Rate of the weak convergence of a stochastic stratified process

Natella V. O'Bryant

Department of Mathematics, University of California at Irvine, 103 Multipurpose Science and Technology Building, Irvine, CA 92697

Abstract

We consider laws of a stochastic process whose trajectories are governed by a two-dimensional weakly dissipative dynamical system with a degenerate Hamiltonian. Using Wasserstein L_p-distance as the shortest distance between the measures on the corresponding canonical probability spaces, we estimate the rate of the weak convergence of the laws of the original process to the unique law of a Markov process evolving on a stratified space. Additionally, for a related (non-stratified) example of a process on a cylinder, the corresponding estimates are obtained using two different methods. One of the methods revealed an explicit expression for the Wasserstein distance.

This is the first explicit attempt to use minimal metrics to estimate the rate of convergence of stochastic *stratified* processes.

Key words: Markov processes, stratified processes, Hamiltonian systems, weak convergence, stochastic averaging, Wasserstein distance
1991 MSC: 60F17, 60J

1 Introduction

During the last decade several stochastic stratified processes were introduced and studied in the probability literature. They include, for example, fiber Brownian motion introduced by Bass and Burdzy in [1] as a process whose evolution switches between a one-dimensional and a two-dimensional one, a Markov process on a 'whiskered sphere' described by Sowers in [16], processes on trees like the Walsh process, the Evans process, spider martingales, and Brownian snakes (see, for example, [3] and [19]).

Email address: natella@member.ams.org (Natella V. O'Bryant).

Preprint submitted to Elsevier Science *6 August 2004*

A Hamiltonian system with a degenerate Hamiltonian and with multiple time scales was considered in [11,12]. Here weak convergence to a stratified Markov process was shown on a space with the same geometry as in [16]. A related non-degenerate class of the single-degree-of-freedom systems excited by both periodic and random perturbations was also considered in [10], resulting in the same conclusion about the structure of the limiting space. Recently, Evans and Sowers [4] made certain generalizations about possible structures of the stratified spaces as evolution spaces for the graph-valued Markov processes.

The purpose of this work is to estimate the rate of convergence of a particular graph-valued Markov process by using the so-called Wasserstein-type metrics, which are traditionally used in mass transportation problems.

Growing applications of Wasserstein metrics demonstrate their power in many issues of nonlinear problem-solving. They have been previously applied to study the rates of convergence of diffusion processes, mainly by constructing embedded discrete-time processes, and almost exclusively for one-dimensional problems which can be reformulated in terms of pseudo-inverse functions. Such reformulations then allow for an application of the Fokker-Planck equation. The connections between the Fokker-Planck and the Wasserstein metric have also been previously explored. For example, it is known [9] that the Fokker-Planck equation can be interpreted as the steepest descent for a free energy related to Boltzmann-Gibbs entropy with respect to the Wasserstein metric.

The difficulty in applying the Fokker-Planck equation in our case is in the complex structure of the stratified space, which does not allow for an explicit reformulation in terms of pseudo-inverse functions. Nevertheless, we can make use of the Fokker-Planck equation in some limited way, and we show that even in the case of a stratified space, it serves as a powerful instrument for analyzing the time evolution of the laws of stochastic processes.

Section 2 of this work is devoted to the heuristic explanation of the dynamics behind the stratified process of our interest. Section 3 contains detailed and rather technical description of the generators of the processes of our interest, and as much as we would like to avoid these technicalities, it does not appear to be possible for the sake of subsequent discussion. The very short Section 4 has the formal definition of the Wasserstein L_p distance and the formulation of our goal, which finally involves the explicit expression for this distance. Section 5 is conceptually the main part of this work, and it contains the key Theorem 2. Sections 6 and 7 contain two estimates of the Wasserstein L_p-distance for the related classical problem. The results of these computations are presented in the form of Theorems 3 and 4. Finally, Section 8 contains some unanswered questions, natural to pose, and to work on, as a result of this paper.

2 The Evolution Space of the Stratified Markov Process

There exist equivalent approaches which we could undertake for analyzing the time evolution of laws of Markov processes. We can study the related martingales on the corresponding canonical spaces [13], [17]. We can also study the dynamics of the trajectories of these Markov processes by means of the corresponding Itô equations [5]. These two approaches have a tight connection, and parallel asymptotic theories are built using each of them. One of the links of this connection is expressed in the form of the Fokker-Planck equation, or forward Kolmogorov equation, which describes the evolution of the probability densities for a stochastic process, and is inherently related to the corresponding Itô equation.

Acknowledging that differential equations provide a clearer intuitive description of a system, we will outline this description of our problem first, and then we will phrase it using the martingale language required for our purposes. We will keep both these approaches in mind, since we would like to explore the connection between them.

Essentially, we consider the following diffusion

$$d\mathbf{X}_t^\varepsilon(\mathbf{x}) = \nabla^\perp H\left(\mathbf{X}_t^\varepsilon(\mathbf{x})\right) dt + \varepsilon\, d\mathbf{W}_t, \quad \mathbf{X}_0^\varepsilon(\mathbf{x}) = \mathbf{x} = (x_1, x_2)\,, \tag{1}$$

where $\nabla^\perp = \left(\frac{\partial}{\partial x_2}, -\frac{\partial}{\partial x_1}\right)$ is chosen to be orthogonal to ∇, $H : \mathbb{R}^2 \to [0,\infty)$ is a smooth enough function, $\lim_{|\mathbf{x}|\to\infty} H(\mathbf{x}) = +\infty$, ε is a small parameter, and $\mathbf{W}$ is a two-dimensional Brownian motion. We can refer to this system as stochastic Hamiltonian, since H turns out to be the first integral of the corresponding deterministic system

$$d\mathbf{X}_t^0(\mathbf{x}) = \nabla^\perp H\left(\mathbf{X}_t^0(\mathbf{x})\right) dt, \quad \mathbf{X}_0^0(\mathbf{x}) = \mathbf{x}. \tag{2}$$

The motion described by the system (1) can be thought of as fast rotation along the trajectories of the deterministic system (2) combined with slow motion across them. In contrast to the deterministic equation, H can take different values on the same trajectory of the stochastic system, and the same value on different trajectories, but it can still serve as the coordinate of a trajectory in a local sense. Our interest is in analyzing the long-term behavior of the slow diffusion across the deterministic trajectories, i.e., the asymptotic behavior of the slow process $H\left(\mathbf{X}^\varepsilon\right)$, as $\varepsilon \to 0$.

We now need to address the properties of function H, vaguely described above as a 'smooth enough function'. In general, the limiting space, i.e., the space where the weak limit of the slow process lives, is stratified [7]. Depending on the properties of H, it may have various staratification patterns, ranging from trivial to very complicated [4].

It turns out that successful analysis of the asymptotics of $H(X^\varepsilon)$ can be done if the process X^ε is rescaled, so that the new time scale becomes the natural time scale for the slow process $H(X^\varepsilon)$ itself. Then, according to a classical result, due to Has'minskiĭ [8], if the Hamiltonian H has one well, then on the new time scale the slow process $H\left(\mathbf{X}^\varepsilon\right)$ weakly converges to the unique Markov process on $\mathbb{R}$. Note that we can treat an open segment of the real line as a stratified space with trivial stratification, and a closed segment of the real line as a space with three strata, an open segment, and two end-points.

We will sketch our calculation for the rate of this convergence below to demonstrate on a simple example some of the methods used.

A Hamiltonian with an even slightly more complicated nature leads to a non-trivial stratification of the limiting space. Freidlin and Wentzell [6] showed that if H is double-well, then the weak limit has strata of two types, four points and three open segments, all joined through one of the points.

A more complicated stratification pattern is exhibited by the space of the limiting process in the case when the critical set of H is allowed to have an interior [16]. The canonical example of such a Hamiltonian is a cut-off paraboloid, $H(\mathbf{x}) = \left(\max\{0, |\mathbf{x}| - 1\}\right)^n,\ n > 2$. In this case, the evolution space of the limiting Markov process resembles a lollipop: its strata are a 2-sphere, a line, and a point. This is the type of stratified system of interest in this work.

3 The Generator of the Limiting Markov Process

In this section we describe the limiting process on the stratified space, in the notation consistent with our problem.

For any functions $f, g \in C^2(\mathbb{R}^2;\mathbb{R})$, we define two second-order differential operators

$$\mathcal{L}f(\mathbf{x}) := \frac{1}{2}\sum_{i,j=1}^{2}\gamma_{ij}(\mathbf{x})\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}), \tag{3}$$

and

$$\langle df, dg\rangle(\mathbf{x}) := \mathcal{L}(fg)(\mathbf{x}) - f(\mathbf{x})\,\mathcal{L}g(\mathbf{x}) - g(\mathbf{x})\,\mathcal{L}f(\mathbf{x}).$$

We assume that the matrix $\{\gamma_{ij}(\mathbf{x})\}, 1 \le i, j \le 2$, is symmetric for every $\mathbf{x} \in \mathbb{R}^2$, and consists of smooth enough measurable functions. We require that $\mathcal{L}$ be uniformly elliptic, i.e., that there exists $\delta > 0$ such that

$$\sum_{i,j=1}^{2}\gamma_{ij}(\mathbf{x}) z_i z_j \geq \delta\,|\mathbf{z}|^2$$

for all $\mathbf{x} \in \mathbb{R}^2$ and all $\mathbf{z} = (z_1, z_2) \in \mathbb{R}^2$.

Let the canonical probability space $\Omega := C\left([0,\infty), \mathbb{R}^2\right)$ be the space of the continuous trajectories from $[0,\infty)$ to $\mathbb{R}^2$ [2]. Denote by $\mathbf{X}_t$ the canonical process and by $\mathcal{F}_t$ the filtration generated by $\{\mathbf{X}_s; 0 \le s \le t\}$.

Let

$$H(\mathbf{x}) := \left(\max\{0, K(\mathbf{x})\}\right)^n,\ n > 2,$$

where K is a smooth real-valued function on $\mathbb{R}^2$, with $\lim_{|\mathbf{x}|\to\infty} K(\mathbf{x}) = +\infty$. Assume that the closure $\overline{\mathbf{V}}$ of the set

$$\mathbf{V} := \left\{\mathbf{x} \in \mathbb{R}^2 : K(\mathbf{x}) < 0\right\}$$

is diffeomorphic to the unit disc, and that K is nonsingular, i.e., that $\nabla K \neq 0$ on $\mathbb{R}^2 \setminus \mathbf{V}$. The nonsingularity property is essential for the problem; it implies that there exists a small $a > 0$ such that the set

$$\mathcal{E} := \left\{\mathbf{x} \in \mathbb{R}^2 : |K(\mathbf{x})| < a\right\},$$

containing the boundary of $\mathbf{V}$, has one component, and that K is nonsingular on $\mathcal{E}$.

Here is an example of a smooth function $\overline{K}$ which does not satisfy these properties,

$$\overline{K}(\mathbf{x}) := \begin{cases} 0 & \text{if } |\mathbf{x}| \le 1, \\ |\mathbf{x}|\, \mathbf{e}^{-\frac{1}{(1-|\mathbf{x}|)^2}} & \text{if } |\mathbf{x}| > 1. \end{cases}$$

It is not suitable for our problem, since it is singular on the set

$$\left\{\mathbf{x} \in \mathbb{R}^2 : \overline{K}(\mathbf{x}) = 0\right\},$$

although the rest of the required properties are satisfied.

For any $f \in C^2\left(\mathbb{R}^2\right)$, define the differential operator

$$\mathcal{L}^\varepsilon f(\mathbf{x}) := \mathcal{L} f(\mathbf{x}) + \frac{1}{\varepsilon^2} \nabla^\perp H(\mathbf{x}) \cdot \nabla f(\mathbf{x}), \tag{4}$$

and assume that its domain $\mathcal{D}^\varepsilon = \mathcal{D}\left(\mathcal{L}^\varepsilon\right)$ contains a dense subset of functions from $C^2\left(\mathbb{R}^2\right)$.

Fix a level set of H of height $H^* > 0$, let $K^* = \left(H^*\right)^{1/n}$, define the open sets

$$\mathbf{I} := \left\{\mathbf{x} \in \mathbb{R}^2 : K(\mathbf{x}) < K^*\right\} \quad \text{and} \quad \mathbf{U} := \mathbf{I} \setminus \overline{\mathbf{V}}, \tag{5}$$

and define the $\mathcal{F}_t$-stopping time τ to be the time of the first exit from $\mathbf{I}$, i.e.,

$$\tau := \inf\left\{s \ge 0 : \mathbf{X}_s \notin \mathbf{I}\right\}.$$

Definition 1 *Given* $\mathbf{x} \in \bar{\mathbf{I}}$, *define* $\mathbb{P}^{\varepsilon} \in \mathcal{P}\left(C\left([0,\infty;\mathbb{R}^2\right)\right)$ *to be a solution of the stopped martingale problem for* $(\mathcal{L}^{\varepsilon}, \delta_{\mathbf{x}}, \mathbf{I})$, *and so is the corresponding coordinate process* $\mathbf{X}$, *i.e., for every* $f \in C^2\left(\mathbb{R}^2\right)$, *and for all* $0 \le s \le t$,

$$f\left(\mathbf{X}_{t\wedge\tau}\right) - f\left(\mathbf{X}_{s\wedge\tau}\right) - \int_{s\wedge\tau}^{t\wedge\tau} \mathcal{L}^{\varepsilon}\left(\mathbf{X}_u\right) du$$

is a $\mathbb{P}^{\varepsilon}$*-martingale.*

For future reference, we also define another $\mathcal{F}_t$-stopping time τ_1 to be the first time of reaching the boundary of the critical set, starting from $\mathbf{U}$, i.e.,

$$\tau_1 := \inf\left\{s \ge 0 : \mathbf{X}_s \in \partial\mathbf{V} \mid \mathbf{X}_0 \in \mathbf{U}\right\}.$$

Let ξ_t be the flow generated by $\nabla^{\perp}H$, i.e.,

$$\dot{\xi}_t(\mathbf{x}) = \nabla^{\perp}H\left(\xi_t(\mathbf{x})\right), \quad \xi_0(\mathbf{x}) = \mathbf{x}.$$

We can form the following quotient space $\mathbf{\Gamma} := \bar{\mathbf{I}}/\sim$ by using the chain equivalence relation $\sim$ based on the flow ξ_t (for details of this construction see [15]). For every $\mathbf{x} \in \bar{\mathbf{I}}$, define the natural map $\pi(\mathbf{x}) := [\mathbf{x}]$, and for $G \subset \bar{\mathbf{I}}$ denote $\mathbf{\Gamma}_G = \pi(G)$. This quotient space

$$\mathbf{\Gamma} = \mathbf{\Gamma}_{\mathbf{V}} \cup \mathbf{\Gamma}_{\partial\mathbf{V}} \cup \mathbf{\Gamma}_{\mathbf{U}} \cup \mathbf{\Gamma}_{\partial\mathbf{I}}$$

turns out to be homeomorphic to a stratified space, a submanifold of $\mathbb{R}^3$, which is essentially a 2-sphere with a line segment attached. Denoting the homeomorphism by $g : \mathbf{\Gamma} \to \mathbb{R}^3$, we can define the metric d on $\mathbf{\Gamma}$ by setting

$$d([\mathbf{x}],[\mathbf{y}]) = |g([\mathbf{x}]) - g([\mathbf{y}])|_{\mathbb{R}^3}$$

for any $[\mathbf{x}],[\mathbf{y}] \in \mathbf{\Gamma}$, and we see that the constructed metric space $(\mathbf{\Gamma}, d)$ is Polish.

Define $\mathbf{Y} := \left[\mathbf{X}_{\cdot\wedge\tau}\right]$, and for each $\varepsilon > 0$, define the probability measure $\mathbb{P}^{\varepsilon,*}$ as the projection of measure $\mathbb{P}^{\varepsilon}$ on $\mathbf{\Gamma}$,

$$\mathbb{P}^{\varepsilon,*}([G]) := \mathbb{P}^{\varepsilon}\left\{\mathbf{Y} \in [G]\right\}, \quad G \in \mathcal{B}\left(C([0,\infty), \bar{\mathbf{I}})\right).$$

Also, let $\mathbf{Y}_1 := \left[\mathbf{X}_{\cdot\wedge\tau\wedge\tau_1}\right]$, and

$$\mathbb{P}_1^{\varepsilon,*}([G]) := \mathbb{P}^{\varepsilon}\left\{\mathbf{Y}_1 \in [G]\right\}, \quad G \in \mathcal{B}\left(C([0,\infty), \overline{\mathbf{U}})\right).$$

Given a function F defined on Γ and a set $[G] \in \Gamma$, we denote the restriction of F to $[G]$ by F_G. In particular, if $F \in C(\Gamma;\mathbb{R})$, then $F \circ \pi \in C(\bar{\mathbf{I}};\mathbb{R})$ and

- $F_{\mathbf{V}} \circ \pi(\mathbf{x}) = F_{\mathbf{V}}(\mathbf{x}), \quad \mathbf{x} \in \mathbf{V};$

- $F_{\overline{\mathbf{U}}} \circ \pi(\mathbf{x}) = (F_{\overline{\mathbf{U}}} \circ K^{-1}) \circ K(\mathbf{x}), \quad \mathbf{x} \in \overline{\mathbf{U}}$;
- $F_K := F_{\overline{\mathbf{U}}} \circ K^{-1}$ maps $[0, K^*]$ into $\mathbb{R}$;
- $F \circ \pi(\mathbf{x}) = F_{\mathbf{V}}(\mathbf{x})\chi_{\mathbf{V}}(\mathbf{x}) + F_K(K(\mathbf{x}))\chi_{\overline{\mathbf{U}}}(\mathbf{x}), \quad \mathbf{x} \in \bar{\mathbf{I}} = \overline{\mathbf{U}} \cup \mathbf{V}$.

Let $\mathcal{D}_{\mathbb{A}_0}$ be the set of all real-valued functions on $\mathbf{\Gamma}$, whose restriction to $\mathbf{\Gamma}_{\mathbf{V}} \cup \mathbf{\Gamma}_{\mathbf{U}}$ is continuous. Define the pre-averaging operator $\mathbb{A}_0 : \mathcal{D}_{\mathbb{A}_0} \to C\left(\mathbf{\Gamma}_{\mathbf{V}} \cup \mathbf{\Gamma}_{\mathbf{U}}\right)$ by

$$\mathbb{A}_0 f([\mathbf{x}]) := \begin{cases} f(\mathbf{x}), & \text{if } [\mathbf{x}] \in \mathbf{\Gamma}_{\mathbf{V}} \\ \frac{\int_{\mathbf{y}\in[\mathbf{x}]} \frac{f(\mathbf{y})}{|\nabla K(\mathbf{y})|}\, dl(\mathbf{y})}{\int_{\mathbf{y}\in[\mathbf{x}]} \frac{1}{|\nabla K(\mathbf{y})|}\, dl(\mathbf{y})}, & \text{if } [\mathbf{x}] \in \mathbf{\Gamma}_{\mathbf{U}}, \end{cases}$$

where dl is the arc-length element. Let

$$\mathcal{D}_{\mathbb{A}} := \left\{ f \in \mathcal{D}_{\mathbb{A}_0} : \lim_{\substack{[\mathbf{y}]\to\mathbf{\Gamma}_{\partial\mathbf{V}} \\ [\mathbf{y}]\in\mathbf{\Gamma}_{\mathbf{V}}\cup\mathbf{\Gamma}_{\mathbf{U}}}} \mathbb{A}_0 f([\mathbf{y}]) \text{ and } \lim_{\substack{[\mathbf{y}]\to\mathbf{\Gamma}_{\partial\mathbf{I}} \\ [\mathbf{y}]\in\mathbf{\Gamma}_{\mathbf{V}}\cup\mathbf{\Gamma}_{\mathbf{U}}}} \left(\mathbb{A}_0 f\right)([\mathbf{y}]) \text{ exist} \right\},$$

and define the averaging operator $\mathbb{A} : \mathcal{D}_{\mathbb{A}} \to C(\mathbf{\Gamma})$ by

$$\mathbb{A} f([\mathbf{x}]) := \lim_{\substack{[\mathbf{y}]\to[\mathbf{x}] \\ [\mathbf{y}]\in\mathbf{\Gamma}_{\mathbf{V}}\cup\mathbf{\Gamma}_{\mathbf{U}}}} \mathbb{A}_0 f([\mathbf{y}]).$$

Notice that if $\mathbf{x} \in \bar{\mathbf{I}} \setminus \overline{\mathbf{V}}$, then $[\mathbf{x}]$ is totally defined by $H(\mathbf{x})$, and therefore by $K(\mathbf{x})$, and

$$\mathbb{A} f([\mathbf{x}]) = \mathbb{A} f\left(K^{-1}(K(\mathbf{x}))\right),$$

so it makes sense to define $\mathbb{A}_K f(h) := \mathbb{A} f\left(K^{-1}(h)\right),\ 0 < h \le K^*$.

If $F \in C^2\left(\Gamma_{\mathbf{V}} \cup \Gamma_{\mathbf{U}}; \mathbb{R}\right)$, then $\mathcal{L}\left(F \circ \pi\right)$ is well-defined on the set $\mathbf{V} \cup \mathbf{U}$. If $\mathcal{L}\left(F \circ \pi\right) \in \mathcal{D}_{\mathbb{A}}$, then for $[\mathbf{x}] \in \mathbf{\Gamma}_{\mathbf{V}} \cup \mathbf{\Gamma}_{\mathbf{U}}$, we set

$$\begin{aligned} \mathbb{A}^* F([\mathbf{x}]) &:= \mathbb{A}\left(\mathcal{L}\left(F \circ \pi\right)\right)(\mathbf{x}) \\ &= \begin{cases} \left(\mathcal{L} F_{\mathbf{V}}\right)(\mathbf{x}) & \text{if } [\mathbf{x}] \in \mathbf{\Gamma}_{\mathbf{V}}, \\ b\left(K(\mathbf{x})\right) F_K'(K(\mathbf{x})) + \frac{1}{2}\sigma^2\left(K(\mathbf{x})\right) F_K''(K(\mathbf{x})) & \text{if } [\mathbf{x}] \in \mathbf{\Gamma}_{\mathbf{U}}. \end{cases} \end{aligned}$$

where

$$b := \mathbb{A}_K\left(\mathcal{L} K\right) \quad \text{and} \quad \sigma^2 := \mathbb{A}_K \langle dK, dK \rangle.$$

We define an 'inner' gluing operator

$$\underline{\mathbb{G}} F := \lim_{\substack{[\mathbf{x}]\to\mathbf{\Gamma}_{\partial\mathbf{V}} \\ [\mathbf{x}]\in\mathbf{\Gamma}_{\mathbf{V}}}} \left(\mathbb{A}_0 \langle dF_{\mathbf{V}}, dK \rangle\right)([\mathbf{x}]) = \lim_{\substack{\mathbf{x}\to\partial\mathbf{V} \\ \mathbf{x}\in\mathbf{V}}} \langle dF_{\mathbf{V}}, dK \rangle(\mathbf{x}) \tag{6}$$

for functions F with $F_{\mathbf{V}} \in C^1(\mathbf{\Gamma_V})$, and an 'outer' gluing operator

$$\overline{\mathbb{G}}F := \lim_{\substack{[\mathbf{x}]\to\mathbf{\Gamma}_{\partial\mathbf{V}}\\ [\mathbf{x}]\in\mathbf{\Gamma_U}}} \left(\mathbb{A}_0\langle dF_K, dK\rangle\right)([\mathbf{x}]) = \lim_{\substack{[\mathbf{x}]\to\mathbf{\Gamma}_{\partial\mathbf{V}}\\ [\mathbf{x}]\in\mathbf{\Gamma_U}}} \left(\int_{\mathbf{y}\in[\mathbf{x}]} \frac{1}{\|\nabla K(\mathbf{x})\|} dl(\mathbf{y})\right)^{-1} \int_{\mathbf{y}\in[\mathbf{x}]} \frac{\langle dF_K, dK\rangle}{\|\nabla K(\mathbf{x})\|} dl(\mathbf{y}) \tag{7}$$

for functions F with $F_{\mathbf{U}} \in C^1(\mathbf{\Gamma_U})$, if these limits exist.

Now we define the domain

$$\mathcal{D}^* := \Big\{ F \in C(\mathbf{\Gamma}) : F|_{\mathbf{\Gamma_V}\cup\mathbf{\Gamma_U}} \in C^2(\mathbf{\Gamma_V}\cup\mathbf{\Gamma_U}), \quad \mathcal{L}(F\circ\pi) \in \mathcal{D}_{\mathbb{A}}, \quad \underline{\mathbb{G}}F = \overline{\mathbb{G}}F, \quad \lim_{[\mathbf{y}]\to\mathbf{\Gamma}_{\partial\mathbf{I}}} (\mathbb{A}^*F)([\mathbf{y}]) = 0 \Big\}, \tag{8}$$

and for any $F \in \mathcal{D}^*$, we define the operator

$$\mathcal{L}^*F([\mathbf{x}]) := \lim_{[\mathbf{y}]\to[\mathbf{x}]} \mathbb{A}^*F([\mathbf{y}]), \quad [\mathbf{x}] \in \mathbf{\Gamma}. \tag{9}$$

Again, we set the coordinate functions $\mathbf{Y}^*_t(\omega) = \omega(t)$, $t \geq 0$, on the corresponding traditionally defined space $\Omega^* = C([0,\infty);\mathbf{\Gamma})$ with the σ-algebra $\mathcal{F}^* = \vee_{t\geq 0}\mathcal{F}^*_t$, where $\mathcal{F}^*_t = \sigma\{\mathbf{Y}^*_s : 0 \leq s \leq t\}$.

By Theorem 2.11 ([16, Theorem 2.11, pg. 861]), the $\mathbb{P}^\varepsilon$-laws of $\mathbf{Y}$ converge to the law of a $\mathbf{\Gamma}$-valued Markov process $\mathbf{Y}^*$. More precisely, the family of measures $\{\mathbb{P}^{\varepsilon,*}\}$ is tight in the Prohorov's topology on $\mathcal{P}(C([0,\infty),\mathbf{\Gamma}))$ [2], and converges to the unique solution $\mathbb{P}^*$ of the martingale problem for $\left(\mathcal{L}^*, \delta_{[\mathbf{x}]}\right)$. Denote the corresponding expectation operator by $\mathbb{E}^*$.

Also, for any $f \in C^2(\mathbb{R})$, define the second order differential operator

$$\mathcal{L}_K f(h) := b(h) f'_K(h) + \frac{1}{2}\sigma^2(h) f''_K(h), \quad \text{if } h \in (0, K^*),$$

and the domain

$$\mathcal{D}^*_1 := \Big\{ f \in C^2((0,K^*)) : \quad \mathcal{L}_K f \in C^2((0,K^*)), \quad \lim_{h\to 0}(\mathcal{L}_K f)(h) = \lim_{h\to K^*}(\mathcal{L}_K f)(h) = 0 \Big\}. \tag{10}$$

For any $f \in \mathcal{D}^*_1$, define

$$\mathcal{L}^*_K f(h) := \lim_{y\to h} \mathcal{L}_K f(y).$$

According to the classical results of Has'minskiĭ, on the corresponding canonical probability space, the family of probability measures $\{\mathbb{P}_1^{\varepsilon,*}\}$ converges to the unique solution $\mathbb{P}_1^*$ of the martingale problem for $(\mathcal{L}_K^*, \delta_{h_0})$, where $h_0 = K(\mathbf{x}_0)$.

4 Martingale Measures and Wasserstein L_p-distances

Denote by Φ the Polish space $(C([0,\infty),\mathbf{\Gamma}),\rho)$ of continuous trajectories on $\mathbf{\Gamma}$, equipped with the sup-norm. Let $M_0(\Phi)$ be the space of all probability measures on Φ, and let $M_p(\Phi)$ be the space of all probability measures with finite moments of order $p \geq 0$, i.e., those measures μ that for some (and therefore, for any) $\phi_0 \in \Phi$,

$$M_p(\Phi) := \left\{\mu \in M_0(\Phi) : \int_\Phi \rho\left(\phi_0, \phi\right)^p d\mu(\phi) < \infty\right\}.$$

Notice that since ρ is bounded, $M_p(\Phi)$ simply coincides with $M_0(\Phi)$, but we will keep varying the indices for consistency.

Let μ^ε be a probability measure on the product space $\Phi \times \Phi$ with marginals $\mathbb{P}^{\varepsilon,*}$ and $\mathbb{P}^*$, defined in Section 3 above, and let

$$\Pi\left(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^*\right) := \left\{\mu :\ \mu\left(A \times \Phi\right) = \mathbb{P}^{\varepsilon,*}(A),\ \mu\left(\Phi \times A\right) = \mathbb{P}^*(A)\right\},$$

with A being any measurable subset of Φ. Define

$$W_p\left(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^*\right) := \left(\inf_{\mu \in \Pi(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^*)} \int_{\Phi\times\Phi} \rho\left(\phi,\psi\right)^p d\mu(\phi,\psi)\right)^{1/p}, \tag{11}$$

$1 \leq p < \infty$. It can be shown that (11) defines a metric on $M_p(\Phi)$, and therefore, on $M_0(\Phi)$. We will refer to W_p as the Wasserstein L_p-distance. The terminology associated with these metrics varies considerably. For other possible names see [14,18,20].

Since Wasserstein L_p-distances metrize weak convergence (see [18] for the proof), in our case it is guaranteed that

$$W_p\left(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^*\right) \to 0 \text{ as } \varepsilon \to 0.$$

This work is one of the steps in order to find explicit positive-valued decaying functions $\alpha_p(\varepsilon), \beta_p(\varepsilon)$ depending on the properties of the generators $\mathcal{L}^\varepsilon$ (4) and $\mathcal{L}^*$ (9), such that for every $\varepsilon > 0$, the following holds: $\alpha_p(\varepsilon) \leq W_p\left(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^*\right) \leq \beta_p(\varepsilon)$. Here we provide initial estimates on the rate of the weak convergence of the laws of the described stratified Markov processes, and investigate how good these estimates are. Here we do not investigate whether numerical simulations may shed light on this question; at this stage, we choose to rely on the analytical methods.

5 Estimates on the Wasserstein L_p-distance for a Stratified Process

Later we will show that when the Wasserstein distances can be represented in terms of pseudo-inverse distribution functions, this presentation can deliver the best possible estimates. In particular, $W_p\left(\mathbb{P}_1^{\varepsilon,*}, \mathbb{P}_1^*\right)$ can be computed explicitly. Unfortunately, the rather complicated geometry of $\mathbf{\Gamma}$ does not allow for an easy expression of the distances in those terms. Nevertheless, what gives us reasonable hope is the fact that these distances depend very little on the underlying geometrical structure, and that many related results hold in extreme generality, i.e., for Polish spaces.

One such result, not relying on the underlying geometric structure of the space is the following famous duality theorem, due to Strassen. We state this result for the Polish space Φ, defined above. Together with any measurable set $A \subset \Phi$, we define

$$A^\alpha := \{\phi \in \Phi : \rho(\phi, A) \leq \alpha\}, \quad A_\alpha := \{\phi \in A : \rho(\phi, \partial A) > \alpha\}$$

Theorem 1 *(Strassen's Theorem) Let $\mathbb{P}, \mathbb{Q} \in M_0(\Phi)$, and $\alpha \geq 0$. Then*

$$\inf_{\mu \in \Pi(\mathbb{P},\mathbb{Q})} \mu\left(\{\rho(\phi,\psi) > \alpha\}\right) = \sup_{\substack{A \subset \Phi \\ A \text{ closed}}} \left\{\mathbb{P}(A) - \mathbb{Q}(A^\alpha)\right\}.$$

This result is one of the main tools which allows us to prove the main theorem for the stratified problem. First, we define the set $\overline{\Phi}$ to be the subset of Φ consisting of those paths which start inside $\mathbf{\Gamma_U}$, and eventually reach $\mathbf{\Gamma_{\partial V}}$.

Theorem 2 *If $p > 1$, then the Wasserstein L_p-distance between the laws $\mathbb{P}^{\varepsilon,*}$ and the limiting law $\mathbb{P}^*$ is at least the distance between the laws $\mathbb{P}_1^{\varepsilon,*}$ and $\mathbb{P}_1^*$ of the two corresponding processes, stopped when reaching the boundary of the critical set, and at most a multiple of the total variation of the laws $\mathbb{P}^{\varepsilon,*}$ and $\mathbb{P}^*$ over all the paths starting inside $\mathbf{\Gamma_U}$ and reaching the boundary of the critical set $\mathbf{\Gamma_{\partial V}}$, i.e.,*

$$W_p\left(\mathbb{P}_1^{\varepsilon,*}, \mathbb{P}_1^*\right) \leq W_p\left(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^*\right) \leq C^p \sup_{\substack{A \subset \overline{\Phi} \\ A \text{ closed}}} \left\{\mathbb{P}^{\varepsilon,*}(A) - \mathbb{P}^*(A)\right\},$$

where C is a universal constant, independent of ε.

Proof: Using Theorem 1 for the measures $\mathbb{P}^{\varepsilon,*}$ and $\mathbb{P}^{*}$, we see that for any $\alpha \geq 0$,

$$\begin{aligned}
&W_p^p\left(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^{*}\right) = \inf_{\mu \in \Pi(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^{*})} \int_{\Phi\times\Phi} \rho\left(\phi, \psi\right)^p d\mu(\phi,\psi) \\
&= \inf_{\mu \in \Pi(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^{*})} \left(\int_{(\Phi\times\Phi)\cap\{\rho(\phi,\psi)\leq\alpha\}} + \int_{(\Phi\times\Phi)\cap\{\rho(\phi,\psi)>\alpha\}} \right) \rho\left(\phi, \psi\right)^p d\mu(\phi,\psi) \\
&\leq \inf_{\mu \in \Pi(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^{*})} \left(\alpha^p + \int_{(\Phi\times\Phi)\cap\{\rho(\phi,\psi)>\alpha\}} \rho\left(\phi, \psi\right)^p d\mu(\phi,\psi) \right) \qquad (12) \\
&= \alpha^p + \inf_{\mu \in \Pi(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^{*})} \int_{(\Phi\times\Phi)\cap\{\rho(\phi,\psi)>\alpha\}} \rho\left(\phi, \psi\right)^p d\mu(\phi,\psi) \\
&\leq \alpha^p + C^p \sup_{\substack{A\subset\Phi \\ A \text{ closed}}} \left\{ \mathbb{P}^{\varepsilon,*}\left(A\right) - \mathbb{P}^{*}\left(A^{\alpha}\right) \right\},
\end{aligned}$$

where C is equal to the finite diameter of $\mathbf{\Gamma}$, i.e.,

$$C := \sup_{[x],[y]\in\mathbf{\Gamma}} d\left(\left[\mathbf{x}\right], \left[\mathbf{y}\right]\right).$$

Letting $\alpha \to 0$ in (12), and using the definitions of $\mathbb{P}^{\varepsilon,*}$ and $\mathbb{P}^{*}$, we obtain

$$W_p^p\left(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^{*}\right) \leq C^p \lim_{\alpha\to 0} \sup_{\substack{A\subset\Phi \\ A \text{ closed}}} \Big[\mathbb{P}^{\varepsilon}\left\{\mathbf{Y} \in A\right\} - \mathbb{P}^{*}\left\{\mathbf{Y}^{*} \in A^{\alpha}\right\} \Big]. \qquad (13)$$

Any set $A \subset \Phi$ can be represented as a disjoint union of two subsets of the space Φ of continuous paths with images in two different non-trivial strata, namely,

$$A = \left(A \cap \Phi_1\right) \dot{\bigcup} \left(A \cap \Phi_2\right),$$

where

$$\Phi_1 := C\left([0,\infty); \mathbf{\Gamma}_{\mathbf{V}}\right), \quad \Phi_2 := C\left([0,\infty); \mathbf{\Gamma}_{\overline{\mathbf{U}}}\right).$$

Here $\overline{\mathbf{U}} = \partial\mathbf{V} \cap \mathbf{U} \cap \partial\mathbf{I}$, and therefore $\mathbf{\Gamma}_{\overline{\mathbf{U}}}$, is closed. The boundary of the critical set $\partial\mathbf{V}$ is included in the set that forms Φ_2, rather than Φ_1. The reason for that lies behind the chain equivalence [15]. According to this equivalence relation, all the points of $\partial\mathbf{V}$ are identified, forming a single equivalence class $\mathbf{\Gamma}_{\partial\mathbf{V}}$, in the same way as other points of $\mathbf{U}$ on the same level also form a separate equivalence class. On the other hand, any point of the open set $\mathbf{V}$ forms its own equivalence class, and only such points are used to define Φ_1. With this in mind,

$$\sup_{\substack{A\subset\Phi\\ A\text{ closed}}}\left[\mathbb{P}^{\varepsilon}\left\{\mathbf{Y}\in A\right\}-\mathbb{P}^{*}\left\{\mathbf{Y}^{*}\in A^{\alpha}\right\}\right]$$
$$\leq\sum_{i=1}^{2}\sup_{\substack{A\subset\Phi\\ A\text{ closed}}}\left[\mathbb{P}^{\varepsilon}\left\{\mathbf{Y}\in A\cap\Phi_{i}\right\}-\mathbb{P}^{*}\left\{\mathbf{Y}^{*}\in A^{\alpha}\cap\Phi_{i}\right\}\right].\quad(14)$$

Considering these two summands separately, notice that the set $\mathbf{\Gamma}_{\overline{\mathbf{U}}}$ is homeomorphic to a closed interval $[0,K^{*}]\subset\mathbb{R}$, and the set $\mathbf{\Gamma}_{\mathbf{V}}$ is essentially an open disc stereographically projected onto a punctured 2-sphere.

Recall that $\mathbb{P}^{\varepsilon}$ and $\mathbb{P}^{*}$ each solve a martingale problem, associated with the generator $\mathcal{L}^{\varepsilon}$ and $\mathcal{L}^{*}$, respectively. Then, notice that $\mathcal{L}^{\varepsilon}$ and $\mathcal{L}^{*}$ coincide on any subset of the open set $\mathbf{\Gamma}_{\mathbf{V}}$, since there is 'no averaging' done inside such a set. More explicitly, for any $F\in\mathcal{D}^{*}\cap\mathcal{D}^{\varepsilon}$, whose restriction onto $\mathbf{V}$ is denoted by $F_{\mathbf{V}}$, and for any $\mathbf{x}\in\mathbf{V}$,

$$\mathcal{L}^{*}F[\mathbf{x}]=\mathcal{L}^{\varepsilon}F_{\mathbf{V}}(\mathbf{x})=\mathcal{L}F_{\mathbf{V}}(\mathbf{x}).$$

Therefore,

$$\begin{aligned}&\sup_{\substack{A\subset\Phi\\ A\text{ closed}}}\left[\mathbb{P}^{\varepsilon}\left\{\mathbf{Y}\in A\cap\Phi_{1}\right\}-\mathbb{P}^{*}\left\{\mathbf{Y}^{*}\in A^{\alpha}\cap\Phi_{1}\right\}\right]\\ &=\sup_{\substack{A\subset\Phi\\ A\text{ closed}}}\left[\mathbb{P}^{\varepsilon}\left\{\mathbf{Y}\in A\cap\Phi_{1}\right\}-\mathbb{P}^{\varepsilon}\left\{\mathbf{Y}\in A^{\alpha}\cap\Phi_{1}\right\}\right]\\ &=\sup_{\substack{A\subset\Phi\\ A\text{ closed}}}\left[-\mathbb{P}^{\varepsilon}\left\{\mathbf{Y}\in(A^{\alpha}\setminus A)\cap\Phi_{1}\right\}\right]=0,\end{aligned}\quad(15)$$

since this supremum is reached when $A\cap\Phi_{1}=A^{\alpha}\cap\Phi_{1}=\Phi_{1}$.

Combining (13)-(15), we obtain

$$W_{p}^{p}\left(\mathbb{P}^{\varepsilon,*},\mathbb{P}^{*}\right)$$
$$\leq C^{p}\lim_{\alpha\to0}\sup_{\substack{A\subset\Phi\\ A\text{ closed}}}\left[\mathbb{P}^{\varepsilon}\left\{\mathbf{Y}\in A\cap\Phi_{2}\right\}-\mathbb{P}^{*}\left\{\mathbf{Y}^{*}\in A^{\alpha}\cap\Phi_{2}\right\}\right].\quad(16)$$

Consider all those sets A for which paths from A^{α} do not intersect $\mathbf{\Gamma}_{\partial\mathbf{V}}$ for any α less than some $\alpha_{0}>0$. Denote this subset of Φ_{2} by Φ_{0}. Since all paths start inside $\mathbf{\Gamma}_{\mathbf{U}}$, it simply means that Φ_{0} consists of those paths which always stay away from the critical set $\overline{\mathbf{V}}$ of the Hamiltonian. Then $\Phi\setminus\Phi_{0}$ consists of such sets A whose any enlargement A^{α} contains paths intersecting the boundary $\mathbf{\Gamma}_{\partial\mathbf{V}}$ of the critical set. This implies that $A\subset\Phi\setminus\Phi_{0}$ itself contains paths intersecting $\mathbf{\Gamma}_{\partial\mathbf{V}}$. Since A is a closed set, if this were not the case, there would exist an *open* enlargement A^β of A not containing any paths intersecting $\mathbf{\Gamma}_{\partial\mathbf{V}}$. But this is not possible, since any closed enlargement $A^{\beta/2}$ always contains at least one such path. From inequality (16), we obtain

$$\begin{aligned} W_p^p\left(\mathbb{P}^{\varepsilon,*},\mathbb{P}^*\right) &\\ \le C^p \max\Bigg\{ &\sup_{\substack{A\subset\Phi_0\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\cap\Phi_2\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A\cap\Phi_2\right\}\Big],\\ &\lim_{\alpha\to 0}\sup_{\substack{A\subset\Phi\setminus\Phi_0\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\cap\Phi_2\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A^\alpha\cap\Phi_2\right\}\Big]\Bigg\}. \end{aligned} \tag{17}$$

Using the observation that $\mathbb{P}^*\left\{Y^*\in A^\alpha\cap\Phi_2\right\}$ is monotone in α, from (17), we see that

$$\begin{aligned} W_p^p\left(\mathbb{P}^{\varepsilon,*},\mathbb{P}^*\right) &\\ \le C^p \max\Bigg\{ &\sup_{\substack{A\subset\Phi_0\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A\right\}\Big],\\ &\lim_{\alpha\to 0}\sup_{\substack{A\subset\Phi_2\setminus\Phi_0\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A^\alpha\cap\Phi_2\right\}\Big]\Bigg\}\\ = C^p \max\Bigg\{ &\sup_{\substack{A\subset\Phi_0\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A\right\}\Big],\\ &\sup_{\substack{A\subset\Phi_2\setminus\Phi_0\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A\right\}\Big]\Bigg\}. \end{aligned} \tag{18}$$

Next, simply notice that the maximum variation cannot be achieved over those paths which are not allowed to intersect $\mathbf{\Gamma}_{\partial\mathbf{V}}$, and so

$$\begin{aligned} W_p^p\left(\mathbb{P}^{\varepsilon,*},\mathbb{P}^*\right) &\le C^p \sup_{\substack{A\subset\Phi_2\setminus\Phi_0\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A\right\}\Big]\\ &\le C^p \sup_{\substack{A\subset\overline{\Phi}\\ A \text{ closed}}}\Big[\mathbb{P}^\varepsilon\left\{\mathbf{Y}\in A\right\}-\mathbb{P}^*\left\{\mathbf{Y}^*\in A\right\}\Big] \end{aligned} \tag{19}$$

To obtain the lower bound, use the definition of the Wasserstein L_p-distance once again, to notice that

$$
\begin{aligned}
W_p^p\left(\mathbb{P}^{\varepsilon,*},\mathbb{P}^*\right) &= \inf_{\mu\in\Pi(\mathbb{P}^{\varepsilon,*},\mathbb{P}^*)} \int_{\Phi\times\Phi} \rho\left(\phi,\psi\right)^p d\mu(\phi,\psi) \\
&\geq \inf_{\mu\in\Pi(\mathbb{P}^{\varepsilon,*},\mathbb{P}^*)} \int_{\Phi_0\times\Phi_0} \rho\left(\phi,\psi\right)^p d\mu(\phi,\psi) \\
&= \inf_{\mu\in\Pi\left(\mathbb{P}_1^{\varepsilon,*},\mathbb{P}_1^*\right)} \int_{\Phi_0\times\Phi_0} \rho\left(\phi,\psi\right)^p d\mu(\phi,\psi) \\
&= W_p^p\left(\mathbb{P}_1^{\varepsilon,*},\mathbb{P}_1^*\right).
\end{aligned}
\tag{20}
$$

This completes the proof of this theorem. ■

The two subsequent sections are devoted to two different examples of estimating the Wasserstein L_p-distance in the classical case of a one-well Hamiltonian. The first method allows us to obtain crude estimates which, after computing the distance explicitly, turn out to be somewhat satisfying.

6 First Example — Hamiltonian Cylinder: Initial Naive Estimates

Let $\mathbf{W}$ be a Brownian motion on a probability space $(\Omega,\mathcal{F},\mathbb{P})$, and let $\varepsilon>0$ be a small parameter. Consider the dynamics of $Z_t^\varepsilon=(\Theta_t^\varepsilon,Y_t^\varepsilon)$ with its trajectories on a cylinder,

$$
\begin{aligned}
d\Theta_t^\varepsilon &= \frac{1}{\varepsilon^2}\,dt, \\
dY_t^\varepsilon &= \sigma\left(\Theta_t^\varepsilon\right)\,d\mathbf{W}_t,
\end{aligned}
\tag{21}
$$

with certain initial conditions $(\Theta_0^\varepsilon,Y_0^\varepsilon)=(\theta_0,y_0)$, where σ is a smooth bounded 1-periodic function. Here we can think of Z^ε as the rescaled graph of Y^ε in cylindrical time.

The corresponding deterministic system is given in its normal form,

$$
d\Theta_t = dt, \quad dY_t = 0,
$$

and is trivially Hamiltonian. The variables Θ and Y are the so-called action-angle variables. In this deterministic case, the action Y takes a constant value y independent of time t, and the angle Θ increases linearly with time, at a constant rate $\omega(y)\equiv 1$. As for the stochastic system (21), the angular rotation is sped-up by a factor of ε^{-2} (or, equivalently, time is appropriately rescaled), and is much faster than the vertical displacement. This vertical slow motion is described by

$$
Y_t^\varepsilon = y_0 + \int_0^t \sigma\left(\theta_0+\frac{s}{\varepsilon^2}\right)\,d\mathbf{W}_s \tag{22}
$$

for all $t\geq 0$.

On the canonical probability space $C([0,\infty),\mathbb{S}^1\times\mathbb{R})$ [2] of the continuous trajectories on the cylinder $\mathbb{S}^1\times\mathbb{R}$, for each small $\varepsilon>0$, and every Borel-measurable set $A\times B\in\mathcal{B}(C([0,\infty),\mathbb{S}^1\times\mathbb{R}))$, we define the following probability measure

$$\mathbb{Q}^\varepsilon(A\times B)=\mathbb{P}\{Z^\varepsilon\in A\times B\}.$$

This measure is Markovian, and uniquely solves the martingale problem for $(\mathcal{L}^\varepsilon,\delta_0)$, where δ_0 is a point-mass at $(\theta_0,y_0)\in\mathbb{S}^1\times\mathbb{R}$. For any $(\theta,y)\in\mathbb{S}^1\times\mathbb{R}$, and $f\in C_b^2(\mathbb{S}^1\times\mathbb{R})$,

$$\mathcal{L}^\varepsilon f(\theta,y):=\frac{1}{\varepsilon^2}\frac{\partial f(\theta,y)}{\partial\theta}+\frac{\sigma^2(\theta)}{2}\frac{\partial^2 f(\theta,y)}{\partial y^2}. \tag{23}$$

In other words, if $0\le s\le t$, and $f\in C_b^2(\mathbb{S}^1\times\mathbb{R})$, then

$$\begin{aligned}&f(Z_t^\varepsilon)-f(Z_s^\varepsilon)-\int_s^t\mathcal{L}^\varepsilon f(Z_u^\varepsilon)\,du\\&=f(Z_t^\varepsilon)-f(Z_s^\varepsilon)-\frac{1}{2}\int_s^t\sigma^2\left(\Theta_u^\varepsilon\right)\frac{\partial^2 f}{\partial y^2}(Z_u^\varepsilon)\,du\\&=\int_s^t\sigma\left(\Theta_u^\varepsilon\right)\frac{\partial f}{\partial y}(Z_u^\varepsilon)\,d\mathbf{W}_u,\end{aligned}$$

is a $\mathbb{Q}^\varepsilon$-martingale.

Also, we construct a projection measure $\mathbb{Q}^{\varepsilon,*}$ on $C([0,\infty),\mathbb{R})$ by setting

$$\mathbb{Q}^{\varepsilon,*}(B)=\mathbb{Q}^\varepsilon(\mathbb{S}^1\times B)=\mathbb{P}\{Y^\varepsilon\in B\}.$$

Notice that although a projection measure of a Markovian measure does not need to be Markovian itself, in this case it is Markovian simply due to the fact that the angular movement of the system on the cylinder is deterministic.

Let $M_t^\varepsilon=\int_0^t\sigma(\theta_0+\frac{s}{\varepsilon^2})d\mathbf{W}_s$, then its quadratic variation is $\langle M^\varepsilon\rangle_t=\int_0^t\sigma^2(\theta_0+\frac{s}{\varepsilon^2})ds$. We can perform a time change, such that given $\varepsilon>0$, there exists a Brownian motion $\mathbf{B}$ such that $M_t^\varepsilon=\mathbf{B}_{\langle M^\varepsilon\rangle_t}$.

Now, we introduce a Markov process Y^* whose trajectories have dynamics described by

$$dY_t^*=\hat\sigma\,d\mathbf{B}_t,\quad Y_0^*=y_0, \tag{24}$$

where $\hat\sigma=\left(\int_0^1\sigma^2(u)\,du\right)^{1/2}$, and we also introduce the Markovian measure

$$\mathbb{Q}^*(B):=\mathbb{P}\{Y^*\in B\}$$

on the canonical probability space $\Big(C([0,\infty),\mathbb{R}),\mathcal{B}(C([0,\infty),\mathbb{R}))\Big)$.

Recall that according to the classical averaging argument, due to Has'minskiĭ, the law of the process Y^ε weakly converges to the law of a Markov process.

The next theorem in particular reveals that this unique limiting measure is the law of Y^*.

Theorem 3 *The Wasserstein L_p-distance between the laws of the processes Y^ε and Y^* is decreasing faster than $\varepsilon^{1-\alpha}$ for any $\alpha > 0$. For any $1 \leq p \leq \infty$, the distance decreases together with the modulus of continuity of the standard Brownian motion, and*

$$W_p(\mathbb{Q}^{\varepsilon,*}, \mathbb{Q}^*) \leq 2\,C\,\varepsilon \left(\log \frac{1}{2\,C^2\varepsilon^2}\right)^{1/2}, \tag{25}$$

where $C = \sup_{0\leq s\leq 1} |\sigma(s)|$.

Proof: Since the Wasserstein L_p-distance is the minimizer over all the product measures from $M_p(C([0,\infty);\mathbb{R}) \times C([0,\infty);\mathbb{R}))$ whose marginals are $\mathbb{Q}^{\varepsilon,*}$ and $\mathbb{Q}^*$, obviously taking any such product measure should deliver an upper bound to the distance. Consider $Q^\varepsilon(A) := \mathbb{P}\left\{(Y^\varepsilon, Y^*) \in A\right\}$, then the W_p-distance between the measures induced by laws of the processes Y^ε and $\mathbb{Q}^*$ on the time interval up to a fixed finite time T can be estimated as follows:

$$\begin{aligned} W_p^p(\mathbb{Q}^{\varepsilon,*}, \mathbb{Q}^*) &\leq \int_{C([0,T];\mathbb{R})\times C([0,T];\mathbb{R})} \rho\,(x,y)^p\, dQ^\varepsilon(x,y) \\ &= \mathbb{E}\left[\sup_{0\leq t\leq T} |Y_t^\varepsilon - Y_t^*|^p\right] \\ &= \mathbb{E}\left[\sup_{0\leq t\leq T} \left|B_{\langle M^\varepsilon\rangle_t} - B_{\hat{\sigma}^2 t}\right|^p\right]. \end{aligned}$$

Using periodic properties of σ and the definition of $\hat{\sigma}$, notice that for any $0 \leq t \leq T$,

$$\begin{aligned} \left|\langle M^\varepsilon\rangle_t - \hat{\sigma}^2 t\right| &= \left|\int_0^t \left(\sigma^2\left(\theta_0 + \frac{s}{\varepsilon^2}\right) - \hat{\sigma}^2\right) ds\right| \\ &= \varepsilon^2 \left|\int_{\theta_0+\lfloor t/\varepsilon^2\rfloor}^{\theta_0+t/\varepsilon^2} \left(\sigma^2\,(s) - \hat{\sigma}^2\right) ds\right| < 2\,C^2\,\varepsilon^2, \end{aligned} \tag{26}$$

where $C = \sup_{0\leq s\leq 1} |\sigma(s)|$. Then, by using the modulus of continuity $\omega_{\mathbf{B}}$ of Brownian motion and (26), we may notice that

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \left|B_{\langle M^\varepsilon\rangle_t} - B_{\hat{\sigma}^2 t}\right|^p\right] \leq \omega_{\mathbf{B}}^p(2\,C^2\,\varepsilon^2),$$

and inequality (25) indeed holds. ∎

Remark 1 *Our argument does not make it clear whether this estimate is sharp. In fact, in the next section we show that it is not. Nevertheless, we chose to present this argument, since this method looks more promising for the stratified case. Unfortunately, the method used below to obtain exact estimates* *is not applicable in the stratified case since it relies on the one-dimensional structure of the image of the path-space. Comparing the two results, we will also observe that in the classical case, the discrepancy between our initial naive estimates and the exact distance is of the smallest possible order.*

7 Second Example — Hamiltonian Cylinder: Straightforward Computation

For simplified presentation, and without loss of generality, assume that the system starts at the point $(\theta_0, y_0) = (0, 0)$. Let $p_\varepsilon = p_\varepsilon(t, y)$ be the probability density of finding the system (21) in a state $(t/\varepsilon^2, y)$ at a given time t. Assume that $p_\varepsilon(t, y) = p_\varepsilon(t + \varepsilon^2, y)$, and $\lim_{y\to\pm\infty} p_\varepsilon(t, y) = 0$. Consider the Fokker-Planck evolution equation of this density:

$$\frac{\partial p_\varepsilon}{\partial t} = (\mathcal{L}^\varepsilon)^* \, p_\varepsilon, \tag{27}$$

where

$$(\mathcal{L}^\varepsilon)^* \, f(\theta, y) = -\frac{1}{\varepsilon^2}\frac{\partial f(\theta, y)}{\partial \theta} + \frac{\sigma^2(\theta)}{2}\frac{\partial^2 f(\theta, y)}{\partial y^2}$$

is the adjoint of $\mathcal{L}^\varepsilon$ defined in (23). Keeping this expression for the adjoint in mind, and that $\partial p_\varepsilon/\partial\theta = 0$, we can rewrite equation (27) as

$$\frac{\partial p_\varepsilon}{\partial t} = \frac{1}{2}\sigma^2\left(\frac{t}{\varepsilon^2}\right)\frac{\partial^2 p_\varepsilon}{\partial y^2}. \tag{28}$$

We may notice that this is also the Fokker-Planck equation for the probability density of finding the system (22) in a state y at time t, since the rotation is deterministic. This density is Gaussian, and we can find its explicit presentation to be

$$p_\varepsilon(t, y) = \begin{cases} \frac{1}{\sqrt{2\pi\int_0^t \sigma^2\left(\frac{s}{\varepsilon^2}\right)ds}} \exp\left\{-\frac{y^2}{2\int_0^t \sigma^2\left(\frac{s}{\varepsilon^2}\right)ds}\right\}, & \text{if } 0 < t \le \varepsilon^2, \\ \frac{1}{\sqrt{2\pi\varepsilon^2\hat{\sigma}^2}} \exp\left\{-\frac{(y-y_0)^2}{2\varepsilon^2\hat{\sigma}^2}\right\}, & \text{if } t = 0, \end{cases}$$
$$p_\varepsilon(t, y) = p_\varepsilon(t + \varepsilon^2, y).$$

Analogously, if $p_* = p_*(t, y)$ is the density of finding the system (24) in a state y at a fixed time t, then

$$\frac{\partial p_*}{\partial t} = \frac{1}{2}\hat{\sigma}^2\frac{\partial^2 p_*}{\partial y^2}, \tag{29}$$

and

$$p_*(t,y) = \begin{cases} \frac{1}{\sqrt{2\pi t\hat{\sigma}^2}} \exp\left\{-\frac{y^2}{2t\hat{\sigma}^2}\right\}, & \text{if } t > 0, \\ \delta_{y_0}(y), & \text{if } t = 0. \end{cases}$$

Let Φ denote the probability distribution induced by the standard normal density. Consider the probability distributions $F_\varepsilon = F_\varepsilon(t,y)$ and $F_* = F_* = F_*(t,y)$ induced respectively by the time-dependent densities $p_\varepsilon(t,y)$ and $p_*(t,y)$, i.e.,

$$F_\varepsilon(t,y) = \int_{-\infty}^{y} p_\varepsilon(t,h)\,dh, \quad \text{and} \quad F_*(t,y) = \int_{-\infty}^{y} p_*(t,h)\,dh.$$

For every t we can define the following two inverse distribution functions,

$$F_\varepsilon^{-1}(t,m) := \inf\{y : F_\varepsilon(t,y) > m\},$$
$$F_*^{-1}(t,m) := \inf\{y : F_*(t,y) > m\},$$

$0 \le m \le 1$. Notice that

$$F_\varepsilon^{-1}(t,m) = \begin{cases} \Phi^{-1}(m)\,\varepsilon\hat{\sigma}, & \text{if } t = 0; \\ \Phi^{-1}(m)\sqrt{\int_0^t \sigma^2(s/\varepsilon^2)\,ds}, & \text{if } t > 0. \end{cases} \tag{30}$$

Also, $F_*^{-1}(0,m) = 0$ for $m < 1$, and for $t > 0$,

$$F_*^{-1}(t,m) = \Phi^{-1}(m)\hat{\sigma}\sqrt{t}. \tag{31}$$

In order to find the Wasserstein L_p-distance between the corresponding laws of Y^ε and Y^*, we will employ the well-known Hoeffding lemma [18], which is most often used in the theory of mass transportation. The lemma implies that the W_p-distance between these two laws can be viewed as the distance between the corresponding pseudo-inverse distribution functions. More explicitly, in our case, the W_p-distance between the laws of Y^ε and Y^* over an interval $[0,T]$ satisfies

$$W_p^p\left(\mathbb{Q}^{\varepsilon,*}, \mathbb{Q}^*\right) = \int_0^1 \sup_{0\le t\le T} \left|F_\varepsilon^{-1}(t,m) - F_*^{-1}(t,m)\right|^p dm. \tag{32}$$

Using expressions (30) and (31),

$$\begin{aligned} &W_p\left(\mathbb{Q}^{\varepsilon,*}, \mathbb{Q}^*\right) \\ &= \max\left\{\varepsilon\,\hat{\sigma}, \sup_{0<t\le T}\left|\sqrt{\int_0^t \sigma^2(s/\varepsilon^2)\,ds} - \hat{\sigma}\sqrt{t}\right|\right\}\left(\int_0^1 |\Phi^{-1}(m)|^p\,dm\right)^{1/p}. \end{aligned}$$

If $0 < t < \varepsilon^2$, then $\int_0^{t/\varepsilon^2} \sigma^2(s)\, ds < \hat{\sigma}$, and

$$0 < \hat{\sigma}\sqrt{t} - \sqrt{\int_0^t \sigma^2(s/\varepsilon^2)\, ds} = \hat{\sigma}\sqrt{t} - \varepsilon\sqrt{\int_0^{t/\varepsilon^2} \sigma^2(s)\, ds}$$
$$< \varepsilon \left(\hat{\sigma} - \sqrt{\int_0^{t/\varepsilon^2} \sigma^2(s)\, ds} \right) < \varepsilon\, \hat{\sigma}.$$

If $t \geq \varepsilon^2$, use $|a-b| = \frac{|a^2-b^2|}{a+b}$ for $a, b > 0$, to notice that

$$\left| \hat{\sigma}\sqrt{t} - \sqrt{\int_0^t \sigma^2(s/\varepsilon^2)\, ds} \right| \leq \frac{\varepsilon^2\, \hat{\sigma}^2}{\hat{\sigma}\left(\sqrt{t} + \sqrt{t - \varepsilon^2}\right)} \leq \varepsilon\, \hat{\sigma}.$$

Therefore,

$$W_p\left(\mathbb{Q}^{\varepsilon,*}, \mathbb{Q}^*\right) = C_p\, \varepsilon$$

with $C_p = \hat{\sigma}\, (\mathbb{E}|Z|^p)^{1/p}$, where Z is a standard normal random variable.

As this computation delivered the explicit expression for the W_p-distance, now we can indeed see that Theorem 3 did not provide the sharp estimate for the rate of convergence of these laws, and the following result holds on a cylinder.

Theorem 4 *If $p > 1$, then the Wasserstein L_p-distance between the laws of the processes Y^ε and Y^* is equal to $C_p\, \varepsilon$, where $C_p = \hat{\sigma}\, (\mathbb{E}|Z|^p)^{1/p}$, with Z having the standard normal law.*

8 Open Questions about the Stratified Problem

As we have computed the L_p-distance in the classical case, we see that $W_p(\mathbb{P}^{\varepsilon,*}, \mathbb{P}^*)$ is at least of the order ε, and at most of the order delivered by the total variation of the measures $\mathbb{P}^{\varepsilon,*}$ and $\mathbb{P}^*$ over the paths crossing the boundary of the critical set of the degenerate Hamiltonian.

How is this total variation connected with the gluing conditions? And is it?

The analysis of the convergence of these measures [16] relies on a finer set of so-called Has'minskiĭ's coordinates, which are used about the boundary of the critical set. Is it possible to obtain an explicit bound on the total variation of these laws using Has'minskiĭ's coordinates? This question is likely to be connected to the previous one, and if answered positively, will reveal the role that gluing plays in this matter.

References

[1] Richard F. Bass and Krzysztof Burdzy, 2000. Fiber Brownian motion and the "hot spots" problem. *Duke Math. J.*, 105(1):25–58.

[2] Stewart N. Ethier and Thomas G. Kurtz, 1986. *Markov processes.* John Wiley & Sons Inc., New York. Characterization and convergence.

[3] Steven N. Evans, 2000. Snakes and spiders: Brownian motion on **r**-trees. *Probab. Theory Related Fields*, 117(3):361–386.

[4] Steven N. Evans and Richard B. Sowers, 2003. Pinching and twisting Markov processes. *Ann. Probab.*, 31(1):486–527.

[5] M. I. Freidlin and A. D. Wentzell, 1998. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, New York.

[6] Mark I. Freidlin and Alexander D. Wentzell, 1994. Random perturbations of Hamiltonian systems. *Mem. Amer. Math. Soc.*, 109(523):viii+82.

[7] Mark Goresky and Robert MacPherson, 1988. *Stratified Morse theory.* Springer-Verlag, Berlin.

[8] R. Z. Has'minskiĭ, 1963. Diffusion processes with a small parameter. *Izv. Akad. Nauk SSSR Ser. Mat.*, 27:1281–1300.

[9] Richard Jordan, David Kinderlehrer, and Felix Otto, 1998. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17 (electronic).

[10] N. Sri Namachchivaya and Richard B. Sowers, 2001. Unified approach for noisy nonlinear Mathieu-type systems. *Stoch. Dyn.*, 1(3):405–450.

[11] Natella V. O'Bryant, 2003. A noisy system with a flattened Hamiltonian and multiple time scales. *Stoch. Dyn.*, 3(1):1–54.

[12] Natella V. O'Bryant, 2004. Doudble-level averaging on a stratified space, in: R.J. Swift A.C. Krinik (Eds), *Stochastic Processes and Functional Analysis*, 238:277–294.

[13] G. C. Papanicolaou and W. Kohler, 1974. Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math.*, 27:641–668.

[14] Svetlozar T. Rachev, 1991. *Probability metrics and the stability of stochastic models.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester.

[15] Clark Robinson, 1999. *Dynamical systems.* CRC Press, Boca Raton, FL, second edition. Stability, symbolic dynamics, and chaos.

[16] Richard B. Sowers, 2002. Stochastic averaging with a flattened Hamiltonian: a Markov process on a stratified space (a whiskered sphere). *Trans. Amer. Math. Soc.*, 354(3):853–900 (electronic).

[17] Daniel W. Stroock and S. R. Srinivasa Varadhan, 1979. *Multidimensional diffusion processes.* Springer-Verlag, Berlin.

[18] Cédric Villani, 2003. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics.* American Mathematical Society, Providence, RI.

[19] Marc Yor, 1997. *Some aspects of Brownian motion. Part II.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel.

[20] Vladimir M. Zolotarev, 1997. *Modern theory of summation of random variables.* Modern Probability and Statistics. VSP, Utrecht.