

# Article A Prehistorical Approach to Optimal Fluctuations for General Langevin Dynamics with Weak Gaussian White Noises

Feng Zhao <sup>1</sup>, Yang Li <sup>2</sup> and Xianbin Liu <sup>1,\*</sup>

- State Key Laboratory of Mechanics and Control of Mechanical Structure, College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
- <sup>2</sup> School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China
- \* Correspondence: xbliu@nuaa.edu.cn

Featured Application: Noise-induced large fluctuations and activation behaviors attract much attention in diverse scientific contexts—for example, involving stochastic resonances and coherent resonances, nucleation in electrochemical systems, passage of ions through open ionic channels in biological membranes, etc. The concept of prehistory probability has been proposed in equilibrium settings to show the focusing effect of the large fluctuational paths on a special one called the optimal path, and recently has been successfully applied to analyze selective ionic conduction through an artificial nanopore in a single graphene sheet. This paper is devoted to showing that such a focusing effect still holds in the stationary, nonstationary, and quasi-stationary settings for both equilibrium and non-equilibrium Langevin systems. The results here shed new light on the underlying mechanisms hidden in the noisy large fluctuation phenomena, and make further studies of other complex stochastic models possible.

Abstract: The dynamics of a stochastic system that exhibits large fluctuations to a given state are almost deterministic due to weak random perturbations. Such large fluctuations occur with overwhelming probability in the vicinity of the so-called optimal path, which is a vital concept in physics, chemistry, and biology, as it uncovers the way in which a physical process escapes from the attractive domain of a metastable state and transitions between different metastable states. In this paper, we examine the prehistorical description of the optimal fluctuation with the unifying framework of Langevin dynamics by means of a quantity called prehistory probability density. We show that the optimal fluctuation has a strong connection with the time reversal of the associated diffusion process. Specifically, in the stationary or quasi-stationary cases, it is found that the prehistory probability density actually acts as the transition probability density of the reversed process. As noise intensity approaches zero, it focuses on the average dynamics of the reversed process due to the law of large numbers, which is then shown to coincide with the time reversal of the optimal path. The local dispersion of the prehistory probability density can thus be reformulated as a Gaussian distribution corresponding to the linearized part of the reversed process. In addition, as an analogue of the original definition in the stationary states, it is proven that the concept of prehistory probability density can also be extended to nonstationary cases where similar properties are still valid. Based on these theoretical results, an algorithm is designed and then successfully applied to a one-dimensional example at the end, showing its accuracy for pinpointing the location of the optimal path and its efficacy in cases where multiple optimal paths coexist.

**Keywords:** large fluctuation; optimal path; prehistory probability density; time reversal of diffusion process

# 1. Introduction

Macroscopic variables characterizing the dynamics of a physical system typically fluctuate continuously due to environmental noises, and such fluctuations tend to become



Citation: Zhao, F.; Li, Y.; Liu, X. A Prehistorical Approach to Optimal Fluctuations for General Langevin Dynamics with Weak Gaussian White Noises. *Appl. Sci.* 2022, *12*, 8569. https://doi.org/10.3390/app12178569

Academic Editors: Ephraim Suhir, Roberto Zivieri and Emin Bayraktar

Received: 1 August 2022 Accepted: 25 August 2022 Published: 26 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). more regular as the intensity of external random perturbations decreases. The minuteness of the noise suggests that the evolution of a physical system is concentrated in its deterministic average dynamics on a finite timescale. The local small-amplitude fluctuations of the stochastic variables around a certain deterministic trajectory can be approximated by a Gaussian distribution whose width is proportional to the strength of the noise. However, when long-term effects of the noise are taken into account, large excursions away from the average dynamics will take place with a probability close to one. The large fluctuations, although infrequent, are responsible for many physical processes, such as chemical reactions [1], stochastic resonances and coherent resonances [2,3], unidirectional motion and energy transduction of molecular motors (also known as Brownian ratchets) [4], enzyme kinetics in living systems [5], etc.

The statistical description of significant fluctuations is a daunting problem. One extensively used technique to describe them relies on the path integral method [6–8], which is essentially a formal representation of the transition probability density in the form of an integral of appropriate functional with respect to a translation-invariant measure in the infinite-dimensional space of all possible stochastic trajectories [9]. An equivalent but mathematically more rigorous formulation is the large deviation theory established by Freidlin and Wentzell [10]. Both theories assert that the probability of an event or the value of an expectation is exponentially dominated by the minimum of the so-called action functional, and that the rare large fluctuations, when occurring, are more likely to be close to the minimizer of the functional. Therefore, two crucial concepts—the optimal paths (OPs) minimizing the functional and the minimum of the action—serve as the central objects of these theories, which reveal the almost deterministic behaviors hidden in complex stochastic systems.

Based on these theories, a great deal of effort has been devoted to the investigation of large fluctuations—especially in non-equilibrium systems. Compared with systems in thermal equilibrium, even a simple non-equilibrium system can exhibit quite different behaviors due to the lack of detailed balance. For example, singularities such as caustics and cusps may arise in the patterns of the extreme paths due to folding structures of the unstable Lagrangian manifold, implying non-differentiability of quasi-potential in some regions, called a switching set [11–13]. A symmetry-breaking bifurcation of the optimal escape path in a double-well system was presented in [14], showing the coexistence of multiple mostprobable escape paths in the cases far away from equilibrium. A phenomenon of cycling of exit from a planar unstable limit cycle was observed in [15], revealing that the distributions of the first escape time and the first escape location both rotate around the unstable orbit periodically with respect to the logarithm of the noise intensity. Moreover, the asymptotic behavior of the exit location's distribution in a general two-dimensional dynamical system has been studied analytically in [16,17]. It was found that the distribution is non-Gaussian and skewed on one side of the saddle, which is a generic feature in non-equilibrium systems. Moreover, the phenomenon of transition between complex invariant sets in twodimensional discrete mapping systems has been explored by many researchers over the last few decades. Graham et al. [18] discussed the multifractal features of the quasi-potential for systems with hyperbolic fractal attractors or repellers. Kraut and Feudel [19,20] noticed the reduction in the activation energy effect due to the existence of a chaotic invariant set of saddle types. Luchinsky et al. [21–23] numerically investigated the transition from a strange attractor to a stable limit cycle and the transition between coexisting chaotic attractors with a fractal boundary. Both results revealed that the most probable escape always occurs by passing through a series of saddle-type invariant sets, such as saddle points and saddle cycles with various periods.

Another new quantitative approach for the investigation of large fluctuations away from a stable fixed point—prehistory probability density (PPD)—was proposed by Dykman et al. [24]. This approach asserts that the OP  $\mathbf{x}_{op}(t)$ , reaching a prescribed state  $\mathbf{x}_f$  at the moment  $t_f$ , is located at the prehistorical peaks of probability density  $p_h(\mathbf{x}, t | \mathbf{x}_f, t_f)$  for  $t < t_f$ , and the occurrence of large fluctuations in the given state focuses on a tube of trajectories

around the optimal one, with its width corresponding to the dispersion of the PPD. Thus, the PPD provides information about the OP and the probability that it can be followed, making it especially suitable for experimental or numerical statistics to calculate the optimal fluctuation, as described in [25–30]. The advantages have been demonstrated numerically in the investigation of the multiple peak features of the distribution [12] and the escape from chaotic attractors [31].

However, the theoretical derivation of the PPD is only displayed in the equilibrium systems. Although numerous experimental and numerical examples also confirm the validity of the PPD in the non-equilibrium cases, its physical meaning is still puzzling. We should note that the PPD can actually be obtained by calculating the laws of the prehistory positions for all possible large fluctuations reaching a given state. This time-reversed statistical approach reminds us of the concept of time reversibility, which plays a crucial role in statistical physics, since whether or not the time-reversed process coincides with the original one determines whether the system is in equilibrium or non-equilibrium. Therefore, it seems quite natural to ask whether there is any connection between the PPD and the time reversal of the original diffusion process, and if so, what the OP and the local Gaussian dispersion represent. Can the results in the stationary cases be extended to nonstationary cases? The answers to these questions form the bulk of this paper.

The remainder of this paper is organized as follows: A brief review of the optimal fluctuations within the framework of general Langevin processes is presented in Section 2. The prehistorical description of the optimal fluctuations in both the stationary and quasistationary cases, and its extension to the nonstationary cases, are formulated in Section 3. A numerical algorithm and its application to a one-dimensional example are exhibited in Section 4. Finally, conclusions are drawn in Section 5.

# 2. Optimal Fluctuations of General Langevin Dynamics

Consider a classical dynamical system subjected to a small multiplicative Gaussian white noise described by Itô's stochastic differential equation:

$$\mathbf{d}\mathbf{x}_t^{\varepsilon} = \mathbf{b}(\mathbf{x}_t^{\varepsilon})\mathbf{d}t + \sqrt{\varepsilon}\boldsymbol{\sigma}(\mathbf{x}_t^{\varepsilon})\mathbf{d}\mathbf{w}(t), \ \mathbf{x}_0^{\varepsilon} = \mathbf{x}_0, \ t \ge 0$$
(1)

where  $\mathbf{b}(\mathbf{x})$  is an *n*-dimensional drift field,  $\mathbf{w}(t)$  is a standard  $\mathbb{R}^n$ -valued Wiener process,  $\varepsilon > 0$  measures the noise intensity,  $\mathbf{\sigma}(\mathbf{x})$  is a smooth matrix-valued function, and  $\mathbf{A}(\mathbf{x}) = \mathbf{\sigma}(\mathbf{x})\mathbf{\sigma}(\mathbf{x})^T$  is the diffusion coefficient. Here, the non-homogeneity of the diffusion coefficient is assumed. It is a common model in disciplines such as physics, chemistry, and biology to simulate molecular dynamics for molecular systems or materials. In physical systems, as the noise is caused by the thermal motion of molecules, the noise intensity  $\varepsilon \propto k_B T$  and, thus, is often small.

In the weak noise limit, solutions to Equation (1) approach its deterministic average (in the absence of noise) in probability on a finite timescale. Occasional large fluctuations take place with small probability, as is rigorously shown by the large deviation theory. According to this theory [10], the probability of observing sample paths close to a given one  $\varphi(s)$  can be estimated as follows:

$$P_{\mathbf{x}_0}\{\sup_{0\le s\le t}|\mathbf{x}_s^{\varepsilon}-\boldsymbol{\varphi}(s)|<\delta\}\sim \exp\left(-\varepsilon^{-1}S_t[\boldsymbol{\varphi}]\right)$$
(2)

The action functional *S* and the Lagrangian *L* are given as follows:

$$S_t[\boldsymbol{\varphi}] = \frac{1}{2} \int_0^t L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) \mathrm{d}s \tag{3}$$

$$L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) = (\dot{\boldsymbol{\varphi}} - \boldsymbol{\varphi}(\boldsymbol{\varphi}))^{T} \mathbf{A}(\boldsymbol{\varphi})^{-1} (\dot{\boldsymbol{\varphi}} - \mathbf{b}(\boldsymbol{\varphi}))$$
(4)

Note that the action is nonzero unless  $\varphi(t)$  is a trajectory of the average system; that is, large fluctuations take place with an exponentially small probability. As the noise intensity approaches zero, it is this exponential convergence rate that highlights such a path with

a smaller value of the action, and tells us which almost unlikely event is more likely to take place and which is more unlikely. As a result, the minimum of the action functional dominates the possibility that a rare event occurs, and the path where the minimum is obtained is the most likely one in this rare event. Thus, the principle of variation could be used to provide the optimal fluctuation as a solution to the Euler–Lagrange equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} \left( \frac{\partial L}{\partial \dot{\boldsymbol{\varphi}}} \right) = \frac{\partial L}{\partial \boldsymbol{\varphi}} \tag{5}$$

with some constraints satisfied. Similarly inspired by classical mechanics, we define a Hamiltonian as the Fenchel–Legendre transform of the Lagrangian, i.e.,

$$H(\mathbf{x},\mathbf{p}) = \mathbf{b}(\mathbf{x})^T \mathbf{p} + \frac{1}{2} \mathbf{p}^T \mathbf{A}(\mathbf{x}) \mathbf{p}$$
(6)

The minimization of Equation (3) is then equivalent to solving a constrained Hamiltonian problem with its governing equation given as follows:

$$\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{b}(\mathbf{x}) + \mathbf{A}(\mathbf{x})\mathbf{p}$$
  
$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{x}} = -(\nabla \mathbf{b}(\mathbf{x}))^T \mathbf{p} - \frac{1}{2}\nabla(\mathbf{p}^T \mathbf{A}(\mathbf{x})\mathbf{p})$$
(7)

A direct result of the large deviation theory is that the transition probability density, if it exits, can be expressed in the following exponential form:

$$p(\mathbf{x}, t | \mathbf{x}_0) \sim \exp\left(-\frac{V(\mathbf{x}_0; t, \mathbf{x})}{\varepsilon}\right)$$
 (8)

which is the same as the formal path integral formulation in the classical work of Onsager and Machlup [32,33]. Here, the rate function  $V(\mathbf{x}_0; t, \mathbf{x})$  is defined as follows:

$$V(\mathbf{x}_0; t, \mathbf{x}) = \min_{\mathbf{x}(0) = \mathbf{x}_0, \mathbf{x}(t) = \mathbf{x}} S[\mathbf{x}(s)]$$
(9)

This indicates that the probability of finding stochastic trajectories stemming from  $\mathbf{x}_0$  and arriving at  $\mathbf{x}$  after time span t can be exponentially estimated by the function  $V(\mathbf{x}_0; t, \mathbf{x})$ . Therefore, the minimizer  $\mathbf{x}_{nop}(\cdot) = \operatorname{argmin}_{\mathbf{x}(0)=\mathbf{x}_0,\mathbf{x}(t)=\mathbf{x}}S[\mathbf{x}(\cdot)]$ , representing the most probable path in this rare event, is defined as the nonstationary optimal path (NOP). By using the principle of variation, the minimization shown in Equation (9) can then be reduced to the following constrained Hamiltonian problem:

$$\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}}, \ \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{x}} \mathbf{x}(0) = \mathbf{x}_0, \ \mathbf{x}(t) = \mathbf{x}$$
(10)

and the value of the action satisfies the Hamilton-Jacobi equation:

$$\frac{\partial V(\mathbf{x}_0; t, \mathbf{x})}{\partial t} + H(\mathbf{x}, \nabla_{\mathbf{x}} V(\mathbf{x}_0; t, \mathbf{x})) = 0$$
(11)

If it can be solved beforehand, the NOP can be gained by integrating along the nonstationary optimal fluctuation field, as follows:

$$\dot{\mathbf{x}} = \mathbf{b}(\mathbf{x}) + \mathbf{A}(\mathbf{x})\nabla_{\mathbf{x}}V(\mathbf{x}_0; t, \mathbf{x})$$
(12)

Moreover, the large deviation theory also permits us to analyze the effect of the random perturbation on infinite timescales, in which the central concept in the large deviation theory is the quasi-potential, defined as follows:

$$V(\mathbf{x}_0; \mathbf{x}) = \inf_{t>0} \min_{\mathbf{x}_0, \mathbf{x}_0(t)=\mathbf{x}} S[\mathbf{x}(\cdot)]$$
(13)

This provides the principal terms of asymptotic expressions about the long-term behaviors of a stochastic system. For example, consider a bounded domain *D* with a smooth non-characteristic boundary  $\partial D$  such that its closure  $D \cup \partial D$  is totally contained in the attractive

domain of a stable fixed point  $\mathbf{x}_{eq}$ . For small  $\varepsilon$ , the stochastic trajectories typically fluctuate around the fixed point, and occasionally hit  $\partial D$ , which is a rare event, as our observation is always performed within a finite timescale. The first exit time  $\tau_{\partial D}^{\varepsilon} = \inf\{t > 0, \mathbf{x}_{t}^{\varepsilon} \in \partial D\}$ , introduced to quantify the timescale of such a rare event, can be estimated as follows:

$$\mathbf{E}_{\mathbf{x}_{0}}\tau_{\partial D}^{\varepsilon} \sim \exp\left(\varepsilon^{-1}V(\mathbf{x}_{eq};\partial D)\right), \ \mathbf{x}_{0} \in D$$
(14)

$$\forall \eta > 0, \lim_{\varepsilon \to 0} \mathsf{P}_{\mathbf{x}_0} \Big\{ e^{\varepsilon^{-1}(V(\mathbf{x}_{eq};\partial D) - \eta)} < \tau^{\varepsilon}_{\partial D} < e^{\varepsilon^{-1}(V(\mathbf{x}_{eq};\partial D) + \eta)} \Big\} \to 1$$
(15)

where  $V(\mathbf{x}_{eq}; \partial D) = \min_{\mathbf{y} \in \partial D} V(\mathbf{x}_{eq}; \mathbf{y})$ . It can be seen that the occurrence of escaping from the domain *D* is concentrated on the timescale of exponential order  $\exp(V(\mathbf{x}_{eq}; \partial D)/\varepsilon)$ . Within this timescale, almost all stochastic trajectories remain in the domain *D*. The system acquires its quasi-stationary state when the time span is close to this timescale. As time goes by further, typical events such as escapes and transitions occur and, thus, stochastic trajectories fluctuate in a much larger region. When the time span reaches infinity, the balance of probabilistic outflows and inflows in any local area is achieved. The system is then in its stationary state if it exists, with its stationary density having the Wentzel– Kramers–Brillouin (WKB)-type approximation

$$\rho_s(\mathbf{x}) \sim \exp\left(-\varepsilon^{-1} V(\mathbf{x}_{eq}; \mathbf{x})\right), \ \mathbf{x} \in D$$
(16)

Thus, the core for estimating the probability of a certain rare event or an expectation about the long-term behavior of the system is to solve the constrained optimization problem shown in Equation (13), where  $\mathbf{x}_0$  is usually a fixed point  $\mathbf{x}_{eq}$ , and  $\mathbf{x}$  is located in its attractive domain. In this case, the quasi-potential is the stationary solution of the Hamilton–Jacobi equation (Equation (11)), i.e.,

$$H(\mathbf{x}, \nabla_{\mathbf{x}} V(\mathbf{x}_{eq}; \mathbf{x})) = 0$$
(17)

The optimal path  $\mathbf{x}_{op}$  (OP), defined as the most probable path connecting  $\mathbf{x}_{eq}$  and  $\mathbf{x}$ , is the projection of the trajectory lying on the unstable manifold of the fixed point ( $\mathbf{x}_{eq}$ , 0) for the Hamiltonian system (Equation (7)). It corresponds to the solution of the constrained Hamiltonian problem:

$$\mathbf{x} = \frac{\partial \mathbf{p}}{\partial \mathbf{p}}, \ \mathbf{p} = -\frac{\partial \mathbf{x}}{\partial \mathbf{x}}$$
$$\lim_{t \to -\infty} \mathbf{x}(t) = \mathbf{x}_{eq}, \ \mathbf{x}(0) = \mathbf{x}$$
(18)

If the quasi-potential is given in advance, the optimal fluctuation field for the OP will be as follows:

$$\dot{\mathbf{x}} = \mathbf{b}(\mathbf{x}) + \mathbf{A}(\mathbf{x})\nabla_{\mathbf{x}}V(\mathbf{x}_{eq};\mathbf{x})$$
(19)

#### 3. Prehistorical Description of the Optimal Fluctuations

In this section, we explore both the OP and NOP by means of another quantity, called prehistory probability density (PPD). We derive the evolution equation of the PPD and then relate it to the time reversal of the original diffusion process. We can see that, as  $\varepsilon \ll 1$ , the PPD has good properties of being capable of pinpointing the locations of both the OP and the NOP, and providing the statistical features of the nearby paths.

#### 3.1. In the Stationary Cases

# 3.1.1. Evolutionary Law of the PPD

Motivated by the original definition of the PPD exhibited in equilibrium systems [24], we here define the PPD for our general Langevin dynamics (Equation (1)) as follows:

 $\langle \rangle$ 

$$p_h(\mathbf{x}, t | \mathbf{x}_f) = \frac{\rho_s(\mathbf{x})}{\rho_s(\mathbf{x}_f)} p(\mathbf{x}_f | \mathbf{x}, t), \ t < 0, \ \mathbf{x}, \mathbf{x}_f \in D$$
(20)

where  $p(\mathbf{x}_f | \mathbf{x}, t) = p(\mathbf{x}_f, 0 | \mathbf{x}, t) = p(\mathbf{x}_f, -t | \mathbf{x}, 0)$  represents the transition probability from  $\mathbf{x}$  to  $\mathbf{x}_f$  over the time span -t. It can be seen that the PPD actually describes the prehistorical

law of the stationary process conditionally on the prescribed state  $\mathbf{x}_f$ . Our purpose here is to show that the PPD will focus on the OP as the noise intensity approaches zero.

Recall that the stationary density  $\rho_s$  and the function  $p(\mathbf{x}_f | \mathbf{x}, t)$  obey the stationary forward equation and the Kolmogorov backward equation, respectively:

$$\mathcal{L}_{\mathbf{x}}^{*}\rho_{s}^{\varepsilon}(\mathbf{x}) = \frac{\varepsilon}{2} \sum_{i,j} \frac{\partial^{2}}{\partial x_{i} \partial x_{j}} \left( A_{ij}(\mathbf{x})\rho_{s}^{\varepsilon}(\mathbf{x}) \right) - \sum_{i} \frac{\partial}{\partial x_{i}} (b_{i}^{\varepsilon}(\mathbf{x})\rho_{s}^{\varepsilon}(\mathbf{x})) = 0, \ \int_{\mathbb{R}^{n}} \rho_{s}^{\varepsilon}(\mathbf{x}) d\mathbf{x} = 1$$
(21)

$$-\frac{\partial p(\mathbf{x}_f|\mathbf{x},t)}{\partial t} = \mathcal{L}_{\mathbf{x}} p(\mathbf{x}_f|\mathbf{x},t) = \frac{\varepsilon}{2} \sum_{i,j} A_{ij}(\mathbf{x}) \frac{\partial^2 p(\mathbf{x}_f|\mathbf{x},t)}{\partial x_i \partial x_j} + \sum_i b_i^{\varepsilon}(\mathbf{x}) \frac{\partial p(\mathbf{x}_f|\mathbf{x},t)}{\partial x_i}$$
(22)

where  $\mathcal{L}_x^*$  represents the Fokker–Planck operator, with  $\mathcal{L}_x$  being its adjoint. These equations can then be applied to derive the evolution of the PPD over time. We have two results; the first is as follows:

$$-\frac{\partial p_h(\mathbf{x},t|\mathbf{x}_f)}{\partial t} = \overline{\mathcal{L}}_{\mathbf{x}}^* p_h(\mathbf{x},t|\mathbf{x}_f) = \frac{\varepsilon}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \left( A_{ij}(\mathbf{x}) p_h(\mathbf{x},t|\mathbf{x}_f) \right) - \sum_i \frac{\partial}{\partial x_i} \left( \overline{b}_i^\varepsilon(\mathbf{x}) p_h(\mathbf{x},t|\mathbf{x}_f) \right), \ t < 0$$
(23)

$$\overline{b}_{i}^{\varepsilon}(\mathbf{x}) = -b_{i}(\mathbf{x}) + \sum_{j} \frac{\varepsilon}{\rho_{s}(\mathbf{x})} \frac{\partial}{\partial x_{j}} \left( A_{ij}(\mathbf{x})\rho_{s}(\mathbf{x}) \right)$$
(24)

After a time-reversed transformation  $t \rightarrow -t$ , we have:

$$\frac{\partial p_h(\mathbf{x}, -t|\mathbf{x}_f)}{\partial t} = \overline{\mathcal{L}}_{\mathbf{x}}^* p_h(\mathbf{x}, -t|\mathbf{x}_f), \ t > 0$$
(25)

which, thus, relates  $p_h(\mathbf{x}, -t | \mathbf{x}_f)$  to a new diffusion process:

$$d\mathbf{y}_{t}^{\varepsilon} = \mathbf{b} \left(\mathbf{y}_{t}^{\varepsilon}\right) dt + \sqrt{\varepsilon} \boldsymbol{\sigma}(\mathbf{y}_{t}^{\varepsilon}) d\mathbf{w}(t) = \left(-b_{i}(\mathbf{y}_{t}^{\varepsilon}) + \sum_{j} \frac{\varepsilon}{\rho_{s}(\mathbf{y}_{t}^{\varepsilon})} \frac{\partial}{\partial y_{j}} \left(A_{ij}(\mathbf{y}_{t}^{\varepsilon})\rho_{s}(\mathbf{y}_{t}^{\varepsilon})\right)\right) dt + \sqrt{\varepsilon} \boldsymbol{\sigma}(\mathbf{y}_{t}^{\varepsilon}) d\mathbf{w}(t), \ t > 0$$

$$(26)$$

Note that for  $p_h(\mathbf{x}, 0 | \mathbf{x}_f) = \delta(\mathbf{x} - \mathbf{x}_f)$ , the PPD actually acts as the transition probability density of this new process with the initial condition  $\mathbf{y}^{\varepsilon}(0) = \mathbf{x}_f$ , i.e.,

$$p_h(\mathbf{y}, -t \left| \mathbf{x}_f \right) = \overline{p}(\mathbf{y}, t \left| \mathbf{x}_f \right), \ t > 0$$
<sup>(27)</sup>

Whereas, the second is:

$$\frac{\partial p_h(\mathbf{x},t \left| \mathbf{x}_f \right)}{\partial t} = \frac{\varepsilon}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \left( A_{ij}(\mathbf{x}) p_h(\mathbf{x},t \left| \mathbf{x}_f \right) \right) - \sum_i \frac{\partial}{\partial x_i} \left( \widetilde{b}_i^{\varepsilon}(\mathbf{x},t) p_h(\mathbf{x},t \left| \mathbf{x}_f \right) \right), \ t < 0$$
(28)

$$\widetilde{b}_{i}^{\varepsilon}(\mathbf{x},t) = b_{i}(\mathbf{x}) + \sum_{j} \varepsilon A_{ij}(\mathbf{x}) \frac{\partial}{\partial x_{j}} \ln p(\mathbf{x}_{f}|\mathbf{x},t)$$
(29)

This relates  $p_h(\mathbf{x}, t | \mathbf{x}_f)$  to another diffusion process:

$$d\mathbf{z}_{t}^{\varepsilon} = \tilde{\mathbf{b}}^{\varepsilon}(\mathbf{z}_{t}^{\varepsilon}, t)dt + \sqrt{\varepsilon}\sigma(\mathbf{z}_{t}^{\varepsilon})d\mathbf{w}(t) = \left(b_{i}(\mathbf{z}_{t}^{\varepsilon}) + \sum_{j}\varepsilon A_{ij}(\mathbf{z}_{t}^{\varepsilon})\frac{\partial}{\partial z_{j}}\ln p(\mathbf{x}_{f}|\mathbf{z}_{t}^{\varepsilon}, t)\right)dt + \sqrt{\varepsilon}\sigma(\mathbf{z}_{t}^{\varepsilon})d\mathbf{w}(t), \ t < 0$$
(30)

In fact, Equations (23) and (28) describe the evolutionary law of the PPD in the negative and positive directions of time, respectively. The first equation has the advantage that only the information of  $\rho_s$  is needed, while for the second, a collection of density  $\left\{ p(\mathbf{x}_f | \mathbf{x}, t), t < 0 \right\}$  is required. We discuss the interpretation of these new processes in

Section 3.4. In the following section, we use the former to see what happens when the noise intensity is close to zero.

## 3.1.2. For the Weak Noise Limit

According to the large deviation theory, the stationary density  $\rho_s(\mathbf{x})$  is exponentially dominated by the quasi-potential. We have the following WKB-type approximation:

$$\rho_{s}^{\varepsilon}(\mathbf{x}) = \frac{C_{s}^{\varepsilon}(\mathbf{x})}{\varepsilon^{n/2}} \exp\left(-\frac{V(\mathbf{x}_{eq}, \mathbf{x})}{\varepsilon}\right), \ \mathbf{x} \in D$$
(31)

where  $C_s^{\varepsilon}(\mathbf{x}) = C_s(\mathbf{x}) + o(\varepsilon)$ . Substituting it into Equation (24), we get:

$$\mathbf{b}^{-\varepsilon}(\mathbf{x}) = -\mathbf{b}(\mathbf{x}) - \mathbf{A}(\mathbf{x})\nabla V(\mathbf{x}_{eq}, \mathbf{x}) + \varepsilon \left(\frac{\mathbf{A}(\mathbf{x})\nabla C_s(\mathbf{x})}{C_s(\mathbf{x})} + \nabla \cdot \mathbf{A}(\mathbf{x})\right) + o(\varepsilon)$$
(32)

For weak noise limit, the  $\varepsilon$ -independent part represents the average dynamics of  $y_t^{\varepsilon}$ :

$$\dot{\mathbf{y}}_{t}^{0} = -\mathbf{b}(\mathbf{y}_{t}^{0}) - \mathbf{A}(\mathbf{y}_{t}^{0}) \nabla_{\mathbf{y}} V(\mathbf{x}_{eq}, \mathbf{y}_{t}^{0})$$
(33)

Compared with Equation (19), we can see that this is exactly the time reversal of the optimal fluctuation field. According to the law of large numbers,

$$\forall T > 0, \forall \delta > 0, \lim_{\varepsilon \to 0} P_{\mathbf{x}_f} \left( \max_{0 \le t \le T} \left| \mathbf{y}_t^{\varepsilon} - \mathbf{y}_t^0 \right| > \delta \right) = 0$$
(34)

As a result,  $\overline{p}(\mathbf{y}, t | \mathbf{x}_f)$  focuses on  $\mathbf{y}_t^0$  if  $\mathbf{y}^0(0) = \mathbf{x}_f$ ; that is, the OP  $\mathbf{x}_{op}(t)$  follows the peak trajectory of the PPD  $p_h(\mathbf{x}, t | \mathbf{x}_f)$ .

If the quasi-potential is smooth, for each point  $\mathbf{x}_f$ , there exists a unique OP connecting it with the fixed point  $\mathbf{x}_{eq}$ . This satisfies Equation (19); thus, the PPD has a single peak trajectory. This is a trivial situation.

The main complexity comes from the case where the quasi-potential is non-differentiable. This is a frequent occurrence for systems far from equilibrium. Since the quasi-potential is a solution to the optimization problem (Equation (13)), the principle of variation is applied to transform it into the constrained Hamiltonian problem (Equation (18)). We can divide the domain *D* into three parts ( $D = D_1 \cup D_2 \cup D_3$ ) as follows:

- 1. We say that  $\mathbf{x}_f \in D_1$  if there exists a unique solution to the constrained Hamiltonian problem with  $\mathbf{x}(0) = \mathbf{x}_f$ . This is the OP; thus, the PPD has a single peak trajectory.
- 2. We say that  $\mathbf{x}_f \in D_2$  if there exist multiple solutions to the constrained Hamiltonian problem with  $\mathbf{x}(0) = \mathbf{x}_f$  and, among them, there exists a unique one that has the minimum action. This is the OP. In this case, the PPD will have multiple peak trajectories. Each of them corresponds to one solution to the constrained Hamiltonian problem. As  $\varepsilon \to 0$ , the exponential convergence rate (Equation (2)) highlights the OP, as it has the minimum action. As a result, the PPD possesses a unique highest peak trajectory located at the OP.
- 3. We say that  $\mathbf{x}_f \in D_3$  if there exist multiple solutions to the constrained Hamiltonian problem with  $\mathbf{x}(0) = \mathbf{x}_f$  and there exists more than one path with the same minimum action. This is the case where multiple OPs coexist, and each of them satisfies Equation (19). Let  $\mathbf{x}_{op1}$  and  $\mathbf{x}_{op2}$  denote two of the OPs.  $(\mathbf{x}_{op1}(t), \nabla V(\mathbf{x}_{eq}, \mathbf{x}_{op1}(t)))$  and  $(\mathbf{x}_{op2}(t), \nabla V(\mathbf{x}_{eq}, \mathbf{x}_{op2}(t)))$  are then the corresponding Hamiltonian trajectories. Since Equation (7) is a smooth vector field, these trajectories are also smooth. This suggests that non-differentiability does not occur along these paths except at the endpoint. At  $\mathbf{x}_f$ , the trajectories must have distinct momentum  $\nabla V(\mathbf{x}_{eq}, \mathbf{x}_{op1}(0)) \neq \nabla V(\mathbf{x}_{eq}, \mathbf{x}_{op2}(0))$ . The level set of the quasi-potential intersects itself transversely there and, consequently, the non-differentiability emerges. In fact,  $D_3$  consists of all of the non-differentiable points of the quasi-potential. At these points, the PPD possesses multiple peak trajectories with the same height, and if we vary the target point  $\mathbf{x}_f$  and let it pass through  $D_3$  transversely, the height of these peak trajectories changes, resulting in the highest

peak trajectory—that is, the OP—switching from one of them to another suddenly. This is why  $D_3$  is often called the switching region.

Furthermore, along the OP, the local deviation  $\zeta = \mathbf{x} - \mathbf{x}_{op}$  is approximatively Gaussian, i.e.,

$$p_h(\mathbf{x}, t | \mathbf{x}_f) = C \exp\left(-\frac{1}{2\varepsilon} \left(\mathbf{x} - \mathbf{x}_{op}(t)\right)^T \mathbf{\kappa}(t)^{-1} \left(\mathbf{x} - \mathbf{x}_{op}(t)\right)\right), \ \mathbf{x} \to \mathbf{x}_{op}(t)$$
(35)

The local Gaussian approximation can be reformulated in terms of the new diffusion process, as follows:

$$- d\boldsymbol{\zeta}_t = \nabla_x \bar{\mathbf{b}}^0(\mathbf{x}_{op}(t))\boldsymbol{\zeta}_t dt + \sqrt{\varepsilon}\boldsymbol{\sigma}(\mathbf{x}_{op}(t))d\mathbf{w}(t), \ t < 0, \ \boldsymbol{\zeta}(0) = 0$$
(36)

The noise-independent dispersion parameter, defined as  $\kappa(t) = (E\zeta_t \zeta_t^T - E\zeta_t E\zeta_t^T)/\varepsilon$ , satisfies the time-varying Lyapunov differential equation:

$$\dot{\boldsymbol{\kappa}} = -\nabla_{\boldsymbol{x}} \bar{\boldsymbol{b}}^{0} (\boldsymbol{x}_{op}(t)) \boldsymbol{\kappa} - \boldsymbol{\kappa} \left( \nabla_{\boldsymbol{x}} \bar{\boldsymbol{b}}^{0} (\boldsymbol{x}_{op}(t)) \right)^{T} + \frac{1}{2} \mathbf{A} (\boldsymbol{x}_{op}(t)), \ t < 0, \ \boldsymbol{\kappa}(0) = 0$$
(37)

In summary, the results shown here provide us with a complete picture of the optimal fluctuation in the stationary states. The method here is based on the evolutionary law of the PPD instead of the complex path integral technique. Consequently, it extends the original derivation of the PPD in equilibrium systems described in [24] (where  $\mathbf{b}(\mathbf{x}) = -\nabla U(\mathbf{x})$  and  $\mathbf{A}(\mathbf{x}) = \mathbf{I}$ ) to more general Langevin dynamics (Equation (1)), without requiring detailed balance. It is also notable that the empirical VG relaxation method proposed in [34] for calculating the dispersion of the PPD in non-equilibrium systems is essentially the same as Equation (37). In addition, this also testifies theoretically that the PPD is a powerful tool for us to illustrate the coexistence of multiple OPs, as shown by the numerical experiments in [11]. Next, we extend the PPD to nonstationary states, to show that similar properties still hold there.

# 3.2. In the Nonstationary Cases

#### 3.2.1. Evolutionary Law of the Nonstationary PPD

Here, we consider the transition from a point  $\mathbf{x}_0$  to  $\mathbf{x}_f$  over the time span *T*. Similar to the stationary cases, the nonstationary PPD is defined as follows:

$$p_h(\mathbf{x}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0) = \frac{p(\mathbf{x}, t | \mathbf{x}_0, 0) p(\mathbf{x}_f, T | \mathbf{x}, t)}{p(\mathbf{x}_f, T | \mathbf{x}_0, 0)}, \ 0 \le t \le T$$
(38)

This describes the prehistorical law of the stochastic process conditionally on the initial state  $\mathbf{x}^{\varepsilon}(0) = \mathbf{x}_0$  and final state  $\mathbf{x}^{\varepsilon}(T) = \mathbf{x}_f$ . The functions  $p(\mathbf{x}, t | \mathbf{x}_0, 0)$  and  $p(\mathbf{x}_f, T | \mathbf{x}, t)$  obey the Kolmogorov forward equation (Fokker–Planck equation) and the Kolmogorov backward equation, respectively:

$$\frac{\partial p(\mathbf{x}, t | \mathbf{x}_0, 0)}{\partial t} = \mathcal{L}_{\mathbf{x}}^* p(\mathbf{x}, t | \mathbf{x}_0, 0)$$
(39)

$$-\frac{\partial p(\mathbf{x}_f, T | \mathbf{x}, t)}{\partial t} = \mathcal{L}_{\mathbf{x}} p(\mathbf{x}_f, T | \mathbf{x}, t)$$
(40)

It is therefore straightforward to obtain

$$\frac{\partial p_h(\mathbf{x},t|\mathbf{x}_f,T;\mathbf{x}_0,0)}{\partial t} = \frac{p(\mathbf{x}_f,T|\mathbf{x},t)}{p(\mathbf{x}_f,T|\mathbf{x}_0,0)} \frac{\partial p(\mathbf{x},t|\mathbf{x}_0,0)}{\partial t} + \frac{p(\mathbf{x},t|\mathbf{x}_0,0)}{p(\mathbf{x}_f,T|\mathbf{x}_0,0)} \frac{\partial p(\mathbf{x}_f,T|\mathbf{x},t)}{\partial t}$$
(41)

Substituting Equations (39) and (40) into Equation (41) and regrouping the terms yields two distinct expressions of the evolutionary law. The first is as follows:

$$-\frac{\partial p_h}{\partial t} = \frac{\varepsilon}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \left( A_{ij}(\mathbf{x}) p_h \right) - \sum_i \frac{\partial}{\partial x_i} \left( \overline{b}_i^{\varepsilon}(\mathbf{x}, T-t) p_h \right), \ 0 \le t \le T$$
(42)

$$\overline{b}_{i}^{\varepsilon}(\mathbf{x}, T-t) = -b_{i}(\mathbf{x}) + \sum_{j} \frac{\varepsilon}{p(\mathbf{x}, t | \mathbf{x}_{0})} \frac{\partial}{\partial x_{j}} (A_{ij}(\mathbf{x}) p(\mathbf{x}, t | \mathbf{x}_{0}))$$
(43)

Whereas the second is as follows:

$$\frac{\partial p_h}{\partial t} = \frac{\varepsilon}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \left( A_{ij}(\mathbf{x}) p_h \right) - \sum_i \frac{\partial}{\partial x_i} \left( \widetilde{b}_i^{\varepsilon}(\mathbf{x}, t) p_h \right), \ 0 \le t \le T$$
(44)

$$\widetilde{b}_{i}^{\varepsilon}(\mathbf{x},t) = b_{i}(\mathbf{x}) + \sum_{j} \varepsilon A_{ij}(\mathbf{x}) \frac{\partial}{\partial x_{j}} \ln p(\mathbf{x}_{f},T|\mathbf{x},t)$$
(45)

The former is related to a new diffusion process  $\mathbf{y}_t^{\varepsilon}$ :

$$\begin{aligned} \mathbf{d}\mathbf{y}_{t}^{\varepsilon} &= \bar{\mathbf{b}}^{\varepsilon}(\mathbf{y}_{t}^{\varepsilon}, t)\mathbf{d}t + \sqrt{\varepsilon}\sigma(\mathbf{y}_{t}^{\varepsilon})\mathbf{d}\mathbf{w}(t) \\ &= \left(-b_{i}(\mathbf{y}_{t}^{\varepsilon}) + \sum_{j} \frac{\varepsilon}{p(\mathbf{y}_{t}^{\varepsilon}, T-t|\mathbf{x}_{0})} \frac{\partial}{\partial y_{j}} \left(A_{ij}(\mathbf{y}_{t}^{\varepsilon})p(\mathbf{y}_{t}^{\varepsilon}, T-t|\mathbf{x}_{0})\right)\right) \mathbf{d}t + \sqrt{\varepsilon}\sigma(\mathbf{y}_{t}^{\varepsilon})\mathbf{d}\mathbf{w}(t) \end{aligned}$$

$$(46)$$

in the sense of:

$$p_h(\mathbf{y}, T - t | \mathbf{x}_f, T; \mathbf{x}_0, 0) = \overline{p}(\mathbf{y}, t | \mathbf{x}_f, 0), \ 0 \le t \le T$$

$$\tag{47}$$

Meanwhile the second equation (Equation (44)) can be related to another diffusion process  $\mathbf{z}_t^{\varepsilon}$ :

$$d\mathbf{z}_{t}^{\varepsilon} = \widetilde{\mathbf{b}}^{\varepsilon}(\mathbf{z}_{t}^{\varepsilon}, t)dt + \sqrt{\varepsilon}\sigma(\mathbf{z}_{t}^{\varepsilon})d\mathbf{w}(t) = \left(b_{i}(\mathbf{z}_{t}^{\varepsilon}) + \sum_{j}\varepsilon A_{ij}(\mathbf{z}_{t}^{\varepsilon})\frac{\partial}{\partial z_{j}}\ln p(\mathbf{x}_{f}, T|\mathbf{z}_{t}^{\varepsilon}, t)\right)dt + \sqrt{\varepsilon}\sigma(\mathbf{z}_{t}^{\varepsilon})d\mathbf{w}(t), \ 0 \le t \le T$$

$$(48)$$

in the sense of:

$$p_h(\mathbf{z}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0) = \widetilde{p}(\mathbf{z}, t | \mathbf{x}_0, 0), \ 0 \le t \le T$$

$$\tag{49}$$

The process shown in Equation (48) is called the generalized Brownian bridge, which has been derived in [35–38] to display transition paths starting at  $\mathbf{x}_0$  at time 0 and conditioned to end at a given point  $\mathbf{x}_f$  at time *T*. It is worth noting that although Equations (42) and (44) have the same meaning in depicting the evolutionary process of the PPD, different prior information is needed in the actual calculation. In Equation (42),  $\{p(\mathbf{x}, t | \mathbf{x}_0), 0 \le t \le T\}$  is needed, while in Equation (44),  $\{p(\mathbf{x}_f, T | \mathbf{x}, t), 0 \le t \le T\}$  is required. Here, we use the former to show that the PPD will focus on the NOP as the noise intensity approaches zero.

# 3.2.2. For the Weak Noise Limit

Note that the conditional probability  $p(x, t | x_0)$  has the WKB approximation of the following form:

$$p(\mathbf{x},t|\mathbf{x}_0) = \frac{C(\mathbf{x},t,\mathbf{x}_0,\varepsilon)}{\varepsilon^{n/2}} \exp\left(-\frac{V(\mathbf{x}_0;t,\mathbf{x})}{\varepsilon}\right)$$
(50)

Substituting it into Equation (43), and expanding for  $\varepsilon \ll 1$ , we can determine the average dynamics of  $y_t^{\varepsilon}$ :

$$\dot{\mathbf{y}}_{t}^{0} = \bar{\mathbf{b}}^{0}(\mathbf{y}_{t}^{0}, t) = -\mathbf{b}(\mathbf{y}_{t}^{0}) - \mathbf{A}(\mathbf{y}_{t}^{0})\nabla_{\mathbf{y}}V(\mathbf{x}_{0}; T - t, \mathbf{y}_{t}^{0})$$
(51)

After a time-reversed transformation  $t \to T - t$ , this will have the same form as Equation (12). In this case,  $\overline{p}(\mathbf{y}, t | \mathbf{x}_f, 0)$  will focus on  $\mathbf{y}_t^0$  if  $\mathbf{y}^0(0) = \mathbf{x}_f$ . Consequently, the NOP will be achieved at the peak trajectory of the nonstationary PPD.

If  $V(\mathbf{x}_0; t, \mathbf{x})$  is a smooth function, there exists a unique path connecting  $\mathbf{x}_0$  with the prescribed point  $\mathbf{x}_f$ ; thus, the PPD has a single peak trajectory. However, what is similar to stationary cases is the troublesome non-differentiability. In nonstationary cases, this takes place more frequently—even in equilibrium systems. Here, we can also divide the space  $\mathbb{R}^n \times \mathbb{R}$  into three parts ( $F_1 \cup F_2 \cup F_3$ ) according to the number of solutions to the constrained Hamiltonian problem (Equation (10)), as well as their value of the action (Equation (3)).

For points  $(\mathbf{x}_f, T)$  located in these sets, the nonstationary PPD possesses a unique peak trajectory, a unique highest peak trajectory, and multiple highest peak trajectories with the same height. This is similar to the stationary situation, which will not be repeated here.

In addition, it is also worth pointing out that the local deviation  $\zeta = \mathbf{x} - \mathbf{x}_{nop}$  along the NOP is still Gaussian, i.e.,

$$p_h(\mathbf{x}_{nop} + \boldsymbol{\zeta}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0) = C \exp\left(-\frac{1}{2\varepsilon} \boldsymbol{\zeta}^T \boldsymbol{\kappa}(t)^{-1} \boldsymbol{\zeta}\right), \ 0 < t < T$$
(52)

Here, the noise-independent dispersion parameter  $\kappa(t)$  is the finite-time solution of the time-varying Lyapunov equation:

$$\dot{\boldsymbol{\kappa}} = -\nabla_{\mathbf{x}} \bar{\mathbf{b}}^{0} (\mathbf{x}_{nop}(t), T-t) \boldsymbol{\kappa} - \boldsymbol{\kappa} \Big( \nabla_{\mathbf{x}} \bar{\mathbf{b}} \big( \mathbf{x}_{nop}(t), T-t \big) \Big)^{T} + \frac{1}{2} \mathbf{A} \big( \mathbf{x}_{nop}(t) \big), \ 0 < t < T$$
(53)  
$$\boldsymbol{\kappa}(0) = 0, \ \boldsymbol{\kappa}(T) = 0$$

Note that

$$\lim_{t \to 0} p_h(\mathbf{x}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0) = \delta(\mathbf{x} - \mathbf{x}_0)$$
(54)

T

$$\lim_{t \to T} p_h(\mathbf{x}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0) = \delta(\mathbf{x} - \mathbf{x}_f)$$
(55)

If one of the initial conditions in Equation (53) is valid, the other will hold automatically. The results here show that the nonstationary PPD has a good property of being able to pinpoint the location of the NOP, and can also be applied to identify points where there are multiple coexisting NOPs. In particular, this negates the assertions made in [39], saying that the PPD is non-Coursian and highly course triad. To illustrate the characteristical

that the PPD is non-Gaussian and highly asymmetrical. To illustrate the above theoretical analysis more intuitively, a numerical algorithm and its application to a simple example, in this case, are presented in Section 4.

#### 3.3. In the Quasi-Stationary Cases

Here, we restrict our investigations to the domain *D* defined previously. Consider the diffusion that is absorbed in the boundary  $\partial D$ . The law of it, represented by

$$p_{abs}(\mathbf{x}, t | \mathbf{x}_0) = P_{\mathbf{x}_0}(\mathbf{x}_t^{\varepsilon} \simeq \mathbf{x}, t < \tau_{\partial D}^{\varepsilon} | \mathbf{x}_0), \ \mathbf{x} \in D$$
(56)

satisfies the Kolmogorov forward equation:

$$\frac{\partial p_{abs}}{\partial t} = \mathcal{L}_{\mathbf{x}}^* p_{abs}, \ t > 0, \ \mathbf{x} \in D 
p_{abs}(\mathbf{x}, t | \mathbf{x}_0) = 0, \ \mathbf{x} \in \partial D$$
(57)

The probability in *D* continues to be lost as time evolves, in the sense of

$$\int_{D} p_{abs}(\mathbf{x}, t | \mathbf{x}_0) \, \mathrm{d}x = P_{\mathbf{x}_0}(t < \tau_{\partial D}^{\varepsilon}) \tag{58}$$

which corresponds to the probability that the stochastic trajectories escape from D. The quasi-stationary distribution is then defined as the invariant probability measure with respect to the law of temporal evolution, conditionally on not having left the domain D up to the moment t, namely [40]:

$$\rho_{qs}(\mathbf{x}) = \frac{\int_D P_{\mathbf{x}_0}(\mathbf{x}_t^{\varepsilon} \simeq \mathbf{x}_t < \tau_{\partial D}^{\varepsilon})\rho_{qs}(\mathbf{x}_0) d\mathbf{x}_0}{\int_D P_{\mathbf{x}_0}(t < \tau_{\partial D}^{\varepsilon})\rho_{qs}(\mathbf{x}_0) d\mathbf{x}_0} 
= \frac{\int_D p_{abs}(\mathbf{x}_t | \mathbf{x}_0)\rho_{qs}(\mathbf{x}_0) d\mathbf{x}_0}{\int_D P_{\mathbf{x}_0}(t < \tau_{\partial D}^{\varepsilon})\rho_{qs}(\mathbf{x}_0) d\mathbf{x}_0}$$
(59)

This is the eigenfunction associated with the principal eigenvalue  $-\lambda_1$  for the Fokker– Planck operator  $\mathcal{L}^*$  with a homogeneous boundary condition [41,42]. For any initial distribution  $p_0$ , the quasi-stationary distribution can be approached by the so-called Yaglom limit of the conditional probability [43]:

$$\rho_{qs}(\mathbf{x}) = \lim_{t \to \infty} \frac{\int_D P_{\mathbf{x}_0}(\mathbf{x}_t^{\varepsilon} \simeq \mathbf{x}, t < \tau_{\partial D}^{\varepsilon}) p_0(\mathbf{x}_0) d\mathbf{x}_0}{\int_D P_{\mathbf{x}_0}(t < \tau_{\partial D}^{\varepsilon}) p_0(\mathbf{x}_0) d\mathbf{x}_0}$$
(60)

Now, consider a timescale in the following range:

$$1 \ll t \ll \exp\left(\frac{V(\mathbf{x}_{eq}, \partial D)}{\varepsilon}\right) \tag{61}$$

A typical example of such a timescale is  $t \sim \exp((V(\mathbf{x}_{eq}, \partial D) - \eta)/\varepsilon)$ , in which the process forgets its initial position and  $P_{\mathbf{x}_0}(t < \tau_{\partial D}^{\varepsilon})$  is still close to one. It follows that

$$p(\mathbf{x},t|\mathbf{x}_0) = P_{\mathbf{x}_0}(\mathbf{x}_t^{\varepsilon} \simeq \mathbf{x}, t < \tau_{\partial D}^{\varepsilon}) + P_{\mathbf{x}_0}(\mathbf{x}_t^{\varepsilon} \simeq \mathbf{x}, t \ge \tau_{\partial D}^{\varepsilon}) = p_{abs}(\mathbf{x},t|\mathbf{x}_0) + o(1)$$
(62)

$$p_{abs}(\mathbf{x}, t | \mathbf{x}_0) \simeq \rho_{qs}(\mathbf{x}) P_{x_0}(t < \tau_{\partial D}^{\varepsilon})$$
(63)

As a consequence, during the range of time (Equation (61)), the distribution  $p(\mathbf{x}, t | \mathbf{x}_0)$  is well approximated by the quasi-stationary distribution  $\rho_{qs}(\mathbf{x})$  and, in particular, it does not depend on the time or the initial position  $\mathbf{x}_0$ .

In the definition of the quasi-stationary distribution, a repulsive effect of the boundary is inevitable due to the conditioning with respect to the event  $\{t < \tau_{\partial D}^{\varepsilon}\}$ . In a domain far away from  $\partial D$ , this effect disappears. Therefore, the quasi-stationary distribution approximately has a WKB-type form, i.e.,

$$\rho_{qs}(\mathbf{x}) = \frac{C_{qs}(\mathbf{x})}{\varepsilon^{n/2}} \exp\left(-\frac{V(\mathbf{x}_{eq}, \mathbf{x})}{\varepsilon}\right)$$
(64)

However, this effect implies that the quasi-stationary density vanishes in the neighborhood of  $\partial D$  and, thus, boundary layer approximation has to be performed to obtain the exact expression of  $\rho_{qs}$ , which is non-negligible in the calculation of the transition rate and the distribution of the exit position. Accordingly, for **x** near the boundary, we have

$$\rho_{qs}(\mathbf{x}) = \frac{C_{bl}(\pi_D(\mathbf{x}), \eta(\mathbf{x})/\varepsilon)}{\varepsilon^{n/2}} \exp\left(-\frac{V(\mathbf{x}_{eq}, \mathbf{x})}{\varepsilon}\right)$$
(65)

where  $\pi_D(\mathbf{x})$  is the orthogonal projection of  $\mathbf{x}$  onto  $\partial D$ , and  $\eta(\mathbf{x})$  is the distance between x and  $\pi_D(\mathbf{x})$ . The boundary layer develops on a characteristic length scale  $\varepsilon$ , and when away from the boundary, this expression return to the usual WKB form (Equation (64)) quickly, i.e.,

$$\lim_{d \to \infty} C_{bl}(\mathbf{y}, d) = C_{qst}(\mathbf{y}), \ \mathbf{y} \in \partial D$$
(66)

Note that only the principal order term of the probability determinates the behaviors of large fluctuation, as indicated above. Assume that the system is initially in the quasistationary state. Substituting Equations (62), (63), and (65) into Equation (46), we get the same average dynamics as those of the stationary cases. This result shows that even if the stationary probability density does not exist, or we have no knowledge of its existence, the same optimal fluctuation in the timescale before escaping can also be displayed in the attractive domain of some fixed points.

## 3.4. Interpretation of the PPD

#### 3.4.1. Time Reversal of the Diffusion Process

In the following section, we fix the time interval as [0, T]. Let  $\{\mathbf{x}_t^{\epsilon}, 0 \le t \le T\}$  be a solution to the stochastic differential Equation (1), with its laws at each moment represented by  $\{p(\mathbf{x}, t), 0 \le t \le T\}$ . The law of the forward evolution process

$$p(\mathbf{x}, t_0) \to p(\mathbf{x}, t_1) \to \dots \to p(\mathbf{x}, t_M), \ 0 = t_0 < t_1 < \dots < t_M = T$$
(67)

is given by the following forward equation:

$$\frac{\partial p(\mathbf{x},t)}{\partial t} = \mathcal{L}_{\mathbf{x}}^* p(\mathbf{x},t)$$
(68)

It is natural to ask how the backward process, i.e.,

$$p(\mathbf{x}, t_0) \leftarrow p(\mathbf{x}, t_1) \leftarrow \dots \leftarrow p(\mathbf{x}, t_M), \ 0 = t_0 < t_1 < \dots < t_M = T$$
(69)

evolves. If we define the reversed process as  $\mathbf{y}_t^{\varepsilon} = \mathbf{x}_{T-t}^{\varepsilon}$ , it is equivalent to asking how the laws  $\overline{p}(\mathbf{y}, t) = p(\mathbf{y}, T - t)$  of the reversed process evolve.

It was proven that the reversed Markov process is also Markovian, while the reversed diffusion is not necessarily a diffusion. A sufficient condition for the reversibility of the diffusion property requires smoothness of the coefficients as well as the uniform non-degeneracy of the diffusion term [44]. Under this condition, the reversed process  $\mathbf{y}_t^{\varepsilon}$  is also a diffusion, whose generator is given as follows [45,46]:

$$\overline{\mathcal{L}}_{\mathbf{y},t} = \sum_{i,j} \frac{\varepsilon}{2} A_{ij}(\mathbf{y}) \frac{\partial^2}{\partial y_i \partial y_j} + \sum_i \overline{b}_i^{\varepsilon}(\mathbf{y},t) \frac{\partial}{\partial y_i}$$
(70)

$$\overline{b}_{i}^{\varepsilon}(\mathbf{y},t) = -b_{i}(\mathbf{y}) + \sum_{j} \frac{\varepsilon}{p(\mathbf{y},T-t)} \frac{\partial}{\partial y_{j}} (A_{ij}(\mathbf{y})p(\mathbf{y},T-t))$$
(71)

In other words,

$$\frac{\frac{\partial p(\mathbf{y}, T-t)}{\partial t}}{\sum_{i,j} \varepsilon} = \overline{\mathcal{L}}_{\mathbf{y},t}^* p(\mathbf{y}, T-t)$$
  
=  $\sum_{i,j} \frac{\varepsilon}{2} \frac{\partial^2}{\partial y_i \partial y_j} (A_{ij}(\mathbf{y}) p(\mathbf{y}, T-t)) + \sum_i \frac{\partial}{\partial y_i} (\overline{b}_i^\varepsilon(\mathbf{y}, t) p(\mathbf{y}, T-t))$ (72)

where  $\overline{\mathcal{L}}_{\mathbf{y},t}^*$  is the adjoint operator of  $\overline{\mathcal{L}}_{\mathbf{y},t}$ . After a time-reversed transformation  $t \to T - t$ , we have

$$\frac{\partial p(\mathbf{y},t)}{\partial t} = \overline{\mathcal{L}}_{\mathbf{y},T-t}^{*} p(\mathbf{y},t) \\
= \sum_{i,j} \frac{\varepsilon}{2} \frac{\partial^{2}}{\partial y_{i} \partial y_{j}} \left( A_{ij}(\mathbf{y}) p(\mathbf{y},t) \right) + \sum_{i} \frac{\partial}{\partial y_{i}} \left( \overline{b}_{i}^{\varepsilon}(\mathbf{y},T-t) p(\mathbf{y},t) \right)$$
(73)

which gives the evolutionary law of the backward process (Equation (69)). It can be seen that the reversed process is generally time-nonhomogeneous, and its evolutionary law depends on the density  $p(\mathbf{x}, t)$ . Different choices of the initial density  $p(\mathbf{x}, 0)$  yield different laws  $\{p(\mathbf{x}, t), 0 \le t \le T\}$  and, thus, different reversed evolution laws.

#### 3.4.2. Interpretation of the Stationary and Nonstationary PPD

If the initial density is chosen as  $p(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{x}_0)$ , we have  $p(\mathbf{x}, t) = p(\mathbf{x}, t | \mathbf{x}_0)$ . Substituting this into Equation (73), we get

$$\frac{\partial p(\mathbf{x},t|\mathbf{x}_{0})}{\partial t} = \overline{\mathcal{L}}_{\mathbf{x},T-t}^{*} p(\mathbf{x},t|\mathbf{x}_{0}) \\
= \sum_{i,j} \frac{\varepsilon}{2} \frac{\partial^{2}}{\partial x_{i} \partial x_{j}} \left( A_{ij}(\mathbf{x}) p(\mathbf{x},t|\mathbf{x}_{0}) \right) + \sum_{i} \frac{\partial}{\partial x_{i}} \left( \overline{b}_{i}^{\varepsilon}(\mathbf{x},T-t) p(\mathbf{x},t|\mathbf{x}_{0}) \right)$$
(74)

$$\overline{b}_{i}^{\varepsilon}(\mathbf{x}, T-t) = -b_{i}(\mathbf{x}) + \sum_{j} \frac{\varepsilon}{p(\mathbf{x}, t | \mathbf{x}_{0})} \frac{\partial}{\partial x_{j}} (A_{ij}(\mathbf{x}) p(\mathbf{x}, t | \mathbf{x}_{0}))$$
(75)

which, thus, gives the evolutionary law of the backward process

$$p(\mathbf{x}, t_0 | \mathbf{x}_0) = \delta(\mathbf{x} - \mathbf{x}_0) \leftarrow p(\mathbf{x}, t_1 | \mathbf{x}_0) \leftarrow \dots \leftarrow p(\mathbf{x}, t_M | \mathbf{x}_0), \ 0 = t_0 < t_1 < \dots < t_M = T$$
(76)

In this case, the nonstationary PPD actually serves as the transition probability density of the process (Equation (76)) in the sense of

$$\int p(\mathbf{x}_f, T | \mathbf{x}_0) p_h(\mathbf{x}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0) d\mathbf{x}_f = \int p(\mathbf{x}, t | \mathbf{x}_0, 0) p(\mathbf{x}_f, T | \mathbf{x}, t) d\mathbf{x}_f = p(\mathbf{x}, t | \mathbf{x}_0)$$
(77)

Thus, it satisfies

$$-\frac{\partial p_h(\mathbf{x}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0)}{\partial t} = \overline{\mathcal{L}}^*_{\mathbf{x}, T-t} p_h(\mathbf{x}, t | \mathbf{x}_f, T; \mathbf{x}_0, 0)$$
(78)

which is the same as Equation (42).

If the system initially lies in its stationary state, we have  $p(\mathbf{x}, t) = \rho_s(\mathbf{x})$ . The expressions in Equations (70) and (71) can be simplified as follows:

$$\overline{\mathcal{L}}_{\mathbf{y}} = \sum_{i,j} \frac{\varepsilon}{2} A_{ij}(\mathbf{y}) \frac{\partial^2}{\partial y_i \partial y_j} + \sum_j \overline{b}_i^{\varepsilon}(\mathbf{y}) \frac{\partial}{\partial y_i}$$
(79)

$$\overline{b}_{i}^{\varepsilon}(\mathbf{y}) = -b_{i}(\mathbf{y}) + \sum_{j} \frac{\varepsilon}{\rho_{s}(\mathbf{y})} \frac{\partial}{\partial y_{j}} (A_{ij}(\mathbf{y})\rho_{s}(\mathbf{y}))$$
(80)

In this case, the backward process  $\rho_s(\mathbf{x}) \leftarrow \rho_s(\mathbf{x}) \leftarrow \cdots \leftarrow \rho_s(\mathbf{x})$  can be described as follows:  $0 = -\frac{\partial \rho_s(\mathbf{x})}{\partial t} = \overline{L}_{**}^* \rho_s(\mathbf{x})$ (81)

$$0 = -\frac{1}{\partial t} = \mathcal{L}_{\mathbf{x}} \rho_s(\mathbf{x})$$
(81)

Similarly, the stationary PPD is the transition probability density of the backward process in the sense of

$$\int \rho_s(\mathbf{x}) p_h(\mathbf{x}, t | \mathbf{x}_f) d\mathbf{x}_f = \int \rho_s(\mathbf{x}) p(\mathbf{x}_f | \mathbf{x}, t) d\mathbf{x}_f = \rho_s(\mathbf{x})$$
(82)

This, of course, satisfies Equation (23).

## 3.4.3. Relationship between $\mathbf{y}_t^{\varepsilon}$ and $\mathbf{z}_t^{\varepsilon}$

In fact, Equation (44) describes the forward evolution process of the PPD, i.e.,

$$\delta(\mathbf{x} - \mathbf{x}_0) = p_h(\mathbf{x}, t_0 | \mathbf{x}_f, T; \mathbf{x}_0, 0) \rightarrow p_h(\mathbf{x}, t_1 | \mathbf{x}_f, T; \mathbf{x}_0, 0) \rightarrow \cdots \rightarrow p_h(\mathbf{x}, t_M | \mathbf{x}_f, T; \mathbf{x}_0, 0) = \delta(\mathbf{x} - \mathbf{x}_f), \quad 0 = t_0 < t_1 < \dots < t_M = T \quad (83)$$

Motivated by the relationship between Equations (68) and (72), we can find that the corresponding backward evolution process should obey

$$-\frac{\partial p_h}{\partial t} = \frac{\varepsilon}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} \left( A_{ij}(\mathbf{x}) p_h \right) - \sum_i \frac{\partial}{\partial x_i} \left( \overline{\tilde{b}}_i^\varepsilon(\mathbf{x}, T-t) p_h \right), \ 0 \le t \le T$$
(84)

where

$$\overline{b}_{i}^{\varepsilon}(\mathbf{x}, T-t) = -\overline{b}_{i}^{\varepsilon}(\mathbf{x}, t) + \sum_{j} \frac{\varepsilon}{p_{h}(\mathbf{x}, t | \mathbf{x}_{f}, T; \mathbf{x}_{0}, 0)} \frac{\partial}{\partial x_{j}} \left( A_{ij}(\mathbf{x}) p_{h}(\mathbf{x}, t | \mathbf{x}_{f}, T; \mathbf{x}_{0}, 0) \right) 
= -b_{i}(\mathbf{x}) - \sum_{j} \varepsilon A_{ij}(\mathbf{x}) \frac{\partial}{\partial x_{j}} \ln p(\mathbf{x}_{f}, T | \mathbf{x}, t) 
+ \sum_{j} \frac{\varepsilon}{p(\mathbf{x}, t | \mathbf{x}_{0}, 0) p(\mathbf{x}_{f}, T | \mathbf{x}, t)} \frac{\partial}{\partial x_{j}} \left( A_{ij}(\mathbf{x}) p(\mathbf{x}, t | \mathbf{x}_{0}, 0) p(\mathbf{x}_{f}, T | \mathbf{x}, t) \right) 
= -b_{i}(\mathbf{x}) + \sum_{j} \frac{\varepsilon}{p(\mathbf{x}, t | \mathbf{x}_{0}, 0)} \frac{\partial}{\partial x_{j}} \left( A_{ij}(\mathbf{x}) p(\mathbf{x}, t | \mathbf{x}_{0}, 0) \right) = \overline{b}_{i}^{\varepsilon}(\mathbf{x}, T - t)$$
(85)

After a time-reversed transformation  $t \to T - t$ , we can determine the drift field of Equation (46). Thus,  $\mathbf{y}_t^e$  and  $\mathbf{z}_t^e$ , conditioned at  $\mathbf{y}_0^e = \mathbf{x}_f$  and  $\mathbf{z}_0^e = \mathbf{x}_0$  respectively, are the time-reversed process of one another, which describe the backward and forward evolution of the nonstationary PPD, respectively.

For  $\mathbf{y}_t^{\varepsilon}$  in Equation (26) and  $\mathbf{z}_t^{\varepsilon}$  in Equation (30) in the stationary cases, the conclusion will be similar, which will not be repeated here.

# 3.4.4. At Equilibrium

We say the system in Equation (1) is at equilibrium if it is stationary and its law is timesymmetrical, i.e.,  $\mathcal{L}_{\mathbf{x}} = \overline{\mathcal{L}}_{\mathbf{x}}$ . Thus, we can obtain the following condition of detailed balance:

$$2\mathbf{A}^{-1}(\mathbf{x})\left(\mathbf{b}(\mathbf{x}) - \frac{\varepsilon}{2}\nabla \cdot \mathbf{A}(\mathbf{x})\right) = \varepsilon \nabla \ln(\rho_{s}(\mathbf{x}))$$
(86)

If expressing the stationary density as  $\rho_s(\mathbf{x}) = \exp(-U(\mathbf{x},\varepsilon)/\varepsilon)$ , we can obtain

$$2\mathbf{A}^{-1}(\mathbf{x})\left(\mathbf{b}(\mathbf{x}) - \frac{\varepsilon}{2}\nabla \cdot \mathbf{A}(\mathbf{x})\right) = -\nabla U(\mathbf{x},\varepsilon)$$
(87)

That is, it requires  $\mathbf{A}^{-1}(\mathbf{x})(\mathbf{b}(\mathbf{x}) - \frac{\varepsilon}{2}\nabla \cdot \mathbf{A}(\mathbf{x}))$  to be a gradient field. We can thus divide it into the following three categories:

1.  $\mathbf{A}(\mathbf{x}) = \mathbf{I}$  and  $\mathbf{b}(\mathbf{x}) = -\nabla u_0(\mathbf{x})$ , such that  $u_0(\mathbf{x}_{eq}) = 0$ . In this case, the quasi-potential  $V(\mathbf{x}_{eq}, \mathbf{x}) = 2u_0(\mathbf{x})$ , as it satisfies the Hamilton–Jacobi equation (Equation (17)). The potential at equilibrium is  $U(\mathbf{x}, \varepsilon) = 2u_0(\mathbf{x}) + const$ , as the normalization condition  $\int \rho_s(\mathbf{x}) d\mathbf{x} = 1$  should be satisfied. Thus, the quasi-potential coincides with the potential up to a constant. The optimal fluctuation field will be as follows:

$$\mathbf{b}(\mathbf{x}) + \mathbf{A}(\mathbf{x})\nabla_{\mathbf{x}}V(\mathbf{x}_{eq};\mathbf{x}) = -\mathbf{b}(\mathbf{x})$$
(88)

That is, the OP will follow the time reversal of the deterministic average dynamics. As a result, time symmetry for the optimal fluctuations and relaxation motions (in the absence of noise) under the time-reversed transformation holds.

- 2.  $\mathbf{A}^{-1}(\mathbf{x})\nabla\cdot\mathbf{A}(\mathbf{x}) = \nabla u_1(\mathbf{x})$  and  $\mathbf{A}^{-1}(\mathbf{x})\mathbf{b}(\mathbf{x}) = -\nabla u_0(\mathbf{x})$ , such that  $u_0(\mathbf{x}_{eq}) = 0$ . In this case, Equation (87) will hold for any choice of  $\varepsilon$ . Similarly, the quasi-potential satisfies  $V(\mathbf{x}_{eq}, \mathbf{x}) = 2u_0(\mathbf{x})$ , and Equation (88) is still valid, but  $U(\mathbf{x}, \varepsilon) = 2u_0(\mathbf{x}) + \varepsilon u_1(\mathbf{x}) + const$ . The deviation of the metastable states (i.e., minimum points of *U*) from the fixed points of the average system will emerge due to the non-homogeneity of the diffusion coefficient.
- 3.  $\mathbf{A}^{-1}(\mathbf{x})\mathbf{b}(\mathbf{x})$  is not a gradient field. In this case, if Equation (87) is valid, it holds only for an isolated point  $\varepsilon = \varepsilon_0$ . This is the case where neither  $\mathbf{A}^{-1}(\mathbf{x})\mathbf{b}(\mathbf{x})$  nor  $\mathbf{A}^{-1}(\mathbf{x})\nabla \cdot \mathbf{A}(\mathbf{x})$  is a gradient field, while their linear combination is. If  $\varepsilon_0$  is small enough, we can also approximate the potential by breaking it down into two parts  $U(\mathbf{x}, \varepsilon_0) = U_0(\mathbf{x}) + \varepsilon_0 U_1(\mathbf{x}, \varepsilon_0) + const$  such that  $U_0(\mathbf{x}_{eq}) = 0$ , where the  $\varepsilon_0$ -independent part  $U_0(\mathbf{x})$  is nothing but the quasi-potential, with  $\exp(-U_1)$  being the prefactor of the WKB approximation (Equation (31)). The deviation of the metastable states from the fixed points of the average system still exists, but the optimal fluctuation field will not coincide with  $-\mathbf{b}(\mathbf{x})$ . This is very different from the case of additive noise.

#### 4. Numerical Algorithm and Results

An advantage of the PPD is that it permits us to analyze the optimal fluctuation statistically. If

$$\mathcal{C}(\mathbf{x},t;\mathbf{x}_f) = \left\{ \mathbf{x}(s), s \le 0 | \mathbf{x}(0) = \mathbf{x}_f, \mathbf{x}(t) = \mathbf{x} \right\}, \ t < 0$$
(89)

then we have

$$p_h(\mathbf{x}, t | \mathbf{x}_f) = P\left\{\mathbf{x}_t^{\varepsilon} \in \mathcal{C}(\mathbf{x}, t; \mathbf{x}_f) | p(\mathbf{x}, s) = \rho_{s(qs)}(\mathbf{x}), s < 0\right\}, \ t < 0$$
(90)

This implies that the stationary PPD can be obtained by calculating the laws of the prehistory positions for all possible large fluctuations reaching the given state  $x_f$ , conditioned in the stationary or quasi-stationary states. Specifically, we need to:

- 1. Simulate the Langevin dynamics for a sufficiently long time until they reach their stationary or quasi-stationary state;
- 2. Extract the sample trajectories that hit  $\mathbf{x}_{f}$ ;
- 3. Obtain the PPD by calculating the law of the prehistorical positions of these trajectories statistically.

Searching for its peak trajectory yields the OP. The idea presented in [12,24,25,31] for calculating the PPD and the OP is essentially the same as Equation (90).

Moreover, when the system in Equation (1) is an equilibrium system, we have

$$\frac{\partial p_h(\mathbf{x}, -t|\mathbf{x}_f)}{\partial t} = \overline{\mathcal{L}}_{\mathbf{x}}^* p_h(\mathbf{x}, -t|\mathbf{x}_f) = \mathcal{L}_{\mathbf{x}}^* p_h(\mathbf{x}, -t|\mathbf{x}_f), \ t > 0$$
(91)

This suggests that the stationary PPD, in this case, can be obtained by solving the Fokker–Planck equation or simulating the Langevin motions, as described in [27–29].

However, unlike the stationary or quasi-stationary cases—where observing large fluctuations is always possible provided that the observation period is long enough—in the nonstationary cases, the occurrence of large excursions from the average dynamics over a finite time span is essentially rare, which means that applying traditional techniques such as the Monte Carlo method to extract the optimal path is almost impossible. In order to address this difficulty, the methods of cell mapping and Gaussian approximation for short time increments are used here to turn the evolution of diffusions to that of the corresponding Markov chains. Then, the nonstationary PPD can be obtained by solving the discrete version of Equation (78) (see below for details).

#### 4.1. Algorithm

For simplicity, we assume the dimension n = 1. In order to track the temporal evolution of the law of the stochastic variable  $x_t^{\epsilon}$  numerically, the discretization of both the state space and time is necessary, which is formulated as follows:

First, we assume that a domain E = [a, b] containing points  $x_0$  and  $x_f$  could be chosen such that the vector field  $\mathbf{b}(\mathbf{x})$  at the boundary  $\partial E$  points towards its interior; meanwhile:

$$\min_{x \in \partial E, 0 \le t \le T} V(x_0; t, x) > V(x_f; T, x_0)$$
(92)

It follows that, for vanishing noise, almost all trajectories will stay in this domain over the interval [0, T], and the event of escaping from *E* becomes more unlikely than that of transition between  $x_0$  and  $x_f$ . In other words, the NOP connecting  $x_0$  and  $x_f$  is wholly contained in the domain *E*, and neglecting all of the trajectories escaping before the moment *T* has almost no effect on the calculation of the NOP. Thus, focusing on the trajectories not having left from the domain *E* up to time *T* will be precise enough.

Next, we divide the domain *E* into *N* subintervals with uniform sizes  $\Delta x = (b - a)/N$ . The *i*-th cell is defined as the interval  $[a + (i - 1)\Delta x, a + i\Delta x]$ . We choose a representative element in each cell as  $x(i) = a + (i - 1/2)\Delta x$ . The whole state space is then partitioned into N + 1 cells, with the (N + 1)-th cell being defined as the complement of *E*.

Similarly, we discretize the time interval [0, T] into M steps of uniform sizes  $\Delta t = T/M$ . A large enough M could be chosen such that the local Gaussian approximation could be applied, i.e.,

$$x_{t+\Delta t} \simeq x_t + b(x_t)\Delta t + \sqrt{\varepsilon\sigma(x_t)}w_{\Delta t}$$
(93)

Let  $\{p_m(i), m = 0, 1, \dots, M, i = 1, 2, \dots, N+1\}$  denote the probability of observing the process in the *i*-th cell at the moment  $m\Delta t$ . A Markov chain can then be defined to describe the discrete-time evolution of the absorbed diffusion, i.e.,

$$p_{m+1}(j) = \sum_{i} p_m(i) P_{i,j}, \ j = 1, 2, \cdots, N+1, \ m = 0, 1, \cdots, M-1$$
(94)

with its transition probability  $\{P_{i,j}, i, j = 1, 2, \dots, N+1\}$  given as follows:

$$P_{i,j} = \frac{1}{\sqrt{2\pi\epsilon\Delta t}} \exp\left(-\frac{(x(j)-x(i)-b(x(i))\Delta t)^2}{2\epsilon\Delta t}\right) \Delta x, \ i, j = 1, 2, \cdots, N$$

$$P_{i,N+1} = 1 - \sum_{j=1}^{N} P_{ij}, \ P_{N+1,i} = 0, \ i = 1, 2, \cdots, N$$

$$P_{N+1,N+1} = 1$$
(95)

To calculate the NOP, we need to explore the reversed evolution of the original process. The evolution  $p_{m+1} \rightarrow p_m$  is generally time-inhomogeneous, with its transition probability having the following form:

$$\overline{P}_{i,j}^{(m)} = \begin{cases} \frac{p_m(j)P_{j,i}}{p_{m+1}(i)}, & p_{m+1}(i) > 0\\ 0, & p_{m+1}(i) = 0 \end{cases}$$
(96)

Thus, if we choose the distribution  $p_0$  and  $\overline{p}_M$  to be the probability concentrated on the cells containing  $x_0$  and  $x_f$ , respectively, the nonstationary PPD is then given as follows:

$$\overline{p}_{m}(j) = \sum_{i} \overline{p}_{m+1}(i) \overline{P}_{i,j}^{(m)}$$
(97)

Searching for the peak of  $\overline{p}_m$  and calculating its local Gaussian coefficient can yield the NOP and its dispersion. The numerical results are displayed next.

## 4.2. Example

We here choose A(x) = 1 and b(x) = -U'(x) with the potential  $U(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2$ . This is a prototypical example of a double-well system. It has two minimum points,  $x_{sl} = -1$  and  $x_{sr} = 1$ , corresponding to the two stable fixed points, and a local maximum point  $x_u = 0$  corresponding to the unstable fixed point of the average system. The global structure of the potential is displayed in Figure 1. It can be seen that restricting the domain to E = [-2, 2] is precise enough for our calculation. In the following section, we partition the domain *E* into N = 1000 cells, and discretize the time interval [-2, 2] into M = 1000 small segments. The algorithm designed above can then be applied to show the effects of various parameters  $x_0$ ,  $x_f$  and *T* on the nonstationary optimal fluctuations.



**Figure 1.** The potential U(x) of the double-well model; its two minima correspond to the two metastable states  $x_{sl} = -1$  and  $x_{sr} = 1$ , with an unstable fixed point  $x_u = 0$  separating them.

First, we fix  $x_0 = -1$  and  $x_f = 1$  to study the noise-induced transition between the two stable states. The global structure of the Hamiltonian system is plotted in Figure 2a. The red curve indicates the OP from  $x_0$  to  $x_f$ , while the green curve corresponds to the OP of the reversed transition. This is the maximum-likelihood event with a minimum cost of the action. Near the fixed points, the momentum p is close to zero, and so is the velocity  $\dot{x}$ , implying that the stochastic trajectories around it always spend almost-infinite time oscillating near the fixed points, and occasionally fluctuate to a bounded domain. However, this is not the case with the NOPs. As a finite time interval [0, T] is required, the NOPs correspond to the Hamiltonian orbits above the heteroclinic orbit, which can be parameterized by the initial condition  $(x_0, p_0)$ . A typical NOP transitioning between these two points is shown by a purple curve in Figure 2a. It can be seen that the momentum here is always positive, meaning that a large bias from the average dynamic always exists. As a result, this is a rare event with exponentially small probability, and can scarcely be observed.



**Figure 2.** (a) The vector field of the Hamiltonian system. The red and green curves denote the OP from  $x_{sl}$  to  $x_{sr}$  and its reverse, respectively. The purple curve is a typical example of the NOP from  $x_{sl}$  to  $x_{sr}$ . (b) The time required for a Hamiltonian orbit to hit the line  $x = x_{sr}$  from the initial point  $(x_{xl}, p_0)$ .

To demonstrate how the nonstationary PPD can pinpoint the location of the NOP, we first calculate the time required for an NOP to reach the target point. As shown in Figure 2b, this is a monotone function of the initial momentum. For any choice of T, a unique NOP exists, and no singular behavior arises consequently. The PPD will always present a single peak trajectory accordingly. Now, let T = 10. The PPDs and their peak trajectories for various noise intensities are illustrated in Figure 3. The results can be compared with those obtained by the minimum action method (MAM) [47] (details of the MAM are presented in Appendix A). This shows that our numerical method could approximate the NOP fairly well when  $\varepsilon$  is small enough. However, the matching effect soon fails as  $\varepsilon$  increases. In particular, the hopping of the peak trajectory of the PPD occurs, as displayed in Figure 3b,c. The law for all of the fluctuation paths connecting  $x_0$  with  $x_f$  is initially concentrated on the domain near  $x_0$ . Larger noise makes significant fluctuation more likely. Consequently, a local peak near  $x_f$  emerges as time goes by, whose height subsequently exceeds that of the original one and dominates the law at the end. This is a result caused by the  $\varepsilon$ -dependent part of Equation (46), including the  $O(\varepsilon)$  part in the drift term as well as the weak Gaussian noise. To ensure that the peak trajectory of the PPD converges to the NOP, both of them require  $T \ll 1/\varepsilon$ . As a result, lowering  $\varepsilon$  or T will reduce the error of our algorithm. As we can see from Figures 3d and 4, for a fixed T, the peak trajectory of the nonstationary PPD coincides with the NOP only if the noise intensity is small enough, while for a fixed  $\varepsilon$ , the deviation of the peak trajectory from the NOP increases as T increases. The larger T is, the smaller the  $\varepsilon$  needed to ensure the matching effect between the two paths. The numerical results are in perfect agreement with our theoretical predictions.

Next, we explore how the  $\varepsilon$ -independent parameter  $\kappa$  evolves when T and  $x_f$  vary. The NOPs for differential target points  $x_f$  and T are shown in Figure 5, with their corresponding dispersion parameters displayed in Figure 6. This shows that  $\kappa$  varies approximately linearly near the initial and the final points, and continuously increases to reach its maximum. For a short time span T = 1,  $\kappa$  has a single peak, while if we choose  $x_0 = 1$  and increase T, two peaks of  $\kappa$  will emerge. Moreover, the value of the dispersion parameter grows dramatically when T increases, as shown in Figure 7. Since the dispersion of the PPD reflects the sensitivity to fluctuations along the NOP, a larger  $\kappa$  implies that stochastic trajectories are more likely to be biased from the optimal one and, thus, a smaller  $\varepsilon$  is needed to ensure the convergence. This is consistent with our previous analysis. In particular, it is worth pointing out that  $\kappa(t^*) = 1/2$  at the moment where the optimal path reaches

the unstable fixed point  $x_{nop}(t^*) = 0$ . This is because the NOP approaches the stationary one as *T* becomes large enough. As a result, the value of  $\kappa$  near the unstable fixed point can be approximately described by the quasi-potential of the reversed process—that is,  $\kappa(t^*) \approx -1/(2U''(0))$ .



**Figure 3.** The PPDs and their peak trajectories; here,  $x_0 = -1$ ,  $x_f = 1$  and T = 10. (a)  $\varepsilon = 0.001$ . (b)  $\varepsilon = 0.01$ . (c)  $\varepsilon = 0.1$ . (d) Comparison of the NOP obtained by the MAM and the peak trajectories of the PPDs with different noise intensities.



**Figure 4.** Comparison of the NOP obtained by the MAM and the peak trajectories of the PPDs with different noise intensities: (a) T = 5. (b) T = 7.



**Figure 5.** The NOP for different target positions  $x_f$ . (a) T = 1,  $\varepsilon = 0.01$ . (b) T = 4,  $\varepsilon = 0.001$ .



**Figure 6.** The noise-independent dispersion parameter  $\kappa$  as a function of t and x for different final positions  $x_f$ . (**a**,**b**) T = 1,  $\varepsilon = 0.01$ ; (**c**,**d**) T = 4,  $\varepsilon = 0.001$ .



**Figure 7.** The emergence of multiple peaks of the dispersion parameter  $\kappa$  for distinct time gaps *T*. (a) The variations of the dispersion parameter  $\kappa$  with respect to time *t*. (b) The variations of the dispersion parameter  $\kappa$  with respect to position *x*. Here,  $x_f = 1$  and  $\varepsilon = 0.001$ .

In the end, we show the multiple-peaks effect of the PPD. Here, we can assume that  $-1 < x_0 < x_f < 0$ . In this case, the circular phase trajectories in the interior of the heteroclinic orbits now serve as the optimal phase orbits of our optimization problem. The time required for an NOP to reach the target point is illustrated in Figure 8a. We can divide them into five scenarios, each of which has a typical example displayed in Figure 8b. This shows that, as  $T \to \infty$ , there are only two principal phase orbits as candidates for the NOP: case B and case C (case A must have a larger action). Which one dominates the PPD depends on the value of the action. Since they approach the heteroclinic orbits, their action could be approximately estimated by  $V_B = 2U(x_f) - 2U(x_{sl})$  and  $V_C = 2U(x_u) - 2U(x_0)$ , respectively. Thus, the switch between different optimal paths occurs when  $x_f$  is close to a point  $x_f^*$  satisfying  $2U(x_f^*) - 2U(x_{sl}) = 2U(x_u) - 2U(x_0)$ . For our choice of  $x_0 = -0.8$ , we get  $x_f^* = -0.2589$ . The PPD and its peak trajectories for  $x_f = -0.24$ , -0.25, -0.26 are presented in Figure 9, confirming the above analysis.



**Figure 8.** (a) The time required for a Hamiltonian trajectory to hit the line  $x = x_f$  from the initial point  $(x_0, p_0)$ . There exist five cases with their Hamiltonian orbits drawn in (b). Here,  $x_0 = -0.8$  and  $x_f = -0.3$ .



**Figure 9.** The coexistence of multiple NOPs and mutation of the peak trajectory of the PPD when varying  $x_f$ . (a)  $x_f = -0.24$ . (b)  $x_f = -0.26$ . (c) $x_f = -0.25$ . (d) The two coexisting paths in (c) corresponding to the transitions  $x_0 \rightarrow x_{sl} \rightarrow x_0 \rightarrow x_f$  (solid green line) and  $x_0 \rightarrow x_f \rightarrow x_u \rightarrow x_f$  (solid red line). Here  $x_0 = -0.8$  and T = 7.

# 5. Conclusions

The prehistorical description of the optimal fluctuations with the framework of the Langevin dynamics was presented in this paper. By means of the so-called prehistory probability density, we showed that both the NOP and the OP are related to the corresponding nonstationary and stationary reversed diffusion motion, respectively. Specifically, we proved that the PPD, in fact, acts as the transition probability density of the reversed process. As the noise intensity is close to zero, the stochastic trajectories focus on their average dynamics due to the law of large numbers, where our optimal fluctuation path is exactly located. This explains why the peaks of the PPD approximately coincide with the OP or NOP for vanishing noise. Moreover, we found that near the OP or NOP, the PPD is approximately Gaussian, with its Gaussian dispersion coefficient being reformulated by the linear part of the reversed process. This is a result of the type of the central limit theorem. Thus, the PPD gives us a complete picture of the optimal fluctuations, since it not only allows us to capture information about the location of the OP and the NOP, but also provides the characteristics of the nearby paths.

Based on these theoretical analyses, we designed an algorithm and performed it with a one-dimensional example to visualize the results, showing that the PPD can be used to precisely position the optimal fluctuation path only if the noise intensity is small enough. In particular, it is effective in identifying the coexistence of multiple optimal fluctuation paths. However, the matching effect soon fails when increasing the noise intensity or the time span. Lowering  $\varepsilon$  or T will reduce the error of our algorithm, but from a computational viewpoint, in order to make our algorithm more precise, noise should neither be set so weak that the exponentially small probability will be lower than the minimum limit that the computer can identify, nor so strong that the incompatibility of the peak trajectory of the PPD and the optimal fluctuation path obtained by large deviation theory will emerge. Moreover, it is worth pointing out that even though our conclusion is valid in arbitrarily finite dimensions, due to the limitations of our computational power, the numerical calculation based on the technique of the PPD is, in fact, only useful for systems with lower dimensions. A higher dimension implies that the optimal fluctuation path can be observed more rarely, and more and more computational effort is required, which is an inevitable difficulty known as "the curse of dimensionality" in computational science.

We should finally note that the results here can also be extended to non-autonomous cases without any additional assumptions. The only difference is that the invariance of the time transition transformation is no longer held. Thus, we must express all of the quantities as functions of two-time parameters. In particular, this technique can also be used to investigate the projection singularity of the Lagrangian manifold for a periodic stationary system. One can refer to [12] as a simple example.

**Author Contributions:** Conceptualization, F.Z., Y.L. and X.L.; methodology, F.Z.; validation, Y.L. and X.L.; formal analysis, F.Z. and Y.L.; investigation, F.Z. and X.L.; writing—original draft, F.Z.; writing—review and editing, F.Z., Y.L. and X.L.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 12172167).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

**Data Availability Statement:** The data used to support the findings of this study are available from the authors upon request.

Acknowledgments: This research was supported by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

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

# Appendix A. Numerical Solutions of the Optimization Problems in Equations (9) and (13)

According to the gradient descent method, in order to obtain the minimizer of a scalar function  $U(\mathbf{x})$ , one needs only to let the variables evolve in the opposite direction of the gradient field, i.e.,

$$\dot{\mathbf{x}} = -\nabla U(\mathbf{x}) \tag{A1}$$

The minimum action method [47] is essentially a gradient descent method to obtain the minimizer of the optimization problem:

$$V(\mathbf{x}_0; T, \mathbf{x}_f) = \min_{\mathbf{x}(0) = \mathbf{x}_0, \mathbf{x}(T) = \mathbf{x}_f} S[\mathbf{x}(s)]$$
(A2)

That is, one needs only to solve the following partial differential equation:

$$\frac{\partial \mathbf{x}(\tau,t)}{\partial \tau} = -\frac{\delta S[\mathbf{x}(\cdot)]}{\delta \mathbf{x}}(t) = -\left[\frac{\mathrm{d}L(\mathbf{x},\dot{\mathbf{x}})}{\mathrm{d}\mathbf{x}} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}L(\mathbf{x},\dot{\mathbf{x}})}{\mathrm{d}\dot{\mathbf{x}}}\right)\right], \ 0 < t < T, \ \tau > 0,$$
(A3)

With the constraints  $\mathbf{x}(0) = \mathbf{x}_0, \mathbf{x}(T) = \mathbf{x}_f$ .

Specifically, we can discretize the time interval [0, T] into M steps of uniform sizes  $\Delta t = T/M$ . If  $\mathbf{x}_m^k = \mathbf{x}(k\Delta\tau, m\Delta t)$ , the optimization problem can now be solved in an iterative manner.

- 1. Starting from the *k*th guess  $\mathbf{x}_m^k$ , compute a descent direction  $\mathbf{d}_m^k = -\frac{\delta S[\mathbf{x}(\cdot)]}{\delta \mathbf{x}}\Big|_{\mathbf{x}=-1}$
- 2. Descend along the descent direction  $\mathbf{x}_m^{k+1} = \mathbf{x}_m^k + \Delta \tau \mathbf{d}_m^k$  to obtain the next guess  $\mathbf{x}_m^{k+1}$ ;
- 3. Finally, iterate until convergence.

As we can see, the algorithm here is fundamentally different from that described in Section 4. The algorithm here is based on the gradient descent method to solve the optimization problem, while the algorithm in Section 4 is designed to display the peak trajectory of the nonstationary PPD. The numerical results for distinct noise intensity and time gap are presented in Figures 3 and 4. It can be seen that the peak trajectory of the nonstationary PPD coincides with the NOP only if the noise intensity is small enough, justifying our theoretical analysis performed in Section 3.2.

In addition, it is notable that the gradient descent method can also be applied to calculate the OP. Since additional minimization over the transition time is required, we first need to transform the optimization problem in Equation (13) to a form similar to Equation (9). Based on the Maupertuis principle,

$$V(\mathbf{x}_{eq}; \mathbf{x}_{f}) = \inf_{T>0_{\mathbf{x}(0)}=\mathbf{x}_{eq}, \mathbf{x}(T)=\mathbf{x}_{f}} S[\mathbf{x}(\cdot)]$$

$$= \min_{\mathbf{x}(0)=\mathbf{x}_{eq}, \mathbf{x}(T)=\mathbf{x}_{f} \vartheta: H(\mathbf{x}, \vartheta)=0} \int_{\mathbf{x}_{eq}}^{\mathbf{x}_{f}} \vartheta^{T} d\mathbf{x}$$

$$= \min_{\mathbf{x}(0)=\mathbf{x}_{eq}, \mathbf{x}(1)=\mathbf{x}_{f} \vartheta: H(\mathbf{x}, \vartheta)=0} \int_{0}^{1} \vartheta^{T} \mathbf{x}' ds$$

$$= \min_{\mathbf{x}(0)=\mathbf{x}_{eq}, \mathbf{x}(1)=\mathbf{x}_{f}} \hat{S}[\mathbf{x}(\cdot)]$$
(A4)

where  $S[\mathbf{x}(\cdot)]$  represents the geometric minimum action. As we can see, the action is invariant under reparametrization. Thus, it enables us to treat the infinite time—as displayed in the constraints in Equation (18)—with finite discretization points. The above algorithm can then be used to calculate the OP, which is known as the geometric minimum action method (gMAM). Since only the NOP is plotted in this paper, we do not repeat the gMAM here. One can refer to [48] for further details.

# References

- 1. Kramers, H.A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **1940**, *7*, 284–304. [CrossRef]
- 2. Gammaitoni, L.; Hänggi, P.; Jung, P.; Marchesoni, F. Stochastic Resonance. Rev. Mod. Phys. 1998, 70, 223. [CrossRef]
- Freidlin, M.I. Quasi-Deterministic Approximation, Metastability and Stochastic Resonance. *Phys. D Nonlinear Phenom.* 2000, 137, 333–352. [CrossRef]
- Zhang, X.-J.; Qian, H.; Qian, M. Stochastic Theory of Nonequilibrium Steady States and Its Applications. Part I. *Phys. Rep.* 2012, 510, 1–86. [CrossRef]
- Ge, H.; Qian, M.; Qian, H. Stochastic Theory of Nonequilibrium Steady States. Part II: Applications in Chemical Biophysics. *Phys. Rep.* 2012, 510, 87–118. [CrossRef]
- 6. Graham, R. Path Integral Formulation of General Diffusion Processes. Z. Phys. B Condens. Matter 1977, 26, 281–290. [CrossRef]
- McKane, A.J.; Luckock, H.C.; Bray, A.J. Path Integrals and Non-Markov Processes. I. General Formalism. *Phys. Rev. A* 1990, 41, 644. [CrossRef]
- Bray, A.J.; McKane, A.J.; Newman, T.J. Path Integrals and Non-Markov Processes. II. Escape Rates and Stationary Distributions in the Weak-Noise Limit. *Phys. Rev. A* 1990, 41, 657. [CrossRef]
- 9. Freidlin, M.I. *Functional Integration and Partial Differential Equations (AM-109)*; Princeton University Press: Princeton, NJ, USA, 2016; Volume 109.
- 10. Ventsel', A.D.; Freidlin, M.I. On Small Random Perturbations of Dynamical Systems. Russ. Math. Surv. 1970, 25, 1–55. [CrossRef]
- 11. Dykman, M.I.; Millonas, M.M.; Smelyanskiy, V.N. Observable and Hidden Singular Features of Large Fluctuations in Nonequilibrium Systems. *Phys. Lett. A* **1994**, *195*, 53–58. [CrossRef]
- Dykman, M.I.; Luchinsky, D.G.; McClintock, P.V.E.; Smelyanskiy, V.N. Corrals and Critical Behavior of the Distribution of Fluctuational Paths. *Phys. Rev. Lett.* 1996, 77, 5229. [CrossRef] [PubMed]
- 13. Smelyanskiy, V.N.; Dykman, M.I.; Maier, R.S. Topological Features of Large Fluctuations to the Interior of a Limit Cycle. *Phys. Rev. E* **1997**, *55*, 2369. [CrossRef]
- Maier, R.S.; Stein, D.L. Effect of Focusing and Caustics on Exit Phenomena in Systems Lacking Detailed Balance. *Phys. Rev. Lett.* 1993, 71, 1783. [CrossRef]

- 15. Berglund, N.; Gentz, B. On the Noise-Induced Passage through an Unstable Periodic Orbit I: Two-Level Model. *J. Stat. Phys.* 2004, 114, 1577–1618. [CrossRef]
- 16. Maier, R.S.; Stein, D.L. Limiting Exit Location Distributions in the Stochastic Exit Problem. *SIAM J. Appl. Math.* **1997**, *57*, 752–790. [CrossRef]
- 17. Luchinsky, D.G.; Maier, R.S.; Mannella, R.; McClintock, P.V.E.; Stein, D.L. Observation of Saddle-Point Avoidance in Noise-Induced Escape. *Phys. Rev. Lett.* **1999**, *82*, 1806. [CrossRef]
- Graham, R.; Hamm, A.; Tél, T. Nonequilibrium Potentials for Dynamical Systems with Fractal Attractors or Repellers. *Phys. Rev. Lett.* 1991, 66, 3089. [CrossRef]
- 19. Kraut, S.; Feudel, U. Enhancement of Noise-Induced Escape through the Existence of a Chaotic Saddle. *Phys. Rev. E* 2003, 67, 15204. [CrossRef]
- Kraut, S.; Feudel, U. Noise-Induced Escape through a Chaotic Saddle: Lowering of the Activation Energy. *Phys. D Nonlinear Phenom.* 2003, 181, 222–234. [CrossRef]
- Luchinski, D.G.; Khovanov, I.A. Fluctuation-Induced Escape from the Basin of Attraction of a Quasiattractor. J. Exp. Theor. Phys. Lett. 1999, 69, 825–830. [CrossRef]
- Silchenko, A.N.; Beri, S.; Luchinsky, D.G.; McClintock, P.V.E. Fluctuational Transitions through a Fractal Basin Boundary. *Phys. Rev. Lett.* 2003, *91*, 174104. [CrossRef] [PubMed]
- Silchenko, A.N.; Beri, S.; Luchinsky, D.G.; McClintock, P.V.E. Fluctuational Transitions across Different Kinds of Fractal Basin Boundaries. *Phys. Rev. E* 2005, 71, 46203. [CrossRef] [PubMed]
- Dykman, M.I.; McClintock, P.V.E.; Smelyanski, V.N.; Stein, N.D.; Stocks, N.G. Optimal Paths and the Prehistory Problem for Large Fluctuations in Noise-Driven Systems. *Phys. Rev. Lett.* 1992, *68*, 2718. [CrossRef] [PubMed]
- 25. Luchinsky, D.G.; McClintock, P.V.E.; Dykman, M.I. Analogue Studies of Nonlinear Systems. *Reports Prog. Phys.* 1998, 61, 889. [CrossRef]
- Hales, J.; Zhukov, A.; Roy, R.; Dykman, M.I. Dynamics of Activated Escape and Its Observation in a Semiconductor Laser. *Phys. Rev. Lett.* 2000, *85*, 78. [CrossRef]
- 27. Morillo, M.; Casado, J.M.; Gómez-Ordónez, J. Brownian Dynamics Simulation of the Prehistory Problem. *Phys. Rev. E* 1997, 55, 1521. [CrossRef]
- 28. Arrayás, M.; Casado, J.M.; Ordóñez, J.G.; McClintock, P.V.E.; Morillo, M.; Stein, N.D. Dispersion of the Prehistory Distribution: Analog Experiments and Numerical Results. *Phys. Rev. Lett.* **1998**, *80*, 2273. [CrossRef]
- 29. Gómez-Ordoñez, J.; Casado, J.M.; Morillo, M. Prehistory Problem for Systems Driven by White Noise. Phys. Rev. E 1996, 54, 2125.
- Guardiani, C.; Gibby, W.; Barabash, M.; Luchinsky, D.; Khovanov, I.; McClintock, P. Prehistory Probability Distribution of Ionic Transitions through a Graphene Nanopore. In Proceedings of the 25th International Conference on Noise and Fluctuations (ICNF 2019), Neuchâtel, Switzerland, 18–21 June 2019.
- Anishchenko, V.S.; Khovanov, I.A.; Khovanova, N.A.; Luchinsky, D.G.; McClintock, P.V.E. Noise-Induced Escape from the Lorenz Attractor. *Fluct. Noise Lett.* 2001, 1, L27–L33. [CrossRef]
- 32. Onsager, L.; Machlup, S. Fluctuations and Irreversible Processes. Phys. Rev. 1953, 91, 1505. [CrossRef]
- Machlup, S.; Onsager, L. Fluctuations and Irreversible Process. II. Systems with Kinetic Energy. *Phys. Rev.* 1953, 91, 1512. [CrossRef]
- Zhu, J.; Wang, J.; Gao, S.; Liu, X. Dispersion of the Prehistory Distribution for Non-Gradient Systems. J. Stat. Mech. Theory Exp. 2020, 2020, 23207. [CrossRef]
- Szavits-Nossan, J.; Evans, M.R. Inequivalence of Nonequilibrium Path Ensembles: The Example of Stochastic Bridges. J. Stat. Mech. Theory Exp. 2015, 2015, P12008. [CrossRef]
- 36. Orland, H. Generating Transition Paths by Langevin Bridges. J. Chem. Phys. 2011, 134, 174114. [CrossRef]
- Majumdar, S.N.; Orland, H. Effective Langevin Equations for Constrained Stochastic Processes. J. Stat. Mech. Theory Exp. 2015, 2015, P06039. [CrossRef]
- Delarue, M.; Koehl, P.; Orland, H. Ab Initio Sampling of Transition Paths by Conditioned Langevin Dynamics. J. Chem. Phys. 2017, 147, 152703. [CrossRef] [PubMed]
- 39. Vugmeister, B.E.; Botina, J.; Rabitz, H. Nonstationary Optimal Paths and Tails of Prehistory Probability Density in Multistablestochastic Systems. *Phys. Rev. E* 1997, *55*, 5338. [CrossRef]
- 40. Champagnat, N.; Villemonais, D. General Criteria for the Study of Quasi-Stationarity. arXiv 2017, arXiv:1712.08092.
- Guillin, A.; Nectoux, B.; Wu, L. Quasi-Stationary Distribution for Strongly Feller Markov Processes by Lyapunov Functions and Applications to Hypoelliptic Hamiltonian Systems. 2020. Available online: https://hal.archives-ouvertes.fr/hal-03068461/ (accessed on 31 July 2022).
- Le Bris, C.; Lelievre, T.; Luskin, M.; Perez, D. A Mathematical Formalization of the Parallel Replica Dynamics. *Monte Carlo Methods and Applications*. 2012, 18, 119–146. [CrossRef]
- 43. Bouchet, F.; Reygner, J. Generalisation of the Eyring–Kramers Transition Rate Formula to Irreversible Diffusion Processes. *Proc. Ann. Henri Poincaré* **2016**, *17*, 3499–3532. [CrossRef]
- 44. Millet, A.; Nualart, D.; Sanz, M. Integration by Parts and Time Reversal for Diffusion Processes. *Ann. Probab.* **1989**, *17*, 208–238. [CrossRef]
- 45. Haussmann, U.G.; Pardoux, E. Time Reversal of Diffusions. Ann. Probab. 1986, 14, 1188–1205. [CrossRef]

- 46. Cattiaux, P.; Conforti, G.; Gentil, I.; Léonard, C. Time Reversal of Diffusion Processes under a Finite Entropy Condition. *arXiv* **2021**, arXiv:2104.07708.
- 47. Weinan, E.; Ren, W.; Vanden-Eijnden, E. Minimum Action Method for the Study of Rare Events. *Commun. Pure Appl. Math.* **2004**, 57, 637–656.
- 48. Heymann, M.; Vanden-Eijnden, E. The Geometric Minimum Action Method: A Least Action Principle on the Space of Curves. *Commun. Pure Appl. Math. A J. Issued By Courant Inst. Math. Sci.* 2008, *61*, 1052–1117. [CrossRef]