\frac{1}{t}\log\mathbb{P}[\tau_c > t]$ as t increases. Therefore, the numerical investigation focuses on computing the exponential decay rate of $\mathbb{P}[\tau_c > t]$ with respect to t, that is,

$$r(\varepsilon) = -\lim_{t\to\infty} \frac{1}{t}\log\mathbb{P}[\tau_c > t],$$

where ε denotes the noise magnitude in (2). Since only a finite number of coupling events can be sampled, an efficient algorithm is required both to provide statistical evidence for the existence of the exponential tail and to estimate its decay rate with reasonable accuracy.

The main challenge is that $\mathbb{P}[\tau_c > t]$ typically does not exhibit exponential decay until t is sufficiently large. It is therefore necessary to identify a suitable threshold t^* such that the tail $\mathbf{1}_{\{\tau_c > t^*\}}(\tau_c - t^*)$ approximately follows an exponential distribution, while keeping t^* as small as possible to ensure that enough samples with $\tau_c > t^*$ are available. However, most exponentiality tests that have been attempted yield a threshold t^* that is too small, resulting in the failure of the log-linear plot of the tail to stabilize into a linear trend. This is likely due to the sensitivity of the plots to small deviations in tail behavior.

The goal of our algorithm is to determine a suitable t^* such that the log-linear plot of $\mathbb{P}[\tau_c > t]$ is approximately linear for all $t > t^*$. That is, the confidence interval of the estimated values of $\mathbb{P}[\tau_c > t]$ should contain a straight line on the logarithmic scale for all $t > t^*$. The algorithm proceeds as follows. First, select a sequence of times $t_0, t_1, \dots, t_N$, where t_N is typically set as the maximum of the sampled coupling times. Let M denote the total sample size, and for each i, let n_i be the number of samples satisfying $\tau_c > t_i$. The Agresti–Coull method [28] provides a confidence interval for each i of the form

$$[\tilde{p}_i^-, \tilde{p}_i^+] := [\tilde{p}_i - z\sqrt{\frac{\tilde{p}_i}{\tilde{M}}(1-\tilde{p}_i)}, \tilde{p}_i + z\sqrt{\frac{\tilde{p}_i}{\tilde{M}}(1-\tilde{p}_i)}],$$

where $\tilde{M} = M + z^2$, $\tilde{p}_i = (n_i + \frac{z^2}{2})/\tilde{M}$, and $z = \Phi^{-1}(1-\alpha/2)$ is the α-quantile of the standard normal distribution. In practice, $z = 1.96$ and $\alpha = 0.05$ are commonly used.

Given any $N_0 \in \{1, \dots, N\}$, a weighted linear regression can be performed to fit the points $(t_i, \log \tilde{p}_i)$ for $i = N_0, \dots, N$, where each point $(t_i, \log \tilde{p}_i)$ is assigned a weight of n_i/M. If the regression yields a linear function of the form $y = at + b$, then N_0 is considered acceptable if it satisfies

$$\left|\{N_0 \le i \le N\} : at_i + b \notin [\tilde{p}_i^-, \tilde{p}_i^+]\right| < \alpha(N - N_0 + 1), \tag{82}$$

ensuring the residuals $\mathbf{1}_{\{\tau_c > t_{N_0}\}}(\tau_c - t_{N_0})$ are statistically consistent with an exponential tail beginning at t_{N_0}. For each candidate N_0, this procedure evaluates whether the tail distribution of τ_c beyond t_{N_0} is approximately exponential. The final choice of N_0 is the smallest index that satisfies the condition (82), which can be efficiently found via binary search over $\{1, \dots, N\}$ in $\mathcal{O}(\log N)$ iterations. The threshold t^* is then defined by t_{N_0}, and the exponential decay rate is given by the slope a of the corresponding weighted regression line.

5.2. Quadratic potential function

The first example considers the quadratic potential function. The primary objective is to numerically verify the theoretical result stated in Theorem 1.1. This example will be revisited in Section 5.7 to examine the consistency of the first passage times between the continuous-time process and its time-h sampled chain as the step size h tends to zero, in accordance with the approximation (45) discussed in Remark 3.2.

Consider the quadratic potential function

$$U(x) = \frac{1}{2}x^T \boldsymbol{A} x, \quad x \in \mathbb{R}^k,$$

where $\boldsymbol{A}$ is a $k \times k$ Lehmer matrix whose entries are given by $\boldsymbol{A}_{ij} = \min(i,j)/\max(i,j)$. The matrix $\boldsymbol{A}$ is symmetric and positive definite [29]. The associated SDE is

$$\mathrm{d}Z_t = -\boldsymbol{A} Z_t \mathrm{d}t + \varepsilon \mathrm{d}W_t, \tag{83}$$

where W_t is a k-dimensional Wiener process, and $\varepsilon > 0$ denotes the noise magnitude.

In the numerical simulations, the time step size h is fixed at 0.001, unless stated otherwise. Fig. 2 displays the probability distribution of the coupling time τ_c. The four panels show $\mathbb{P}[\tau_c > t]$ versus t on a log-linear scale for Lehmer matrices of size 2×2, 4×4, 6×6, and 8×8, respectively. For each case, the noise magnitude ε is set to 0.02, 0.1, 0.5, and 1.5. The slopes and linear fitting in the log-linear plots are determined using the algorithm described in Section 5.1. The smallest eigenvalue of $\boldsymbol{A}$ is indicated in the subtitle of each subplot in Fig. 2.

In all four cases, although the probability distribution of τ_c varies significantly with the noise magnitude, the slopes of the exponential tails remain unchanged. Moreover, the smallest eigenvalue of $\boldsymbol{A}$, which can be computed explicitly, closely approximates the slope of the corresponding exponential tail, with an error of at most 0.01. This observation is consistent with Theorem 1.1, which asserts that the slope of the exponential tail is determined by the convexity of the potential function and is independent of the noise magnitude.

5.3. 1D double-well potential

This subsection considers an asymmetric one-dimensional double-well potential given by

$$U(x) = x^4 - 2x^2 + 0.2x, \quad x \in \mathbb{R}.$$

The potential U has two local minima located at $x = 0.9740$ and $x = -1.0241$. The barrier height that a trajectory must overcome to transition from the left well to the right is approximately 1.2074, while the reverse transition requires overcoming a barrier of approximately 0.8076; see the bottom left panel of Fig. 3.

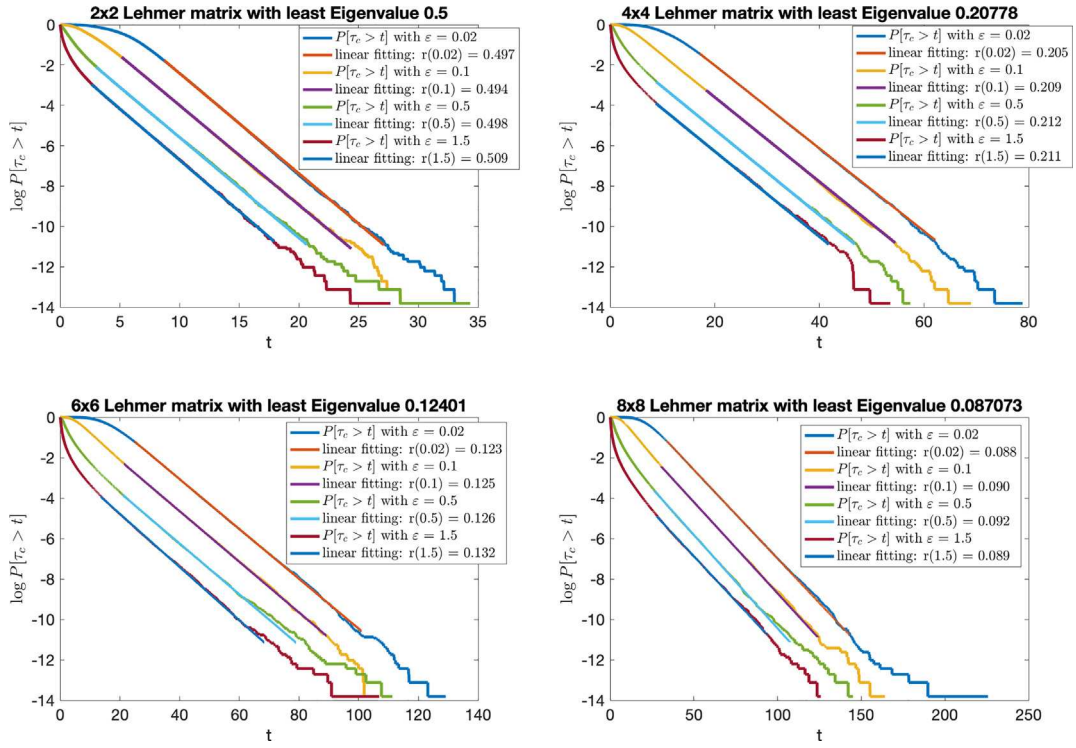


Fig. 2. Log-linear plots of $\mathbb{P}[\tau_c > t]$ versus t and their exponential tails. The four panels correspond to Lehmer matrices of sizes $2, 4, 6$, and 8. The smallest eigenvalue of each matrix is indicated in the title of the corresponding subplot.

The purpose of this example is to numerically verify the theoretical result of Theorem 1.2, which asserts that the exponential tail of the coupling time distribution is determined by the *lower* of the two barrier heights. The time step size and coupling method are the same as those used in the previous examples. The noise magnitudes ε are chosen as $0.32, 0.36, 0.4, 0.45, 0.5, 0.6$, and 0.7. For each value of ε, the exponential tail $r(\varepsilon)$ is estimated using the weighted linear regression algorithm described in Section 5.1. The corresponding results are shown in top panels of Fig. 3. It is observed that the exponential decay rate $r(\varepsilon)$ varies significantly with respect to ε.

In the bottom right panel of Fig. 3, the quantity $y(\varepsilon) := -\varepsilon^2 \log r(\varepsilon)$ is plotted against ε^2, revealing an approximately linear relationship. A linear extrapolation of $y(\varepsilon)$ as $\varepsilon \to 0$ yields the limiting value $y(0) = 1.617$, which closely agrees with the theoretical value $y(0) = 2H_U = 1.615$, where H_U denotes the lower barrier height of the potential. This confirms the validity of Theorem 1.2 in the asymmetric double-well setting.

5.4. Interacting particle system in the double-well potential

This subsection considers a variation of the double-well potential introduced in the previous subsection. Let

$$V(x) = x^4 - 2x^2 + 0.2x, \quad x \in \mathbb{R},$$

denote the double-well potential. Consider three particles moving along V under overdamped Langevin dynamics, with additional pairwise interactions. The total energy potential is given by

$$U(x_1, x_2, x_3) = \sum_{i=1}^{3} V(x_i) + \sigma \sum_{i,j=1,2,3,\, i \neq j} (x_i - x_j)^2,$$

where $\sigma > 0$ is the interaction strength.

The function U has two trivial local minima at $x_1 = x_2 = x_3 = 0.9740$ and $x_1 = x_2 = x_3 = -1.0241$, corresponding to all three particles occupying the same basin of V. For sufficiently small $\sigma > 0$, U also admits six additional local minima, corresponding to configurations in which the particles are distributed across different basins; see the top panel of Fig. 4 for a sample trajectory.

Two extreme regimes of interactions are notable. When $\sigma = 0$, i.e., when there are no interactions among the three particles, the particles move independently, resulting in a barrier height of the energy landscape identical to that of V. When $\sigma \to \infty$, the interaction is strong enough so that the three particles must move together as a single unit, making the barrier height of the energy

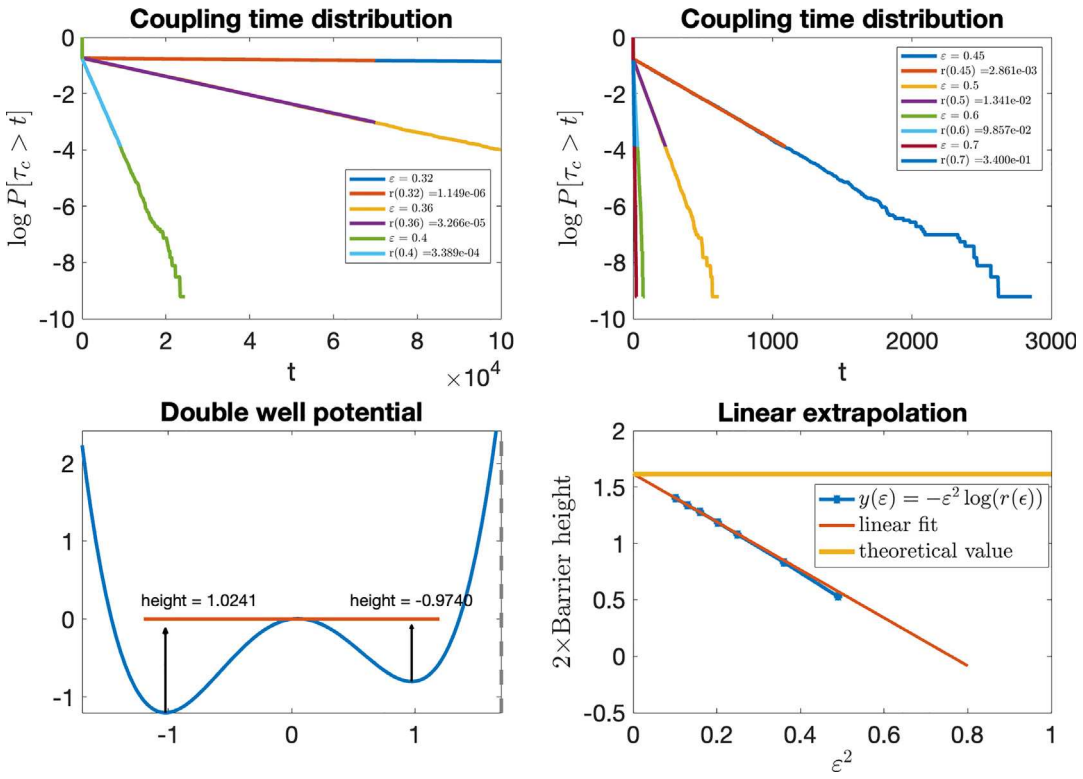


Fig. 3. Top: Coupling time distributions for different noise magnitudes. Bottom left: Asymmetric double-well potential. Bottom right: Linear extrapolation of the essential barrier height.

potential U three times that of V. For any fixed $\sigma > 0$, the essential barrier height H_U lies between the barrier heights of the two extreme cases, i.e., $0.8076 \le H_U \le 3 \times 0.8076 = 2.4228$, with H_U increasing as σ increases.

The estimate H_U, and the distribution of the coupling time τ_c is computed for various values of ε. For $\sigma = 0.05$, values of ε are chosen as 0.4, 0.41, 0.42, 0.43, 0.45, 0.47, 0.5, 0.55, 0.6, 0.7; for $\sigma = 0.1$, values of ε are 0.41, 0.42, 0.43, 0.44, 0.45, 0.47, 0.5, 0.55,0.6, 0.7. The decay rate $r(\varepsilon)$ of the exponential tails is estimated in both cases using linear weighted regression. The relationship between $r(\varepsilon)$ and ε exhibits a similar trend to that observed for the double-well potential in the previous subsection. A linear extrapolation of $y(\varepsilon) := -\varepsilon^2 \log r(\varepsilon)$ provides an estimate of the essential barrier height. As shown in the middle right panel of Fig. 4, the linear extrapolation yields $y(0) = 1.7374$ for $\sigma = 0.05$ and $y(0) = 1.9598$ for $\sigma = 0.1$, both of which are expected to be approximately twice the barrier height $2H_U$, which will be computed using the String method below. As expected, the barrier height increases with the interaction strength among the three particles.

To validate the essential barrier height inferred from the coupling approach, the String method (see, e.g., [30]) is employed to compute the heights of various barriers between the local minima $(0.9740, 0.9740, 0.9740)$ and $(-1.0241, -1.0241, -1.0241)$ in the energy landscape. As shown in Fig. 5, the essential barriers, defined as the highest barrier that a trajectory must overcome to enter the basin of the global minimum, correspond to the leftmost barrier in the lower left panels of Fig. 5: (A) for $\sigma = 0.01$ and (B) for $\sigma = 0.1$, respectively.

In this example, since the three particles are indistinguishable, the energy potential exhibits significant symmetry: the eight local minima can be classified into two types, each consisting of four specific cases. These cases correspond to configurations where (i) all three particles reside in the same basin (global or local), or (ii) two of the three particles lie in one basin (global or local), while the remaining particle resides in the other basin. Fig. 5 shows that the minimal energy path (MEP) connecting the two minima $(0.9740, 0.9740, 0.9740)$ and $(-1.0241, -1.0241, -1.0241)$ passes through all four cases. Thus, the essential barrier height H_U can be attained along such an MEP although, in principle, it should be determined by taking the supremum over all paths connecting any local minima to the global minima. In Fig. 5, the computed values are $H_U = 0.8961$ for $\sigma = 0.05$ and $H_U = 0.9916$ for $\sigma = 0.1$, which correspond to the theoretical values $y(0) = 1.7922$ for $\sigma = 0.05$ and $y(0) = 1.9832$ for $\sigma = 0.1$, respectively.

The result from the String method is further validated using the equivalent characterization (21) by numerically computing all 27 critical points of U, including all the minima and saddle points. The essential barrier heights obtained through this approach are $H_U = 0.8962$ for $\sigma = 0.05$ and $H_U = 0.9916$ for $\sigma = 0.1$, which are nearly identical to those computed using the String method. As shown in Fig. 4, both values also closely match $y(0)/2$, the estimate obtained via linear extrapolation from the exponential tails of coupling times.

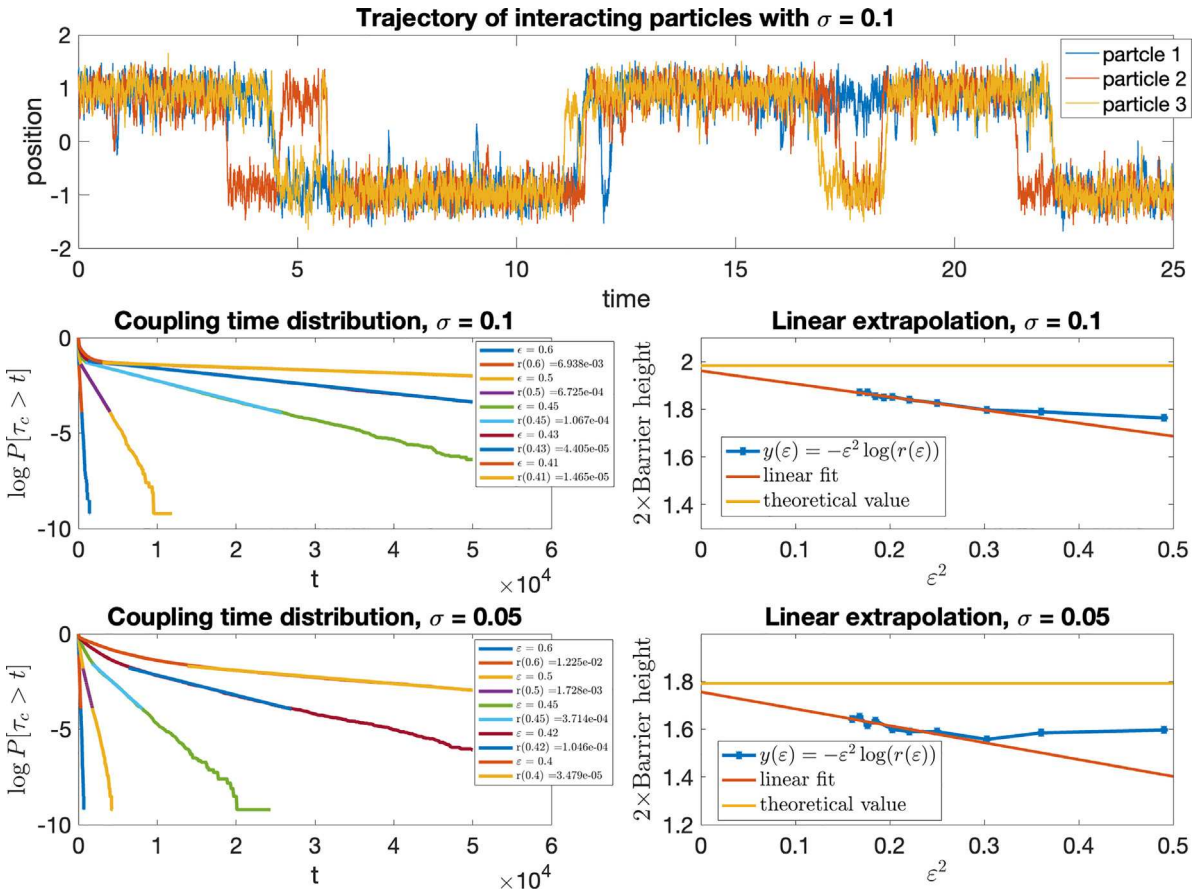


Fig. 4. Top: Sample trajectory of a three-particle interacting system in a double-well potential. Middle: Coupling time distributions and $y(\varepsilon) = -\varepsilon^2 \log r(\varepsilon)$ versus ε^2 for $\sigma = 0.1$. Bottom: Coupling time distributions and $y(\varepsilon) = -\varepsilon^2 \log r(\varepsilon)$ versus ε^2 for $\sigma = 0.05$. Theoretical values of $y(0)$ in the middle-right and bottom-right panels are obtained from the minimum energy path.

5.5. Rosenbrock function

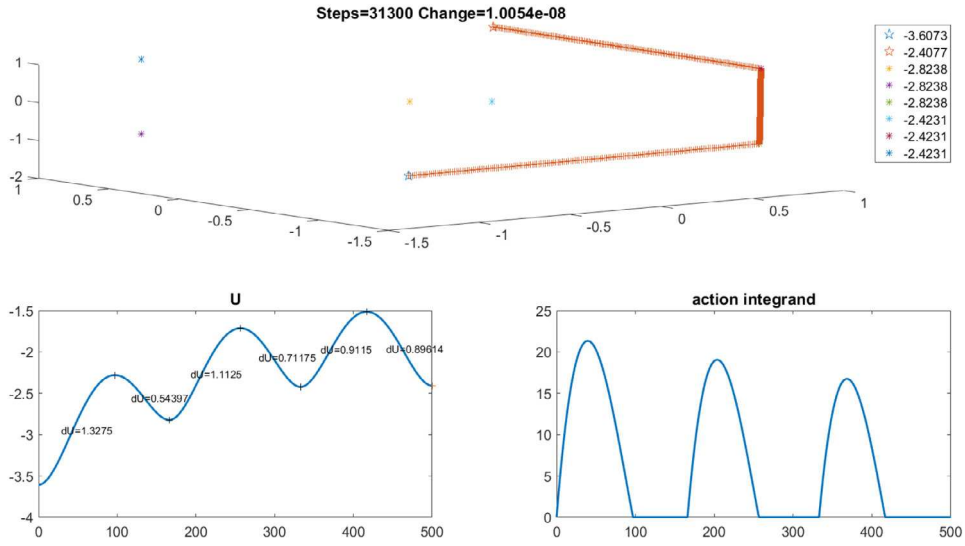
This example examines the well-known non-convex landscape of the Rosenbrock function in both two- and four-dimensional cases. For $N \in \mathbb{N}_+$, the Rosenbrock function is defined as

$$R_N(\boldsymbol{x}) = \sum_{i=1}^{N-1} [b(x_{i+1} - x_i^2)^2 + (a - x_i)^2], \quad \boldsymbol{x} \in \mathbb{R}^N,$$

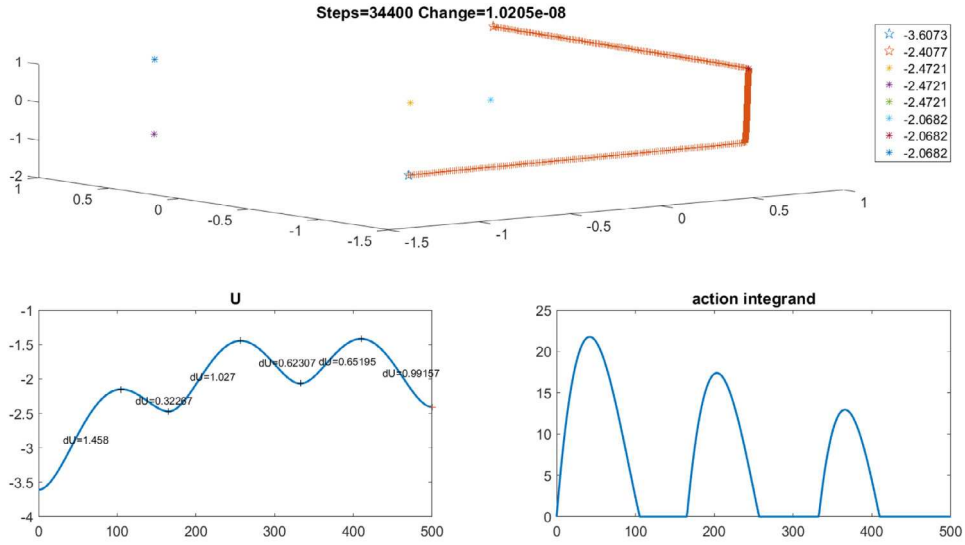
where a and b are constants. In this study, the parameters are chosen as $a = 1$ and $b = 20$. For $N = 2$, the function R_N admits a unique minimum at $(1, 1)$, while for $N = 4$, it possesses a global minimum at $(1, 1, 1, 1)$ and a local minimum at $(-1, 1, 1, 1)$. Fig. 6 illustrates the function landscape: the top-left panel displays $\log R_2(\boldsymbol{x})$, and the bottom-left panel shows a slice of $\log R_4(\boldsymbol{x})$ at $x_3 = x_4 = 1$. A logarithmic scale is used to better visualize the detailed structure near each minimum. In the vicinity of each minimum, the function exhibits a valley-like shape, remaining convex only within a very small neighborhood. The landscape of R_4 cannot be fully captured by a single heat map slice; however, it is straightforward to verify that the convex region of R_4 is relatively small.

The noise magnitude is set as $\varepsilon = 0.001$, 0.01, 0.1, 1.0, 1.5, and 2.0 for R_2, and as $\varepsilon = 0.001$, 0.003, 0.01, 0.03, 0.1, 0.3, and 1.0 for R_4. The corresponding coupling time distributions are shown in the two right panels of Fig. 6. It can be observed that for both cases, when the noise is sufficiently small, the tails of the coupling time distributions appear parallel in the log-linear plot. This behavior arises because the coupling time is primarily determined by the local convexity near the global minimum, in agreement with the result of Theorem 1.1. However, as the noise increases, trajectories are more likely to explore the entire valley rather than remaining confined to the neighborhood of the global minimum. Consequently, the coupling time distributions are altered.

Another interesting phenomenon is that for the potential function R_4, the coupling time distribution does not exhibit an exponentially small tail with respect to the noise magnitude, even when $\varepsilon = 0.001$. This contrasts with the theoretical results for the double-well potential. Moreover, even when one of the coupled processes is initialized at the local minimum $(-1, 1, 1, 1)$, the tail of the coupling time distribution remains largely unchanged, as shown in the plot labeled "$\varepsilon = 0.001$ fixed". This phenomenon



(A) $\sigma = 0.05$



(B) $\sigma = 0.1$

Fig. 5. Minimum Energy Path (MEP) computed using the String method [30] at high numerical resolution. The MEP represents the most likely transition path between two metastable states in the zero-temperature limit of overdamped Langevin dynamics. It is known (e.g., [31]) to reveal barrier heights and descent depths along the transition path, which are labeled by dU values in the bottom-left panel. The top panel visualizes the MEP in three-dimensional space (x_1, x_2, x_3), where the legend lists the values of the potential U at each local minimum. The bottom-right panel displays the integrand of the Freidlin–Wentzell action functional as a function of the arc-length parameterization of the path, serving as a sanity check to verify the correctness of the computed MEP.

occurs because the basin of the local minimum is shallow and separated by a low barrier, which can be easily crossed by a trajectory allowing it to quickly reach the valley of the global minimum.

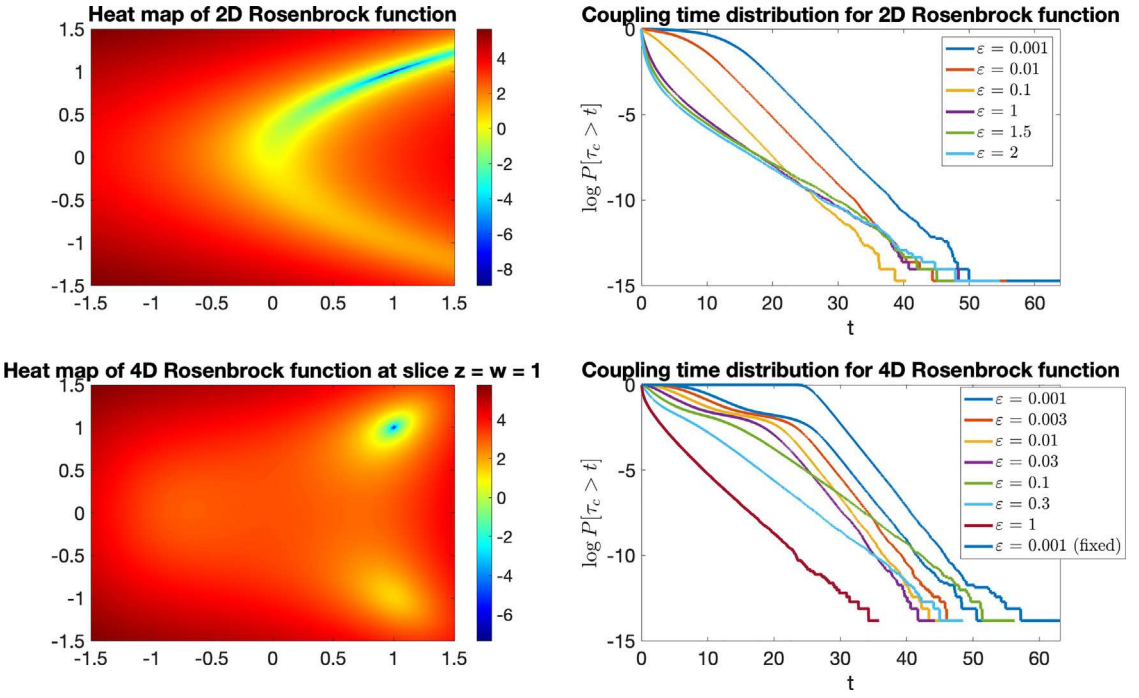


Fig. 6. Top left: Landscape of R_2. Top right: Coupling time distribution for R_2 under different noise magnitudes. Bottom left: Landscape of R_4. Bottom right: Coupling time distribution for R_4 under different noise magnitudes.

5.6. Loss functions of artificial neural networks

This subsection investigates the performance of the coupling method in a high-dimensional setting. Specifically, the training process of an artificial neural network (ANN) with two hidden layers is considered, where the first and second layers contain N_1 and N_2 neurons, respectively. Let $\mathrm{ReLU}(z) = \max\{z, 0\}$ denote the rectified linear unit activation function. The ANN considered here is defined by the following structure

$$\mathbf{h}_1 = \mathrm{ReLU}(W_1 \boldsymbol{x} + \mathbf{b}_1) \tag{84}$$

$$\mathbf{h}_2 = \mathrm{ReLU}(W_2 \mathbf{h}_1 + \mathbf{b}_2) \tag{85}$$

$$y = W_3 \mathbf{h}_2 + b_3\,, \tag{86}$$

where $\boldsymbol{x} \in \mathbb{R}^2$ is the input and $y \in \mathbb{R}$ is the output. The vectors $\mathbf{b}_1 \in \mathbb{R}^{N_1}, \mathbf{b}_2 \in \mathbb{R}^{N_2}$ and $b_3 \in \mathbb{R}$ denote the bias terms. The weight matrices W_1, W_2, and W_3 have dimensions $N_1 \times 2$, $N_2 \times N_1$, and $1 \times N_2$, respectively. Let $\boldsymbol{\theta}$ denote the collection of all trainable parameters, including the entries of $W_1, W_2, W_3, \mathbf{b}_1, \mathbf{b}_2$, and b_3. The total number of parameters is given by

$$\dim \boldsymbol{\theta} = (N_1 N_2 + 3N_1 + 2N_2 + 1).$$

For notational convenience, the ANN defined by (84)–(86) is denoted by $y = \mathrm{NN}(\boldsymbol{\theta}, \boldsymbol{x})$.

The objective of the training process is to approximate the quadratic function $y = |\boldsymbol{x}|^2$ using the ANN. Given a training set $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_M; y_1, \dots, y_M\}$, the loss function is defined by

$$L(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \big(y_i - \mathrm{NN}(\boldsymbol{\theta}, \boldsymbol{x}_i)\big)^2\,,$$

where the training set size is fixed at $M = 100$. The input points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{100}$ are uniformly sampled from $[-1, 1]^2$, and the corresponding target values are given by $y_i = |\boldsymbol{x}_i|^2$. The first column of Fig. 8 illustrates the distribution of the collocation points and the target function $y = |\boldsymbol{x}|^2$. The goal is to analyze the structure of the loss function $L(\boldsymbol{\theta})$.

The coupling method is applied to three ANNs with different hidden layer sizes: $N_1 = 4, N_2 = 3$ (referred to as the "small ANN"), $N_1 = N_2 = 10$ (the "medium ANN"), and $N_1 = N_2 = 20$ (the "large ANN"). In this example, the small ANN is under-parameterized, while the large ANN is over-parameterized. It is often believed that over-parameterization tends to reduce barrier heights in the loss landscape of ANNs (see, e.g., [32–38]). However, rigorous justification remains elusive due to the complex structure of high-dimensional loss functions. The coupling-based approach proposed here may offer a viable tool in this regard by computing the essential barrier height of such loss functions.

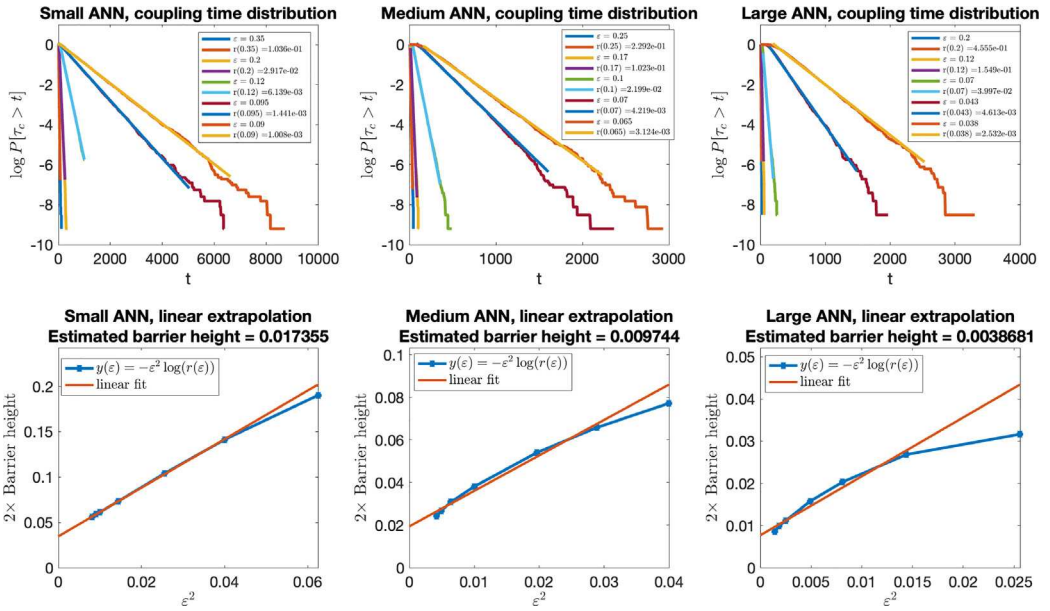


Fig. 7. Coupling time distributions and linear extrapolation of $y(\varepsilon) = -\varepsilon^2 \log r(\varepsilon)$ versus ε^2. Left: small ANN; middle: medium ANN; right: large ANN.

Fig. 7 presents the coupling time distributions for the three neural networks under ten different noise magnitudes. For visual clarity, only five noise levels are shown. As in previous examples, the slopes are estimated via weighted linear regression. The six smallest values of ε^2 are used for the linear extrapolation of $y(\varepsilon) := -\varepsilon^2 \log r(\varepsilon)$ versus ε^2, as displayed in the lower panels. It is observed that the large ANN exhibits a lower essential barrier height. More precisely, no significant barrier is detected within the region explored by the coupling method. This observation is consistent with the findings in [38], which adopts a different approach based on computing the MEPs between the local minima of the loss surface. Although it is theoretically possible that a high-barrier local minimum exists in a remote region not reached by the coupling trajectories, such cases have not been reported to the best of our knowledge. Moreover, practical ANN training is typically regularized, which prevents $|\theta|$ from becoming excessively large.

The small ANN in this example is under-parameterized, as it contains only 31 parameters to be learned, whereas the training set comprises 100 samples. As illustrated in Fig. 7, the loss function of the small ANN exhibits a much larger essential barrier height compared to both the medium and large ANNs. Regarding the training performance, when initialized randomly, the small ANN may converge to a "bad" local minimum that fails to accurately approximate the target function (see the middle panels of Fig. 8). In contrast, for all tested initial conditions, both the medium and large ANNs consistently converge to a "good" local minimum of the training loss function, yielding satisfactory approximations of the target function (see the right panels of Fig. 8). This finding aligns with existing studies on the loss landscapes of ANNs [1,39,40].

5.7. Numerical verification of assumptions

In this subsection, assumptions **(H1)–(H3)** proposed in Section 4 are numerically verified. In addition, the consistency between the first passage times of the continuous-time process and its discrete-time counterpart is examined, as discussed in Remark 3.2.

5.7.1. Numerical verification of Remark 3.2 .

We first numerically verify the consistency of the first passage times defined in Remark 3.2 as two reflection-coupled trajectories approach each other. Specifically, we examine the assumption (45)

$$\lim_{h\to 0} |\tau_h^0 - \tau_h^h| = 0, \quad \mathbb{P}\text{-a.s.}$$

where $\tau_h^0 = \inf_{t>0}\{|X_t - Y_t| = 2\sqrt{h}\}$ denotes the continuous-time first passage time, and $\tau_h^h = h \cdot \inf_{n>0}\{|X_{nh} - Y_{nh}| = 2\sqrt{h}\}$ is its discrete-time counterpart based on a time-h sampled chain.

This verification can be conducted using an extrapolation argument. Let $h_1 = h/n$ for some integer n, and define the first passage time of the time-h_1 sampled chain by

$$\tau_h^{h_1} = h_1 \cdot \inf_{n>0}\{|X_{nh_1} - Y_{nh_1}| = 2\sqrt{h}\}.$$

By the strong approximation property of the Euler–Maruyama scheme for SDEs,

$$\lim_{h_1\to 0} \tau_h^{h_1} = \tau_h^0, \quad \mathbb{P}\text{-a.s.},$$

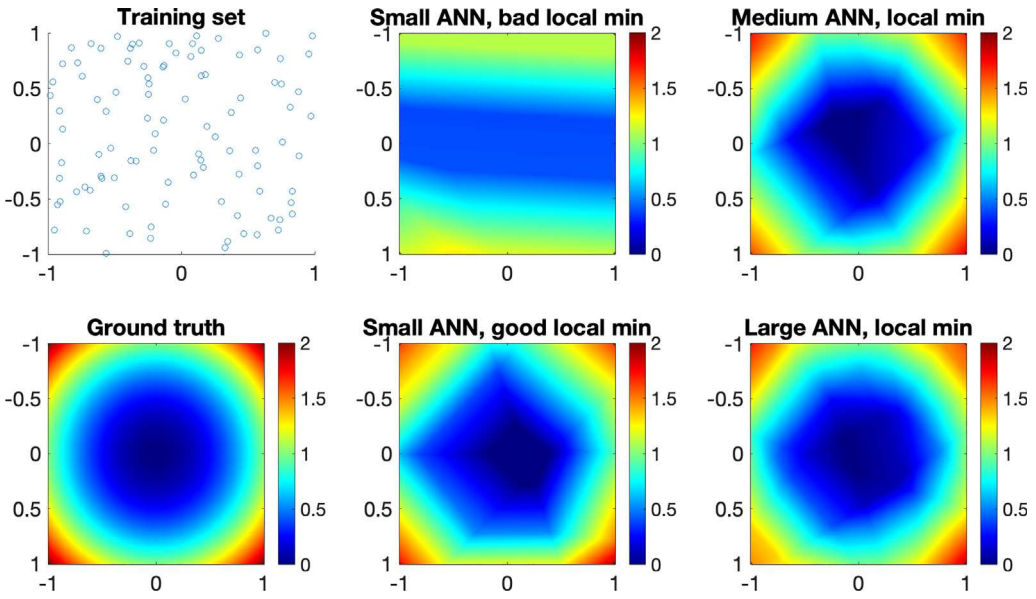


Fig. 8. Left column: Training set and target function. Middle column: $y = \mathrm{NN}(\boldsymbol{\theta}, \boldsymbol{x})$ for $\boldsymbol{\theta}$ at the "bad" and "good" local minima of the small neural network. Right column: $y = \mathrm{NN}(\boldsymbol{\theta}, \boldsymbol{x})$ for the medium and large neural networks.

so it suffices to compare τ_h^h and $\tau_h^{h_1}$, with the latter serving as an approximation of τ_h^0 for the same trajectory.

We apply the extrapolation method to both the quadratic potential function in Section 5.2 and the interacting particle system in Section 5.4. In the quadratic potential case, the initial values of X_t and Y_t are set to $(0.5, 0.7)$ and $(-0.5, -0.6)$, respectively. For the interacting particle system, X_t and Y_t are initialized at $(1, 1, 1)$ and $(-1, -1, -1)$, respectively, indicating that they belong to the basins of different local minima.

In the top-left and bottom-left panels of Fig. 9, the quantity $(\tau_h^h - \tau_h^{h_1})$ is plotted against $\sqrt{h_1}$ for five different values of h. In both examples, this difference exhibits approximately linear behavior as $\sqrt{h_1} \to 0$. An extrapolation at $h_1 = 0$ provides an estimate of $(\tau_h^h - \tau_h^0)$. The top-right and bottom-right panels of Fig. 9 display $(\tau_h^h - \tau_h^0)$ versus $\sqrt{h}$ for $h = 0.0002, 0.0005, 0.001, 0.005$, and 0.01, respectively. The results show that this error decreases as $h \to 0$, and a linear fit suggests that $(\tau_h^h - \tau_h^0)$ is approximately proportional to $\sqrt{h}$, consistent with the findings of [41,42]. Although the error in estimating τ_h is larger for the interacting particle system due to the presence of multiple local minima, the numerical results still exhibit the expected convergence behavior as $h \to 0$.

In the following subsections, the interacting particle system described in Section 5.4 is used to numerically verify assumptions **(H1)–(H3)**. The coupling strength is set to $\sigma = 0.05$.

5.7.2. *Numerical verification of* **(H1)**

Let $X_0 = (1, 1, 1)$ and $Y_0 = (-1, -1, -1)$, ensuring that the trajectory Y_t is initiated near the global minimum. Based on the barrier heights illustrated in Fig. 5 and the definitions of $\mathcal{I}$, the set $\boldsymbol{B}_1$ is identified as the complement of the basin containing $(1, 1, 1)$. The simulation is performed under four different noise magnitudes: $\varepsilon = 0.6, 0.65, 0.7$, and 0.75. At each step of the Euler–Maruyama scheme,it is numerically checked whether X_t and Y_t lie in $\boldsymbol{B}_1$. The criterion for determining whether a point $\boldsymbol{x} = (x_1, x_2, x_3)$ belongs to $\boldsymbol{B}_1$ is as follows: for each $i = 1, 2, 3$, if either $x_i > 0.11$, or $0 \le x_i < 0.11$ while $-\partial U/\partial x_i > 0$, then $\boldsymbol{x} \notin \boldsymbol{B}_1$. This condition is sufficient for all samples in our numerical simulation.

Remarkably, across tens of millions of samples, Y_t was *never* observed to exit $\boldsymbol{B}_1$ before X_t entered it. A similar phenomenon is observed in the one-dimensional double-well potential, where Y_t remains in B_1 until X_t enters. This can be explained by noting that reflection-coupled Brownian motions have the same action functionals. When X_t exits the basin B_2 and enters $\boldsymbol{B}_1$, the action functional of the associated driving Brownian motion of X_t is highly likely to be close to H_U. Consequently, the action functional corresponding to the Brownian motion term in Y_t is unlikely to be sufficiently large to drive Y_t out of $\boldsymbol{B}_1$. Therefore, assumption **(H1)** is numerically verified with the even stronger conclusion that

$$\mathbb{P}\left[Y_s \in \boldsymbol{B}_1 \text{ for all } s \in [0, t] \middle| \kappa_{\boldsymbol{X}}(\boldsymbol{B}_1) > t\right] \approx 1.$$

5.7.3. *Numerical verification of* **(H2)**

Let B_2 denote the basin of attraction containing $(1, 1, 1)$. According to Proposition 4.3 and Remark 4.4, it suffices to verify condition (a) therein. Specifically, this involves numerically estimating the probability that a trajectory enters the interior of B_2. The approximate boundary of B_2 is depicted in the left panel of Fig. 10. It suffices to consider initial points from the boundary, as the probability is expected to be higher when starting from the interior.

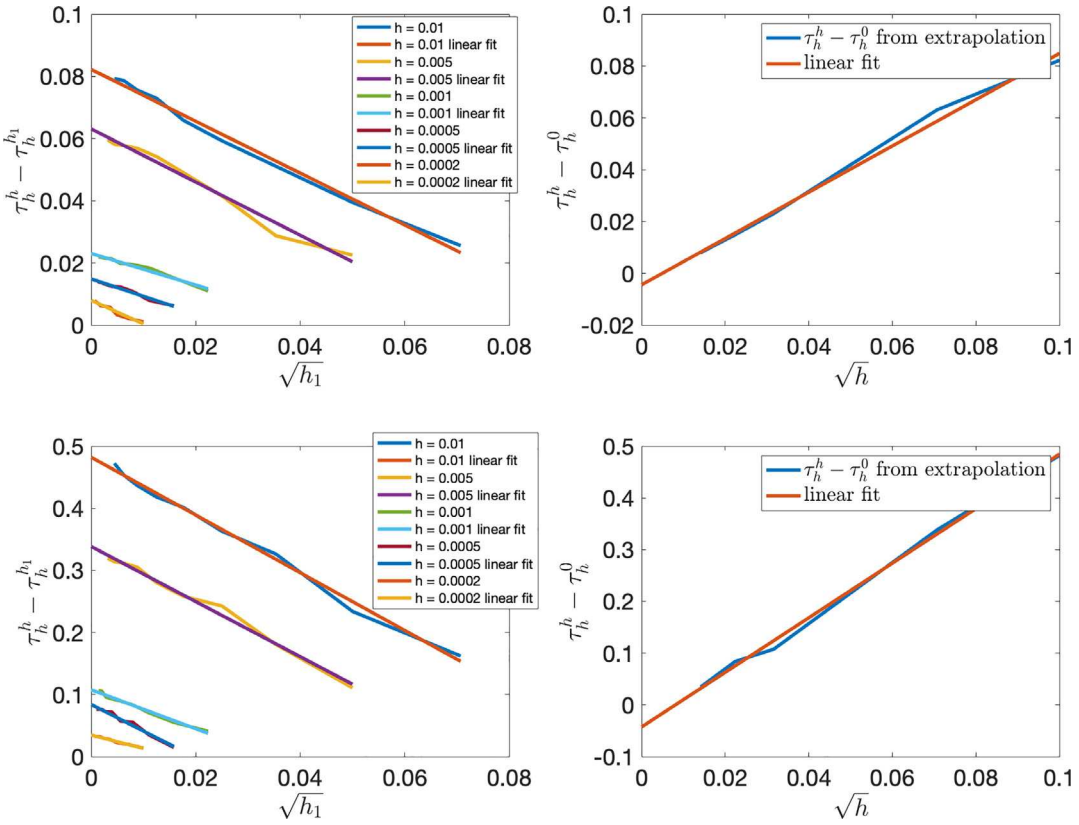


Fig. 9. Left: $(\tau_h^h - \tau_h^{h_1})$ vs. $\sqrt{h_1}$ for five different values of h. Top left: Quadratic potential function. Bottom left: Interacting particle system. Right: $(\tau_h^h - \tau_h^0)$ vs. $\sqrt{h}$ with a linear fit. Top right: Quadratic potential function. Bottom right: Interacting particle system.

Three initial values of X_0 are selected from a corner, an edge, and a face of ∂B_2, respectively, as marked in red in Fig. 10 (Left). The initial value of Y_0 is fixed at $(1, 1, 1)$. For each case, the probability that the coupled process (X_t, Y_t) remains in the δ-interior of $B_2 \times B_2$ (with $\delta = 0.01$) throughout the time interval $[h, T_0]$ is computed, where $h = 0.05$, $T_0 = -\log \varepsilon$, and ε varies from 0.001 to 0.01. As shown in Fig. 10, this probability remains uniformly bounded from below as $\varepsilon \to 0$. This numerically confirms condition (a) of Proposition 4.3, thereby verifying assumption **(H2)**.

5.7.4. Numerical verification of **(H3)**

Assumption **(H3)** is numerically verified by computing the overshoot time. The criterion for determining whether a trajectory enters the basin B_1 is the same as that used in Section 5.7.2. The noise magnitudes are set to $0.5, 0.55$, and 0.6. For each value of ε, the probability distribution of the overshoot time, given by $\xi_1 - \max\{\kappa_{\boldsymbol{X}}, \kappa_{\boldsymbol{Y}}\}$, is estimated using 1×10^7 samples.

As illustrated in Fig. 11, the tail distribution of $\xi_1 - \max\{\kappa_{\boldsymbol{X}}, \kappa_{\boldsymbol{Y}}\}$ exhibits a two-phase behavior. The second phase corresponds to the scenario where one of the trajectories, X_t or Y_t, makes an excursion to other basins after entering B_1 and subsequently returns, while the other trajectory remains within B_1. Due to the low probability of such an event, a large number of samples are required to capture the exponential tail. In Fig. 11, the distributions of the overshoot time and coupling time are compared. In the log-linear plot, the slope of the overshoot time decreases rapidly as the noise magnitude decreases, yet it remains steeper than that of the coupling time. As the theoretical result indicates that the tail of the coupling time distribution is close to the essential barrier height H_U, this numerical observation thereby verifies assumption **(H3)**.

6. Conclusion and further discussions

This paper investigates the relationship between the geometry of a multi-dimensional potential landscape and the distributions of coupling time for the overdamped Langevin system associated with the potential. This study is motivated by the fact that the exponential tail of the coupling time distribution provides a lower bound for the spectral gap of the Fokker–Planck operator governing the Langevin dynamics. It has long been believed that certain geometric properties of a region can be inferred from the spectrum of an associated differential operator, as famously illustrated by Kac's question [43], "Can one hear the shape of a drum?" In a similar spirit, this work takes a preliminary step toward understanding the structure of a potential landscape by establishing connections between its geometry and the statistical properties of coupling times.

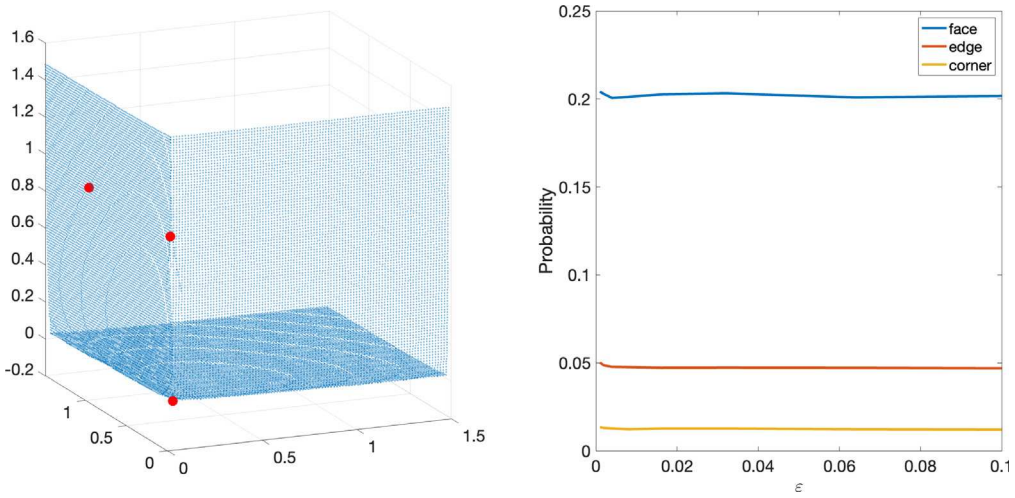


Fig. 10. Left: Approximate boundary of $\boldsymbol{B}_1$, with three initial values of X_0 marked in red. Right: Probability that (X_t, Y_t) remains in the δ-interior of $\boldsymbol{B}_1$ over the time interval $[h, T_0] = [0.05, -\log\varepsilon]$.

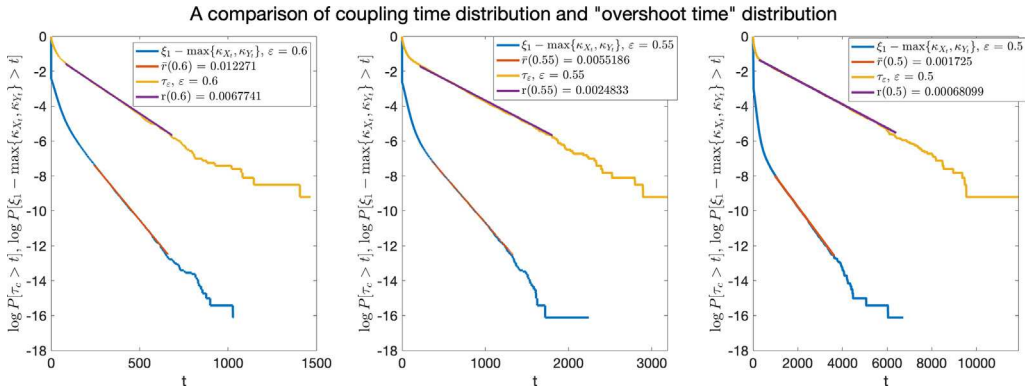


Fig. 11. Comparison of probability distributions of the overshoot time $\xi_1 - \max\{\kappa_{\boldsymbol{X}}, \kappa_{\boldsymbol{Y}}\}$ and the coupling time. The left, middle, and right panels correspond to $\varepsilon = 0.6, 0.55$, and 0.5 respectively. The asymptotic slope of the overshoot time distribution in the log-linear plot is denoted by $\bar{r}$.

It is shown that, in the limit of vanishing noise, the exponential tails of the coupling time distributions exhibit *qualitatively* distinct behaviors in the single-well potential and multi-well settings. Specifically, for a strongly convex single-well potential, the rate of exponential tail is uniformly bounded below by a constant that depends on the convexity of the potential. In contrast, for a multi-well potential, the rate of the exponential tail decays exponentially as the noise strength tends to zero. These results are supported by both theoretical analysis and numerical verification.

The coupling scheme used in this paper combines reflection coupling and maximal coupling to improve efficiency. It is observed that the upper bound on the tail distribution obtained through this scheme is close to optimal, in the sense that it nearly achieves equality in the coupling inequality. To estimate the exponential decay rate of the tail in the small noise regime, a linear extrapolation is employed. This decay rate is governed by the essential barrier height, a concept introduced in this paper to capture the global structural features of the potential landscape. In particular, the essential barrier height is applied to analyze the loss landscape of artificial neural networks, and the corresponding numerical observations are consistent with findings from related studies employing alternative methodologies.

Although this work focuses on the distribution of coupling times, further information about the potential landscape is expected to be extracted from the coupling-based analyses. For instance, the distribution of coupling locations may provide additional insights into the geometry of the underlying landscape. Furthermore, the present study only concerns the tail of the coupling time distribution, which is associated with the principal eigenvalue of the Fokker–Planck operator. An investigation of conditional coupling times – specifically, those conditioned on avoiding coupling in the deepest well – could reveal spectral information associated with non-principal eigenvalues, which correspond to the lower energy barriers. These extensions represent promising directions for future research.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yao Li reports financial support was provided by National Science Foundation. Molei Tao reports financial support was provided by National Science Foundation. Molei Tao reports financial support was provided by US Department of Energy. Shirou Wang reports financial support was provided by National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are grateful to the anonymous referees for their valuable comments, which significantly improved the manuscript.

Appendix A. Proof of Lemma 2.6

To establish the exponential tail, more refined estimates of the eigenvalues of the Dirichlet operator are required. Specifically, let $\mathcal{L}_D^\epsilon$ denote the infinitesimal generator of the process defined in (2), and let $D \subset \mathbb{R}^k$ be an open set with a regular boundary ∂D. Based on potential theory, [18,19] provide sharp estimates for the low-lying eigenvalues and the corresponding eigenfunctions of the Dirichlet problem

$$\begin{aligned} \mathcal{L}_D^\epsilon u - \lambda u = 0, \quad &\text{in} \quad D \\ u = 0, \quad &\text{in} \quad D^c \end{aligned} \tag{87}$$

The following Proposition A.1 summarizes the sharp bounds of the principal eigenvalues, as established in Proposition 3.2 of [19] and Theorem 3.1 of [18]. The corresponding bound on the principal eigenfunction is stated in Proposition A.2, which follows from Proposition 3.3 in [19].

Proposition A.1 (*Sharp Bound on Eigenvalues [18,19]*). *Assume that $D \subseteq \mathbb{R}^d$ is open and let $U : D \to \mathbb{R}$ be a potential function satisfying* **(U3)**. *Suppose that D contains $L \geq 1$ local minima of U, and that there exists a unique minimum $x \in D$ such that*

$$U(z^*(x, D^c)) - U(x) = \max_{1 \leq i \leq L} \{U(z^*(x_i, D^c)) - U(x_i)\}.$$

Let $B \equiv B_x$ denote a neighborhood of x, and denote the first entrance time of X_t into any subset $A \subseteq \mathbb{R}^d$ by τ_A. Then there exist constants $\alpha > 0$, $C < \infty$, and $\delta > 0$, independent of ϵ, such that the principal eigenvalue $\lambda_1 < 0$ of $\mathcal{L}_D^\epsilon$ satisfies

$$\frac{\mathrm{cap}_B(D^c)}{\|h_{B,D^c}\|_2^2}(1 - C\epsilon^\alpha)(1 - e^{-\delta/\epsilon^2}) \leq |\lambda_1| \leq \frac{\mathrm{cap}_B(D^c)}{\|h_{B,D^c}\|_2^2}(1 + C\epsilon^\alpha)(1 + e^{-\delta/\epsilon^2}),$$

where $h_{B,D^c}(z) := \mathbb{P}_z[\tau_B < \tau_{D^c}]$, the norm $\|\cdot\|_2$ is taken with respect to the invariant probability measure π^ϵ of (2), and capacity

$$\mathrm{cap}_B(D^c) = e^{-2U(z^*)/\epsilon^2} \frac{(2\pi\epsilon)^d}{2\pi} \frac{|\lambda_1^*(z^*)|}{\sqrt{|det(\nabla^2 U(z^*))|}}(1 + O(\epsilon|\ln\epsilon|))$$

for $z^* = z^*(B, D^c)$. Here, $\lambda_1^*(z^*)$ denotes the negative eigenvalue of the Hessian of U at z^*.

Proposition A.2 (*Sharp Bounds on Eigenfunctions [18,19]*). *Under the assumptions of Proposition A.1, let ϕ_1 be the eigenfunction of $\mathcal{L}_D^\epsilon$ corresponding to λ_1, normalized such that $\inf_{x \in \partial B} \phi_1 = 1$. Then*

$$h_{B,D^c}(z) \leq \phi_1(z) \leq h_{B,D^c}(z)(1 + C\epsilon^\alpha)(1 + e^{-\delta/\epsilon^2}).$$

These estimates on the principal eigenvalue and eigenfunction yield the exponential tail for the first hitting time in Lemma 2.6.

Proof of Lemma 2.6. The argument is a modification of Theorem 1.4 from [19]. Assume $z \notin B_1$, since the bound is trivial otherwise. Set $D = B_1^c$ and $B = B_2$. Then

$$\mathrm{cap}_B(D^c) = e^{-2U(z^*)/\epsilon^2} \frac{(2\pi\epsilon)^d}{2\pi} \frac{|\lambda_1^*(z^*)|}{\sqrt{|\det(\nabla^2 U(z^*))|}}(1 + O(\epsilon|\ln\epsilon|)),$$

where $z^* = z^*(x_1, x_2)$.

Note that for sufficiently small $\epsilon > 0$, if $h_{B,D^c}(z) \simeq 1$ for $z \in B_i$, then it follows that $\Phi(x_i, x_2) < \Phi(x_i, x_1)$. This, in turn, implies $U(x_i) > U(x_2)$; otherwise, we would have

$$\Phi(x_2, x_1) - U(x_2) < \Phi(x_i, x_1) - U(x_i),$$

which contradicts (29). In particular, $h_{B,D^c}(z) \simeq 1$ for $z \in B_2$. Therefore, integrating against π^ϵ, it follows that

$$\|h_{B,D^c}\|_2^2 \simeq e^{-2U(x_2)/\epsilon^2}.$$

Noting that $H_U = U(z^*) - U(x_2)$, it follows from Proposition A.1 that

$$|\lambda_1| = e^{-2H_U/\varepsilon^2}(1 + O(\varepsilon^\alpha))(1 + O(e^{-\delta/\varepsilon^2})).$$

Now observe that

$$\mathbb{P}_z[\kappa_{\boldsymbol{Z}}(B_1) > t] = \left(e^{t\mathcal{L}_D^\varepsilon}\mathbf{1}_{B_1^c}\right)(z),$$

so that

$$\mathbb{P}_z[\kappa_{\boldsymbol{Z}}(B_1) > t] \le A_{z,\varepsilon}e^{-\lambda_1 t}$$

for some constant $A_{z,\varepsilon}$ depending on z and ε.

For the lower bound, note that $\mathcal{L}_D^\varepsilon$ is self-adjoint in the weighted space $L^2_{\pi^\varepsilon}$, and thus its eigenfunctions form an orthogonal basis. It follows that

$$\mathbb{P}_z[\kappa_{\boldsymbol{Z}}(B_1) > t] = \left(e^{t\mathcal{L}_D^\varepsilon}\mathbf{1}_{B_1^c}\right)(z) \ge \mathbb{E}_{\hat{\pi}^\varepsilon}[\phi_1\mathbf{1}_{B_1^c}] \cdot e^{-\lambda_1 t} \cdot \phi_1(z),$$

where $\hat{\pi}^\varepsilon$ denotes the normalized restriction of π^ε to D. By Proposition A.2,

$$\phi_1(z) \simeq h_{B,D^c}(z)(1 + O(\varepsilon^\alpha)).$$

Thus, for $z \in B$, one has $h_{B,D^c}(z) \simeq 1$, and hence

$$\mathbb{P}_\mu[\kappa_{\boldsymbol{Z}}(B_1) > t] \ge \mathbb{E}_{\hat{\pi}^\varepsilon}[\phi_1\mathbf{1}_{B_1^c}] \cdot e^{-\lambda_1 t} \cdot \mathbb{E}_\mu(\phi_1).$$

Since μ is fully supported, the term $\mathbb{E}_\mu(\phi_1) \simeq \mathbb{E}_\mu(h_{B,D^c}) > 0$, yielding the desired lower bound. □

Remark A.3. The leading-order term of $A_{z,\varepsilon}$ is proportional to $\phi_1(z)$. Furthermore, if z lies in the interior of B_2, then $\mathbb{P}_z[\tau_{B_1} > \tau_{B_\varepsilon(x_2)}] \simeq 1$. Consequently, the leading-order term of $A_{z,\varepsilon}$ can be bounded by a constant that is independent of both ε and z. Any dependence on ε and z arises exclusively through the coefficients associated with higher-order eigenfunctions in the spectral decomposition of the semigroup. On the time scale $\kappa_{\boldsymbol{Z}}(B_1) = O(e^{2H_U/\varepsilon^2})$, these higher-order terms are of order $O(e^{-\delta/\varepsilon^2})$, making the prefactors $A_{z,\varepsilon}$ and $A_{\mu,\varepsilon}$ effectively independent of ε.

Remark A.4. The result in Lemma 2.6 still holds if B_1 is replaced by $\boldsymbol{B}_1$, as defined in (34). This follows from (37), which ensures that, when $D = \boldsymbol{B}_1^c$, the deepest local minimum remains to be x_2 and the height of the saddle $z^*(B, D^c)$ remains to be $U(z^*(x_1, x_2))$. Hence, the proof of Lemma 2.6 continues to hold.

Appendix B. Proof of Proposition 4.3

We first establish that the event in which X_t and Y_t remain within the same basin and couple within a finite time interval of order $\mathcal{O}(-\log\epsilon)$ occurs with a strictly positive probability. This directly implies **(H2)**(i).

By assumption (a), for any initial value $(X_0, Y_0) \in \overline{B_i \times B_i}$, the pair (X_{T_0}, Y_{T_0}) belongs to the δ-interior $B_\delta^i \times B_\delta^i$ with probability γ_0. Since B_i is the basin of attraction of x_i, denote by $\mathbf{x}_t$ the deterministic gradient flow $\dot{\mathbf{x}}_t = -\nabla U(\mathbf{x}_t)$. Then there exists a constant $T_1 = \mathcal{O}(1)$ such that for any $\mathbf{x}_0 \in B_\delta^i$, the deterministic trajectory satisfies $\mathbf{x}_{T_1} \in B_{c,1}^i \subset B_c^i$, where $B_{c,1}^i$ is an open subset in the interior of B_c^i.

By the standard small random perturbation argument (see, for instance, Chapter 4, Lemma 2.1 of [24]), for any $\varepsilon > 0$ sufficiently small and any finite time interval, both processes X_t and Y_t remain close to the deterministic trajectory $\mathbf{x}_t$ with high probability, say at least 0.9. Thus, combining this with assumption (a), define the event

$$E_0 := \left\{(X_t, Y_t) \in B_i \times B_i \text{ for all } t \in [h, T_0 + T_1),\ \text{and } (X_{T_0+T_1}, Y_{T_0+T_1}) \in B_{c,1}^i \times B_{c,1}^i\right\}.$$

Then

$$\mathbb{P}[E_0 \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \ge 0.9\gamma_0.$$

Let $\tilde{U}$ be a strongly convex potential satisfying **(U1)** such that $U = \tilde{U}$ on B_c^i. Denote by $(\tilde{X}_t, \tilde{Y}_t)$ the coupled process associated with $\tilde{U}$. Then (X_t, Y_t) coincides with $(\tilde{X}_t, \tilde{Y}_t)$ as long as $(\tilde{X}_t, \tilde{Y}_t)$ remains in $B_c^i \times B_c^i$. By Theorem 1.1, for any $\gamma_2 \in (0,1)$, there exists $T_2 = \mathcal{O}(1)$ such that

$$\mathbb{P}[\tilde{X}_t \text{ and } \tilde{Y}_t \text{ couple before } T_2 \mid (\tilde{X}_0, \tilde{Y}_0) \in B_{c,1}^i \times B_{c,1}^i] \ge \gamma_2.$$

Moreover, since $(\tilde{X}_t, \tilde{Y}_t)$ remains in $B_{c,1}^i \times B_{c,1}^i$ when $\varepsilon = 0$, the small random perturbation argument yields that for any $\varepsilon > 0$ sufficiently small and any $\gamma_3 \in (0,1)$, there holds

$$\mathbb{P}[(\tilde{X}_t, \tilde{Y}_t) \in B_c^i \times B_c^i \text{ for all } t \le T_2 \mid (\tilde{X}_0, \tilde{Y}_0) \in B_{c,1}^i \times B_{c,1}^i] \ge \gamma_3.$$

Choose γ_2 and γ_3 such that $\gamma_2 + \gamma_3 > 1$, and define the events

$$E_1 := \{\tilde{X}_t \text{ and } \tilde{Y}_t \text{ couple before } T_2\}, \quad E_2 := \{(\tilde{X}_t, \tilde{Y}_t) \in B_c^i \times B_c^i \text{ for all } t \le T_2\}.$$

Then

$$\begin{aligned}\mathbb{P}[E_1 \cap E_2 \mid (\tilde{X}_0, \tilde{Y}_0) \in B^i_{c,1} \times B^i_{c,1}] &\geq \mathbb{P}[E_1 \mid (\tilde{X}_0, \tilde{Y}_0) \in B^i_{c,1} \times B^i_{c,1}] \\ &\quad + \mathbb{P}[E_2 \mid (\tilde{X}_0, \tilde{Y}_0) \in B^i_{c,1} \times B^i_{c,1}] - 1 \\ &\geq \gamma_2 + \gamma_3 - 1 > 0.\end{aligned}$$

Since (X_t, Y_t) coincides with $(\tilde{X}_t, \tilde{Y}_t)$ on $E_1 \cap E_2$, it follows that

$$\mathbb{P}[X_{T_0+T_1+t} = Y_{T_0+T_1+t} \text{ for some } t \in [0, T_2] \mid (X_{T_0+T_1}, Y_{T_0+T_1}) \in B^i_{c,1} \times B^i_{c,1}] \geq \gamma_2 + \gamma_3 - 1 > 0.$$

Combining all estimates above, there exists a constant $T := T_0 + T_1 + T_2 = \mathcal{O}(-\log \varepsilon)$ such that

$$\begin{aligned}&\mathbb{P}[(X_t, Y_t) \in B_i \times B_i \text{ for all } t \in (h, T], \text{ and } X_t = Y_t \text{ for some } t \in (h, T] \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \\ &\geq \mathbb{P}[E_0 \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \cdot \mathbb{P}[X_{T_0+T_1+t} = Y_{T_0+T_1+t} \text{ for some } t \in [0, T_2] \mid (X_{T_0+T_1}, Y_{T_0+T_1}) \in B^i_{c,1} \times B^i_{c,1}] \\ &\geq 0.9\gamma_0(\gamma_2 + \gamma_3 - 1) := \gamma > 0. \end{aligned} \tag{88}$$

This completes the verification of **(H2)**(i), where $\gamma_1 = 1 - \gamma_0$.

To prove **(H2)**(ii), observe that τ_ε denotes the first time at which either X_t and Y_t couple or one of them exits the basin B_i. Hence, prior to time τ_ε, the processes X_t and Y_t remain in the same basin and have not yet coupled. Consequently, for any $t \geq T$, it follows that

$$\mathbb{P}[\tau_\varepsilon > t \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \leq \prod_{n=1}^{\lfloor t/T \rfloor} \mathbb{P}[\tau_\varepsilon \circ \theta^{(n-1)T} > T \mid (X_{(n-1)T}, Y_{(n-1)T}) \in \overline{B_i \times B_i}].$$

Recalling that $T = \mathcal{O}(-\log \varepsilon)$ and applying (88), one obtains

$$\mathbb{P}[\tau_\varepsilon \circ \theta^{(n-1)T} > T \mid (X_{(n-1)T}, Y_{(n-1)T}) \in \overline{B_i \times B_i}] \leq \gamma_1,$$

uniformly for all $(X_{(n-1)T}, Y_{(n-1)T}) \in \overline{B_i \times B_i}$. Therefore, for any $t > T$,

$$\mathbb{P}[\tau_\varepsilon > t \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \leq \gamma_1^{\lfloor t/T \rfloor} \leq e^{-r_0(\varepsilon)t},$$

where $r_0(\varepsilon) = \mathcal{O}(T^{-1}) = \mathcal{O}(-1/\log \varepsilon)$.

For $t \in [0, T]$, since $\mathbb{P}[\tau_\varepsilon > t \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \leq 1$, there exists a constant $C_0 = \mathcal{O}(1)$ independent of t and ε such that

$$\mathbb{P}[\tau_\varepsilon > t \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \leq C_0 e^{-r_0(\varepsilon)t}.$$

Thus, for all $t > 0$,

$$\mathbb{P}[\tau_\varepsilon > t \mid (X_0, Y_0) \in \overline{B_i \times B_i}] \lesssim e^{-r_0(\varepsilon)t}.$$

Finally, since the number of basins is finite, **(H2)**(ii) follows by applying this estimate over all L basins.

References

[1] Mahdi Soltanolkotabi, Adel Javanmard, Jason D. Lee, Theoretical insights into the optimization landscape of over-parameterized shallow neural networks, IEEE Trans. Inform. Theory 65 (2) (2018) 742–769.
[2] Zeyuan Allen-Zhu, Yuanzhi Li, Yingyu Liang, Learning and generalization in overparameterized neural networks, going beyond two layers, in: NeurIPS, vol. 32, 2019.
[3] Simon Du, Jason Lee, Haochuan Li, Liwei Wang, Xiyu Zhai, Gradient descent finds global minima of deep neural networks, ICML (2019) 1675–1685.
[4] Yaim Cooper, Global minima of overparameterized neural networks, SIAM J. Math. Data Sci. 3 (2) (2021) 676–691.
[5] Yuqing Wang, Minshuo Chen, Tuo Zhao, Molei Tao, Large learning rate tames homogeneity: Convergence and balancing effect, 2022, ICLR.
[6] Chaoyue Liu, Libin Zhu, Mikhail Belkin, Loss landscapes and optimization in over-parameterized non-linear systems and neural networks, Appl. Comput. Harmon. Anal. 59 (2022) 85–116.
[7] Erich Baur, Metastabilität Von Reversiblen Diffusionsprozessen, Diploma Thesis, Bonn University, 2011.
[8] Anton Bovier, Frank Den Hollander, Metastability: A Potential-Theoretic Approach, Springer, 2016.
[9] Torgny Lindvall, L.C.G. Rogers, Coupling of multidimensional diffusions by reflection, Ann. Probab. 14 (3) (1986) 860–872.
[10] Mu-Fa Chen, Shao-Fu Li, Coupling methods for multidimensional diffusion processes, Ann. Probab. (1989) 151–177.
[11] Torgny Lindvall, Lectures on the Coupling Method, Courier Corporation, 2002.
[12] Andreas Eberle, Arnaud Guillin, Raphael Zimmer, Couplings and quantitative contraction rates for Langevin dynamics, Ann. Probab. 47 (4) (2019) 1982–2010.
[13] Martin Hairer, Convergence of markov processes (lecture note), 2010, www.hairer.org/notes/Convergence.pdf.
[14] Yao Li, Shirou Wang, Numerical computations of geometric ergodicity for stochastic dynamics, Nonlinearity 33 (12) (2020) 6935–6970.
[15] Wolfgang Doeblin, Exposé de la theorie des chaînes simples constantes de Markov à un nombre fini d'états, Rev. Math. Union Interbalkanique 2 (1938) 77–105.
[16] Andreas Eberle, Mateusz B. Majka, Quantitative contraction rates for Markov chains on general state spaces, Electron. J. Probab. 24 (26) (2019) 1–36.
[17] Alexander Lipton, Vadim Kaushansky, On the first hitting time density of an Ornstein-Uhlenbeck process, 2018, arXiv preprint arXiv:1810.02390.
[18] Anton Bovier, Michael Eckhoff, Vèronique Gayrard, Markus Klein, Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times, J. Eur. Math. Soc. 6 (4) (2004) 399–424.
[19] Anton Bovier, Vèronique Gayrard, Markus Klein, Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues, J. Eur. Math. Soc. 7 (1) (2005) 69–99.

[20] Boris Leblanc, Renault Olivier, Olivier Scaillet, A correction note on the first passage time of an Ornstein-Uhlenbeck process to a boundary, Finance Stoch. 4 (1) (2000) 109–111.
[21] David Aldous, Random walks on finite groups and rapidly mixing Markov chains, in: Séminaire de Probabilités XVII 1981/82, Springer, 1983, pp. 243–297.
[22] Igal Sason, Entropy bounds for discrete random variables via maximal coupling, IEEE Trans. Inform. Theory 59 (11) (2013) 7118–7131.
[23] Martin Day, On the exponential exit law in the small parameter exit problem, Stochastics 8 (4) (1983) 297–323.
[24] Mark Freidlin, Alexander Wentzell, Random Perturbations of Dynamical Systems, Springer, 1998.
[25] Esa Nummelin, Pekka Tuominen, Geometric ergodicity of harris recurrent Markov chains with applications to renewal theory, Stochastic Process. Appl. 12 (2) (1982) 187–202.
[26] Sergio Angel Almada Monter, Yuri Bakhtin, Normal forms approach to diffusion near hyperbolic equilibria, Nonlinearity 24 (6) (2011) 1883.
[27] Yuri Bakhtin, Noisy heteroclinic networks, Probab. Theory Related Fields 150 (1) (2011) 1–42.
[28] Alan Agresti, Brent A. Coull, Approximate is better than "exact" for interval estimation of binomial proportions, Am. Stat. 52 (2) (1998) 119–126.
[29] Morris Newman, John Todd, The evaluation of matrix inversion programs, J. Soc. Indust. Appl. Math. 6 (1958) 466–476.
[30] Weinan E, Weiqing Ren, Eric Vanden-Eijnden, Simplified and improved string method for computing the minimum energy paths in barrier-crossing events, J. Chem. Phys. 126 (16) (2007) 164103.
[31] Weinan E, Weiqing Ren, Eric Vanden-Eijnden, String method for the study of rare events, Phys. Rev. B 66 (5) (2002) 052301.
[32] Arthur Jacot, Franck Gabriel, Clément Hongler, Neural tangent kernel: Convergence and generalization in neural networks, Adv. Neural Inf. Process. Syst. 31 (2018).
[33] Song Mei, Andrea Montanari, Phan-Minh Nguyen, A mean field view of the landscape of two-layers neural networks, Proc. Natl. Acad. Sci. 115 (33) (2018) E7665–E7671.
[34] Lenaic Chizat, Edouard Oyallon, Francis Bach, On lazy training in differentiable programming, Adv. Neural Inf. Process. Syst. 32 (2019).
[35] Song Mei, Theodor Misiakiewicz, Andrea Montanari, Mean-field theory of two-layers neural networks: dimension-free bounds and kernel limit, in: Conference on Learning Theory, PMLR, 2019, pp. 2388–2464.
[36] Grant M. Rotskoff, Eric Vanden-Eijnden, Trainability and accuracy of neural networks: An interacting particle system approach, 2018, arXiv preprint arXiv:1805.00915.
[37] Justin Sirignano, Konstantinos Spiliopoulos, Mean field analysis of neural networks: A central limit theorem, Stochastic Process. Appl. 130 (3) (2020) 1820–1852.
[38] Felix Draxler, Kambis Veschgini, Manfred Salmhofer, Fred Hamprecht, Essentially no barriers in neural network energy landscape, in: International Conference on Machine Learning, PMLR, 2018, pp. 1309–1318.
[39] Anna Choromanska, Mikael Henaff, Michael Mathieu, Gérard Ben Arous, Yann LeCun, The loss surfaces of multilayer networks, in: Artificial Intelligence and Statistics, PMLR, 2015, pp. 192–204.
[40] Kenji Kawaguchi, Jiaoyang Huang, Leslie Pack Kaelbling, Every local minimum value is the global minimum value of induced model in nonconvex machine learning, Neural Comput. 31 (12) (2019) 2293–2323.
[41] Emmanuel Gobet, Stéphane Menozzi, Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme, Stochastic Process. Appl. 112 (2) (2004) 201–223.
[42] Emmanuel Gobet, Stéphane Menozzi, Stopped diffusion processes: boundary corrections and overshoot, Stochastic Process. Appl. 120 (2) (2010) 130–162.
[43] Mark Kac, Can one hear the shape of a drum? Amer. Math. Monthly 73 (4 Part II) (2020) 1–23.