On the long–time behavior of a perturbed conservative system with degeneracy.

Wenqing Hu * .

Abstract

We consider in this work a model conservative system subject to dissipation and Gaussian–type stochastic perturbations. The original conservative system possesses a continuous set of steady states, and is thus degenerate. We characterize the long–time limit of our model system as the perturbation parameter tends to zero. The degeneracy in our model system carries features found in some partial differential equations related, for example, to turbulence problems.

Keywords: Random perturbations of dynamical system, group symmetry, invariant measure, nonlinear dynamics, irreversibility.

2010 Mathematics Subject Classification Numbers: 37L40, 60H10, 76F20.

1 Introduction.

Many Hamiltonian systems that arise in mechanics, mechanical engineering, as well as hydrodynamics are subject to group symmetry. As an example, in the study of the motion of an ideal incompressible fluid, V.I.Arnold had proposed (see [1], [2], [3], [44]) a beautiful picture that describes the dynamics of ideal incompressible fluid as geodesic flows on the group of all diffeomorphisms of a certain domain (see also the author's related work [32] in this direction). The studies of random perturbations of Hamiltonian systems, or general dynamical systems with symmetry, in particular the long–time dynamics and problems about invariant measures of these systems are of interest (see also the author's related work [31], [18], [17]). Schematically, the general problem can be formulated as follows. We are given a dynamical system

$$\dot{x} = b(x) \tag{1}$$

in an ambient space $x \in M$ (M can be a Riemanian manifold). Usually we assume $b(x)$ preserves the energy. Then we assume that for some group G the system (1) has

*Department of Mathematics and Statistics, Missouri University of Science and Technology (formerly University of Missouri, Rolla). Email: huwen@mst.edu

some symmetry with respect to G. The last sentence about symmetry of the system (1) with respect to the group G is a bit vague and could be understood in many different ways. It can be understood in a strict way so that the group can act on the space M (in particular, it is such case when $G = M$) and the dynamics of (1) is invariant with respect to G–action. It can also be understood as a more "rough" symmetry, in the sense for example that the stable attractors of (1) has equivalent dynamical properties under G–action (in [16] such dynamical property is in the sense of equivalence of logarithmic asymptotics of transition probabilities when we add a small noise to (1), this is related to the notion of "quasi–potential", see [28], [30]). Our goal is to describe the effect of adding a small noise to (1). That is, we study systems of type

$$\dot{\mathcal{X}}^\varepsilon = b(\mathcal{X}^\varepsilon) + \xi^\varepsilon \tag{2}$$

where ξ^ε is a deterministic and/or stochastic perturbation depending on the small parameter(s) $\varepsilon = (\varepsilon_1, ..., \varepsilon_k)$. Recent progresses in this direction have shown that an effective description of the long–time behavior of (2) is the motion on the cone of invariant measures of the unperturbed system (1) (see [16]). Several examples of such description are recently demonstrated in [16], [23], [21], [24], [22].

The above paradigm is only a general scheme. In this work we are interested in studying a model problem that falls under the above general paradigm. Let us consider the following system (see [7], [4, Section 4.4]) corresponding to (1):

$$\begin{cases} dx_t = -x_t y_t dt \ , \\ dy_t = x_t^2 dt \ . \end{cases} \tag{3}$$

A phase picture of system (3) can be seen in Figure 1(a). We see that the whole line Oy_A contains stable equilibriums and the whole line Oy_B contains unstable equilibriums. This is different from the cases considered in [28], [25]. In this case we can understand the symmetry of (3) in a more rough way: the stable and unstable equilibriums are symmetric with respect to shifts in the directions of Oy_A and Oy_B, respectively. The unperturbed system (3) preserves the energy $E(x, y) = x^2 + y^2$. The driving vector field $b(x, y) = (-xy, x^2)$ is degenerate on $x = 0$. Let us add a perturbation to system (3) that consists of a deterministic friction and a random noise:

$$\begin{cases} d\mathcal{X}_t^\varepsilon = -\mathcal{X}_t^\varepsilon \mathcal{Y}_t^\varepsilon dt - \varepsilon \mathcal{X}_t^\varepsilon dt + \sqrt{\varepsilon} dW_t^1 \ , \ \mathcal{X}_0^\varepsilon = x_0 \ , \\ d\mathcal{Y}_t^\varepsilon = (\mathcal{X}_t^\varepsilon)^2 dt - \varepsilon \mathcal{Y}_t^\varepsilon dt + \sqrt{\varepsilon} dW_t^2 \ , \ \mathcal{Y}_0^\varepsilon = y_0 \ . \end{cases} \tag{4}$$

Here W_t^1 and W_t^2 are two independent standard 1–dimensional Brownian motions; the small parameter $\varepsilon > 0$ is the intensity of the friction, and the small parameter $\sqrt{\varepsilon} > 0$ represents the intensity of the noise. System (4) is a two–dimensional nonlinear stochastic equation involving a non–potential force. It is this non–potential force that has the essential effect of creating a line of stable fixed points (attracting line Oy_A)

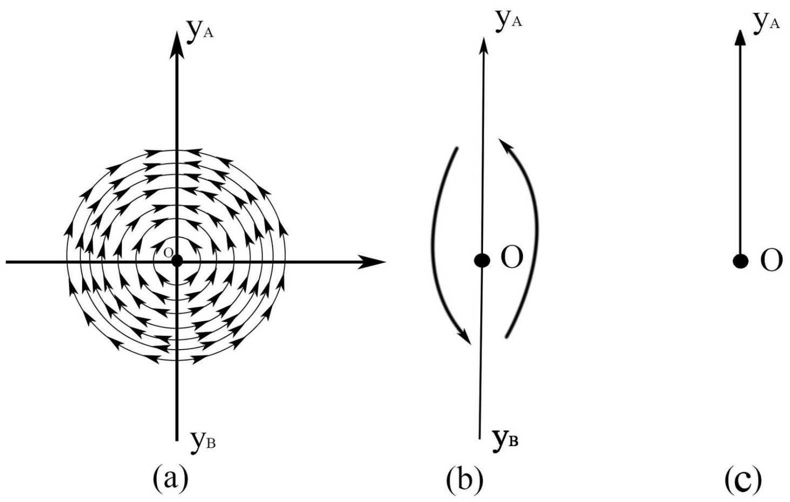

Figure 1: The AB model.

touching a line of unstable fixed points (repelling line Oy_B). In the subsequent text we sometimes refer to this model as the AB–model.

Our goal in this paper is to study the long–time behavior of system (4) as $\varepsilon \downarrow 0$. By further developing results in [16], [23], [21], [24], [22], we will characterize the limiting process as a diffusion process Y_t on the positive–y semi–axis. The limiting diffusion process Y_t behaves as a 2–dimensional radial Bessel process with linear damping, and henceforce we call it a damped 2–d radial Bessel process, abbreviated as *damped–BES(2)* (for Bessel process in arbitrary dimension see [39, Chapter XI, §1]). The origin O is an inaccessible point for damped–BES(2). Diffusion processes on singular 1–dimensional manifolds as the limit of averaging procedure has been considered in [26], [27], among many other literature. The major contribution in our work is that we consider the manifold of unstable equilibria touching the manifold of stable equilibria. This results in non–trivial analysis that leads to our limiting process Y_t as well as the inaccessibility of the origin O. We will describe the limiting Markov diffusion process Y_t via its infinitesimal generator, and we show the weak convergence by making use of tightness and the classical martingale problem method.

In a certain sense, our model problem here differs from the set–up in the classical Freidlin–Wentzell theory (see [25]) in that the point–like asymptotically stable attractor is replaced by a manifold. We can view our limiting process Y_t on Oy_A, the damped–BES(2) process, as a "process–level attractor" of our system. For $\varepsilon > 0$, the dynamics of the system as $\varepsilon \downarrow 0$ corresponds to the "metastable" behavior (see [15]). We will show that under this scenario the "metastable" behavior of the system is characterized by jumps between points on Oy_A and Oy_B.

We are motivated by finite dimensional models for the inviscid stochastic 2–d Navier–Stokes equations written in vorticity form (see [33], [45, Lecture 39])

$$\frac{\partial \omega}{\partial t} + (u \cdot \nabla)\omega - \nu\Delta\omega = \sqrt{\nu}\eta(t,x) \ , \ u = \mathcal{K}\omega \ , \ \omega(0,x) = \omega_0(x) \ , \tag{5}$$

in which $\mathcal{K} = \nabla^{\perp}\Delta^{-1}$ is the Biot–Savart operator, $\eta(t,x)$ is a noise, and the viscosity parameter $\nu \to 0$. An unsolved issue here targets at studying the vanishing noise limit of stationary measures of the 2–d stochastic Navier–Stokes system (see open problem 3 in the last section of the survey [33]). The difficulty there is that one has to put a rather restrictive hypothesis, namely the unperturbed dynamics has to be globally asymptotically stable. To remove this restriction, in the finite dimensional case this problem is rather well–understood, and one can establish the so–called Freidlin–Wentzell asymptotics for stationary measures (see Section 6.4 in [30]). As for stochastic PDEs, similar results can be proved, provided that the global attractor for the unperturbed dynamics has a "regular structure". The latter means that the attractor consists of finitely many steady–states and the heteroclinic orbits joining them. A result in this direction has been proved in [35] for the case of a damped nonlinear wave equation. However, the global attractor for the 2–d Euler system does not have a regular structure, and in fact it has continuous sets of steady states (see [45, Lecture 68]). More generally, systems that arise in hydrodynamics, such as in the context of Euler's equation, typically possess equilibrium points that belong to an infinite dimensional "manifold" of other equilibria. These has been found in experiments (see [42], [43]), in numerical simulations (see [41]), explained using arguments based on statistical mechanics (see [8], [40], [36], [6]), as well as explained theoretically (see [5], [38]). Our system (3) is a very simple finite–dimensional example of such type, in which the attractor is a semi–line Oy_A. When we add a damping to (3), we obtain for fixed $\varepsilon > 0$ the model system (4) without the stochastic noise, which admits only one single attractor O. Of course, the situation will be much more complicated for the Euler and the Navier–Stokes equations. For example, in low dimensions a good example is the famous Lorenz attractor (see [46]). However, a surprising geometric connection is that our system (3) can be viewed as the Euler–Arnold equation (see [44], [2, Appendix 2]) for the group of all affine transformations of a line ℓ (see [37] for more on this group), while the 2–d Euler equation is the Euler–Arnold equation for the group of all diffeomorphisms transforming the domain in which the fluid is moving (see [1] and [2, Appendix 2]). The formulation of our system (3) as the Euler–Arnold equation will be discussed in Section 6.

The paper is organized as follows. In Section 2 we will explain the heuristics of the limiting mechanism. In Section 3 we demonstrate the main convergence theorem as well as its proof. In Section 4 we prove auxiliary lemmas that are needed in Section 3. In Section 5 we describe the dynamics of our model system for small but nonzero $\varepsilon > 0$. In Section 6 we discuss the formulation of our system (3) as the Euler–Arnold equation for the group of all affine transformations of a line. Some remarks and generalizations are provided in Section 7.

2 Heuristic description of the limiting mechanism.

To describe the limiting motion as $\varepsilon \downarrow 0$, we can first do a time rescaling $t \to \dfrac{t}{\varepsilon}$. Let $(X_t^\varepsilon, Y_t^\varepsilon) = (\mathcal{X}_{t/\varepsilon}^\varepsilon, \mathcal{Y}_{t/\varepsilon}^\varepsilon)$. Then we have

$$\begin{cases} dX_t^\varepsilon = -\dfrac{1}{\varepsilon}X_t^\varepsilon Y_t^\varepsilon dt - X_t^\varepsilon dt + dW_t^1 \ , \ X_0^\varepsilon = x_0 \ , \\ dY_t^\varepsilon = \dfrac{1}{\varepsilon}(X_t^\varepsilon)^2 dt - Y_t^\varepsilon dt + dW_t^2 \ , \ Y_0^\varepsilon = y_0 \ . \end{cases} \tag{6}$$

In this way, we see the separation of a "fast" motion which is governed by the non–potential force term, and a "slow" motion which is due to the random perturbation. Due to the effect of the fast motion, starting from anywhere (x_0, y_0) that is not lying on the semi–axis Oy_B, the process $(X_t^\varepsilon, Y_t^\varepsilon)$ will come close to the attracting line Oy_A in a relatively short time. Let π denote this hitting operator, so that we have the following definition.

Definition 2.1. *We define a projection operator $\pi : \mathbb{R}^2\backslash Oy_A \to Oy_A$, or equivalently $y^\pi(x_0, y_0) : \mathbb{R}^2\backslash Oy_A \to \mathbb{R}_+$, such that $\pi(x_0, y_0) = (0, y^\pi(x_0, y_0))$ as follows: when $(x_0, y_0) \in \mathbb{R}^2\backslash(Oy_A \cup Oy_B)$, we set $y^\pi(x_0, y_0) = \lim\limits_{t\to\infty} y(t)$ where $(x(t), y(t))$ is the deterministic flow in (3) with initial condition $(x(0), y(0)) = (x_0, y_0)$; when $(0, y_0)$ in Oy_B (i.e. $y_0 < 0$), we can then naturally extend the operator π onto the Oy_B axis, so that $y^\pi(0, y_0) = y^\pi(|y_0|\sin\kappa, -|y_0|\cos\kappa)$ for some small $\kappa > 0$; finally, we define $y^\pi(0, 0) = 0$.*

In the limit as $\varepsilon \downarrow 0$, the process $(X_t^\varepsilon, Y_t^\varepsilon)$ is pushed by the flow onto Oy_A, and will be close to $\pi(x_0, y_0)$ in short time. There, the Y–component Y_t^ε behaves as a 2–dimensional linearly damped radial Bessel process (damped–BES(2)) on Oy_A:

$$dY_t = \left(\frac{1}{2Y_t} - Y_t\right) dt + dW_t^2 \ , Y_0 = y^\pi(x_0, y_0) \ . \tag{7}$$

Indeed, when Y_t is close to O, the large positive drift term $\dfrac{1}{2Y_t}$ comes from the limit of the positive drift $\dfrac{(X_t^\varepsilon)^2}{\varepsilon}$ in the Y–equation of (6) as $\varepsilon \downarrow 0$ (which is illustrated as Corollary 4.3). This makes the origin O an inaccessible point for Y_t. However, for small $\varepsilon > 0$, the process $(X_t^\varepsilon, Y_t^\varepsilon)$ may still enter a thin strip around the half–line Oy_B through O. Due to the strong Markov property of the process $(X_t^\varepsilon, Y_t^\varepsilon)$, once it enters the domain $\mathbb{R}^2\backslash Oy_B$, it will move along the fast flow to hit somewhere on Oy_A. For any fixed $\varepsilon > 0$, the probability of hitting the level $Y = -a$ for some $a > 0$ before moving along the fast flow and hit somewhere on Oy_A decays to 0 as $\varepsilon \downarrow 0$. As the process Y_t^ε

is closer to the origin O, the positive drift term $\dfrac{(X_t^\varepsilon)^2}{\varepsilon}$ pushes the process Y_t^ε to bounce back to positive y–axis. Thus our limiting Y–process, the damped–BES(2), only lives on the positive Y–axis (see Figure 1(c)).

The above scenario can be roughly seen by considering the radial process $r_t^\varepsilon = \sqrt{(X_t^\varepsilon)^2 + (Y_t^\varepsilon)^2}$. In fact, by applying Itô's formula to (6) we see that

$$\begin{aligned} dr_t^\varepsilon &= \frac{X_t^\varepsilon}{r_t^\varepsilon}\left[\left(-\frac{1}{\varepsilon}X_t^\varepsilon Y_t^\varepsilon - X_t^\varepsilon\right)dt + dW_t^1\right] \\ &\qquad + \frac{Y_t^\varepsilon}{r_t^\varepsilon}\left[\left(\frac{1}{\varepsilon}(X_t^\varepsilon)^2 - Y_t^\varepsilon\right)dt + dW_t^2\right] + \frac{1}{2}\frac{(Y_t^\varepsilon)^2}{(r_t^\varepsilon)^3}dt + \frac{1}{2}\frac{(X_t^\varepsilon)^2}{(r_t^\varepsilon)^3}dt \\ &= \left(\frac{1}{2r_t^\varepsilon} - r_t^\varepsilon\right)dt + dW_t^r \ , \ r_0^\varepsilon = \sqrt{(X_0^\varepsilon)^2 + (Y_0^\varepsilon)^2} \end{aligned} \tag{8}$$

where W_t^r is a standard Brownian motion on $\mathbb{R}$. When the process X_t^ε is pushed by the flow to be close to the Y–axis, we have that Y_t^ε is close to r_t^ε, and thus (8) indicates the limiting Y–dynamics (7).

However, for fixed $\varepsilon > 0$, at a subexponential time scale, excursions of the process $(X_t^\varepsilon, Y_t^\varepsilon)$ moving from Oy_A towards a level set $y = -a$ will be observed. These excursions are directly crossing through a neighborhood of O. Due to the repelling nature of Oy_B and the random perturbation, the process will not strictly lie on Oy_B but it will move along the fast flow and come close to somewhere on Oy_A. This induces jumps from points in Oy_B to points in Oy_A (see Figure 1(b)). At even larger time scale, such as an exponentially long time scale, large deviation effect makes the process $(X_t^\varepsilon, Y_t^\varepsilon)$ move from the attracting line Oy_A to the repelling line Oy_B. Such moves are not through O but are directed motions against the fast flow. Again the instability of Oy_B and the random perturbation will make the process quickly jump back to Oy_A. This induces back and forth jumps between points in Oy_A and those in Oy_B (see Figure 1(b)). As ε becomes smaller, motions of the process $(X_t^\varepsilon, Y_t^\varepsilon)$ to Oy_B and jumping back become more and more rare, and in the limit no more such jumps appear, so that we come to the limiting process Y_t which cannot penetrate through O. Thus as $\varepsilon > 0$ is close to 0, the description of the "metastable" behavior of system (4) involves both a diffusion part and a jump part.

Figure 2 shows sample pathes of the X_t^ε and Y_t^ε processes, as well as the limiting Y–process (driven by the same Brownian motion as the driving Brownian motion for Y_t^ε) starting from $(X, Y) = (0, 2)$ in 15000 steps for stepsize= 0.0001, with all steps rescaled to $[0, 1]$. In Figure 2(a), (b), the red curves are the sample pathes for Y_t, and the blue curves are for sample pathes Y_t^ε when $\varepsilon = 0.1$ (Figure 2(a)) and $\varepsilon = 0.01$ (Figure 2(b)). In Figure 2(c), (d), the black curves are the sample pathes for X_t^ε when $\varepsilon = 0.1$ (Figure 2(c)) and $\varepsilon = 0.01$ (Figure 2(d)). One can see that the process X_t^ε is mainly localized near 0, and the process Y_t^ε behaves similarly as the process Y_t, especially when the

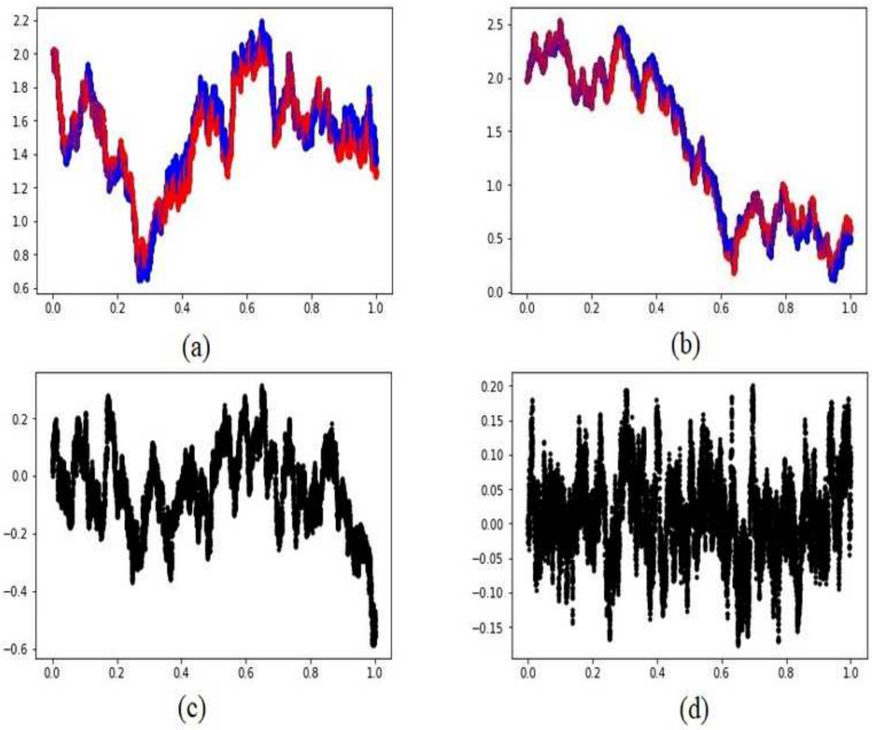

Figure 2: Sample pathes of the X_t^ε and Y_t^ε processes, as well as the limiting Y–process (driven by W_t^2) starting from $(X, Y) = (0, 2)$ in 15000 steps for stepsize= 0.0001, that is rescaled to $[0, 1]$. (a) $\varepsilon = 0.1$; (b) $\varepsilon = 0.01$; the red curves are the sample pathes for Y_t, the blue curves are the sample pathes for Y_t^ε. (c) $\varepsilon = 0.1$; (d) $\varepsilon = 0.01$; the black curves are the sample pathes for X_t^ε.

parameter $\varepsilon > 0$ is small.

Let us also notice that, the cone formed by the set of extremal invariant measures of the unperturbed system (3) consists of both the lines Oy_A and Oy_B. And according to [16] the description of the limiting process shall be given by a Markov process on this cone. Our result is in a sense a specific example of this general paradigm. What we are demonstrating here is that the part Oy_B of this cone is simply inaccessible, and the limiting process just lives on Oy_A. This agrees with the heuristic that Oy_A is the "stable" half–line of equilibriums and Oy_B is the "unstable" half–line of equilibriums.

3 The limiting process and weak convergence theorem.

Let Y_t be defined as the diffusion process on $\mathbb{R}$ with infinitesimal generator given by the operator A and domain of definition $D(A)$ (see [11]). For any continuous function

$f:\mathbb{R}\to\mathbb{R}$ that is twice continuously differentiable in $y\geq 0$ we have

$$Af(y)=\frac{1}{2}\frac{d^2f}{dy^2}(y)+\left(\frac{1}{2y}-y\right)\frac{df}{dy}(y)\ , \quad \text{for all } y>0\ , \tag{9}$$

and

$$Af(O)=\lim_{y\to 0+}Af(y)\ . \tag{10}$$

For $y<0$ we further define

$$Af(y)=0 \text{ for all } y<0\ . \tag{11}$$

The domain of definition of the operator A is given by the set of continuous functions $f:\mathbb{R}\to\mathbb{R}$ such that $f(y)$ are twice continuously differentiable in $y\geq 0$, with the limit of $\dfrac{d^+f}{dy}(y)=\lim\limits_{z\to y, z>y}\dfrac{f(z)-f(y)}{z-y}$ exist and is equal to zero as $y\to 0+$, i.e.

$$\lim_{y\to 0+}\frac{d^+f}{dy}(y)=0\ . \tag{12}$$

By (10) and (12) we infer further that

$$\lim_{y\to 0+}\frac{1}{y}\frac{d^+f}{dy}(y) \tag{13}$$

exists.

The existence of such a process Y_t is guaranteed by the Hille–Yosida theorem (see [14], [34]). The closure $\overline{A|_{D(A)}}$ of the operator A in the space of continuous functions on $\mathbb{R}$ exists and it actually defines a Markov process on $\{y\geq 0\}$, which is a 2–dimensional radial Bessel process with linear damping on $\mathbb{R}_+$, that is inaccessible to the origin O, and it contains isolated points on $\{y<0\}$. Our main theorem can be stated as follows.

Theorem 3.1. *Let $T>0$ and initial condition $(x_0,y_0)\in\mathbb{R}^2$. Then*

(a) For any bounded continuous function $F:\mathbb{R}^2\to\mathbb{R}$ that is uniformly Lipschitz continuous with a Lipschitz constant $Lip(F)<\infty$ we have

$$\lim_{\varepsilon\downarrow 0}\mathbf{E}\left[F(X_T^\varepsilon,Y_T^\varepsilon)-F(0,Y_T^\varepsilon)\right]=0\ . \tag{14}$$

(b) The measures on $\mathbf{C}_{[0,T]}(\mathbb{R})$ induced by the process Y_t^ε converge weakly as $\varepsilon\downarrow 0$ to the measure induced by Y_t with $Y_0=y^\pi(x_0,y_0)$.

Proof. Let $\delta=\delta(\varepsilon)=\varepsilon^\alpha>0$ with $\delta\to 0$ as $\varepsilon\downarrow 0$. We pick $\alpha=\dfrac{1}{10}$. Set $\sigma_0=0$ and

$$\tau_k=\inf\{t\geq\sigma_{k-1}, |Y_t^\varepsilon|=\delta\}\ ,\ \sigma_k=\inf\{t\geq\tau_k, |Y_t^\varepsilon|=2\delta\}\ ,\ k=1,2,....$$

Our proof intuitively goes as follows:

Step 1. We show that if $Y_t^\varepsilon \geq \delta$, then as $\varepsilon \downarrow 0$ the process X_t^ε is very close to the Y–axis. This is proved in Lemma 4.1. We then show in Lemma 4.2 and Corollary 4.3 that as X_t^ε is small, the quantity $\dfrac{(X_t^\varepsilon)^2}{\varepsilon}$ is close to $\dfrac{1}{2Y_t^\varepsilon}$. In particular, this makes the process Y_t^ε behaves close to a 2–dimensional radial Bessel process with linear damping when $Y_t^\varepsilon \geq \delta$.

Step 2. We show that during the time $\tau_k \leq t \leq \sigma_k$ we have $|X_t^\varepsilon| \leq 3\delta$ with high probability. This is because whenever $|X_t^\varepsilon| \geq 2\delta$ the flow (6) with small $\varepsilon > 0$ will quickly bring the particle back to the region $Y \geq \delta$, and during this process the $|X|$–value is less or equal than 3δ. This is done in Lemma 4.4.

Step 3. We show that $\mathbf{P}(Y_{\sigma_k}^\varepsilon = 2\delta) \to 1$ as $\varepsilon \downarrow 0$ and therefore $\delta(\varepsilon) \to 0$. This is because if $Y_t^\varepsilon \leq -1.5\delta$, then the flow of (6) with small $\varepsilon > 0$ will quickly bring the particle back to $Y \geq \delta$, and during this process the Y–coordinate is $\geq -1.99\delta$ with probability $\to 1$ as $\varepsilon \downarrow 0$. This is done in Lemma 4.5.

Step 4. We then estimate $\mathbf{E}(\sigma_k - \tau_k) \lesssim \mathcal{O}(\delta^2)$ in Lemma 4.6. By making use of the fact that $|X_t^\varepsilon|$ will be close to 0 for $\sigma_k \leq t \leq \tau_{k+1}$, we estimate $\mathbf{E}(\tau_{k+1} - \sigma_k) \gtrsim \mathcal{O}(\delta) \to 0$ as $\varepsilon \downarrow 0$ in Lemma 4.7. The asymptotic lower bound for $\mathbf{E}(\tau_{k+1} - \sigma_k)$ provides us with an upper bound on the number of up–crossings $N(\varepsilon) \lesssim \mathcal{O}(\delta^{-1})$ from δ to 2δ before time T. This is done in Lemma 4.8. Combining Lemmas 4.8 and 4.6 we obtain that $N(\varepsilon) \cdot \mathbf{E}(\sigma_k - \tau_k) \to 0$ as $\varepsilon \downarrow 0$.

Steps 1 and 2 together help us to settle (14) so part (a) of this Theorem. To prove part (b) of this Theorem, we shall make use of a modification of Lemma 3.1 in [28, Chapter 8]. This has been used in the works [23], [22], [26], [9], [10], [20], [19]. First, in Lemma 4.9 we show that the family of processes Y_t^ε is tight in $\mathbf{C}_{[0,T]}(\mathbb{R})$. Secondly, we show that for every continuous function $f : \mathbb{R}_+ \to \mathbb{R}$ such that $f \in D(A)$ and every $T > 0$, having bounded derivatives up to the third order, uniformly in the initial condition $(x_0, y_0) \in \mathbb{R}^2$ we have

$$\mathbf{E}_{(x_0,y_0)}\left[f(Y_T^\varepsilon) - f(y^\pi(X_0^\varepsilon, Y_0^\varepsilon)) - \int_0^T Af(Y_t^\varepsilon)dt\right] \to 0 \tag{15}$$

as $\varepsilon \downarrow 0$. The desired convergence in (b) then follows from (15) by the argument using martingale problem formulation of Markov processes (see [13, Chapter 4]). We are left with proving (15). To this end, we decompose

$$\begin{aligned}
& \mathbf{E}\left[f(Y_T^\varepsilon) - f(y^\pi(X_0^\varepsilon, Y_0^\varepsilon)) - \int_0^T Af(Y_t^\varepsilon)dt\right] \\
= & \sum_{k=1}^N \mathbf{E}\left[f(Y_{\tau_k}^\varepsilon) - f(Y_{\sigma_{k-1}}^\varepsilon) - \int_{\sigma_{k-1}}^{\tau_k} Af(Y_t^\varepsilon)dt\right] + \sum_{k=1}^N \mathbf{E}\left[f(Y_{\sigma_k}^\varepsilon) - f(Y_{\tau_k}^\varepsilon) - \int_{\tau_k}^{\sigma_k} Af(Y_t^\varepsilon)dt\right] \\
& \qquad -\mathbf{E}\left[f(Y_{\sigma_N}^\varepsilon) - f(Y_T^\varepsilon) - \int_T^{\sigma_N} Af(Y_t^\varepsilon)dt\right] \\
= & (I) + (II) - (III) .
\end{aligned}$$

Let us first estimate (I). In fact, we can estimate, by Lemma 4.2, that

$$\left|\mathbf{E}\left[f(Y^\varepsilon_{\tau_k})-f(Y^\varepsilon_{\sigma_{k-1}})-\int_{\sigma_{k-1}}^{\tau_k}Af(Y_t^\varepsilon)dt\right]\right|\leq CT\varepsilon^{1-4\alpha}\ .$$

This helps us to conclude, by further making use of Lemma 4.8 that

$$\left|\sum_{k=1}^N\mathbf{E}\left[f(Y^\varepsilon_{\tau_k})-f(Y^\varepsilon_{\sigma_{k-1}})-\int_{\sigma_{k-1}}^{\tau_k}Af(Y_t^\varepsilon)dt\right]\right|\leq CT^2\varepsilon^{1-5\alpha}\to 0$$

as $\varepsilon\downarrow 0$, for $0<\alpha<\dfrac{1}{5}$ say $\alpha=\dfrac{1}{10}$.

To estimate (II), we notice that $Y^\varepsilon_{\sigma_k}=2\delta$ and $Y^\varepsilon_{\tau_k}=\delta$. Thus by using the fact that $f'(0)=0$ we obtain

$$f(Y^\varepsilon_{\sigma_k})-f(0)\approx 4f''(0)\delta^2+\mathcal{O}(\delta^3)\ ,\ f(Y^\varepsilon_{\tau_k})-f(0)\approx f''(0)\delta^2+\mathcal{O}(\delta^3)\ ,$$

so that

$$\left|\mathbf{E}\left[f(Y^\varepsilon_{\sigma_k})-f(Y^\varepsilon_{\tau_k})-\int_{\tau_k}^{\sigma_k}Af(Y_t^\varepsilon)dt\right]\right|\leq C_1|f''(0)|\delta^2+C_2\mathbf{E}(\sigma_k-\tau_k)\ .$$

This combined with the fact that $N(\varepsilon)\cdot\mathbf{E}(\sigma_k-\tau_k)\to 0$ and $N\cdot\delta^2\to 0$ as $\varepsilon\downarrow 0$ from Lemmas 4.8 and 4.6, help us to conclude that $|(II)|\to 0$ as $\varepsilon\downarrow 0$.

Finally it is easy to see that $|(III)|\to 0$ as $\varepsilon\downarrow 0$. Thus (15) is proved. ☐

4 Proof of auxiliary lemmas.

Recall that by (6), we have

$$\begin{aligned} dX_t^\varepsilon&=\left(-\frac{1}{\varepsilon}X_t^\varepsilon Y_t^\varepsilon-X_t^\varepsilon\right)dt+dW_t^1 \quad , \quad X_0^\varepsilon=x_0\ ,\\ dY_t^\varepsilon&=\left(\frac{1}{\varepsilon}(X_t^\varepsilon)^2-Y_t^\varepsilon\right)dt+dW_t^2 \quad , \quad Y_0^\varepsilon=y_0\ .\end{aligned}$$

Lemma 4.1. *For any $\delta=\delta(\varepsilon)$ such that $\delta=\varepsilon^\alpha\to 0$ as $\varepsilon\downarrow 0$ for some $0<\alpha<1$, there exist some $t_0=t_0(\varepsilon)$ which can be picked as $t_0(\varepsilon)=\varepsilon^{(1-\alpha)/2}$, such that as $t\geq t_0(\varepsilon)$ and while $Y_s^\varepsilon\geq\delta$ for $0\leq s\leq t$, we have*

$$\mathbf{E}(X_t^\varepsilon)^2\leq C\varepsilon^{1-\alpha} \tag{16}$$

for some $C>0$.

Proof. Let $Y_s^\varepsilon\geq\delta$ for $0\leq s\leq t$. Let $s\in[0,t]$ and we consider applying Itô's formula to $(X_s^\varepsilon)^2$. In this way, we obtain from (6) that

$$\begin{aligned} d(X_s^\varepsilon)^2 &= 2X_s^\varepsilon dX_s^\varepsilon + (dX_s^\varepsilon)^2 \\ &= 2\left(-\frac{Y_s^\varepsilon}{\varepsilon} - 1\right)(X_s^\varepsilon)^2 ds + 2X_s^\varepsilon dW_s^1 + ds \ . \end{aligned} \tag{17}$$

Therefore taking expectation in (17) we obtain

$$d\mathbf{E}(X_s^\varepsilon)^2 = 2\mathbf{E}\left(-\frac{Y_s^\varepsilon}{\varepsilon} - 1\right)(X_s^\varepsilon)^2 ds + ds \ . \tag{18}$$

As we have $Y_s^\varepsilon \geq \delta$ and $(X_s^\varepsilon)^2 \geq 0$, we can estimate

$$\left(-\frac{Y_s^\varepsilon}{\varepsilon} - 1\right)(X_s^\varepsilon)^2 \leq \left(-\frac{\delta}{\varepsilon} - 1\right)(X_s^\varepsilon)^2 \ ,$$

so that (18) becomes

$$d\mathbf{E}(X_s^\varepsilon)^2 \leq 2\left(-\frac{\delta}{\varepsilon} - 1\right)\mathbf{E}(X_s^\varepsilon)^2 ds + ds \ .$$

Thus

$$\begin{aligned} d\left[e^{2(\frac{\delta}{\varepsilon}+1)s}\mathbf{E}(X_s^\varepsilon)^2\right] &\leq e^{2(\frac{\delta}{\varepsilon}+1)s}\left(2\left(-\frac{\delta}{\varepsilon} - 1\right)\mathbf{E}(X_s^\varepsilon)^2 ds + ds + 2\left(\frac{\delta}{\varepsilon} + 1\right)\mathbf{E}(X_s^\varepsilon)^2 ds\right) \\ &= e^{2(\frac{\delta}{\varepsilon}+1)s} ds \ . \end{aligned}$$

Integrating the above differential inequality in the argument s from 0 to t we see that we have

$$e^{2(\frac{\delta}{\varepsilon}+1)t}\mathbf{E}(X_t^\varepsilon)^2 - \mathbf{E}(X_0^\varepsilon)^2 \leq \frac{1}{2(\frac{\delta}{\varepsilon}+1)}\left(e^{2(\frac{\delta}{\varepsilon}+1)t} - 1\right) \ ,$$

i.e.

$$\mathbf{E}(X_t^\varepsilon)^2 \leq e^{-2(\frac{\delta}{\varepsilon}+1)t}\mathbf{E}(X_0^\varepsilon)^2 + \frac{1}{2(\frac{\delta}{\varepsilon}+1)}(1 - e^{-2(\frac{\delta}{\varepsilon}+1)t}) \ .$$

So finally we obtain the estimate

$$\mathbf{E}(X_t^\varepsilon)^2 \leq e^{-2(\frac{\delta}{\varepsilon}+1)t}\mathbf{E}(X_0^\varepsilon)^2 + \frac{1}{2(\frac{\delta}{\varepsilon}+1)} \ . \tag{19}$$

As we have $\delta = \varepsilon^\alpha$, the above estimate (19) implies that we have

$$\mathbf{E}(X_t^\varepsilon)^2 \leq e^{-2(\varepsilon^{-(1-\alpha)}+1)t}\mathbf{E}(X_0^\varepsilon)^2 + \frac{1}{2}\varepsilon^{1-\alpha} \ .$$

From here we infer that as $t \geq t_0(\varepsilon)$ and $\varepsilon > 0$ sufficiently small we have

$$\mathbf{E}(X_t^\varepsilon)^2 \leq C\varepsilon^{1-\alpha}$$

for some $C > 0$. In particular, we can pick $t_0(\varepsilon) = \varepsilon^{(1-\alpha)/2}$. □

The above estimate (16) cannot provide a precise estimate for $\frac{(X_t^\varepsilon)^2}{\varepsilon}$, which enters as the first term in the right–hand side of the equation for Y_t^ε. In fact, this estimate can be obtained by first noticing the following Lemma.

Lemma 4.2. *There exist some constant $C>0$ so that for small $\varepsilon>0$ and any function $f\in D(A)$ with bounded derivatives up to third order, uniformly in $k=1,2,...,N$ we have*

$$\left|\mathbf{E}\left[f(Y^\varepsilon_{\tau_k})-f(Y^\varepsilon_{\sigma_{k-1}})-\int_{\sigma_{k-1}}^{\tau_k}Af(Y_t^\varepsilon)dt\right]\right|\leq CT\varepsilon^{1-4\alpha}\ . \tag{20}$$

Here the constant $C>0$ depends on the bounds for the derivatives of f.

Proof. Let us assume that there exist uniform constants $M_1>0$, $M_2>0$ and $M_3>0$ such that $|f'(y)|\leq M_1$, $|f''(y)|\leq M_2$ and $|f'''(y)|\leq M_3$. In fact, as $Y^\varepsilon_{\tau_k}=\delta$ and $Y^\varepsilon_{\sigma_k}=2\delta$ for $k=1,2,...,N$, we have

$$\begin{array}{ll}
& \left|\mathbf{E}\left[f(Y^\varepsilon_{\tau_k})-f(Y^\varepsilon_{\sigma_{k-1}})-\int_{\sigma_{k-1}}^{\tau_k}Af(Y_t^\varepsilon)dt\right]-\mathbf{E}\left[f(r^\varepsilon_{\tau_k})-f(r^\varepsilon_{\sigma_{k-1}})-\int_{\sigma_{k-1}}^{\tau_k}Af(r_t^\varepsilon)dt\right]\right| \\
\leq & M_1\left[\mathbf{E}|Y^\varepsilon_{\tau_k}-r^\varepsilon_{\tau_k}|+\mathbf{E}|Y^\varepsilon_{\sigma_{k-1}}-r^\varepsilon_{\sigma_{k-1}}|\right] \\
& +\mathbf{E}\left|\int_{\sigma_{k-1}}^{\tau_k}\left[\left(\frac{1}{2Y_t^\varepsilon}-Y_t^\varepsilon\right)f'(Y_t^\varepsilon)+\frac{1}{2}f''(Y_t^\varepsilon)\right]dt-\int_{\sigma_{k-1}}^{\tau_k}\left[\left(\frac{1}{2r_t^\varepsilon}-r_t^\varepsilon\right)f'(r_t^\varepsilon)+\frac{1}{2}f''(r_t^\varepsilon)\right]dt\right| \\
= & M_1\left[\mathbf{E}|Y^\varepsilon_{\tau_k}-r^\varepsilon_{\tau_k}|+\mathbf{E}|Y^\varepsilon_{\sigma_{k-1}}-r^\varepsilon_{\sigma_{k-1}}|\right] \\
& +\mathbf{E}\left|\int_{\sigma_{k-1}}^{\tau_k}\left[\left(\frac{1}{2Y_t^\varepsilon}-Y_t^\varepsilon\right)f'(Y_t^\varepsilon)+\frac{1}{2}f''(Y_t^\varepsilon)\right]dt-\int_{\sigma_{k-1}}^{\tau_k}\left[\left(\frac{1}{2r_t^\varepsilon}-r_t^\varepsilon\right)f'(r_t^\varepsilon)+\frac{1}{2}f''(r_t^\varepsilon)\right]dt\right| \\
= & M_1\left[\mathbf{E}|Y^\varepsilon_{\tau_k}-r^\varepsilon_{\tau_k}|+\mathbf{E}|Y^\varepsilon_{\sigma_{k-1}}-r^\varepsilon_{\sigma_{k-1}}|\right] \\
& +\mathbf{E}\left|\int_{\sigma_{k-1}}^{\tau_k}\left[\left(\frac{1}{2Y_t^\varepsilon}-Y_t^\varepsilon\right)\left(f'(Y_t^\varepsilon)-f'(r_t^\varepsilon)\right)+\left(\frac{1}{2Y_t^\varepsilon}-Y_t^\varepsilon-\frac{1}{2r_t^\varepsilon}+r_t^\varepsilon\right)f'(r_t^\varepsilon)\right.\right. \\
& \qquad\left.\left.+\frac{1}{2}\left(f''(Y_t^\varepsilon)-f''(r_t^\varepsilon)\right)\right]dt\right| \\
\leq & CM_1\left[\mathbf{E}|Y^\varepsilon_{\tau_k}-r^\varepsilon_{\tau_k}|+\mathbf{E}|Y^\varepsilon_{\sigma_{k-1}}-r^\varepsilon_{\sigma_{k-1}}|\right]+M_2\left(\frac{1}{\delta}+\delta\right)\mathbf{E}\int_{\sigma_{k-1}}^{\tau_k}|Y_t^\varepsilon-r_t^\varepsilon|dt \\
& \qquad +M_1\left(1+\frac{1}{\delta^2}\right)\mathbf{E}\int_{\sigma_{k-1}}^{\tau_k}|Y_t^\varepsilon-r_t^\varepsilon|dt+M_3\mathbf{E}\int_{\sigma_{k-1}}^{\tau_k}|Y_t^\varepsilon-r_t^\varepsilon|dt\ .
\end{array} \tag{21}$$

As we have $Y_t^\varepsilon\geq\delta$ and $r_t^\varepsilon=\sqrt{(X_t^\varepsilon)^2+(Y_t^\varepsilon)^2}\geq Y_t^\varepsilon\geq\delta$ for $\sigma_{k-1}\leq t\leq\tau_k$, we infer that

$$|Y^\varepsilon_{\tau_k}-r^\varepsilon_{\tau_k}|\leq\frac{|(Y_t^\varepsilon)^2-(r_t^\varepsilon)^2|}{|Y_t^\varepsilon+r_t^\varepsilon|}\leq\frac{1}{2\delta}(X_t^\varepsilon)^2\ . \tag{22}$$

From (21) and (22), taking into account Lemma 4.1, we know that, as $\varepsilon>0$ is small, for some constant $M>0$ we have

$$
\begin{aligned}
& \left|\mathbf{E}\left[f(Y^\varepsilon_{\tau_k}) - f(Y^\varepsilon_{\sigma_{k-1}}) - \int_{\sigma_{k-1}}^{\tau_k} Af(Y^\varepsilon_t)dt\right] - \mathbf{E}\left[f(r^\varepsilon_{\tau_k}) - f(r^\varepsilon_{\sigma_{k-1}}) - \int_{\sigma_{k-1}}^{\tau_k} Af(r^\varepsilon_t)dt\right]\right| \\
\leq\ & CM_1\left[\frac{1}{2\delta}\mathbf{E}(X^\varepsilon_{\tau_k})^2 + \frac{1}{2\delta}\mathbf{E}(X^\varepsilon_{\sigma_{k-1}})^2\right] \\
& \qquad + \left[M_2\left(\frac{1}{\delta}+\delta\right) + M_1\left(1+\frac{1}{\delta^2}\right)\frac{1}{2\delta} + M_3\frac{1}{2\delta}\right]\cdot \mathbf{E}\int_{\sigma_{k-1}}^{\tau_k}(X^\varepsilon_t)^2dt \\
\leq\ & \frac{M}{\delta}\left[\mathbf{E}(X^\varepsilon_{\tau_k})^2 + \mathbf{E}(X^\varepsilon_{\sigma_{k-1}})^2\right] + \frac{M}{\delta^3}\cdot\int_0^T \mathbf{E}(X^\varepsilon_t)^2dt \\
\leq\ & C[\varepsilon^{1-2\alpha} + T\varepsilon^{1-4\alpha}] \leq CT\varepsilon^{1-4\alpha} \ .
\end{aligned}
\tag{23}
$$

As we have, by martingale formulation of Markov processes, that

$$
\mathbf{E}\left[f(r^\varepsilon_{\tau_k}) - f(r^\varepsilon_{\sigma_{k-1}}) - \int_{\sigma_{k-1}}^{\tau_k} Af(r^\varepsilon_t)dt\right] = 0 \ ,
$$

we see that the claim (20) follows from (23). □

Corollary 4.3. *For any $1 \leq k \leq N$ and any $\sigma_{k-1} \leq t_1 \leq t_2 \leq \tau_k$, we have*

$$
\left|\mathbf{E}\int_{t_1}^{t_2}\left(\frac{1}{\varepsilon}(X^\varepsilon_t)^2 - \frac{1}{2Y^\varepsilon_t}\right)dt\right| \leq C[\varepsilon^{1-4\alpha} + \varepsilon^{1-6\alpha}(t_2-t_1)] \ . \tag{24}
$$

Proof. Let us consider a function $f \in D(A)$ having bounded derivatives up to the third order. We can apply Itô's formula to the Y–dynamics of (6) and we obtain, for any $\sigma_{k-1} \leq t_1 \leq t_2 \leq \tau_k$, that

$$
\begin{aligned}
& f(Y^\varepsilon_{t_2}) - f(Y^\varepsilon_{t_1}) \\
=\ & \int_{t_1}^{t_2} f'(Y^\varepsilon_t)dY^\varepsilon_t + \frac{1}{2}\int_{t_1}^{t_2} f''(Y^\varepsilon_t)dt \\
=\ & \int_{t_1}^{t_2} f'(Y^\varepsilon_t)\left(\frac{1}{\varepsilon}(X^\varepsilon_t)^2 - Y^\varepsilon_t\right)dt + \int_{t_1}^{t_2} f'(Y^\varepsilon_t)dW^2_t + \frac{1}{2}\int_{t_1}^{t_2} f''(Y^\varepsilon_t)dt \ .
\end{aligned}
$$

This gives

$$
\mathbf{E}\left[f(Y^\varepsilon_{t_2}) - f(Y^\varepsilon_{t_1}) - \int_{t_1}^{t_2} Af(Y^\varepsilon_t)dt\right] = \mathbf{E}\int_{t_1}^{t_2}\left(\frac{1}{\varepsilon}(X^\varepsilon_t)^2 - \frac{1}{2Y^\varepsilon_t}\right)f'(Y^\varepsilon_t)dt \ . \tag{25}
$$

From the proof of Lemma 4.2 we see that the estimate (20) is valid also for the integral from t_1 to t_2. In fact, a finer estimate can be obtained by improving (23) via the estimate

$$
\mathbf{E}\int_{t_1}^{t_2}(X^\varepsilon_t)^2dt \leq C\varepsilon^{1-2\alpha}(t_2-t_1) \ .
$$

So

$$
\left|\mathbf{E}\left[f(Y^\varepsilon_{t_2}) - f(Y^\varepsilon_{t_1}) - \int_{t_1}^{t_2} Af(Y^\varepsilon_t)dt\right]\right| \leq C[\varepsilon^{1-2\alpha} + \varepsilon^{1-4\alpha}(t_2-t_1)] \ .
$$

Thus by (25) we see that

$$\left|\mathbf{E}\int_{t_1}^{t_2}\left(\frac{1}{\varepsilon}(X_t^\varepsilon)^2-\frac{1}{2Y_t^\varepsilon}\right)f'(Y_t^\varepsilon)dt\right|\leq C[\varepsilon^{1-2\alpha}+\varepsilon^{1-4\alpha}(t_2-t_1)]\ .$$

We can pick a function $f\in D(A)$ with bounded derivatives up to third order, such that $f'(y)\geq\delta^2$ for $y\geq\delta$. From here we derive (24). □

Lemma 4.4. *For any $\delta=\delta(\varepsilon)$ such that $\delta=\varepsilon^\alpha\to 0$ as $\varepsilon\downarrow 0$ for $\alpha=\frac{1}{10}$, for any initial condition $|X_0^\varepsilon|\geq 2\delta$, the flow will quickly bring the particle back to the region $Y\geq\delta$, and during this process the $|X|$–value is less or equal than 3δ. In particular, this implies that $\mathbf{P}\left(|X_t^\varepsilon|\leq 3\delta \text{ for } 0\leq t\leq T\right)\to 1$ as $\varepsilon\downarrow 0$.*

Proof. Let us introduce the angular variable $\theta_t^\varepsilon=\arctan\left(\dfrac{Y_t^\varepsilon}{X_t^\varepsilon}\right)$. Here we take the principal branch of the function $\tan\theta$ as $\theta\in\left[-\dfrac{\pi}{2},\dfrac{\pi}{2}\right]$. Due to symmetry of the system (6) with respect to the Y–axis, if the point $\theta\in\left[\dfrac{\pi}{2},\dfrac{3\pi}{2}\right]$, then we can equivalently consider $\widetilde{\theta}=\pi-\theta$ as a replacement of θ. In this way, if $\theta_t^\varepsilon=\dfrac{\pi}{2}$, then the diffusion particle is on the Oy_A axis, and if $\theta_t^\varepsilon=-\dfrac{\pi}{2}$, then the diffusion particle is on the Oy_B axis. Let us apply Itô's formula from (6) to θ_t^ε and we obtain

$$\begin{aligned} d\theta_t^\varepsilon &= -\frac{Y_t^\varepsilon}{(X_t^\varepsilon)^2+(Y_t^\varepsilon)^2}\left(-\frac{1}{\varepsilon}X_t^\varepsilon Y_t^\varepsilon dt - X_t^\varepsilon dt + dW_t^1\right)\\ &\quad+\frac{X_t^\varepsilon}{(X_t^\varepsilon)^2+(Y_t^\varepsilon)^2}\left(\frac{1}{\varepsilon}(X_t^\varepsilon)^2dt - Y_t^\varepsilon dt + dW_t^2\right)\\ &\quad-\frac{1}{2}\left(\frac{2X_t^\varepsilon Y_t^\varepsilon}{[(X_t^\varepsilon)^2+(Y_t^\varepsilon)^2]^2}dt-\frac{2X_t^\varepsilon Y_t^\varepsilon}{[(X_t^\varepsilon)^2+(Y_t^\varepsilon)^2]^2}dt\right)\\ &= \frac{1}{\varepsilon}X_t^\varepsilon dt+\frac{-Y_t^\varepsilon dW_t^1+X_t^\varepsilon dW_t^2}{(X_t^\varepsilon)^2+(Y_t^\varepsilon)^2}\\ &= \frac{1}{\varepsilon}X_t^\varepsilon dt+\frac{1}{(r_t^\varepsilon)^2}dW_t^\theta\ . \end{aligned}\tag{26}$$

Here W_t^θ is another standard Brownian motion on $\mathbb{R}$. Comparing (26) with (8) we see that we have the system

$$\left\{\begin{aligned} d\theta_t^\varepsilon &= \frac{1}{\varepsilon}X_t^\varepsilon dt+\frac{1}{(r_t^\varepsilon)^2}dW_t^\theta\ ,\theta_0^\varepsilon=\arctan\left(\frac{Y_0^\varepsilon}{X_0^\varepsilon}\right)\ ,\\ dr_t^\varepsilon &= \left(\frac{1}{2r_t^\varepsilon}-r_t^\varepsilon\right)dt+dW_t^r\ ,r_0^\varepsilon=\sqrt{(X_0^\varepsilon)^2+(Y_0^\varepsilon)^2}\ . \end{aligned}\right.\tag{27}$$

The processes W_t^θ and W_t^r are two driving standard Brownian motions on $\mathbb{R}$. Set the slow time clock $\mathrm{t}=(\delta/\varepsilon)t$ and let us consider a time–rescaled pair of processes $\Theta_\mathrm{t}^\varepsilon=\theta_{(\varepsilon/\delta)\mathrm{t}}^\varepsilon$ and $R_\mathrm{t}^\varepsilon=r_{(\varepsilon/\delta)\mathrm{t}}^\varepsilon$. Then the stochastic differential equations satisfied by $(\Theta_\mathrm{t}^\varepsilon,R_\mathrm{t}^\varepsilon)$ are given by

$$\begin{cases} d\Theta_{\mathrm{t}}^{\varepsilon} = \dfrac{X_{(\varepsilon/\delta)\mathrm{t}}^{\varepsilon}}{\delta}d\mathrm{t} + \sqrt{\dfrac{\varepsilon}{\delta}} \cdot \dfrac{1}{(R_{\mathrm{t}}^{\varepsilon})^2}dW_{\mathrm{t}}^{\theta} \ , \ \Theta_0^{\varepsilon} = \theta_0^{\varepsilon} \\ dR_{\mathrm{t}}^{\varepsilon} = \dfrac{\varepsilon}{\delta}\left(\dfrac{1}{2R_{\mathrm{t}}^{\varepsilon}} - R_{\mathrm{t}}^{\varepsilon}\right)d\mathrm{t} + \sqrt{\dfrac{\varepsilon}{\delta}}dW_{\mathrm{t}}^{r} \ , R_0^{\varepsilon} = r_0^{\varepsilon} \ . \end{cases} \tag{28}$$

Without loss of generality, let us start the process from some $(X_0^\varepsilon, Y_0^\varepsilon)$ such that $X_0^\varepsilon \geq 2\delta$ and $Y_0^\varepsilon \leq \delta$. In this case we have $R_0^\varepsilon \geq 2\delta$. Consider the stopping time

$$T_X^\varepsilon = \inf\{t \geq 0 : X_t^\varepsilon \leq \delta\} \tag{29}$$

and let $\mathrm{T}_X^\varepsilon = (\delta/\varepsilon)T_X^\varepsilon$. We see that for $t \in [0, T_X^\varepsilon]$ we have $X_t^\varepsilon \geq \delta$ and thus $\dfrac{X_t^\varepsilon}{\delta} \geq 1$. We pick $\delta = \varepsilon^\alpha$ with $\alpha = \dfrac{1}{10}$. Since $\sqrt{\dfrac{\varepsilon}{\delta}} = \varepsilon^{9/20}$ and $\dfrac{\varepsilon}{\delta^2} = \varepsilon^{4/5} = \varepsilon^{16/20}$, it is seen from the R–equation in (28) that for finite t,

$$\mathbf{P}\left(|R_{\mathrm{t}}^\varepsilon - 2\delta| \leq C\varepsilon^{9/20}\right) = 1 \ . \tag{30}$$

In this case, $\sqrt{\dfrac{\varepsilon}{\delta}} \cdot \dfrac{1}{(R_{\mathrm{t}}^\varepsilon)^2} \sim \mathcal{O}\left(\dfrac{\varepsilon^{1/2}}{\delta^2}\right) = \mathcal{O}(\varepsilon^{1/2-1/5}) = \mathcal{O}(\varepsilon^{3/10})$. Therefore, the Θ–equation in (28) can be viewed as a perturbation of the dynamical equation

$$d\Theta_{\mathrm{t}} = \frac{X_{(\varepsilon/\delta)\mathrm{t}}^\varepsilon}{\delta}d\mathrm{t} \ , \ \Theta_0 = \theta_0^\varepsilon \ , \tag{31}$$

such that

$$\mathbf{P}\left(|\Theta_{\mathrm{t}}^\varepsilon - \Theta_{\mathrm{t}}| \leq C\varepsilon^{3/10}\right) = 1 \tag{32}$$

in finite t. From (29), (30), (31) and (32) we know that T_X^ε is finite, and thus $T_X^\varepsilon \sim \mathcal{O}(\varepsilon^{9/10})$ and $Y_{T_X^\varepsilon}^\varepsilon \geq \delta$. From here we know that whenever $X_0^\varepsilon \geq 2\delta$, the flow will quickly bring the particle to the region $Y \geq \delta$, and during this process $X_t^\varepsilon \leq 3\delta$. Thus we see that with high probability, we have $X_t^\varepsilon \leq 3\delta$. The other–side estimate $X_t^\varepsilon \geq -3\delta$ is obtained in a same fashion. □

Lemma 4.5. *For any $\delta = \delta(\varepsilon)$ such that $\delta = \varepsilon^\alpha \to 0$ as $\varepsilon \downarrow 0$ for $\alpha = \dfrac{1}{10}$, for any initial condition $Y_0^\varepsilon \leq -1.5\delta$, the flow will quickly bring the particle back to the region $Y \geq \delta$, and during this process the Y–coordinate is $\geq -1.99\delta$ with probability $\to 1$ as $\varepsilon \downarrow 0$.*

Proof. This is proved in the same way as the proof for Lemma 4.4. □

Lemma 4.6. *We have $\mathbf{E}(\sigma_k - \tau_k) \leq C\delta^2 \to 0$ as $\varepsilon \downarrow 0$ for some constant $C > 0$.*

Proof. Let us introduce the auxiliary OU–process

$$d\widehat{Y}_t = -\widehat{Y}_t dt + dW_t^2 \ , \ \widehat{Y}_0 = Y_0^\varepsilon \ . \tag{33}$$

By Lemma 4.5, we know that as ε is small, with probability close to 1 we have $Y^\varepsilon_{\tau_k} = 2\delta$. Taking this into account, as we have $\dfrac{(X_t^\varepsilon)^2}{\varepsilon} \geq 0$, we can estimate by comparison that

$$\mathbf{E}(\sigma_k - \tau_k) \leq \mathbf{E}\left(\sigma | \widehat{Y}_\sigma = 2\delta\right) \ .$$

Here σ is the first time that the OU–process $\widehat{Y}_t$ starting from $\widehat{Y}_0 = \delta$ hits $Y = \pm 2\delta$. As we have

$$\begin{aligned}
\mathbf{E}\sigma & = \mathbf{E}\left(\sigma | \widehat{Y}_\sigma = 2\delta\right)\mathbf{P}(\widehat{Y}_\sigma = 2\delta) + \mathbf{E}\left(\sigma | \widehat{Y}_\sigma = -2\delta\right)\mathbf{P}(\widehat{Y}_\sigma = -2\delta) \\
& \geq \mathbf{E}\left(\sigma | \widehat{Y}_\sigma = 2\delta\right)\mathbf{P}(\widehat{Y}_\sigma = 2\delta) \\
& = \frac{3}{4}\mathbf{E}\left(\sigma | \widehat{Y}_\sigma = 2\delta\right) \ ,
\end{aligned}$$

we can further estimate

$$\mathbf{E}(\sigma_k - \tau_k) \leq \frac{4}{3}\mathbf{E}\sigma \ . \tag{34}$$

We denote $u(\delta) = \mathbf{E}\sigma$. By the standard theory of stochastic differential equations we know that $u(y)$, $y \in [-2\delta, 2\delta]$ is the solution to the ODE

$$\begin{cases} -yu'(y) + \dfrac{1}{2}u''(y) = -1 \ , \\ u(2\delta) = u(-2\delta) = 0 \ . \end{cases}$$

Solving the above ODE system, we obtain that

$$u(y) = -2\int_{-2\delta}^y e^{z^2}dz \int_{-2\delta}^z e^{-u^2}du + 2\int_{-2\delta}^y e^{z^2}dz \frac{\displaystyle\int_{-2\delta}^{2\delta} e^{z^2}dz \int_{-2\delta}^z e^{-u^2}du}{\displaystyle\int_{-2\delta}^{2\delta} e^{z^2}dz} \ .$$

It is easy to see that as $y \in [-2\delta, 2\delta]$ we have $0 \leq \dfrac{\int_{-2\delta}^y e^{z^2}dz}{\int_{-2\delta}^{2\delta} e^{z^2}dz} \leq 1$. Thus

$$0 \leq u(\delta) \leq \int_\delta^{2\delta} e^{z^2}dz \int_{-2\delta}^z e^{-u^2}du \ .$$

In particular, this implies that $u(\delta) \leq C\delta^2$ for some $C > 0$. Taking into account (34), we obtain the statement of this Lemma. □

Lemma 4.7. *We have $\mathbf{E}(\tau_{k+1} - \sigma_k) \geq C\delta$ as $\varepsilon \downarrow 0$ for some constant $C > 0$.*

Proof. Recall that the Y–equation in (6) has the form

$$dY_t^\varepsilon = \left(\frac{1}{\varepsilon}(X_t^\varepsilon)^2 - Y_t^\varepsilon\right)dt + dW_t^2 \ .$$

Thus by comparison, we know that

$$Y_t^\varepsilon \geq \widehat{Y}_t \ ,$$

in which $\widehat{Y}_t$ is an OU–process defined by

$$d\widehat{Y}_t = -\widehat{Y}_t dt + dW_t^2 \ , \widehat{Y}_0 = Y_0^\varepsilon \ .$$

From here, we know that we have

$$\mathbf{E}(\tau_{k+1} - \sigma_k) \geq \mathbf{E}\tau \ ,$$

where τ is the first time that the process $\widehat{Y}_t$ starting from 2δ hits δ.

Set $u(2\delta) = \mathbf{E}\tau$. From the standard theory of stochastic differential equations we infer that $u(y)$, $y \in [\delta, \infty)$ is the solution to the ODE

$$\begin{cases} -yu'(y) + \dfrac{1}{2}u''(y) = -1 \ , \\ u(\delta) = u(\infty) = 0 \ . \end{cases}$$

Solving the above ODE system, we obtain, for $y \in [\delta, \infty)$, that $u(y) = \lim\limits_{M\to\infty} u_M(y)$, where

$$u_M(y) = -2\int_M^y e^{z^2} dz \int_M^z e^{-u^2} du + 2\int_M^y e^{z^2} dz \frac{\displaystyle\int_M^\delta e^{z^2} dz \int_M^z e^{-u^2} du}{\displaystyle\int_M^\delta e^{z^2} dz} \ .$$

Again, as $M \to \infty$ we have $\lim\limits_{M\to\infty} \dfrac{\displaystyle\int_M^y e^{z^2} dz}{\displaystyle\int_M^\delta e^{z^2} dz} = 1$. Thus in the limit we have

$$\mathbf{E}\tau = u(2\delta) = 2\int_\delta^{2\delta} e^{z^2} dz \int_z^\infty e^{-u^2} du \geq 2\delta e^{\delta^2} \int_{2\delta}^\infty e^{-u^2} du \geq C\delta$$

for some constant $C > 0$. □

Lemma 4.8. *The number of up–crossings $N(\varepsilon)$ from δ to 2δ before time T has the asymptotic $N(\varepsilon) \leq CT\delta^{-1}$ for some constant $C > 0$.*

Proof. This follows from Lemma 4.7. □

Lemma 4.9. *The process Y_t^ε is weakly compact in $\mathbf{C}_{[0,T]}(\mathbb{R})$.*

Proof. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be the probability space for Y_t^ε, $0 \leq t \leq T$, such that for any $\omega \in \Omega$ the sample path $Y_t^\varepsilon(\omega)$, $0 \leq t \leq T$ is a trajectory in $\mathbf{C}_{[0,T]}(\mathbb{R})$. We would like to show that from any sequence $\varepsilon_k \downarrow 0$, $k = 1, 2, ...$ as $k \to \infty$ one can extract a further subsequence $\varepsilon_{k_j} \downarrow 0$, $j = 1, 2, ...$ as $j \to \infty$ such that for any bounded continuous functional F on $\mathbf{C}_{[0,T]}(\mathbb{R})$ we have

$$\mathbf{E}F(Y_t^{\varepsilon_{k_j}}(\omega)) \to \mathbf{E}F(Y_t^0(\omega)) \tag{35}$$

for some $j \to \infty$ and some random element Y_t^0 in $\mathbf{C}_{[0,T]}(\mathbb{R})$. Here $\mathbf{E}$ is the expectation with respect to $\mathbf{P}$.

Unlike any of the previous Lemmas, here we will pick some *fixed* $\delta > 0$. It is easy to see that if we replace $\delta = \varepsilon^\alpha$ by a fixed δ, then Lemmas 4.5, 4.6, 4.7 remain valid (The stopping times σ_k and τ_k can also be defined in a same way as for $\delta = \varepsilon^\alpha$), while the estimate (16) in Lemma 4.1 shall be modified into

$$\mathbf{E}(X_t^\varepsilon)^2 \leq C\frac{\varepsilon}{\delta} \ . \tag{36}$$

Henceforce we will make use of Lemmas 4.5, 4.6, 4.7 in below by directly adapting it to a fixed $\delta > 0$.

Let for any small $\varepsilon > 0$ the family of sample pathes

$$\Omega_{\text{bad}}^{\varepsilon,\delta} = \{\omega : \min_{0\leq t\leq T} Y_t^\varepsilon(\omega) \leq -2\delta\} \ . \tag{37}$$

By Lemma 4.5 we know that $\mathbf{P}(\Omega_{\text{bad}}^{\varepsilon,\delta}) \to 0$ as $\varepsilon \downarrow 0$.

Let us introduce a new probability measure $\widehat{\mathbf{P}}$ on $(\Omega, \mathcal{F}, \mathbf{P})$ as follows. For any event $A \in \mathcal{F}$ we define

$$\widehat{\mathbf{P}}(A) = \frac{\mathbf{P}(A\backslash\Omega_{\text{bad}}^{\varepsilon,\delta})}{\mathbf{P}(\Omega\backslash\Omega_{\text{bad}}^{\varepsilon,\delta})} \ . \tag{38}$$

Let the corresponding expectation be defined by $\widehat{\mathbf{E}}$. As we have $\mathbf{P}(\Omega_{\text{bad}}^{\varepsilon,\delta}) \to 0$ as $\varepsilon \downarrow 0$, we have that $\widehat{\mathbf{E}}X \to \mathbf{E}X$ for any random variable X as $\varepsilon \downarrow 0$. From here we see that to show (35) it suffices to show that

$$\widehat{\mathbf{E}}F(Y_t^{\varepsilon_{k_j}}(\omega)) \to \widehat{\mathbf{E}}F(Y_t^0(\omega)) \tag{39}$$

for some $j \to \infty$ and some random element Y_t^0 in $\mathbf{C}_{[0,T]}(\mathbb{R})$. We then understand (39) is just saying that Y_t^ε is weakly–compact under $\widehat{\mathbf{P}}$. We will then make use of Lemma 5.1 in [29]. In fact, Lemma 5.1 in [29] indicates that in order to show weak–compactness of the family of sample paths in Y_t^ε in $\mathbf{C}_{[0,t]}(\mathbb{R})$ under the measure $\widehat{\mathbf{P}}$, it suffices to show, for each $\delta > 0$, weak–compactness of the family of sample paths $\widetilde{Y}_t^{\varepsilon,\delta}$, where $\widetilde{Y}_t^{\varepsilon,\delta} = Y_t^\varepsilon$ for $\sigma_{k-1} \leq t \leq \tau_k$, $k = 1, 2, ..., N$ and

$$\widetilde{Y}_t^{\varepsilon,\delta} = \delta\frac{\tau_k - t}{\tau_k - \sigma_k} + 2\delta\frac{t - \sigma_k}{\tau_k - \sigma_k}$$

for $\tau_k \leq t \leq \sigma_k$. This is because we have $|Y_t^\varepsilon(\omega) - \widetilde{Y}_t^{\varepsilon,\delta}(\omega)| \leq 4\delta$ for each $\delta > 0$ on $\omega \in \Omega\backslash\Omega_{\text{bad}}^{\varepsilon,\delta}$.

By the classical Prokhorov's theorem, to show weak–compactness of the process $\widetilde{Y}_t^{\varepsilon,\delta}$, it suffices to check tightness of the family of processes $\widetilde{Y}_t^{\varepsilon,\delta}$, $0 \leq t \leq T$. Since $\widetilde{Y}_t^{\varepsilon,\delta}$ is a linear interpolation between $\tau_k \leq t \leq \sigma_k$, we just have to check that, for any $\sigma_{k-1} \leq s_1 \leq s_2 \leq \tau_k$ so that $|s_2 - s_1|$ is small,

$$\widehat{\mathbf{E}}|\widetilde{Y}_{s_2}^{\varepsilon,\delta} - \widetilde{Y}_{s_1}^{\varepsilon,\delta}|^a \leq C|s_1 - s_2|^{1+b} \ , \tag{40}$$

for some $a, b > 0$ and $C > 0$. Since $\widetilde{Y}_s^{\varepsilon,\delta} = Y_s^\varepsilon$ for $\sigma_{k-1} \leq s \leq \tau_k$, and $\mathbf{P}(\Omega_{\text{bad}}^{\varepsilon,\delta}) \to 0$ as $\varepsilon \downarrow 0$, we just have to check (41) for Y_s^ε and $\widehat{\mathbf{E}}$ replaced by $\mathbf{E}$, i.e.

$$\mathbf{E}|Y_{s_2}^\varepsilon - Y_{s_1}^\varepsilon|^a \leq C|s_1 - s_2|^{1+b} \ . \tag{41}$$

Notice that, for any $\sigma_{k-1} \leq s_1 \leq s_2 \leq \tau_k$, we have

$$Y_{s_2}^\varepsilon - Y_{s_1}^\varepsilon = \frac{1}{\varepsilon}\int_{s_1}^{s_2} (X_s^\varepsilon)^2 ds - \int_{s_1}^{s_2} Y_s^\varepsilon ds + (W_{s_2}^2 - W_{s_1}^2) \ . \tag{42}$$

From here, we see that (41) follows from (36).

□

5 "Metastable" behavior of the system as $\varepsilon \downarrow 0$.

The previous section considered the case when $\varepsilon \downarrow 0$. In this case, one can roughly understand that the coupled process $(X_t^\varepsilon, Y_t^\varepsilon)$ converges weakly to $(0, Y_t)$. We can then let $t \to \infty$, so that the damped–BES(2) process Y_t in (8) converges to an invariant measure μ^Y on Oy_A. In this case, ignoring the topology with respect to which we speak about convergence, one can say very vaguely that

$$\lim_{t\to\infty} \lim_{\varepsilon\downarrow 0} (X_t^\varepsilon, Y_t^\varepsilon) = (0 \ , \ \mu^Y \text{ on } Oy_A) \ .$$

It is in this sense that we can understand the measure μ^Y on Oy_A as a global "attractor" of our system $(X_t^\varepsilon, Y_t^\varepsilon)$. One can also consider the case when the two limits are inverted, namely for any given measurable set $\Gamma \subseteq \mathbb{R}^2$ we have the convergence of the form

$$\lim_{\varepsilon\downarrow 0} \lim_{t\to\infty} \mathbf{P}\left((X_t^\varepsilon, Y_t^\varepsilon) \in \Gamma\right) = \mu_0(\Gamma) \ .$$

The limiting measure $\mu_0(\Gamma)$ has been studied in [7] via invariant measure and Kolmogorov (Fokker–Plank) equation, and has been shown to concentrate on Oy_A. In the classical theory regarding random perturbations of dynamical systems (see [30, Section 6.6]), one is interested in considering the above two limits in a coordinated way. Namely we consider the case when $t = t(\varepsilon) \to \infty$ as $\varepsilon \downarrow 0$, and the asymptotic distribution of $(X_{t(\varepsilon)}^\varepsilon, Y_{t(\varepsilon)}^\varepsilon)$. In the classical case such as those demonstrated in [28], [30], the ω–limit sets of the unperturbed system consists of isolated compactum. In this case, if $t(\varepsilon)$

increases sufficiently slowly, then over time $t(\varepsilon)$ the trajectory of $(X^\varepsilon_{t(\varepsilon)}, Y^\varepsilon_{t(\varepsilon)})$ cannot move far from that stable compactum in whose domain of attraction the initial point is. Over larger time intervals there are passages from the neighborhood of this compactum to neighborhoods of others: first to the "closest" compactum (in the sense of the action functional) and then to more and more "far away" ones. Such a phenomenon has been quantitatively characterized as the "metastable" behavior of the system.

The particular feature of the system (4) that we consider here has been in that the unperturbed system admits a continuum of stable attractors. At the level of time–rescaled process (6), this leads to possible "jumps" of $(X^\varepsilon_{t(\varepsilon)}, Y^\varepsilon_{t(\varepsilon)})$ between Oy_A and Oy_B. To illustrate this, let us imagine that we start our process $(X^\varepsilon_t, Y^\varepsilon_t)$ in (6) from $(X^\varepsilon_0, Y^\varepsilon_0)$ such that $Y^\varepsilon_0 \geq 0$.

As ε is small, in very short time $\sim \mathcal{O}(\varepsilon)$, the process $(X^\varepsilon_t, Y^\varepsilon_t)$ first comes close to the Y–axis along the deterministic flow, and it hits a neighborhood of $(0, y^\pi(X^\varepsilon_0, Y^\varepsilon_0))$ [1]. For any $a > 0$, let the stopping time

$$T(a;\varepsilon) = \inf\{t \geq 0; Y^\varepsilon_t \leq -a\} \ . \tag{43}$$

We then define

$$p(a,t;\varepsilon) = \mathbf{P}_{(X^\varepsilon_0, Y^\varepsilon_0)}\left(T(a;\varepsilon) \leq t\right) \tag{44}$$

to be the probability that the trajectory $\{(X^\varepsilon_s, Y^\varepsilon_s)\}_{0\leq s\leq t}$ ever reached below $Y = -a$ on the Oy_B axis. By Lemma 4.5, we have that $p(a,t;\varepsilon) \to 0$ as $\varepsilon \downarrow 0$. We set $t(\varepsilon) \sim \dfrac{1}{p(a,t;\varepsilon)} \to \infty$ as $\varepsilon \downarrow 0$. Then we see that at time scale $\sim t(\varepsilon)$ the process $(X^\varepsilon_{t(\varepsilon)}, Y^\varepsilon_{t(\varepsilon)})$ may demonstrate an excursion to $Y \leq -a$. By combining Lemma 4.1 and the instability of the flow near Oy_B, we see that this excursion happens along the Y–axis and will hit in a neighborhood of $(0, -a)$. In fact, within the half space for $Y > 0$ the process $(X^\varepsilon_t, Y^\varepsilon_t)$ will be pushed by the deterministic flow to be close to the Y–axis. When the excursion diffuses to the half–space with $Y < 0$ but $|X| \neq 0$, the deterministic flow will quickly bring the process $(X^\varepsilon_t, Y^\varepsilon_t)$ back to the half–space with positive Y–value. Therefore the excursion to $(0, -a)$ within the half–space for $Y < 0$ should happen along the Y–axis. At time $t \sim t(\varepsilon)$, the process $(X^\varepsilon_t, Y^\varepsilon_t)$ will be close to $(0, -a)$ and is fluctuating in a neighborhood of this point. Due to instability of the flow near Oy_B axis, the process $(X^\varepsilon_t, Y^\varepsilon_t)$ will then be quickly (at time scale $\sim \mathcal{O}(\varepsilon)$) brought back to a neighborhood of $(0, a)$.

Under the above mechanism, as $\varepsilon > 0$ is small, what we actually see is that the process $(X^\varepsilon_t, Y^\varepsilon_t)$, although mostly stays within the half–plane of positive Y–value, being close to the Y–axis, makes rare excursions to $(0, -a)$ along Y–axis, and after that quickly jumps back to $(0, a)$. As $\varepsilon > 0$ becomes smaller and smaller, the excursion to $(0, -a)$ becomes rarer and rarer, so that in the limit $\varepsilon \downarrow 0$, the process $(X^\varepsilon_t, Y^\varepsilon_t)$ will not

[1] Recall the definition of $y^\pi(x_0, y_0)$ in Definition 2.1.

enter Oy_B any more, and we arrive at the "process level stable attractor" $(0, Y_t)$. This characterizes the metastable behavior of the system (6), and when changed back to the slow time, the perturbed system (4).

6 Formulation of the system (3) as the Euler–Arnold equation for the group of all affine transformations of a line.

In a beautiful paper from 1966 (see [1], also [2, Appendix 2] and [44]), V.I.Arnold observed that many basic equations in physics, including the Euler equations of the motion of a rigid body, and also the Euler equations describing the fluid dynamics of an inviscid incompressible fluid, can be viewed (formally, at least) as geodesic flows on a (finite or infinite dimensional) Riemannian manifold G. This Riemannian manifold G is also a Lie group equipped with a right–invariant metric. Equivalently, these geodesic flows can be written as the solution of an ordinary differential equation at the co–tangent space to the origin of the Lie group G (the dual space of the Lie algebra of G), describing the evolution of the angular momentum (more precisely, the pull back of the angular momentum to the origin). Such an ordinary differential equation has been thereafter named *the Euler–Arnold equation*. Below let us first briefly discuss the background of the Euler–Arnold equation in one subsection, and then in another subsection we will formulate our system (3) as the Euler–Arnold equation for the group of all affine transformations of a line.

6.1 Background of the Euler–Arnold equation.

Let G be an n–dimensional real Lie group. Let $\mathfrak{g}$ be its Lie algebra, i.e., the tangent space of G at the identity element e associated with a commutator relation $[,]$. The commutator relation is defined in the standard way: For two tangent vectors ξ and η the Lie bracket is defined as $[\xi, \eta] = \left.\dfrac{\partial^2}{\partial s \partial t}\right|_{s=t=0} e^{t\xi}e^{s\eta}e^{-t\xi}e^{-s\eta}$. In a coordinate dependent language if $e_1, ..., e_n$ be a basis of $\mathfrak{g}$ so that c_{ij}^k are structure constants, then $[e_i, e_j] = \sum\limits_{k=1}^{n} c_{ij}^k e_k$.

Consider the actions of left and right shifts of G on itself:

$$L_g : G \to G \ , \ L_g h = gh \ ; R_g : G \to G \ , \ R_g h = hg \ .$$

The induced maps on the tangent space at every $h \in G$ are

$$L_{g*} : T_h G \to T_{gh} G \ , \ R_{g*} : T_h G \to T_{hg} G \ .$$

Consider the diffeomorphism $R_{g^{-1}} L_g$, which is an inner automorphism of the group G. This diffeomorphism preserves the identity element e and its derivative at the identity element e is the so called *adjoint representation* Ad_g of the group G. That is to say,

$$Ad_g : \mathfrak{g} \to \mathfrak{g} \ , \ Ad_g = (R_{g^{-1}} L_g)_{*e} \ .$$

The mapping Ad_g satisfies $Ad_g[\xi, \eta] = [Ad_g \xi, Ad_g \eta]$, $\xi, \eta \in \mathfrak{g}$ as well as $Ad_{gh} = Ad_g Ad_h$. We could view the mapping Ad as a mapping from the group to the space of linear operators on $\mathfrak{g}$:

$$Ad(g) = Ad_g \ .$$

The derivative of the mapping Ad at the identity element e of the group G is a linear mapping ad from $\mathfrak{g}$ to the space of linear operators on $\mathfrak{g}$. We have

$$ad = Ad_{*e} : \mathfrak{g} \to \mathrm{End}\mathfrak{g} \ , \ ad_\xi = \left.\frac{d}{dt}\right|_{t=0} Ad_{e^{t\xi}} \ .$$

We see that $ad_\xi \eta = \left.\frac{d}{dt}\right|_{t=0} Ad_{e^{t\xi}} \eta = \left.\frac{d}{dt}\right|_{t=0} (R_{e^{-t\xi}} L_{e^{t\xi}})_{*e} \eta = \left.\frac{\partial^2}{\partial t \partial s}\right|_{t=s=0} e^{t\xi} e^{s\eta} e^{-t\xi} e^{-s\eta} = [\xi, \eta]$.

Let us now consider the dual space $\mathfrak{g}^*$ of the Lie algebra $\mathfrak{g}$. The space $\mathfrak{g}^*$ consists of all real linear functionals on $\mathfrak{g}$: $\mathfrak{g}^* = T_e^* G$. Let us denote the pairing of $\xi \in T_g^* G$ and $\eta \in T_g G$ in the cotangent/tangent spaces at $g \in G$ by the bracket

$$(\xi, \eta) \in \mathbb{R} \ , \xi \in T_g^* G \ , \eta \in T_g G \ .$$

It is natural to define the dual operator $Ad_g^* : \mathfrak{g}^* \to \mathfrak{g}^*$ by the identity

$$(Ad_g^* \xi, \eta) = (\xi, Ad_g \eta) \ .$$

The operator Ad_g^* is the *co-adjoint representation* of the group G.

Correspondingly, one can define

$$ad_\xi^* : \mathfrak{g}^* \to \mathfrak{g}^* \ , \xi \in \mathfrak{g} \ , \ ad_\xi^* = \left.\frac{d}{dt}\right|_{t=0} Ad_{e^{t\xi}}^* \ ,$$

such that

$$(ad_\xi^* \eta, \zeta) = (\eta, ad_\xi \zeta) \ , \eta \in \mathfrak{g}^* \ , \zeta \in \mathfrak{g} \ , \xi \in \mathfrak{g} \ .$$

We may denote

$$ad_\xi^* \eta = \{\xi, \eta\} \ , \ \xi \in \mathfrak{g} \ , \ \eta \in \mathfrak{g}^* \ .$$

We have an identity

$$(\{\xi, \eta\}, \zeta) = (\eta, [\xi, \zeta]) \quad \text{for} \quad \xi \in \mathfrak{g} \ , \ \eta \in \mathfrak{g}^* \ , \ \zeta \in \mathfrak{g} \ .$$

Let us turn to coordinate–dependent language. If $e^1, ..., e^n$ is a basis dual to $e_1, ..., e_n$ in $\mathfrak{g}^*$: $(e^i, e_j) = \delta_j^i$. Then we can calculate $\{e_i, e^j\} = \sum_{k=1}^n c_{ik}^j e^k$.

Let $A : \mathfrak{g} \to \mathfrak{g}^*$ be a symmetric and positive definite linear operator: for any $\xi, \eta \in \mathfrak{g}$ we have $(A\xi, \xi) > 0$ and $(A\xi, \eta) = (A\eta, \xi)$. Let $A_g : T_gG \to T_g^*G$ be defined by $A_g\xi = L_g^* A L_{g^{-1}*}\xi$, $\xi \in T_gG$. In mechanical applications the operator A_g gives the *moment of inertia*. Consider a metric on G defined by an inner product

$$\langle \xi, \eta \rangle_g = (A_g\xi, \eta) = (A_g\eta, \xi) = \langle \eta, \xi \rangle_g$$

for $\xi, \eta \in T_gG$. This metric is a *left invariant metric* on G, i.e., $\langle \xi, \eta \rangle_e = \langle L_{g*}\xi, L_{g*}\eta \rangle_g$, and it makes the Lie group G into a Riemannian manifold. We shall denote the corresponding inner product $\langle , \rangle_e$ at $T_eG = \mathfrak{g}$ simply as $\langle , \rangle$. We shall also denote the operator A_e simply as A.

The above introduced inner product also induces an inner product on T_e^*G. Let $\zeta \in T_e^*G$ and $\mu \in T_e^*G$. We can define $\langle \zeta, \mu \rangle = \langle \zeta, \mu \rangle_e = (\zeta, A^{-1}\mu)$. Such an inner product on T_e^*G makes T_e^*G into an inner product space.

Consider a geodesic curve $g = g(t)$ on the group G, with respect to the metric given by $\langle , \rangle_g$. The trajectory $g = g(t)$ complies with *the principle of least action.* The Lagrangian here is the *kinetic energy* $T(t) = E(t) = \dfrac{1}{2}\langle \dot{g}(t), \dot{g}(t) \rangle_{g(t)}$ and the *action* is $S_{0t}(g) = \displaystyle\int_0^t \frac{1}{2}\langle \dot{g}(s), \dot{g}(s) \rangle_{g(s)} ds$. The trajectory $g = g(t)$ is such that the first variation of the action vanishes.

The *angular velocity* is $\omega = \dot{g}$. Let

$$\omega_c = L_{g^{-1}*}\dot{g} \in \mathfrak{g} \ , \ \omega_s = R_{g^{-1}*}\dot{g} \in \mathfrak{g} \ .$$

These are the so called “angular velocity in the body” (ω_c) and “angular velocity in the space” (ω_s).

The *angular momentum* is defined as

$$M = A_g\dot{g} \ .$$

We see that $M \in T_g^*G$. We consider

$$M_c = L_{g^{-1}}^* M \in \mathfrak{g}^* \ , \ M_s = R_{g^{-1}}^* M \in \mathfrak{g}^* \ .$$

These can be viewed as “angular momentum in the body” (M_c) and “angular momentum in the space” (M_s).

The kinetic energy can be rewritten as

$$T = E = \frac{1}{2}\langle \dot{g}, \dot{g} \rangle_g = \frac{1}{2}\langle \omega_c, \omega_c \rangle = \frac{1}{2}(A\omega_c, \omega_c) = \frac{1}{2}(A_g\dot{g}, \dot{g}) = \frac{1}{2}(M_c, \omega_c) = \frac{1}{2}(M, \dot{g}) \ .$$

Theorem 6.1 (Euler's equation)**.** *We have*

$$\frac{dM_c}{dt} = \{\omega_c, M_c\} \ . \tag{45}$$

Proof. The proof of this Theorem can be found in [2, Appendix 2, Theorem 2]. □

Theorem 6.2 (Euler–Arnold equation). *We have*

$$\frac{dM_c}{dt} = \{A^{-1}M_c, M_c\} \ . \tag{46}$$

Proof. We notice that $\omega_c = L_{g^{-1}*}\dot{g} = A^{-1}AL_{g^{-1}*}\dot{g} = A^{-1}L^*_{g^{-1}}L^*_g AL_{g^{-1}*}\dot{g} = A^{-1}L^*_{g^{-1}}A_g g = A^{-1}M_c$. Thus (46) follows from (45). □

One can see that the evolution of the angular momentum in the body M_c is described by an ordinary differential equation (46) which is the *the Euler–Arnold equation*. The dynamics of this equation is an equivalent way of forming the geodesic flows on the Riemannian manifold G.

6.2 Formulation of the system (3) as the Euler–Arnold equation.

Let G be the group of all affine transformations of a line ℓ (see [37]). We can represent G in terms of the following matrices:

$$G = \left\{ g = g_{a,b} = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} ; a > 0, b \in \mathbb{R} \right\} \ .$$

The group multiplication is then just matrix multiplications: $g_{a_2,b_2}g_{a_1,b_1} = g_{a_1a_2,a_2b_1+b_2}$. The inverse is given by $g^{-1}_{a,b} = g_{\frac{1}{a},-\frac{b}{a}}$. The identity element $e = g_{1,0}$.

The Lie algebra

$$T_eG = \left\{ \begin{pmatrix} x & y \\ 0 & 0 \end{pmatrix} ; x, y \in \mathbb{R} \right\} \ .$$

If $g \in G$ and $M \in T_eG$ then $L_{g*}M = gM$ and $R_{g*}M = Mg$ in which the multiplication is understood as matrix multiplications. This is because we have $\left.\frac{d}{dt}\right|_{t=0} \exp(tM) = gM$ and $\left.\frac{d}{dt}\right|_{t=0} \exp(tM)g = Mg$.

Let us use the inner product $\left(\begin{pmatrix} \xi_1 & \xi_2 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} \eta_1 & \eta_2 \\ 0 & 0 \end{pmatrix} \right) = \xi_1\eta_1 + \xi_2\eta_2$ for $\xi, \eta \in T_eG$. In this way we can identify T_eG with T^*_eG. Let $A : \mathbb{R}^2 \to \mathbb{R}^2$ be the identity matrix. We can introduce a metric on G via A: for any $\xi, \eta \in T_eG$ we introduce $\langle \xi, \eta \rangle = (\xi, \eta)$.

Let $g = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$, $\eta = \begin{pmatrix} \eta_1 & \eta_2 \\ 0 & 0 \end{pmatrix}$ and $\xi = \begin{pmatrix} \xi_1 & \xi_2 \\ 0 & 0 \end{pmatrix}$. Then

$$Ad_g\eta = g\eta g^{-1} = \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 & \eta_2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{a} & -\frac{b}{a} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \eta_1 & -b\eta_1 + a\eta_2 \\ 0 & 0 \end{pmatrix} \ .$$

By definition $(Ad_g^*\xi, \eta) = (\xi, Ad_g\eta) = \eta_1\xi_1 - b\eta_1\xi_2 + a\eta_2\xi_2$. Thus we see that $Ad_g^*\xi = \begin{pmatrix} \xi_1 - b\xi_2 & a\xi_2 \\ 0 & 0 \end{pmatrix}$.

For any $\xi = \begin{pmatrix} \xi_1 & \xi_2 \\ 0 & 0 \end{pmatrix} \in T_eG$ we can calculate

$$\exp(t\xi) = \begin{pmatrix} e^{t\xi_1} & \xi_2 f(\xi_1) \\ 0 & 1 \end{pmatrix}$$

where

$$f(\xi_1) = \begin{cases} \dfrac{e^{t\xi_1} - 1}{\xi_1} & \text{if } \xi_1 \neq 0 \ , \\ t & \text{if } \xi_1 = 0 \ . \end{cases}$$

From here it is readily checked that

$$[\xi, \eta] = \frac{d}{dt}\bigg|_{t=0} Ad_{e^{t\xi}}\eta = \frac{d}{dt}\bigg|_{t=0} \begin{pmatrix} \eta_1 & -\dfrac{\xi_2}{\xi_1}(e^{t\xi_1} - 1)\eta_1 + e^{t\xi_1}\eta_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \xi_1\eta_2 - \xi_2\eta_1 \\ 0 & 0 \end{pmatrix} .$$

Moreover, if $\zeta \in T_e^*G$ we have

$$\{\xi, \zeta\} = \frac{d}{dt}\bigg|_{t=0} Ad^*_{e^{t\xi}}\zeta = \frac{d}{dt}\bigg|_{t=0} \begin{pmatrix} \zeta_1 - \dfrac{\xi_2}{\xi_1}(e^{t\xi_1} - 1)\zeta_2 & e^{t\xi_1}\zeta_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -\xi_2\zeta_2 & \xi_1\zeta_2 \\ 0 & 0 \end{pmatrix} . \quad (47)$$

Theorem 6.3. *The Euler–Arnold equation for the group G of all affine transformations of a line ℓ is equivalent to (3).*

Proof. Set $M_c(t) = (M_{c,1}(t), M_{c,2}(t))$, using (47), the Euler–Arnold equation (46) in Theorem 6.2 is given by

$$(\dot{M}_{c,1}(t), \dot{M}_{c,2}(t)) = \{M_c(t), M_c(t)\} = (-M_{c,2}^2(t), M_{c,1}(t)M_{c,2}(t)) \ .$$

Set $x(t) = M_{c,2}(t)$ and $y(t) = -M_{c,1}(t)$, then the above equation is

$$(-\dot{y}, \dot{x}) = (-x^2, -xy) \ ,$$

which is the same as (3). □

We have seen that our unperturbed system (3) is nothing but the Euler–Arnold equation for the group G of all affine transformations of a line ℓ.

7 Remarks and Generalizations.

1. Let us introduce the elliptic operator

$$L_\varepsilon = \frac{1}{\varepsilon}\left(-xy\frac{\partial}{\partial x} + x^2\frac{\partial}{\partial y}\right) - x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{1}{2}\frac{\partial^2}{\partial y^2} \ . \tag{48}$$

The above elliptic operator can be written as

$$L_\varepsilon = \frac{1}{\varepsilon}L_0 + L_1 \ ,$$

in which

$$L_0 = -xy\frac{\partial}{\partial x} + x^2\frac{\partial}{\partial y} \ , \tag{49}$$

and

$$L_1 = -x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial x^2} + \frac{1}{2}\frac{\partial^2}{\partial y^2} \ . \tag{50}$$

In this way, the operator L_0 degenerates on $x = 0$. One can consider a corresponding Cauchy problem

$$\frac{\partial u^\varepsilon}{\partial t} = L_\varepsilon u^\varepsilon \ , \ u^\varepsilon(0,x,y) = f(x,y) \ , \tag{51}$$

where $f(x,y)$ is a bounded continuous function in $(x,y) \in \mathbb{R}^2$. The solution is represented by

$$u^\varepsilon(t,x,y) = \mathbf{E}_{(x,y)} f(X_t^\varepsilon, Y_t^\varepsilon) \ .$$

By our Theorem 3.1, we infer that $\lim\limits_{\varepsilon\downarrow 0} \mathbf{E}_{(x,y)} f(X_t^\varepsilon, Y_t^\varepsilon) = \lim\limits_{\varepsilon\downarrow 0} \mathbf{E}_{(x,y)} f(0, Y_t^\varepsilon) = \mathbf{E}_{(0,y^\pi(x,y))} f(0, Y_t)$. This gives the following

Corollary 7.1. *Let the initial condition $f(x,y)$ be a bounded continuous function of (x,y). Then as $\varepsilon \to 0$ we have $u^\varepsilon(t,x,y) \to u(t, y^\pi(x,y))$ where $u(t,y)$ is the solution of the equation*

$$\frac{\partial u}{\partial t} = \left(\frac{1}{2y} - y\right)\frac{\partial u}{\partial y} + \frac{1}{2}\frac{\partial^2 u}{\partial y^2} \ , \ u(0,y) = f(0,y) \ \textit{for} \ y \geq 0 \ , \ \frac{\partial u}{\partial y}(0+) = 0 \ . \tag{52}$$

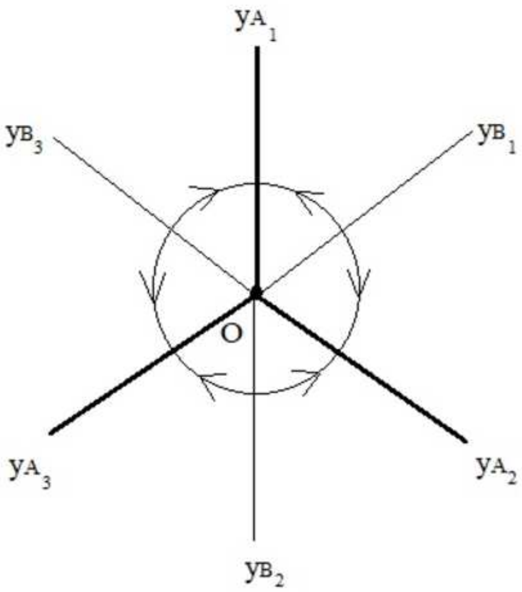

Figure 3: A more general problem.

2. One can consider a more general system such as the one shown in Figure 3. Here the 3 axes Oy_{A_1}, Oy_{A_2} and Oy_{A_3} consist of stable equilibriums and the other 3 axes Oy_{B_1}, Oy_{B_2}, Oy_{B_3} consist of unstable equilibriums. One can analyze this system in a similar fashion as we did in this work, so that we expect to see the limiting process as a diffusion process on a tree Γ (see [26]). The tree $\Gamma = Oy_{A_1} \cup Oy_{A_2} \cup Oy_{A_3}$ consists of 3 edges that are the semi–axes Oy_{A_1}, Oy_{A_2}, Oy_{A_3}. On each edge the limiting process is a Bessel–like process and the interior vertex O is inaccessible. The proof of these facts follows from the method we adopted in this paper as well as the techniques used in [28, Chapter 8], [27], [26]. One can first obtain "localization" type of results as we showed in Lemmas 4.4, 4.5. With such localization results at hand, we then show that the process localized onto Γ converges weakly to a diffusion process on the graph Γ, similarly as we did in the current work.

3. If the system (6) do not have the dissipative terms, so that it looks like

$$\begin{cases} d\mathrm{X}_t^\varepsilon = -\dfrac{1}{\varepsilon}\mathrm{X}_t^\varepsilon \mathrm{Y}_t^\varepsilon dt + dW_t^1 \quad , \quad \mathrm{X}_0^\varepsilon = x_0 \ , \\ d\mathrm{Y}_t^\varepsilon = \dfrac{1}{\varepsilon}(\mathrm{X}_t^\varepsilon)^2 dt + dW_t^2 \quad , \quad \mathrm{Y}_0^\varepsilon = y_0 \ . \end{cases} \tag{53}$$

Then the argument of the Lemmas 4.1–4.9 and the proof of Theorem 3.1 still go through, with minor changes in the estimates. The limiting Y–process will be a process of the form

$$d\mathrm{Y}_t = \frac{1}{\mathrm{Y}_t}dt + dW_t^2 \ , \ \mathrm{Y}_0 = y^\pi(x_0, y_0) \ . \tag{54}$$

In particular, this implies that the Y_t process keeps growing in the direction Oy_A. That is to say, the energy grows in the direction of the stable manifold Oy_A. Geometrically, this phenomenon comes from the fact that the energy constraint given by the conservative flow $b(x,y) = (-xy, x^2)$ provides a positive force around the stable line Oy_A. Thus the energy can keep growing at Oy_A due to the random noise. Such a geometric phenomenon might be related to some problems in 2–d turbulence (see [12]).

Acknowledgement. The author would like to thank Professor Vladimír Šverák from University of Minnesota, USA for fruitful discussions on the formulation of his system (3) as the Euler–Arnold equation on the group of all affine transformations of a line, as well as its relation with fluid mechanics. He also would like to thank the anonymous referee, Professor Yong Liu from Peking University, Beijing, China and Professor Yong Ren from Anhui Normal University, Wuhu, Anhui, China for their valuable comments that improve the first version of this work.

References

[1] V.I. Arnold. Sur la géométrie différentielle des groups de lie de dimension infinite et ses applications à l'hydrodynamique des fluids parfaits. *Ann. Inst. Fourier*, 16:316–361, 1966.

[2] V.I. Arnold. *Mathematical methods of classical mechanics*. Springer, 1978.

[3] V.I. Arnold and B. Khesin. *Topological methods in hydrodynamics*. Springer, 1998.

[4] N. Berglund. Kramers' law: Validity, derivations and generalizations. *Markov Processes and Related Fields*, 19:459–490, 2013.

[5] F. Bouchet and H. Morita. Large–time behavior and asymptotic stability of the 2D Euler and linerized Euler equations. *Physica D*, 239:948–966, 2010.

[6] F. Bouchet and J. Sommeria. Emergence of intense jets and Jupiter's Great Red Spot as maximum–entropy structures. *Journal of Fluid Mechanics*, 464:165–207, 2002.

[7] F. Bouchet and H. Touchette. Non–classical large deviations for a noisy system with non–isolated attractors. *Journal of Statistical Mechanics*, May 2012.

[8] F. Bouchet and A. Venallie. Statistical mechanics of two–dimensional and geophysical flows. *Physics Reports*, 515:227–295, 2012.

[9] D. Dolgopyat and L. Koralov. Averaging of Hamiltonian flows with an ergodic component. *Annals of Probability*, 36:1999–2049, 2008.

[10] D. Dolgopyat and L. Koralov. Averaging of incompressible flows on two dimensional surfaces. *Journal of American Mathematical Society*, 26(2):427–449, 2013.

[11] E.B. Dynkin. One–dimensional continuous strong Markov processes. *Theory of Probability and Its Applications*, IV(1):1–52, 1959.

[12] T. Elgindi, W. Hu, and V. Šverák. On 2d incompressible Euler equations with partial damping. *Communications in Mathematical Physics*, 355(1):145–159, October 2017.

[13] S.N. Ethier and T.G. Kurtz. *Markov processes, characterization and convergence.* John Wiley & Sons, 2005.

[14] W. Feller. Generalized second-order differential operators and their lateral conditions. *Illinois Journal of Mathematics*, 1:459–504, 1957.

[15] M. Freidlin. Sublimiting Distributions and Stabilization of Solutions of Parabolic Equations with a Small Parameter. *Soviet Math Doklady*, 235(5):1042–1045, 1977.

[16] M. Freidlin. On stochastic perturbations of dynamical systems with a "rough" symmetry: Hierarchy of Markov chains. *Journal of Statistical Physics*, 157(6):1031–1045, December 2014.

[17] M. Freidlin and W. Hu. On perturbations of the generalized Landau–Lifschitz dynamics. *Journal of Statistical Physics*, 144:978–1008, 2011.

[18] M. Freidlin and W. Hu. On stochasticity in Nealy–Elastic Systsms. *Stochastics and Dynamics*, 12(3), 2012.

[19] M. Freidlin and W. Hu. On second order elliptic equations with a small parameter. *Communications in Partial Differential Equations*, 38(10):1712–1736, 2013.

[20] M. Freidlin, W. Hu, and A. Wentzell. Small mass asymptotic for the motion with vanishing friction. *Stochastic Processes and their Applications*, 123:45–75, 2013.

[21] M. Freidlin and L. Koralov. Metastable distributions of markov chains with rare transitions. *Journal of Statistical Physics*, 167(6):1355–1375, June 2017.

[22] M. Freidlin, L. Koralov, and A. Wentzell. On diffusions in media with pockets of large diffusivity. `arXiv:1710.03555v1[math.PR]`.

[23] M. Freidlin, L. Koralov, and A. Wentzell. On the behavior of diffusion processes with traps. *Annals of Probability*, 45(5):3202–3222, 2017.

[24] M. Freidlin and L. Korlaov. On stochastic perturbations of slowly changing dynamical systems. *Nonlinearity*, 30(1), December 2016.

[25] M. Freidlin and A. Wentzell. On small random perturbations of dynamical systems. *Russian Mathematical Surveys*, 25(1):1–56, 1970.

[26] M. Freidlin and A. Wentzell. Diffusion processes on graphs and the averaging principle. *Annals of Probability*, 21(4):2215–2245, 1993.

[27] M. Freidlin and A. Wentzell. Random Perturbations of Hamiltonian systems. *Memoirs of the AMS*, 1994.

[28] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 2nd edition, 1998.

[29] M. Freidlin and A. Wentzell. On the Neumann problem for PDE's with a small parameter and the corresponding diffusion processes. *Probability Theory and Related Fields*, 152(1–2):101–140, 2012.

[30] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, 3rd edition, 2012.

[31] W. Hu. On metastability in nearly-elastic systems. *Asymptotic Analysis*, 79(1-2), 2012.

[32] W. Hu and V. Šverák. Dynamics of geodesic flows with random forcing on lie groups with left–invariant metrics. *Journal of Nonlinear Science*, online first, January 25, 2018.

[33] S. Kuksin and A. Shirikyan. Rigorous results in space–periodic two–dimensional turbulence. *Physics of Fluids*, 29:125106, 2017.

[34] P. Mandl. *Analytical Treatment of One–dimensional Markov Processes*. Springer, Berlin, 1968.

[35] D. Martiosyan. Large deviations for stationary measures of stochastic non–linear wave equations with smooth white noise. *Communications in Pure and Applied Mathematics*, to appear, 2017.

[36] J. Miller. Statistical mechanis of Euler equations in two–dimensions. *Physical Review Letters*, 65:2137–2140, 1990.

[37] S.A. Molchanov. Martin boundary for invariant markov processes on a solvable group. (english translation). *Theory of Probability and its Applications*, 12:310–314, 1967.

[38] C. Mouhot and C. Villani. On Landau damping. *Acta Mathematica*, 207:29–201, 2011.

[39] D. Revuz and M. Yor. *Continuous Martingales and Brownian motion, Third Edition.* Springer, 1999.

[40] R. Robert and J. Sommeria. Statistical equilibrium states for two–dimensional flows. *Journal of Fluid Mechanics*, 229:291–310, 1991.

[41] K. Schneider and M. Farge. Final states of decaying 2–d turbulence in bounded domains: influence of the geometry. *Physica D*, 237:2228–2233, 2008.

[42] J. Sommeria. Two dimendional turbulence. *New Trends in Turbulence, Les Houches Summer School, New York Springer*, 74:385–447, 2001.

[43] P. Tabling. Two–dimensional turbulence, a physicist approach. *Physics Reports*, 362(1):1–62, 2002.

[44] T. Tao. The Euler–Arnold equation. *available at* `https://terrytao.wordpress.com/2010/06/07/the-euler-arnold-equation/`.

[45] V. Šverák. *Lecture notes of Selected Topics in Fluid Mechanics.* University of Minnesota, 2011–2012.

[46] R.F. Willams. The structure of Lorentz attractors. *Publications Mathématiques de l'I.H.É.S*, tome 50:73–99, 1979.