ABSTRACT

Title of Dissertation:	RARE EVENTS AND EXIT PROBLEMS FOR STOCHASTIC EQUATIONS: THEORY AND NUMERICS
	Nicholas Francis Schafer Paskal Doctor of Philosophy, 2021
Dissertation Directed by:	Professor Sandra Cerrai Professor Maria Cameron Department of Mathematics

This dissertation is concerned with the small-noise asymptotics of stochastic differential equations and stochastic partial differential equations. In the first part of the manuscript, we present an overview of large deviations theory in the context of stochastic differential equations, with a particular focus on describing the long time behavior of the system in the presence of point attractors. We then present results describing a novel algorithm for computing the *quasi-potential*, a key quantity in large deviations theory, for two-dimensional stochastic differential equations. Our solver, the Efficient Jet Marcher, computes the quasi-potential $U(x)$ on a mesh by propagating U and ∇U outward away from the attractor. By using higher-order interpolation schemes and approximations for the minimum action paths, we are able to achieve 2nd order accuracy in the mesh spacing h.

In the second part of the manuscript, we consider two important problems in large deviations theory for stochastic partial differentiable equations. First, we consider stochastic reaction-diffusion equations posed on a bounded domain, which contain both a large diffusion term and a small noise term. We prove that in the joint small noise and large diffusion limits, the system satisfies a large deviations principle with respect to an action functional that is finite only on paths that are constant in the spatial variable. We then use this result to compute asymptotics of the first exit time of the solution from bounded domains in function spaces. Second, we consider the two-dimensional stochastic Navier-Stokes equations posed on the torus. In the simultaneous limit as the noise magnitude and noise regularization are both sent to 0, the solutions converge to the deterministic Navier-Stokes equations. We prove that the invariant measures, which converge to a Dirac mass at 0, also satisfy a large deviation principle with action functional given by the enstrophy.

RARE EVENTS AND EXIT PROBLEMS FOR STOCHASTIC EQUATIONS: THEORY AND NUMERICS

by

Nicholas Francis Schafer Paskal

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2021

Advisory Committee:
Professor Sandra Cerrai, Chair/Advisor
Professor Maria Cameron, Co-Chair/Co-Advisor
Professor Mark Freidlin
Professor Jacob Bedrossian
Professor Leonid Koralov
Professor Johan Larsson

Dedication

Dedicated to my mother.

Acknowledgments

First, I thank my advisor, Dr. Sandra Cerrai, for introducing me to the field of probability theory and guiding my way through graduate school, and my co-advisor Dr. Maria Cameron, for introducing me to numerics and helping me end strong when graduation was in sight. I thank the members of my dissertation committee: Dr. Mark Freidlin, Dr. Jacob Bedrossian, Dr. Leonid Koralov and Dr. Johan Larsson for all of the indispensable help they have given along the way. Lastly, I thank all of my friends for providing comic relief and, most of all, my family for supporting me throughout all the ups and downs.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the dissertation

The mathematical modeling of rare events in physical systems is a difficult, yet practically important problem. In many real-world applications, it is often the unexpected outcomes that have the greatest impact – whether it is the spontaneous failure of an engine or a short-term flash crash of the stock market [52]. Both data-based and simulation-based methods for estimating rare event probabilities typically suffer from the issue of small observation counts, which tend to yield estimates of very low statistical power. Significant research is conducted into fields like importance sampling [11] to improve our ability to observe rare events via simulation, but the presence of small sample sizes remains as a significant computational hurdle.

Instead, this dissertation takes the perspective of quantifying probabilities of rare events by applying analytic tools directly to the probabilistic mathematical models themselves. These types of results are typically placed under the umbrella of *large deviations theory*, a general framework for quantifying rare events in probabilistic systems. The earliest ideas in the theory date back to Harald Cramer in the early 1900s, who priced insurance premiums by estimating the probabilities and potential amounts of future insurance claims [45]. The general abstract framework for the theory is first attributed to Varadhan in 1966 [65], and the ideas have since been applied to all sorts of asymptotic problems in probability theory.

The content of this manuscript concerns the theory of large deviations applied to the setting of dynamical systems perturbed by small amounts of random noise, a field referred to as Freidlin-Wentzell theory as it was pioneered by Freidlin and Wentzell in the 1960s and 1970s [34, 35, 37]. This particular branch of probability theory and differential equations has attracted immense attention over the last 40 years and a number of sophisticated techniques have been developed. In Chapter 2 we present a brief summary of some of the main results, which form the foundation for our work in the later chapters.

The original work in Chapters 3, 4 and 5 can be broken down into two categories within the Freidlin-Wentzell theory umbrella. In Chapter 3, we consider the *computational* problem of finding numerically some of the important quantities in the theory. In particular, we design a novel technique for numerically evaluating the quasi-potential, an energy function that describes the leading order asymptotics of rare event probabilities in the small-noise limit.

On the other hand, Chapters 4 and 5 consider *theoretical* problems of proving large deviations type results in the small-noise limit for *multi-scale* stochastic partial differential equations. The goal of our work there is to study the interaction between the small-noise limit and other asymptotic features of infinite-dimensional equations. Specifically, in Chapter 4 we consider the interaction between the small-noise limit and *fast transport* limit for stochastic reaction-diffusion equations, while in Chapter 5 we consider the simultaneous small-noise and *singular* noise limits for the stochastic Navier-Stokes equations in dimension 2 on the torus. In the remainder of this chapter, we present an executive summary of the body of this manuscript.

1.1 Large deviations theory for stochastic equations

In Chapter 2, we survey many of the key ideas in Freidlin-Wentzell theory, upon whose framework the work in the later chapters is built. There, we discuss asymptotic properties of the solution X_t^ϵ to the *stochastic* differential equation in $\mathbb{R}^d$

$$dX_t^\epsilon = b(X_t^\epsilon)dt + \sqrt{\epsilon}W_t, \qquad X_0 = x, \tag{1.1.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field, W_t is a d-dimensional Brownian motion, $x \in \mathbb{R}^d$ and $\epsilon > 0$ is a small parameter. The solution X_t^ϵ can be viewed as a perturbation of the solution X_t to the *ordinary* differential equation

$$dX_t = b(X_t)dt, \qquad X_0 = x. \tag{1.1.2}$$

Indeed, one can prove that for any $T > 0$ the solution X_t^ϵ of the stochastic equation (1.1.1) converges in probability as $\epsilon \downarrow 0$ in the space $C([0, T]; \mathbb{R}^d)$ to the solution X_t of the deterministic equation (1.1.2).

The primary goal of large deviations theory in this context is to compute leading order asymptotics of the probabilities that X_t^ϵ experiences significant deviations from its "expected" behavior X_t. These probabilities will in general decay exponentially in ϵ^{-1} as $\epsilon \to 0$ with an exponential rate that can be determined by a *least action* principle. For example, suppose that

$$\mathcal{A} := \{\varphi \in C([0,T]) : \max_{t\in[0,T]} \|\varphi_t - X_t\| > \delta\},$$

represents the set of "rare" trajectories consisting of paths φ that deviate from the expected path X_t by at least δ for some $t \in [0,T]$. Due to the aforementioned convergence in probability $\epsilon \to 0$, the probability $\mathbb{P}(X_t^\epsilon \in \mathcal{A})$ certainly decreases to 0, but in addition its leading order asymptotics can be determined via large deviation theory. Namely,

$$\mathbb{P}(X_t^\epsilon \in \mathcal{A}) \asymp \exp\Big(-\frac{1}{\epsilon}\inf_{\varphi\in\mathcal{A}} \frac{1}{2}\int_0^T \|\dot{\varphi}_t - b(\varphi_t)\|^2\, dt\Big), \tag{1.1.3}$$

where $\|\cdot\|$ denotes Euclidean norm on $\mathbb{R}^d$, while $\asymp$ denotes logarithmic equivalence in the $\epsilon \to 0$ limit (see Chapter 2). An important observation here is that the probability of X_t^ϵ realizing the rare event $\mathcal{A}$ *only* depends on the particular paths $\varphi \in \mathcal{A}$ that minimize the action function (the integral in (1.1.3)). In other words, the leading order term of the probability of event $\mathcal{A}$ is determined entirely by its least unlikely member, which can be determined by finding the solution to a least action problem.

The process X_t^ϵ is said to satisfy a large deviations principle if statements of the form (1.1.3) hold for a general class of events $\mathcal{A}$ (see Section 2.1). The majority of Chapter 2 discusses techniques for proving large deviations principles for equation (1.1.1). Considerable focus is given towards viewing large deviations from the weak convergence perspective, an approach that is utilized several times in our main theoretical results in Chapters 4 and 5.

Considerable attention is also given to the long-time dynamics of equation (1.1.1) in the case where the drift field b possesses a stable, globally-attracting equilibrium $\mathcal{O}$. In such a case, the solution X_t^ϵ will typically stay very close to the attractor $\mathcal{O}$ for exponentially in ϵ^{-1} long periods of time, before occasionally experiencing $\mathcal{O}(1)$ sized excursions. The central tool in the analysis of such systems is the time-stationary probability measure μ^ϵ of equation (1.1.1). While the measures μ^ϵ may not individually be Gibbs measure for fixed $\epsilon > 0$, it turns out that in the $\epsilon \to 0$ limit, the measures μ^ϵ asymptotically resemble Gibbs measures

$$\frac{d\mu^\epsilon}{dx} \asymp \exp\Big(-\frac{U(x)}{\epsilon}\Big), \tag{1.1.4}$$

for some function $U(x)$.

This near-*Gibbsian* structure will allow for the statement and proof of powerful estimates of the first escape time of the process X_t^ϵ from an arbitrary bounded set. In a similar fashion, one can make asymptotic estimates on the frequency of noise-induced transitions between different attractors of b in the case where b has multiple attractors, known as *meta-stable* transitions. The exponent $U(x)$ in (1.1.4), called the *quasi-potential*, features prominently in all of these asymptotic estimates, and much of the later part of Chapter 2 is devoted to providing a thorough description of $U(x)$ and its important properties. In short, this $U(x)$ is the minimum of the action integral (in equation (1.1.3)) over all times $T > 0$ and paths $\varphi \in C([0,T];\mathbb{R}^d)$ such that $U(0) = \mathcal{O}$ and $U(T) = x$. In practical terms, the quasi-potential $U(x)$ provides the cheapest cost of moving from the attractor $\mathcal{O}$ to the point x.

We emphasize that *none* of the work in Chapter 2 is original work, but rather is intended purely to be expository preparation for the remaining chapters. Much of the chapter is attributable to Freidlin and Wentzell, and most of the results can be found in more detail in Chapter 3, 4 and 6 of their book [37]. The weak convergence approach is attributable to Dupuis and Ellis, among others, and can be found in the book [30] and articles [7, 12, 13].

1.2 *Efficient Jet Marcher for computing the quasi-potential*

Chapter 3 is concerned with the development of numerical tools for computing the quasi-potential $U(x)$ for equation (1.1.1) in 2-dimensions. The quasi-potential is important for quantifying much of the long-time asymptotic behavior of (1.1.2) (see Section 2.3.4 and Ch. 4 of [37]), and is of particular interest in biological and ecological systems exhibiting *metastability*, a phenomenon in which there occasionally occur noise-induced transitions between stable attractors. For example, in [49], they use a 2-dimensional nonlinear SDE to model the populations of two interacting plankton species. Their model supports multiple stable population configurations so that metastable transitions may occur. In order to quantify the expected transition rate, they numerically compute the quasi-potential in the relevant region of phase space.

In practice, there are two primary families of methods for solving for the quasi-potential. The first category of methods, which we refer to as path-based techniques, attempt to numerically minimize the action integral directly in order to compute the *minimum action path* (MAP) from the attractor $\mathcal{O}$ to a given point x. The quasi-potential $U(x)$ is then computed by numerical integration of the action along this MAP. Prominent path-based methods include the Adaptive Minimum Action Method [67] and Geometric Minimum Action Method [38].

The second family of quasi-potential solvers consists of what we refer to as *mesh-based* solvers. These solvers attempt to solve for the quasi-potential directly on an entire region of space by numerically finding the quasi-potential solution to the Hamilton-Jacobi equation (see Section 2.3.4)

$$\|\nabla U(x)\|^2 + \frac{1}{2} b(x) \cdot \nabla U(x) = 0, \quad U(\mathcal{O}) = 0. \tag{1.2.1}$$

Mesh-based quasi-potential solvers typically treat equation (1.2.1) like an eikonal equation and possess a structure similar to Sethian's Fast Marching method [57]. These methods work by discretizing the domain into a mesh and computing $U(x)$ at a mesh point x from the values of $U(y)$ at nearby mesh points y by locally approximating a portion of the MAP from $\mathcal{O}$ to x. Since these techniques solve for $U(x)$, they are largely restricted to 2 and 3 dimension in practice. We describe in complete detail the structure of these two families of methods in Chapter 3.

Our main novel result in Chapter 3 is the development of a efficient mesh-based quasi-potential solver in two-dimensions that is second order accurate $\mathcal{O}(h^2)$ in the mesh spacing h. This algorithm, the Efficient Jet Marcher (EJM), is the first such 2nd order method for computing the quasi-potential, to our knowledge. Previous quasi-potential solvers [21, 22] in general displayed $\mathcal{O}(h)$ convergence rates with superconvergence on highly rotational drift fields.

Our solver contains two distinguishing features. The first, inspired by the Jet solver [54], involves an approximation of the MAP between mesh points as a cubic curve, paired with a cubic interpolation of the quasi-potential $U(x)$ *mesh points*. Past methods have been restricted to the use of linear approximations and linear interpolations, respectively. The higher order interpolation, which is the key to overcoming the $\mathcal{O}(h)$ bottleneck, is possible by letting ∇U be part of the solution along with U and by using the following geometric relation between $U(x)$ and the MAP ϕ passing from $\mathcal{O}$ to x (see Section 2.3.4):

$$\nabla U(x) = \|b(x)\| \, \dot{\phi} - b(x). \tag{1.2.2}$$

The second key feature is the use of anisotropic stencils, inspired by [47], to determine *which* neighboring mesh points of x one should use to search for the MAP. These anisotropic stencils form a smaller and more targeted neighborhood than traditionally used neighborhoods, and they allow for a significant reduction in the number of MAP searches and hence, runtime.

1.3 Exit problem for fast transport Reaction-Diffusion equations

In Chapter 4, we move to the infinite dimensional setting of stochastic partial differential equations. We consider the following multi-scale stochastic reaction-diffusion equation posed on a bounded domain $\mathcal{D}$ in $\mathbb{R}^d$ with Neumann boundary conditions,

$$\begin{cases} \partial_t u(t,x) = \dfrac{1}{\epsilon}\Delta u(t,x) + f(u(t,x)) + \sqrt{\epsilon} g(u(t,x))\partial_t W(t,x), \\ \nabla u(t,x) \cdot \hat{n}|_{\partial\mathcal{D}} = 0, \qquad u(0,x) = u_0(x), \end{cases} \tag{1.3.1}$$

where $\epsilon \ll 1$ is a small parameter and W is a Wiener process in $L^2(\mathcal{D})$ with covariance operator Q. Actually, in Chapter 4 we consider a more general version of equation (1.3.1) containing an arbitrary elliptic operator $\mathcal{A}$ as well as additional noise acting only on the boundary; however, we restrict here to the simpler version (1.3.1) for expository purposes.

The most notable feature of this model (1.3.1) is the presence of three distinct scales: the *large* fast transport term Δ/ϵ, the *moderate* non-linear reaction term $f(u)$, and the *small* stochastic reaction term $\sqrt{\epsilon}g(u)dW$. Such a scaling is relevant for chemical systems where local changes in concentration diffuse extremely rapidly and are almost immediately "averaged" across the domain $\mathcal{D}$. In fact, the interaction between this averaging effect and the small-noise limit is the key mathematical phenomena we investigate in this chapter.

One can view the rapid averaging as an agent that reduces the complex infinite dimensional dynamics into something more akin to a finite dimensional problem. Indeed, one may be interested in approximating the SPDE with a simpler ordinary differential equation that treats the concentration u as constant in space. It can be shown [16] that the solutions converge to the solution of the 1-dimensional ordinary differential equation

$$\frac{d\bar{u}}{dt}(t) = \bar{f}(\bar{u}), \qquad \bar{u}(0) = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} u_0(x)dx,$$

where $\bar{f}(r) = \frac{1}{|\mathcal{D}|}\int_{\mathcal{D}} f(u)dx$ is the averaged version of reaction function f. We remark that the integration in the averaging is performed with respect to the invariant measure of the semi-group $e^{t\Delta}$, which happens to be the scaled Lebesgue measure on $\mathcal{D}$. However, for the case of a general elliptic operator $\mathcal{A}$ which we will consider, the invariant measure for the semi-group $e^{t\mathcal{A}}$ will be some other measure on $\mathcal{D}$ that is absolutely continuous with respect to Lebesgue measure.

Our first main result in Chapter 4 (formally from [18]) is that the solutions to (1.3.1) satisfy a large deviations principle in the space $C([\delta, T]; L^2(\mathcal{D}))$ for any $T > 0$ and any positive $\delta > 0$. The action functional for the large deviations principle, which is spelled out explicitly in Chapter 4, is only finite on paths $\varphi \in C([\delta, T]; L^2(\mathcal{D}))$ taking values in the 1-dimensional subspace of $L^2(\mathcal{D})$ consisting of *constant* functions. The requirement of considering intervals $[\delta, T]$ rather than $[0, T]$ is to allow the diffusion time to "dissipate" all of the non-constant modes of the initial condition. In fact, if the initial condition u_0 is a constant function, then the large deviation principle will hold in $C([0, T]; L^2(\mathcal{D}))$.

We then consider the problem of the exit of the solution of (1.3.1) from a bounded subdomain G of $L^2(\mathcal{D})$, supposing that the "averaged" nonlinearity $\bar{f}$ contains a stable globally attracting equilibrium. Our second main result is the proof of the statement

$$\lim_{\epsilon\to 0}\epsilon\log\mathbb{E}\tau^\epsilon = \inf_{x\in G} U(x),$$

where U is an appropriately defined quasi-potential that is finite *only* on constant functions in $L^2(\mathcal{D})$. To prove this, we follow a modified version of the finite-dimensional strategy given in [37] and [28] (see Section 2.3.1 for a sketch). A typical issue when extending this strategy into infinite dimensions is the requirement that the limits in the large deviations principle must be uniform with respect to any initial conditions of the SPDE in a bounded set of $L^2(\mathcal{D})$. In general, there exist a number approaches for obtaining this uniformity with respect to compact sets (see [13], [25]), but it is trickier to obtain for bounded sets. For (1.3.1), we are able to solve this issue by utilizing the fact that the spatial averaging turns bounded sets in $L^2(\mathcal{D})$ into "nearly" compact sets in a very short amount of time.

1.4 Large deviations for the invariant measure of the stochastic Navier-Stokes equations

In the final chapter, we study the following 2-D stochastic Navier-Stokes equations on the periodic box $[0,2\pi]^2$:

$$\begin{cases}\partial_t u(t,x) + (u(t,x)\cdot\nabla)u(t,x) = \Delta u(t,x) + \nabla p(t,x) + \sqrt{\epsilon}\partial_t W_{\delta(\epsilon)}(t,x), \\ \text{div } u(t,x) = 0, \qquad u(0,x) = u_0(x), \qquad u \text{ is periodic.}\end{cases} \tag{1.4.1}$$

Here, $\epsilon \ll 1$ is a small parameter and W_ϵ is a Wiener process in $H = (L^2([0,2\pi]^2))^2$ with spatial correlation on the scale of δ. As the correlation scale $\delta = \delta(\epsilon)$ is sent to 0, the driving force $\partial_t W_{\delta(\epsilon)}$ tends to the space-time white noise. Since the Navier-Stokes equations driven by space-time white noise are not well-posed in $C([0,T];H)$ in dimension greater than 1, problem (1.4.1) is a singular limit problem. Imposing the spatial correlation $\delta(\epsilon)$ allows us to study this singular limit by instead considering the regularized problem in the space $C([0,T];H)$ and then taking ϵ and $\delta(\epsilon)$ to 0.

As $\epsilon \to 0$, one can prove that the solutions to (1.4.1) converge in $C([0,T];H)$ as $\epsilon \to 0$ to the solution to the corresponding deterministic Navier-Stokes equations, provided that $\delta(\epsilon)$ does not decay to 0 too slowly. Moreover, in [15], it is shown that the solutions also satisfy a large deviations principle in $C([0,T];H)$ for any $T > 0$. Following the strategy of [60], it is often possible to use the large deviations principle for the paths to prove a large deviations principle for the invariant probability measures of the system. This was done for equation (1.4.1) with fixed smooth noise, independent of ϵ in [8].

This is particularly interesting for the case of $\delta(\epsilon)$ converging to 0 because the limiting noise is isotropic, which allows the quasi-potential to be written down explicitly. In fact, due to the orthogonality relation on the torus between the Laplacian and Navier-Stokes nonlinearity,

$$0 = \langle \Delta u, (u \cdot \nabla) u \rangle_H,$$

this equation can be considered an infinite-dimensional example of the transverse drift decomposition (2.3.9) discussed in Section 2.3.1. Therefore, the quasi-potential $U : H \to [0, +\infty]$ corresponding to equation (1.4.1) is simply the function

$$U(h) = \|\nabla h\|_H^2 .$$

By sharpening the results of [15] and [8] we are able to prove that the invariant probability measures μ^ϵ of equation (1.4.1) satisfy a large deviation principle in H with respect to the action functional $U(h)$. Effectively, this is a generalization of the statement (1.1.4) to an infinite-dimensional setting where the global attractor is $\mathcal{O} \equiv 0 \in H$, and the attracting component of the drift field is provided by viscous diffusion in the Navier-Stokes equations.

Chapter 2: Large deviations theory for stochastic differential equations

This chapter provides a brief introduction to some of the main topics in large deviations theory in the context of stochastic differential equations. Both the theoretical and numerical material in the later chapters relies heavily on the theory introduced here. We do not provide full proofs in this chapter, but instead try to include intuitive explanations of the main ideas and techniques, while providing references to where full proofs may be found. The chapter is designed for the audience possessing basic understanding of measure theoretic probability theory and stochastic differential equations.

Much of the work discussed in this chapter is attributable to the pioneering work of Mark Freidlin and Alexander Wentzell, beginning in the 1960s. Many of the key results in the field can be found in their book [37], and the majority of the topics discussed in this chapter, in particular, are covered in more detail in Chapters 3 and 4 of [37]. Another excellent overview of the large deviations theory field is given by Amir Dembo and Ofer Zeitouni in the book [28]. A third resource which we rely heavily on in this chapter is the book [30] by Paul Dupuis and Richard Ellis, which provides an alternative perspective on the theory of large deviations from the vantage point of weak convergence of probability measures. This is a perspective we take several times in Chapters 4 and 5 in order to prove large deviations principles for solutions to multi-scale stochastic partial differential equations.

In order to avoid obfuscating the main ideas of the theory, we consider throughout this chapter only the following simple finite-dimensional stochastic differential equation. Nonetheless, most of the results and techniques will apply to more complicated equations as well. For a fixed $\epsilon > 0$, let $X_t^\epsilon \subset \mathbb{R}^d$ be the solution to the stochastic differential equation

$$dX_t^\epsilon = b(X_t^\epsilon)dt + \sqrt{\epsilon}dW_t, \qquad X_0 = x_0, \tag{2.0.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field, W_t is a standard d-dimensional Brownian motion on some probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ and $x_0 \in \mathbb{R}^d$. Throughout this chapter, we will always assume that b is Lipschitz continuous so that equation (2.0.1) is well-posed in the strong probabilistic sense in the space of continuous paths $C([0,T];\mathbb{R}^d)$ for any $T > 0$.

We are interested in what happens to the solution X_t^ϵ as the magnitude of the noise ϵ is taken to 0. Intuitively, one would expect that X_t^ϵ should behave very similarly to the solution X_t of the *ordinary* differential equation

$$\frac{dX_t}{dt} = b(X_t), \qquad X_0 = x_0, \tag{2.0.2}$$

and that the size of the deviations of X_t^ϵ from X_t should decrease as $\epsilon \downarrow 0$, as we see in Figure 2.1. This is true of course, and much of the time the behavior of system (2.0.1) is well approximated by the behavior of system (2.0.2). For example, convergence in probability of X_t^ϵ to X_t as $\epsilon \to 0$ is immediate due to the Lipschitz condition on b. Large deviations theory, on the other hand, is concerned with understanding the regime where this approximation of (2.0.1) by (2.0.2) fails. In particular, it answers the questions of how often and in what ways the random process X_t^ϵ experiences significant deviations from its expected behavior, given by X_t.

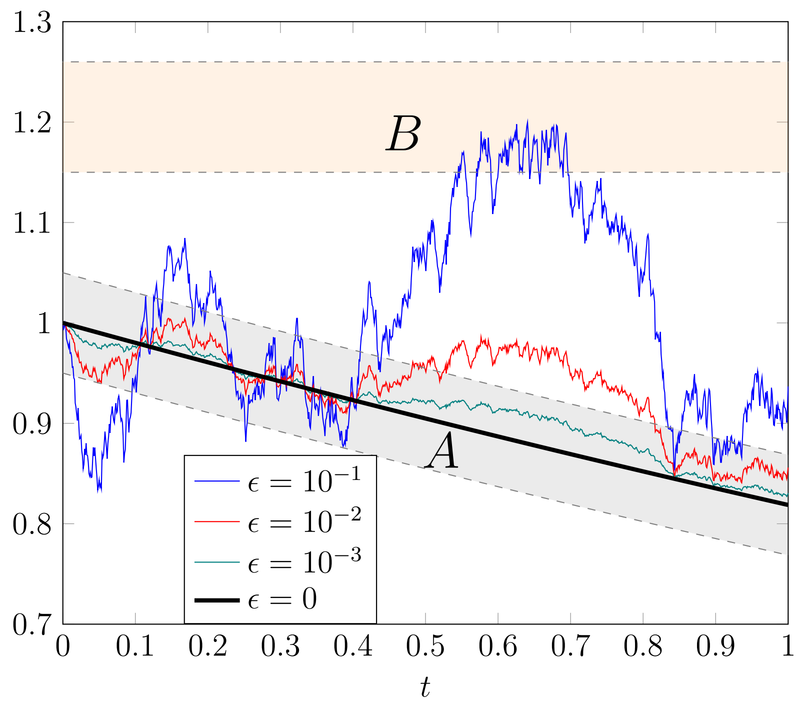


Fig. 2.1: Sample trajectories of equation (2.0.1) in $\mathbb{R}^1$ for $b(x) = -x/5$ and $x_0 = 1$. As $\epsilon \downarrow 0$, the probability of X_t^ϵ remaining in region A approaches 1, while the probability of X_t^ϵ crossing into region B approaches 0.

A simple question one can ask is: how probable is it for the solution X_t^ϵ to lie in a given set of trajectories $\mathcal{A}$? For instance, suppose $\mathcal{A} \subset C([0,T];\mathbb{R}^d)$ is a particular subset of the continuous paths in $\mathbb{R}^d$. If $\mathcal{A}$ contains an open ball in $C([0,T];\mathbb{R}^d)$ (or a tube, in this case) around the solution to the deterministic limit (2.0.2), then one certainly expect that $\mathbb{P}(X^\epsilon \in \mathcal{A})$ should approach 1 as $\epsilon \downarrow 0$. On the other hand, if the set $\mathcal{A}$ fails to contain such an open set, one would expect the probability to converge to 0. For instance, using the sets A and B defined in Figure 2.1, we have

$$\lim_{\epsilon\downarrow 0} \mathbb{P}(X_t^\epsilon \in A \text{ for all } t \in [0,1]) = 1, \tag{2.0.3}$$

$$\lim_{\epsilon\downarrow 0} \mathbb{P}(X_t^\epsilon \in B \text{ for some } t \in [0,1]) = 0. \tag{2.0.4}$$

Large deviations theory is concerned with the latter type of statement. In particular, the theory aims to more *precisely* quantify the rate at which limit (2.0.4) occurs. Refinement of the former statement (2.0.3) falls more under the domain of central limit-type theorems, which aim to more precisely quantify probabilities of the most *likely* outcomes. The study of extremely *unlikely* events on the other hand typically falls far outside the domain of applicability of the central limit theorem.

2.1 Large deviations for sample paths

Before trying to more precisely refine (2.0.4) for stochastic differential equations, it is helpful to first consider the very simple case of a family of Gaussian random variables. Let $\xi^\epsilon := \sqrt{\epsilon}\xi$ where ξ is a standard normal $\mathcal{N}(0,1)$ random variable. As we send ϵ to 0 it is clear that ξ^ϵ converges in probability to 0. Instead, suppose one was interested in the probability $P(\xi^\epsilon > a)$ for some $a > 0$. For small $\epsilon > 0$, one can compute explicitly

$$\begin{aligned}\mathbb{P}(\xi^\epsilon > a) &= \frac{1}{\sqrt{\epsilon}\sqrt{2\pi}}\int_a^\infty \exp\Big(-\frac{x^2}{2\epsilon}\Big)dx \\ &= \frac{\sqrt{\epsilon}}{a\sqrt{2\pi}}\exp\Big(-\frac{a^2}{2\epsilon}\Big) - \frac{\sqrt{\epsilon}}{\sqrt{2\pi}}\int_a^\infty \frac{1}{x^2}\exp\Big(-\frac{x^2}{2\epsilon}\Big)dx \\ &\approx \frac{\sqrt{\epsilon}}{a\sqrt{2\pi}}\exp\Big(-\frac{a^2}{2\epsilon}\Big), \end{aligned} \tag{2.1.1}$$

by integrating by parts in the second line and noting the integral term is negligible for small $\epsilon > 0$. Hence, we see that the probabilities of the rare event $\{\xi^\epsilon > a\}$ decay exponentially in ϵ^{-1}, which is entirely unsurprising due to the exponential density of the Gaussian distribution.

It is helpful to rephrase (2.1.1) in the following way. Let $\mathcal{A} := [a, \infty)$ so that (2.1.1) can be re-interpreted in the form

$$\lim_{\epsilon\to 0} \epsilon \log \mathbb{P}(\xi^\epsilon \in \mathcal{A}) = -\frac{a^2}{2} = \min_{x\in\mathcal{A}} I_\xi(x) \tag{2.1.2}$$

where $I_\xi(x) := x^2/2$. Equation (2.1.2) relates the exponential decay rate of the probability of ξ^ϵ landing in set $\mathcal{A}$ to the minimum of a non-negative action functional I_ξ on that set. The purpose of this re-framing is to illustrate the point that $\mathbb{P}(\xi^\epsilon \in \mathcal{A})$ is really only determined by one member of the set $\mathcal{A}$, namely its most *likely* member a. Thus, in the $\epsilon \to 0$ limit the event of ξ^ϵ landing in $\mathcal{A}$ is effectively indistinguishable from the event of ξ^ϵ landing in an arbitrarily small neighborhood of $\mathcal{A}$'s most likely element. In other words, when a rare event occurs, it will occur with overwhelming probability in the *least* unlikely way possible, where the unlikeliness is in general measured by an nonnegative action functional.

2.1.1 Definitions

The primary goal of the majority of this chapter is to justify large deviations statements of the form of (2.1.2) for solutions to stochastic differential equations. The idea will be fundamentally the same as in the toy example above – namely, the probability of the rare event will decay exponentially fast in ϵ^{-1} with an exponential rate determined by the least unlikely way for the rare event to happen.

We begin by stating the definition of the large deviations principle. We note that there exists a number of definitions, several of which are used in this manuscript, which all formalize statements of the form (2.1.2) in slightly different ways. A very thorough survey of these different formulations and discussion of their equivalences is given in [55]. The assumptions under which the equivalences hold are of some importance in our works in Chapters 4 and 5 of this manuscript.

Definition 2.1.1. Let $\mathcal{X}$ be a Polish space, i.e. a complete separable metric space. A lower-semicontinuous function $I : \mathcal{X} \to [0, +\infty]$ is called a good rate function if the level sets, $\Phi(s) := \{x \in \mathcal{X} : I(x) \leq s\}$ are compact for arbitrary $s \in [0, \infty)$.

Definition 2.1.2. Let $\{\xi^\epsilon\}_{\epsilon>0}$ be a collection of $\mathcal{X}$-valued random variables. The family ξ^ϵ is said to satisfy a large deviations principle in $\mathcal{X}$ with good rate function I and speed ϵ provided that

(a) for any open set $G \in \mathcal{B}(\mathcal{X})$,

$$\liminf_{n\to\infty} \epsilon \log \mathbb{P}(\xi^\epsilon \in G) \geq -I(G), \tag{2.1.3}$$

(b) for any closed set $F \in \mathcal{B}(\mathcal{X})$,

$$\limsup_{n\to\infty} \epsilon \log \mathbb{P}(\xi^\epsilon \in F) \leq -I(F). \tag{2.1.4}$$

Remark 2.1.1. Throughout the remainder, we will use the notation $I(A)$ to denote $\inf_{x\in A} I(x)$ for a given set A. Moreover, we will also talk about large deviations principle satisfied by collections of measures $\{\mu^\epsilon\}_{\epsilon\geq 0}$, in which it is implied that the same definition is intended, but with $\mathbb{P}(\xi^\epsilon \in G)$ and $\mathbb{P}(\xi^\epsilon \in F)$ replaced by $\mu^\epsilon(G)$ and $\mu^\epsilon(F)$, respectively.

□

Another definition of the large deviation principle that we will use is the following.

Definition 2.1.3. Let $\{\xi^\epsilon\}_{\epsilon>0}$ be a collection of $\mathcal{X}$-valued random variables. The family ξ^ϵ is said to satisfy a large deviations principle in $\mathcal{X}$ with good rate function I and speed ϵ provided that

(a) For every $x \in \mathcal{X}$ and $\delta > 0$,

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbb{P}(\xi^\epsilon \in B(x,\delta)) \geq -I(x).$$

(b) For every $s \geq 0$ and $\delta > 0$

$$\limsup_{\epsilon\to 0} \epsilon \log \mathbb{P}(\xi^\epsilon \in B^c(\Phi(s),\delta)) \leq -s,$$

where $B^c(\Phi(s),\delta) := \{h \in \mathcal{X} : \mathrm{dist}_{\mathcal{X}}(h,\Phi(s)) \geq \delta\}$.

Remark 2.1.2. Definitions 2.1.2 and 2.1.3 can be shown to be equivalent. In particular, (a) of Definition 2.1.2 is equivalent to (a) of Definition 2.1.3 while (b) of Definition 2.1.2 is equivalent to (b) of Definition 2.1.3. These definitions are *not* equivalent, however, if I is not a good rate function, i.e. does not have compact level sets.

$\square$

2.1.2 Schilder's Theorem

We hope to prove the validity of a large deviations principle for the solutions to (2.0.1). The first step in achieving this is to consider the simple case of $b \equiv 0$; namely, the case of Brownian motion $X_t^\epsilon = \sqrt{\epsilon} W_t$.

Theorem 2.1.1 (Schilder's Theorem)**.** *For any $T > 0$, the family $\sqrt{\epsilon} W_t$ satisfies a large deviations principle in $C([0,T];\mathbb{R}^d)$ with good rate function*

$$I(\varphi) := \begin{cases} \dfrac{1}{2}\displaystyle\int_0^T \|\dot{\varphi}_t\|^2 \, dt, & \text{if } \varphi \text{ is absolutely continuous and } \varphi(0) = 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.1.5}$$

Remark 2.1.3. Before discussing the proof of Schilder's theorem, we remark on the structure of any typical proof of a large deviations principle. In general, the lower bound (2.1.3) is obtained by *shifting* the probability measure to a new measure where the rare event is now an expected event. After a law of large numbers type argument, the resulting likelihood ratio from the change of measures gives a lower bound on the original probability. The "best" lower bound will then come from the particular change of measures with the *largest* likelihood ratio; that is, from the new measure that is closest to the original measure (formally this is measured by relative entropy, see Section 2.2). The upper bound of (2.1.3) is then shown by verifying that the aforementioned change of measure actually gives the best possible lower bound to exponential order. Typically this is done by the use of exponential Chebyshev inequalities.

We illustrate this abstract discussion in the simple example of the family of Gaussian random variables introduced above. To compute $\mathbb{P}(\sqrt{\epsilon}\xi \in [a, \infty))$, a natural thing to do would be to shift probability measures so that $\sqrt{\epsilon}\xi$ has an expected value lying in the interval $[a, \infty)$ under the new measure. Indeed, if $\sqrt{\epsilon}\xi \sim N(b, \epsilon)$ for any $b \geq a$, then the event $\sqrt{\epsilon}\xi \in [a, \infty)$ becomes likely in the $\epsilon \to 0$ limit. Each of these change of measures corresponding to a value of $b \geq a$ provides a lower bound, since

$$\begin{aligned}
\mathbb{P}(\sqrt{\epsilon}\xi \in [a, \infty)) &= \frac{1}{\sqrt{2\pi\epsilon}} \int_a^\infty \exp\Big(-\frac{x^2}{2\epsilon}\Big) dx \\
&= \frac{1}{\sqrt{2\pi\epsilon}} \exp\Big(-\frac{b^2}{2\epsilon}\Big) \int_{-(b-a)}^\infty \exp\Big(-\frac{y^2}{2\epsilon}\Big) \exp\Big(-\frac{by}{\epsilon}\Big) dy \\
&= \frac{1}{\sqrt{2\pi\epsilon}} \exp\Big(-\frac{b^2}{2\epsilon}\Big) \mathbb{E}\Big[\exp\Big(-\frac{b\xi}{\epsilon}\Big) \mathbb{1}_{\sqrt{\epsilon}\xi \geq -(b-a)}\Big] \\
&\geq \frac{1}{2\sqrt{2\pi\epsilon}} \exp\Big(-\frac{b^2}{2\epsilon}\Big).
\end{aligned}$$

where the final inequality occurs because the event $\{\sqrt{\epsilon}\xi \geq -(b-a)\}$ is asymptotically likely as $\epsilon \to 0$ and the exponential contribution to the integral can be bounded below by 1 by a convexity argument. Of course, the *sharpest* (and correct) lower bound comes from the value $b = a$, which corresponds to the closest such measure $N(b, 1)$ to the original $N(0, 1)$.

$\square$

Outline of proof. A full proof of Schilder's Theorem can be found, for instance in Theorem 2.1 and 2.2 of Chapter 3 of [37] or Theorem 5.2.3 of [28]. In light of the above remark, we see that the constructive step in the proof will occur in the lower bound, and we thus restrict our focus to this direction. This will be a recurring theme throughout *all* of the main results mentioned in this chapter, and we will correspondingly restrict most of our attention to the proofs of the lower bounds.

(Lower Bound) We are interested in proving (2.1.3). Since we are considering Brownian motion, the natural tool to shift probability measures will be the Girsanov Theorem, quoted below in Theorem 2.1.2. Suppose $\varphi \in C([0,T];\mathbb{R}^d)$ is such that $I(\varphi) < \infty$. We will consider the sets $B(\varphi,\delta) = \{h \in C([0,T];\mathbb{R}^d) : \sup_{t\in[0,T]} \|h_t - \varphi_t\| < \delta\}$ for some $\delta > 0$ and show the lower bound corresponding to the second definition of the large deviations principle, Definition 2.1.3. As mentioned above, to bound the probability of this event we shift measures to transform $B(\varphi,\delta)$ into a *likely* event. Defining the new measure $\mathbb{P}^\varphi$ by its density

$$\frac{d\mathbb{P}^\varphi}{d\mathbb{P}} = R^\varphi := \exp\Big(\frac{1}{\sqrt{\epsilon}}\int_0^T \dot{\varphi}_t dW_t - \frac{1}{2\epsilon}\int_0^T \|\dot{\varphi}_t\|^2\, dt\Big),$$

we have by Girsanov's theorem that the shifted process $W^\varphi := W - \dfrac{\varphi}{\sqrt{\epsilon}}$ is a Brownian motion under $\mathbb{P}^\varphi$. Denoting $\mathbb{E}^\varphi$ as the expectation with respect to the measure $\mathbb{P}^\varphi$, we then have

$$
\begin{aligned}
\mathbb{P}(\sqrt{\epsilon}W \in B(\varphi,\delta)) &= \mathbb{P}(\sqrt{\epsilon}W^{\varphi} \in B(0,\delta)) \\
&= \mathbb{E}\mathbb{1}_{\sqrt{\epsilon}W^{\varphi}\in B(0,\delta)} && (2.1.6)\\
&= \mathbb{E}^{\varphi}\big[(R^{\varphi})^{-1}\mathbb{1}_{\sqrt{\epsilon}W^{\varphi}\in B(0,\delta)}\big] && (2.1.7)\\
&= \mathbb{E}^{\varphi}\Big[\exp\Big(-\frac{1}{\sqrt{\epsilon}}\int_0^T \dot{\varphi}_t\cdot dW_t + \frac{1}{2\epsilon}\int_0^T \|\dot{\varphi}_t\|^2\,dt\Big)\mathbb{1}_{W^{\varphi}\in B(0,\delta)}\Big]\\
&= \exp\Big(-\frac{1}{2\epsilon}\int_0^T \|\dot{\varphi}_t\|^2\,dt\Big)\mathbb{E}^{\varphi}\Big[\exp\Big(-\frac{1}{\sqrt{\epsilon}}\int_0^T \dot{\varphi}_t\cdot dW_t^{\varphi}\Big)\mathbb{1}_{\sqrt{\epsilon}W^{\varphi}\in B(0,\delta)}\Big]. && (2.1.8)
\end{aligned}
$$

Indeed, the probability of the event $W^{\varphi} \in B(0,\delta)$ will converge to 1. Moreover, the other term in the expectation can be shown via exponential Chebyshev inequality (or alternatively, by using the fact that $-W^h$ is also a Brownian motion) to decay *slower* than $\exp(-\epsilon^{-1})$. Hence it follows from (2.1.6) that

$$
\liminf_{\epsilon\to 0}\epsilon\log\mathbb{P}(\sqrt{\epsilon}W \in B(\varphi,\delta)) \geq -\frac{1}{2}\int_0^T \|\dot{\varphi}_t\|^2\,dt = -I(\varphi).
$$

(Upper Bound) The upper bound (2.1.4) can be established by discretizing the Brownian motion in time, and using the large deviations principle upper bounds for families of Gaussian random variables. □

Theorem 2.1.2 (Girsanov Theorem)**.** *Let W_t be a Brownian motion with respect to some probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. Let h_t be a measurable $\mathbb{R}^d$-valued process adapted to $\mathcal{F}_t$ such that $h_0 = 0$. Define a new probability measure $\mathbb{P}^h$ by the formula*

$$
\frac{d\mathbb{P}^h}{d\mathbb{P}} := \exp\Big(\int_0^T h_t\cdot dW_t - \frac{1}{2}\int_0^T \|h_t\|^2\,dt\Big),
$$

Then, the process $W_t - \int_0^t h_s ds$ is a Brownian motion with respect to the probability measure $\mathbb{P}^h$.

2.1.3 SDEs

Next, we build large deviations principles for the solutions to (2.0.1) by transferring the large deviations principle for Brownian motion. This can be done by the contraction principle, which asserts that large deviations principles are transferable through continuous mappings.

Theorem 2.1.3. *Suppose that the family X^ϵ satisfies a large deviations principle in Polish space $\mathcal{X}$ with action functional I_X. Let $f : \mathcal{X} \to \mathcal{Y}$ be a continuous mapping between $\mathcal{X}$ and Polish space $\mathcal{Y}$. Then the family $Y^\epsilon := f(X^\epsilon)$ satisfies a large deviations principle in $\mathcal{Y}$ with good rate function*

$$I_Y(\varphi) := \inf\{I_X(\phi) : \quad \phi \in \mathcal{X}, \quad \varphi = f(\phi)\}.$$

Proof. Let $G \subset \mathcal{Y}$ be an open set. Then since f is continuous, $f^{-1}(G) \subset \mathcal{X}$ is an open set. Thus

$$\liminf_{\epsilon\to 0} \epsilon \log \mathbb{P}(Y^\epsilon \in G) = \liminf_{\epsilon\to 0} \epsilon \log \mathbb{P}(X^\epsilon \in f^{-1}(G)) \geq - \inf_{x\in f^{-1}(G)} I_X(x) = - \inf_{y\in G} I_Y(y).$$

The upper bound is similar. Moreover, it is immediate to show that I_Y is a good rate function. $\square$

Now we suppose that X_t^ϵ is the solution to (2.0.1) with Lipschitz continuous drift $b : \mathbb{R}^d \to \mathbb{R}^d$. If we denote the transformation $\mathcal{G} : C([0,T];\mathbb{R}^d) \to C([0,T];\mathbb{R}^d)$

defined by $\mathcal{G}(\phi) = \psi$ where ψ is the solution to

$$\psi(t) = x_0 + \int_0^t b(\psi(s))ds + \phi(t), \tag{2.1.9}$$

then we can immediately apply the contraction principle. Note here that Lipschitz continuity of b implies Lipschitz continuity of $\mathcal{G}$, since if $\psi_1 = \mathcal{G}(\phi_1)$ and $\psi_2 = \mathcal{G}(\phi_2)$, then for any $t \in [0, T]$

$$\begin{aligned}
\|\psi_1(t) - \psi_2(t)\| &\leq c_T \int_0^t \|b(\psi_1(s)) - b(\psi_2(s))\| \, ds + \|\phi_1(t) - \phi_2(t)\| \\
&\leq c_{T,b} \int_0^t \|\psi_1(s) - \psi_2(s)\| \, ds + \|\phi_1(t) - \phi_2(t)\| \, .
\end{aligned}$$

This gives Lipschitz continuity of $\mathcal{G}$ in $C([0, T]; \mathbb{R}^d)$ when combined with the Gronwall inequality.

Therefore, by the contraction principle, X_t^ϵ satisfies a large deviations principle in $C([0, T]; \mathbb{R}^d)$ with action functional

$$\begin{aligned}
I(\varphi) &= \inf \left\{ \frac{1}{2} \int_0^T \|\phi(t)\|^2 \, dt : \ \phi \in C([0, T]; \mathbb{R}^d), \quad \varphi = \mathcal{G}(\phi) \right\} \\
&= \begin{cases} \frac{1}{2} \int_0^T \|\dot{\varphi}_t - b(\varphi_t)\|^2 \, dt, & \text{if } \varphi \text{ is absolutely continuous and } \varphi(0) = x_0, \\ +\infty, & \text{otherwise,} \end{cases}
\end{aligned} \tag{2.1.10}$$

where the 2nd line holds by noting the definition of $\mathcal{G}$ (2.1.9).

2.1.4 Drawbacks

In situations such as equation (2.0.1), the large deviations principle can be established by obtaining a large deviations principle for an underlying Gaussian measure

(Schilder's theorem), and then transferred by a contraction principle to a nonlinear SDE. This is convenient when one can construct the object of interest as a continuous mapping from a simpler measure (in our case a Gaussian measure) where a large deviations result is known. In Chapters 2 and 3, we require large deviations principles for situations where this is not so simple because the scale parameter ϵ shows up in multiple locations. In Chapter 3, we consider a version of the stochastic heat equation

$$du(x,t) = \frac{1}{\epsilon}\Delta u(x,t)dt + b(u(x,t))dt + \sqrt{\epsilon}dW,$$

where the Laplacian term scales with $1/\epsilon$. It is now no longer possible to view the solution as a linear from a simple Gaussian measure $\sqrt{\epsilon}dW$. Similarly, in Chapter 4, we consider the Navier-Stokes equations perturbed by a Gaussian-noise whose spatial covariance operator Q^ϵ also scales with ϵ.

$$du(t,x) + (u \cdot \nabla u)(t,x)dt = \Delta u(t,x)dt + \sqrt{\epsilon}Q^\epsilon dW(t).$$

Due to the presence of these additional ϵ, often the strategy described earlier may become infeasible. There is an alternative approach, the weak convergence approach [30], relying on a different set of tools. The technique will still rely on appropriate mappings between the Gaussian objects and solutions to the SPDEs, however it will allow for solution mappings that themselves are allowed to depend on the parameter ϵ. This is achieved by certain compactness arguments. We outline this approach finite dimensions, and make comments about its extension to infinite dimensional settings.

The weak convergence approach relies fundamentally on a variational representation of functionals of random variables by way of the Kullback–Leibler divergence, also known as the relative entropy.

2.2 The weak convergence approach

In this section, we look at large deviations through the lens of Laplace asymptotics and stochastic control theory. Our main goal is the intuitive description of the weak convergence approach towards large deviations analysis. The first use of weak convergence and stochastic control theory ideas in the setting of large deviations for diffusion processes came from Fleming in 1978 [33]. The general framework for a weak convergence approach to large deviations, however, was developed largely by Dupuis and Ellis [30], and is based on several ideas of Varadhan [65]. This framework was applied to the setting of sample path large deviations first in [7], with extensions to the infinite dimensional case in [12, 13]. A complete description of these methods with full proofs is given there.

To motivate the approach, let us return to the simple example of the family of Gaussian random variable $\xi^\epsilon := \sqrt{\epsilon}\xi$ for $\xi \sim N(0,1)$. As before, we are concerned with the asymptotic evaluation of probabilities of ξ^ϵ lying in some set $\mathcal{A}$. Note that we can always express the probability in the form

$$\epsilon \log \mathbb{P}(\xi^\epsilon \in \mathcal{A}) = \epsilon \log \mathbb{E} \exp\Big(-\frac{h_{\mathcal{A}}(\xi^\epsilon)}{\epsilon}\Big), \tag{2.2.1}$$

where

$$h_{\mathcal{A}}(x) := \begin{cases} 0 & \text{if } x \in \mathcal{A}, \\ +\infty & \text{if } x \notin \mathcal{A}. \end{cases}$$

The main idea of the weak convergence approach is to *approximate* $h_{\mathcal{A}}$ by bounded, continuous functions, and then evaluate the right hand side of (2.2.1) on the approximations of $h_{\mathcal{A}}$ using Laplace method techniques. In our particular simple example of Gaussian random variables, this expectation becomes the integral of an exponential function, so that if $h \in C_b(\mathbb{R})$, we have

$$\begin{aligned} \lim_{\epsilon \to 0} \epsilon \log \mathbb{E} \exp\Big(-\frac{h(\xi^\epsilon)}{\epsilon} \Big) &= \lim_{\epsilon \to 0} \epsilon \log \int_{\mathbb{R}} \exp\Big(-\frac{h(x) + \frac{x^2}{2}}{\epsilon} \Big) dx \\ &= - \inf_{x \in \mathbb{R}} \Big[\frac{x^2}{2} + h(x) \Big], \end{aligned} \tag{2.2.2}$$

where the 2nd line is obtained by using Laplace's method. Thus, we see that if (2.2.2) were also true for $h = h_{\mathcal{A}}$, then we would reproduce the large deviations statements (2.1.4) and (2.1.3) since

$$\inf_{x \in \mathbb{R}} \Big[\frac{x^2}{2} + h_{\mathcal{A}}(x) \Big] = \inf_{x \in \mathcal{A}} \frac{x^2}{2}.$$

This is, in fact, the case (up to topological considerations of the set $\mathcal{A}$ that we ignored) and we can obtain the desired statements (2.1.4) and (2.1.3) by taking appropriate limits of (2.2.2) for a sequence of continuous approximations of $h_{\mathcal{A}}$.

This is the main crux of the weak convergence approach. Rather than proving the asymptotic probability statements (2.1.3) and (2.1.4) for the family of random variables directly, one proves Laplace method type statements of the form (2.2.2) for an arbitrary bounded continuous function h. As we will shortly see, the Laplace's method type statement (2.2.2) will be provable for general random variables by utilizing a variational representation of expectations of exponential functions.

2.2.1 Definitions

Let us first formalize statements (2.2.2) for a general family of random variables on a Polish space $\mathcal{X}$.

Definition 2.2.1. Let I be a good rate function on $\mathcal{X}$. Then the family X^ϵ of random variables is said to satisfy the Laplace principle on $\mathcal{X}$ with rate function I if for all $h \in C_b(\mathcal{X})$,

$$\lim_{\epsilon \to 0} \epsilon \log \mathbb{E} \exp\Big(-\frac{h(X^\epsilon)}{\epsilon}\Big) = -\inf_{x \in \mathcal{X}} \{I(x) + h(x)\}. \tag{2.2.3}$$

As motivated by the previous section, the space of continuous bounded functions is sufficiently rich that the Laplace principle implies the large deviations principle. In fact, the converse is also true when the space $\mathcal{X}$ is Polish. For weaker topological spaces, only the implication *Laplace principle* $\Longrightarrow$ *LDP* holds.

Theorem 2.2.1. *Suppose that $\mathcal{X}$ is a Polish space and I is a good rate function on $\mathcal{X}$. Then the family X^ϵ of random variables satisfies a Laplace principle on $\mathcal{X}$ with rate function I if and only if it satisfies a large deviations principle on $\mathcal{X}$ with rate function I.*

Outline of proof. A full proof of this result can be found for instance in Chapter 1 of [30]. The backward direction (*LDP* $\Longrightarrow$ *Laplace principle*) is attributed to Varadhan

[65]. We restrict our remarks here to those concerning the forward direction, since this is the implication relevant for our purposes.

Both the upper bound (and lower bound) in Definition 2.1.2 can be proven by taking suitable continuous approximations of h_F (and h_G). For the upper bound, one may take for instance functions $h_j(x) = j(d(x, F) \wedge 1)$ where d is the metric on $\mathcal{X}$ and $d(x, F) := \inf_{y \in F} d(x, y)$. It is easy to show that the upper bound follows by taking advantage of the lower semicontinuity of I. For the lower bound, the construction is a function heavily peaked on a ball $B(x^*, \delta)$ where x^* is a point in the set G near the minimizer of action I on G. □

In the simple Gaussian example discussed above, the Laplace principle was seen to be immediately true by using Laplace method for evaluating the asymptotics of a single exponential integral over $\mathbb{R}$. This in general will not be possible. Instead, the main tool towards proving the validity of Laplace principles is the *relative entropy*, also known as the Kullback-Leibler divergence.

Definition 2.2.2. For $\theta \in \mathcal{P}(\mathcal{X})$, the space of probability measures on $\mathcal{X}$, the relative entropy $R(\cdot||\theta)$ is a mapping from $\mathcal{P}(X) \to [0, \infty]$ defined by

$$R(\gamma||\theta) = \begin{cases} \displaystyle\int_X \log \frac{d\gamma}{d\theta} d\gamma & \text{if } \gamma \text{ is absolutely continuous with respect to } \theta, \\ \infty & \text{otherwise.} \end{cases}$$

Remark 2.2.1. Non-negativity of relative entropy is seen by noting the inequality $x \log x - x \geq -1$ for all $x \geq 0$. Thus, for an arbitrary measures γ that is absolutely continuous with respect to measure θ, it follows that

$$R(\gamma||\theta) = \int_X \frac{d\gamma}{d\theta} \log \frac{d\gamma}{d\theta} d\theta \geq \int_X \Big(\frac{d\gamma}{d\theta} - 1\Big) d\theta = 0.$$

Moreover, we see that $R(\gamma||\theta) = 0$ if and only if $\gamma = \theta$ so that $\frac{d\gamma}{d\theta} \equiv 1$.

$\square$

Remark 2.2.2. Relative entropy $R(\gamma||\theta)$ provides a measure of how "far apart" the measures γ and θ are, although it is not quite a metric on the space of probability measures. However, a consequence of Pinsker's inequality is that convergence in relative entropy implies convergence in total variation norm on $\mathcal{P}(\mathcal{X})$.

$\square$

In the context of large deviations, relative entropy will play the role of the action functional. This connection is immediately seen in view of the following representation formula [30].

Proposition 2.2.1. *[Proposition 1.4.2 of [30]] Suppose that $h : X \to \mathbb{R}$ is a measurable, bounded function. Then for any $\theta \in \mathcal{P}(X)$,*

$$-\log \int_X e^{-h(x)} d\theta(x) = \inf_{\gamma \in \mathcal{P}(X)} \Big\{ R(\gamma||\theta) + \int_X h(x) d\gamma(x) \Big\}. \tag{2.2.4}$$

Proof. Let h be a bounded, measurable function. Define the measure γ_0 by

$$\frac{d\gamma_0}{d\theta}(x) = \frac{e^{-h(x)}}{\int_X e^{-h(x)} d\theta(x)}.$$

Because of the normalization, γ_0 is a probability measure. Moreover, because the density is strictly positive, $\gamma_0 \ll \theta$ and $\gamma_0 \gg \theta$. By construction, we then have that

$$R(\gamma_0||\theta) + \int_X h(x) d\gamma_0(x) = -\log \int_X \exp^{-h(x)} d\theta(x).$$

Moreover, for any other $\gamma \in \mathcal{P}(X)$ that is absolutely continuous with respect to θ, we have

$$\begin{aligned}
R(\gamma||\theta) + \int_X h(x) d\gamma(x) &= \int_X \log \frac{d\gamma}{d\theta} d\gamma + \int_X h d\gamma \\
&= \int_X \log \frac{d\gamma}{d\gamma_0} d\gamma + \int_X \log \frac{d\gamma_0}{d\theta} d\gamma + \int_X h d\gamma \\
&= R(\gamma||\gamma_0) - \log \int_X e^{-h(x)} d\theta(x) \\
&\geq -\log \int_X e^{-h(x)} d\theta(x).
\end{aligned}$$

with equality in the last line if and only if $\gamma \equiv \gamma_0$. $\square$

2.2.2 Another route to Schilder's Theorem

In this section we apply the variational formula (2.2.4) to Brownian motion in $\mathbb{R}^d$ to provide an alternate route to Schilder's Theorem. This route will open the door for additional flexibility when trying to prove large deviations results for stochastic equations.

We begin with the variational representation for Brownian motion. The proof given in finite dimensions can be found in [7], while an analogous prove for infinite dimensional Wiener processes can be found in [13].

Proposition 2.2.2. *Let W_t be a Brownian motion in $\mathbb{R}^d$ on probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ and let $h : C([0,T];\mathbb{R}^d) \to \mathbb{R}$ be a bounded, Borel measureable function. Then*

$$-\log \mathbb{E}e^{-h(W)} = \inf_{v\in\mathcal{A}} \mathbb{E}\Big[\frac{1}{2}\int_0^T \|v_t\|^2\, dt + h(W_\cdot + \int_0^\cdot v_s ds)\Big], \tag{2.2.5}$$

where $\mathcal{A}$ is the space of progressively measurable random processes v such that

$$\mathbb{E}\int_0^T \|v_t\|^2\, dt < \infty.$$

Here, we are using the notation $h(W_\cdot + \int_0^\cdot v_s ds)$ to emphasize that h is a function acting on each entire path on $[0,T]$ of the continuous random process $W_t + \int_0^t v_s ds$.

Outline of proof. As with the proof of Theorem 2.1.1, we discuss only the proof of the lower bound which constitutes the more constructive step. We seek to apply the variational formula (2.2.4) to the Wiener measure θ on $C([0,T];\mathbb{R}^d)$.

(Lower bound) The main idea of the proof of the $\geq$ direction in formula (2.2.5) is that the set of measures on $C([0,T];\mathbb{R}^d)$ that are absolutely continuous with respect to θ can be approximated by measures γ_v of the form

$$\frac{d\gamma^v}{d\theta} = \exp\Big(\int_0^T v_t dW_t - \frac{1}{2}\int_0^T \|v_t\|^2\, dt\Big),$$

for some $v \in \mathcal{A}$. For these Girsanov shifted measures, the right hand side of (2.2.4) can be computed for a bounded Borel function $h : C([0,T];\mathbb{R}^d) \to \mathbb{R}$. We let $\mathbb{E}^v$ denote the expectation with respect to this measure. As before, we know by Girsanov's

Theorem that $W_t^v := W_t - \int_0^t v(s)ds$ is a Brownian motion under γ^v. Hence,

$$
\begin{aligned}
R(\gamma^v||\theta) &+ \int_{C([0,T];\mathbb{R}^d)} h(x)d\gamma^v(x) \\
&= \mathbb{E}^v\Big[\int_0^T v_t dW_t - \frac{1}{2}\int_0^T \|v_t\|^2\, dt\Big] + \mathbb{E}^v h(W) \\
&= \mathbb{E}^v\Big[\int_0^T v_t dW_t^v + \frac{1}{2}\int_0^T \|v_t\|^2\, dt\Big] + \mathbb{E}^v h(W^v + \int_0^\cdot v_s ds) \\
&= \mathbb{E}^v\Big[\frac{1}{2}\int_0^T \|v_t\|^2\, dt\Big] + \mathbb{E}^v h(W^v + \int_0^\cdot v_s ds),
\end{aligned}
$$

where the expectation of the stochastic integral vanishes due to the martingale property of the stochastic integral. In view of Proposition 2.2.1, this gives

$$
-\log \mathbb{E}e^{-h(W)} \leq \inf_{v\in\mathcal{A}} \mathbb{E}^v\Big[\frac{1}{2}\int_0^T \|v_t\|^2\, dt + h(W^v + \int_0^\cdot v(s)ds)\Big]. \tag{2.2.6}
$$

This is not quite in the form presented in the statement of theorem. Indeed, the expectations are computed with respect to the measures γ^v rather than θ. But in fact, for each $v \in \mathcal{A}$, it is possible to construct a new $\tilde{v} \in \mathcal{A}$ such that

$$
\mathbb{E}^v\Big[\frac{1}{2}\int_0^T \|v_t\|^2\, dt + h(W^h + \int_0^\cdot v_s ds)\Big] = \mathbb{E}\Big[\frac{1}{2}\int_0^T \|\tilde{v}_t\|^2\, dt + h(W + \int_0^\cdot \tilde{v}_s ds)\Big],
$$

so that the infimum in (2.2.5) is the same as the infimum in (2.2.6). We omit the construction, which is somewhat complicated, but it can be found in the proof of Theorem 3.1 in [7].

(Upper Bound) The fact that the inequality is indeed an equality can be shown using the martingale representation theorem. Indeed, it can be shown that the unique measure γ_0 in the representation formula (2.2.4) does correspond to a Girsanov shift of θ by some $v \in \mathcal{A}$. $\square$

Theorem 2.2.2. *The family $\sqrt{\epsilon}W$ satisfy a Laplace principle in $C([0,T];\mathbb{R}^d)$ with good rate function given by* (2.1.5).

Outline of proof. In view of the similarity between this variational formulation (2.2.5) and the desired Laplace principle (2.2.3) for the family $\sqrt{\epsilon}W$, the desired result is in sight. We discuss only the proof of the upper bound of the limit in (2.2.3), which is the more difficult direction that also provides better insight into the required conditions discussed in the subsequent sections.

(Upper bound) Indeed, to show the validity of a Laplace principle upper bound for the family $\sqrt{\epsilon}W$, we note that

$$\begin{aligned}
-\epsilon \log \mathbb{E} \exp\Big(-\frac{h(\sqrt{\epsilon}W)}{\epsilon}\Big) &= \inf_{u\in\mathcal{A}} \mathbb{E}\Big[\frac{\epsilon}{2}\int_0^T \|u_t\|^2\,dt + h(\sqrt{\epsilon}W + \sqrt{\epsilon}\int_0^\cdot u_s ds)\Big] \\
&= \inf_{u\in\mathcal{A}} \mathbb{E}\Big[\frac{1}{2}\int_0^T \|u_t\|^2\,dt + h(\sqrt{\epsilon}W + \int_0^\cdot u_s ds)\Big],
\end{aligned}$$

where the second line follows by absorbing the ϵ into the infimum. The necessary upper bound can obtained by for each $\epsilon > 0$ taking a $u^\epsilon \in \mathcal{A}$ such that

$$-\epsilon \log \mathbb{E} \exp\Big(-\frac{h(\sqrt{\epsilon}W)}{\epsilon}\Big) \geq \mathbb{E}\Big[\frac{1}{2}\int_0^T \|u_t^\epsilon\|^2\,dt + h(\sqrt{\epsilon}W + \int_0^\cdot u_s^\epsilon ds)\Big] - \epsilon. \quad (2.2.7)$$

We would like the right hand side to ideally convergence to something useful. Indeed, since inequality (2.2.7) provides a uniform bound in $L^2(\Omega; L^2(0,T;\mathbb{R}^d))$ on u^ϵ, it follows by Prokhorov's theorem we can extract a subsequence u^{ϵ_k} converging to some $u \in \mathcal{A}$ in distribution with respect to the *weak* topology of $L^2(0,T;\mathbb{R}^d)$ (since unit balls in $L^2(0,T;\mathbb{R}^d)$ are compact in the weak topology). In effect, our main consideration then turns to the h term in (2.2.7). If we could show that

$$\lim_{\epsilon\to 0} \mathbb{E}h\Big(\sqrt{\epsilon_k}W + \int_0^\cdot u_s^{\epsilon_k} ds\Big) = \mathbb{E}h\Big(\int_0^\cdot u_s ds\Big), \tag{2.2.8}$$

then it would follow rather easily that

$$\begin{aligned}\liminf_{\epsilon\to 0} \epsilon \log \mathbb{E} \exp\Big(-\frac{h(\sqrt{\epsilon}W)}{\epsilon}\Big) &= -\limsup_{\epsilon\to 0} -\epsilon \log \mathbb{E} \exp\Big(-\frac{h(\sqrt{\epsilon}W)}{\epsilon}\Big) \\ &\leq -\inf_{v\in L^2(0,T;\mathbb{R}^d)} \Big[\frac{1}{2}\int_0^T \|u_t\|^2 dt + h(\int_0^\cdot u_s ds)\Big],\end{aligned} \tag{2.2.9}$$

which is precisely the desired upper bound.

The convergence (2.2.8) is where the method's namesake comes into play since this is really a statement about weak convergence of random variables. For the case of Brownian motion, we see that the convergence in (2.2.8) is equivalent to saying (dropping the subsequence notation) that the solution $X_t^{\epsilon,u^\epsilon}$ to the controlled stochastic equation

$$dX_t^{\epsilon,u^\epsilon} = \sqrt{\epsilon}W_t + u_t^\epsilon dt, \qquad X_0^{\epsilon,u^\epsilon} = x_0,$$

converges weakly as a $C([0,T];\mathbb{R}^d)$ random variable to the solution X_t^u to the controlled random equation

$$dX_t^u = u_t dt, \qquad X_0^u = x_0,$$

whenever u^ϵ is a sequence converging in distribution to u with respect to the weak topology of $L^2(0,T;\mathbb{R}^d)$. However, this is indeed immediate since

$$X_t^{\epsilon,u^\epsilon} - X_t^u = \sqrt{\epsilon}W_t + \int_0^t (u_s^\epsilon - u_s) ds,$$

and process $\sqrt{\epsilon}W$ converges pointwise to 0 in $C([0,T];\mathbb{R}^d)$ while $\int_0^\cdot (u_s^\epsilon - u_s)ds$ converges weakly to 0 in the Sobolev space $W^{1,2}(0,T;\mathbb{R}^d)$, which is compactly embedded in $C([0,T];\mathbb{R}^d)$.

□

2.2.3 SDEs

In the previous section we outlined the proof of a large deviations principle for Brownian motion via the weak convergence route. Now, we want to prove a large deviations principle for the solution X_t^ϵ to the SDE (2.0.1). Of course, we could simply use the contraction principle. However, it will prove advantageous to *instead* prove the Laplace principle directly for X_t^ϵ in the exact same manner that we did for Brownian motion.

Indeed, the utility of the weak convergence approach will be very much evident in the step (2.2.8). As before, let us denote the solution map $\mathcal{G}$ that maps a trajectory of the driving noise $\sqrt{\epsilon}W$ to a trajectory of solution X_t^ϵ. Then, for arbitrary continuous bounded function $h \in C_b([0,T];\mathbb{R}^d)$ we have that $h \circ \mathcal{G}$ is bounded and Borel measurable so that

$$-\epsilon \log \mathbb{E} \exp\Big(-\frac{h(X_t^\epsilon)}{\epsilon}\Big) = \inf_{u\in\mathcal{A}} \Big[\frac{1}{2}\int_0^T \|u_t\|^2\, dt + h \circ \mathcal{G}(\sqrt{\epsilon}W + \int_0^\cdot u_s ds)\Big],$$

due to the representation formula (2.2.5). In particular, the solution map $\mathcal{G}$ need *not* be continuous. The proof of Theorem 2.2.2 would then proceed the same way. The only difference would then be in showing the step (2.2.8). In this case, we would instead need to show the convergence

$$\mathcal{G}(\sqrt{\epsilon}W + \int_0^{\cdot} u^\epsilon_s ds) \rightharpoonup \mathcal{G}(\int_0^{\cdot} u_s ds) \tag{2.2.10}$$

in distribution in $C([0,T];\mathbb{R}^d)$ as $\epsilon \to 0$ whenever u^ϵ is a sequence that converges in distribution to u with respect to the weak topology of $L^2(0,T;\mathbb{R}^d)$. Or equivalently, the solution $X^{\epsilon,u}_t$ to the controlled stochastic equation

$$dX^{\epsilon,u^\epsilon}_t = b(X^{\epsilon,u^\epsilon}_t)dt + \sqrt{\epsilon}dW_t + u^\epsilon_t dt, \qquad X^{\epsilon,u^\epsilon}_0 = x_0,$$

converges in distribution in $C([0,T];\mathbb{R}^d)$ to the solution to the random equation

$$dX^u_t = b(X^u_t)dt + u_t dt, \qquad X^u_0 = x_0.$$

This convergence is easy to show in the case where b is Lipschitz continuous. We note also that the right hand side of (2.2.9) that we would get by proceeding through the proof of Theorem 2.2.2 reproduces the action functional (2.1.10). Indeed,

$$\begin{aligned}
\lim_{\epsilon\to 0} \epsilon \log \mathbb{E} \exp\Big(-\frac{h(X^\epsilon_t)}{\epsilon}\Big) &= -\inf_{u\in L^2(0,T;\mathbb{R}^d)} \Big[\frac{1}{2}\int_0^T \|u_t\|^2 dt + h\circ\mathcal{G}(\int_0^{\cdot} u_s ds)\Big] \\
&= -\inf_{u\in L^2(0,T;\mathbb{R}^d)} \Big\{\frac{1}{2}\int_0^T \|u_t\|^2 dt + h(\varphi) : \qquad \varphi = \mathcal{G}\Big(\int_0^{\cdot} u_s ds\Big)\Big\} \\
&= -\inf_{\varphi\in C([0,T];\mathbb{R}^d)} \Big\{\frac{1}{2}\int_0^T \|\dot{\varphi}_t - b(\varphi_t)\|^2 dt + h(\varphi) : \qquad \varphi(0) = x_0\Big\}
\end{aligned}$$

Remark 2.2.3. The key utility here is that we did *not* require the solution map to be continuous like we did with the contraction principle. Instead, we needed only to show the relation (2.2.10). Moreover, there is no reason why one cannot consider solution maps $\mathcal{G}^\epsilon$ that depend on ϵ, so long as the convergence result holds. For example, one could consider equations of the form

$$dX_t = b^\epsilon(X_t)dt + \sigma^\epsilon\sqrt{\epsilon}dW_t.$$

Then, provided there exists some limiting solution map $\mathcal{G}$ such that

$$\mathcal{G}^\epsilon(\sqrt{\epsilon}W + \int_0^\cdot u_s^\epsilon ds) \rightharpoonup \mathcal{G}(\int_0^\cdot u_s ds),$$

a Laplace principle will follow. This will prove extremely useful when studying the models (1.3.1) and (1.4.1), which both display dependence on ϵ in multiple places. We provide a rigorous description of all the required conditions in the most general case in Section 2.2.4.

$\square$

2.2.4 Formalizations

In this section, we formalize the statements of the weak convergence approach in the version suited for infinite dimensional problems from [13]. First we formally define a version of the Laplace principle that is *uniform* with respect to initial conditions.

Definition 2.2.3 (ULP)**.** Let $\{I^x\}_{x\in\mathcal{E}_0}$ be a family of good rate functions on a Polish space $\mathcal{X}$, parametrized by a parameter x in some Polish space $\mathcal{E}_0$ and let A be some Borel subset of $\mathcal{E}_0$. The family of $\mathcal{X}$-valued random variables $\{\xi_x^\epsilon\}_{\epsilon>0,x\in A}$ is said to satisfy the Laplace principle on $\mathcal{X}$ with rate functions I^x, uniformly for $x \in A$, if for any continuous and bounded $h : \mathcal{X} \to \mathbb{R}$

$$\lim_{\epsilon\to 0}\sup_{x\in\mathcal{E}_0}\left|\gamma(\epsilon)\log\mathbb{E}\exp\left(-\frac{h(\xi_x^\epsilon)}{\gamma(\epsilon)}\right) + \inf_{y\in\mathcal{X}}\left(I^x(y)+h(y)\right)\right| = 0. \tag{2.2.11}$$

For the remainder of this section, we assume here as background the basics of Gaussian measure theory in infinite dimensional Hilbert spaces (for an introduction, see [26]). Let H be a separable Hilbert space H, and suppose W is an H-valued Q-Wiener process, where the covariance operator Q is a strictly positive, trace class operator on H. Let $H_0 = Q^{1/2}H$ be its reproducing kernel space endowed with the inner product $\langle h, k\rangle_{H_0} = \langle Q^{-1/2}h, Q^{-1/2}k\rangle_H$.

Assume that $\mathcal{X}$ is a Polish space (of paths) and let $\mathcal{E}_0$ be a Polish space (of initial conditions). Suppose that $\mathcal{G}^\epsilon : \mathcal{E}_0 \times C([0,T] : H) \to \mathcal{X}$ is a family of Borel measurable mappings. We define the spaces

$$S^N(H_0) := \Big\{ u \in L^2(0,T;H_0) : \int_0^T \|u_t\|_{H_0}^2 \, dt \le N \Big\},$$

$$\mathcal{A}^N(H_0) := \{u \in \mathcal{A} : u(\omega) \in S^N(H_0), \mathbb{P} - \text{a.s.}\}.$$

Moreover, the space $S^N(H_0)$ is a compact metric space when endowed with the metric,

$$d(x,y) = \sum_{i=1}^{\infty} \frac{1}{2^i} \Big| \int_0^T \langle x_s - y_s, e_i(s)\rangle_{H_0} ds \Big|.$$

Hypothesis 2.2.1. There exists a measureable map $\mathcal{G}^0 : \mathcal{E}_0 \times C([0,T] : H) \to \mathcal{X}$ satisfying the following conditions.

1. For every $M < \infty$ and compact set $K \subset \mathcal{E}_0$, the set

   $$\Gamma_{M,K} := \Big\{ \mathcal{G}^0\Big(x, \int_0^\cdot u_s ds\Big) : u \in S^M(H_0),\ x \in K \Big\},$$

   is a compact subset of $\mathcal{X}$.

2. For any $M < \infty$, any sequence $x^\epsilon \to x$ in $\mathcal{E}_0$ and any sequence $\{u^\epsilon\}_{\epsilon\in(0,1]} \subset \mathcal{A}^M(H_0)$ with u^ϵ converging in to u in distribution as $S^M(H_0)$ valued random elements (endowed with the topology generated by the metric above), we have

$$\mathcal{G}^\epsilon\Big(x^\epsilon, \sqrt{\epsilon}W + \int_0^\cdot u_s^\epsilon ds\Big) \to \mathcal{G}^0\Big(x, \int_0^\cdot u_s ds\Big),$$

where the convergence is in distribution in the strong topology of $\mathcal{X}$.

Next, define the functional $I : \mathcal{X} \to [0,\infty]$,

$$I^x(\varphi) := \inf\Big\{\frac{1}{2}\int_0^T \|u_t\|_0^2\, dt : u \in L^2(0,T;H_0), \varphi = \mathcal{G}^0\Big(x, \int_0^\cdot u_s ds\Big)\Big\} \quad (2.2.12)$$

Theorem 2.2.3 (Theorem 5 in [13]). *Let $X^\epsilon = \mathcal{G}^\epsilon(x, \sqrt{\epsilon}W)$ and suppose that the assumptions above hold and that for all $\varphi \in \mathcal{X}$, $I^x(\varphi)$ is lower semi-continuous mapping from $\mathcal{E}_0$ to $[0,+\infty$. Then for each $x \in \mathcal{E}_0$, $I^x(f)$ is a good rate function and the family $\{X^\epsilon\}$ satisfies the Laplace principle on $\mathcal{X}$ with rate functions $I^x(\varphi)$, uniformly for x in any compact subset of $\mathcal{E}_0$.*

2.2.5 Equivalences of uniform large deviations principles.

When studying the asymptotics of the long-time dynamics of equation (2.0.1), it will be critical that the limits (2.1.3) and (2.1.4) in the large deviations principle are uniform with respect to initial conditions. Theorem 2.2.3 provides the mechanism for obtaining a Laplace principle that is uniform with respect to initial conditions. However, unlike the corresponding non-uniform case, validity of the uniform Laplace principle does *not* necessarily imply validity of the uniform large deviations principle.

Moreover, the two equivalent large deviations principle definitions (Definitions 2.1.2 and 2.1.3) are no longer necessarily equivalent when one considers their suitably modified uniform versions. The article [55] investigates the necessary conditions for equivalence between the different formulations and provides examples justifying these conditions. In this section, we merely quote some of their main results, which we will be using in Chapters 4 and 5.

First, we define the uniform versions of Definition 2.1.2 and Definition 2.1.3. Sticking with the convention of [55], we refer to the uniform version of Definition 2.1.2 as the *Dembo-Zeitouni Uniform Large Deviations Principle* (DZULDP) and the uniform version of Definition 2.1.3 as the *Freidlin-Wentzell Uniform Large Deviations Principle* (FWULDP). In many typical finite-dimensional settings, such as that given by (2.0.1), the three definitions ULP, DZULDP and FWULDP will typically be equivalent. However, this will not be the case in general in infinite dimensions.

Definition 2.2.4 (DZULDP)**.** Let $\{I^x\}_{x\in\mathcal{E}_0}$ be a family of good rate functions on a Polish space $\mathcal{X}$, parametrized by a parameter x in some Polish space $\mathcal{E}_0$ and let A be some Borel subset of $\mathcal{E}_0$. The family of $\mathcal{X}$-valued random variables $\{\xi^\epsilon_x\}_{\epsilon>0,x\in A}$ is said to satisfy the Dembo-Zeitouni large deviations principle on $\mathcal{X}$ with rate functions I^x, uniformly for $x \in A$, if the following statement holds.

(i) For any $\gamma > 0$ and open set $G \subset \mathcal{X}$, there exists $\epsilon_0 > 0$ such that

$$\inf_{x\in A} \mathbb{P}(\xi^\epsilon_x \in G) \geq \exp\left(-\frac{1}{\epsilon}\left[\sup_{x\in A}\inf_{u\in G} I^x(u) + \gamma\right]\right), \quad \epsilon \leq \epsilon_0.$$

(ii) For any $\gamma > 0$ and closet set $F \subset E$, there exists $\epsilon_0 > 0$ such that

$$\sup_{x \in A} \mathbb{P}(\xi_x^\epsilon \in F) \leq \exp\left(-\frac{1}{\epsilon}\left[\inf_{x \in A} \inf_{u \in F} I^x(u) - \gamma\right]\right), \qquad \epsilon \leq \epsilon_0.$$

Definition 2.2.5 (FWULDP). Let $\{I^x\}_{x \in \mathcal{E}_0}$ be a family of good rate functions on a Polish space $\mathcal{X}$, parametrized by a parameter x in some Polish space $\mathcal{E}_0$ and let A be some Borel subset of $\mathcal{E}_0$. The family of $\mathcal{X}$-valued random variables $\{\xi_x^\epsilon\}_{\epsilon > 0, x \in A}$ is said to satisfy the Freidlin-Wentzell large deviations principle on $\mathcal{X}$ with rate functions I^x, uniformly for $x \in A$, if the following statement holds.

(i) For any $s \geq 0$, $\delta > 0$ and $\gamma > 0$, there exists $\epsilon_0 > 0$ such that

$$\inf_{x \in A}\left(\mathbb{P}\big(\xi_x^\epsilon \in B(x, \delta)\big) - \exp\Big(-\frac{I^x(\varphi) + \gamma}{\epsilon}\Big)\right) \geq 0, \qquad \epsilon \leq \epsilon_0,$$

for any $y \in \Phi^x(s)$, where $\Phi^x(s) := \{h \in \mathcal{X} : I^x(h) \leq s\}$.

(ii) For any $s_0 \geq 0$, $\delta > 0$ and $\gamma > 0$, there exists $\epsilon_0 > 0$ such that

$$\sup_{x \in A} \mathbb{P}\Big(\xi_x^\epsilon \in B^c(\Phi^x(s), \delta)\Big) \leq \exp\Big(-\frac{s - \gamma}{\epsilon}\Big), \quad \epsilon \leq \epsilon_0,$$

for any $s \leq s_0$, where

$$B^c(\Phi^x(s), \delta) = \{h \in \mathcal{X} : \mathrm{dist}_{\mathcal{X}}(h, \Phi^x(s)) \geq \delta\}.$$

The following two proposition of [55] shows necessary conditions in order to have the equivalences ULDP $\iff$ FWULDP and FWULDP $\iff$ DZULDP, respectively.

Proposition 2.2.3 (Theorem 2.5 of [55]). *Let $\{I^x\}_{x\in\mathcal{E}_0}$ be a family of good rate functions on a Polish space $\mathcal{X}$. Suppose that A is a set such that for any $s \geq 0$ the set*

$$\Lambda_{s,A} := \bigcup_{x\in A} \Phi^x(s)$$

is compact in $\mathcal{X}$. Then the family of $\mathcal{X}$-valued random variables $\{\xi_x^\epsilon\}_{\epsilon>0,x\in A}$ satisfies the Laplace principle on $\mathcal{X}$ with rate functions I^x, uniformly for $x \in A$, if and only if it satisfies the Freidlin-Wentzell large deviations principle on $\mathcal{X}$ with rate functions I^x, uniformly for $x \in A$.

Proposition 2.2.4 (Theorem 2.7 of [55]). *Let $\{I^x\}_{x\in\mathcal{E}_0}$ be a family of good rate functions on a Polish space $\mathcal{X}$. Suppose that A is a compact subset of a Polish space $\mathcal{E}_0$ and that the mapping $x \mapsto \Phi^x(s)$ from A to $\mathcal{B}(\mathcal{X})$ is continuous in the Hausdorff metric for any $s \geq 0$, i.e.*

$$x_n \to x, \quad \text{as } n \to \infty \implies \lim_{n\to\infty} \lambda(\Phi^{x_n}(s), \Phi^x(s)) = 0, \tag{2.2.13}$$

where for any $A_1, A_2 \in \mathcal{B}(\mathcal{X})$,

$$\lambda(A_1, A_2) := \max\left\{ \sup_{y\in A_1} \mathrm{dist}_{\mathcal{X}}(y, A_2), \sup_{y\in A_2} \mathrm{dist}_{\mathcal{X}}(y, A_1) \right\}.$$

Then the family of $\mathcal{X}$-valued random variables $\{\xi_x^\epsilon\}_{\epsilon>0,x\in A}$ satisfies the Dembo-Zeitouni large deviations principle on $\mathcal{X}$ with rate functions I^x, uniformly for $x \in A$, if and only if it satisfies the Freidlin-Wentzell large deviations principle on $\mathcal{X}$ with rate functions I^x, uniformly for $x \in A$.

Remark 2.2.4. It turns out that in the case where $\mathcal{E}_0$ is a reflexive space, then the requirement that A be a compact set can be weakened to the requirement that A be a closed and bounded set, as long as (2.2.13) still holds for any sequence x_n that converges to x *weakly* in $\mathcal{E}_0$. We will use this refinement of Proposition 2.4 in Chapter 4.

□

Finally, we also include the following uniform version of the contraction principle, which we will use in Chapter 5.

Theorem 2.2.4 (Uniform Contraction Principle)**.** *Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be two Polish spaces and let J be a good rate function on $\mathcal{X}_1$. Assume the family $\{\zeta^\epsilon\}$ of $\mathcal{X}_1$-valued random variables satisfies the large deviations principle on $\mathcal{X}_1$ with good rate function J. Suppose that $\{\mathcal{G}^x\}_{x\in A}$ is a family of continuous mappings from $\mathcal{X}_1$ to $\mathcal{X}_2$, indexed by x lying in some set A. Moreover, assume that the $\mathcal{G}^x$ are Lipschitz continuous, uniformly over $x \in A$, i.e.*

$$\sup_{x\in A} \sup_{\phi_1\neq\phi_2} \frac{\|\mathcal{G}^x(\phi_1) - \mathcal{G}^x(\phi_2)\|_{\mathcal{X}_2}}{\|\phi_1 - \phi_2\|_{\mathcal{X}_1}} =: L < \infty.$$

Then the family $\{\mathcal{G}^x(\xi^\epsilon\}_{\epsilon>0,x\in A}\}$ of $\mathcal{X}_2$-valued random variables satisfies a Freidlin-Wentzell large deviations principle in $\mathcal{X}_2$ with good rate functions I^x, uniformly for $x \in A$, where I^x is given by

$$I^x(\varphi) := \inf\{J(\psi) : \psi \in \mathcal{X}_1,\ \varphi = \mathcal{G}^x(\psi)\}.$$

Proof. (*Lower Bound.*) Fix $s \geq 0$, $\delta > 0$ and $\gamma > 0$. For each $x \in A$, let $\varphi^x \in \mathcal{X}_2$ be such that $I^x(\varphi^x) \leq s$. Therefore, for each $x \in A$ there exists $\psi^x \in \mathcal{X}_1$ such that $\varphi^x = \mathcal{G}^x(\psi^x)$ and $J(\psi^x) \leq I^x(\varphi^x) + \gamma/2$. Since $J(\psi^x) \leq s + \gamma/2$, for each $x \in A$, we have

$$\begin{aligned}\mathbb{P}(\xi_x^\epsilon \in B_{\mathcal{X}_2}(\varphi^x, \delta)) &= \mathbb{P}(\zeta_x^\epsilon \in \{f \in \mathcal{X}_1 : \|\mathcal{G}^x(f) - \varphi^x\|_{\mathcal{X}_2} < \delta\}) \\ &\geq \mathbb{P}\Big(\zeta_x^\epsilon \in \Big\{f \in \mathcal{X}_1 : \|f - \psi^x\|_{\mathcal{X}_1} < \frac{\delta}{L}\Big\}\Big) \\ &\geq \exp\Big(-\frac{J(\psi^x) + \gamma/2}{\epsilon}\Big) \geq \exp\Big(-\frac{I^x(\varphi^x) + \gamma}{\epsilon}\Big),\end{aligned}$$

for any $\epsilon \leq \epsilon_0$ with $\epsilon_0 > 0$ only depending on s, γ, δ and L.

(*Upper Bound.*) Fix $s_0 \geq 0$, $\delta > 0$ and $\gamma > 0$ and observe that

$$\mathbb{P}(\xi_x^\epsilon \in B^c_{\mathcal{X}_2}(\Phi^x(s), \delta)) = \mathbb{P}\Big(\zeta_x^\epsilon \in \Big\{f \in \mathcal{X}_1 : \inf_{\varphi \in \mathcal{X}_2 : I^x(\varphi) \leq s} \|\mathcal{G}^x(f) - \varphi\|_{\mathcal{X}_2} \geq \delta\Big\}\Big).$$

Note that for a given $f \in \mathcal{X}_1$, if there exists $\psi \in \mathcal{X}_1$ such that $J(\psi) \leq s$ and $\|f - \psi\|_{\mathcal{X}_2} < \frac{\delta}{L}$, then $I^x(\mathcal{G}^x(\psi)) \leq J(\psi) \leq s$ and $\|\mathcal{G}^x(f) - \mathcal{G}^x(\psi)\| < \delta$ for any $x \in A$. Hence, there exists some $\epsilon_0 > 0$ such that

$$\begin{aligned}\mathbb{P}(\xi_x^\epsilon \in B^c_{\mathcal{X}_2}(\Phi^x(s), \delta)) &\leq \mathbb{P}\Big(\zeta_x^\epsilon \in \Big\{f \in \mathcal{X}_1 : \inf_{\psi \in \mathcal{X}_1 : J(\psi) \leq s} \|f - \psi\|_{\mathcal{X}_1} \\ &\geq \frac{\delta}{L}\Big\}\Big) \leq \exp\Big(-\frac{s - \gamma}{\epsilon}\Big),\end{aligned}$$

for any $\epsilon \leq \epsilon_0$. □

2.3 Long time asymptotics

Thus far, we have considered rare events for stochastic equations in the space of trajectories $C([0,T];\mathbb{R}^d)$ for fixed time horizon $[0,T]$. We now seek to understand the small noise asymptotics of the long-time behavior of the dynamical system.

2.3.1 Exit problem

Let X_t^ϵ be the solution to SDE (2.0.1) in $\mathbb{R}^d$. We now assume that the drift field $b : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth vector field that admits a stable, globally attracting equilibrium at the point $\mathcal{O}$, as in Figure 2.2.

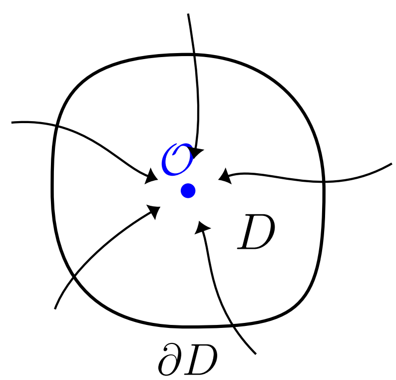


Fig. 2.2: Drift field with global attractor $\mathcal{O}$.

Let us consider what happens to the solution X_t^ϵ when ϵ is small. Since $\mathcal{O}$ is an attractor of the deterministic system (2.0.2), the solution X_t^ϵ will follow closely the deterministic trajectories until very close to the attractor $\mathcal{O}$. Once there, X_t^ϵ will typically remain close to the attractor $\mathcal{O}$ for large periods of time, taking only minor noise-induced excursions away from $\mathcal{O}$. However, when observed over an exponentially large in ϵ^{-1} time horizon, larger noise-induced excursions away from $\mathcal{O}$ will eventually happen.

To study this phenomenon, we consider the problem of the exit of X_t^ϵ from a domain $D \subset \mathbb{R}^d$. As in the figure, we suppose $D \subset \mathbb{R}^d$ is a connected and compact set with smooth boundary that contains the attractor $\mathcal{O}$. Moreover, we suppose that every trajectory of the deterministic system (2.0.2) that starts inside D stays inside D for all future times, and we assume that the initial condition x_0 of (2.0.1) also lies in D. We are interested in quantifying the first exit time of the solution X_t^ϵ from D. Thus, we let the stopping time

$$\tau^\epsilon := \inf\{t > 0 : X_t^\epsilon \notin \bar{D}\},$$

denote the time of first exit of X_t^ϵ from D. Under certain growth assumptions on b, it can be shown that for any $\epsilon > 0$ the exit time τ^ϵ is finite almost surely. The escape times τ^ϵ will of course grow to $+\infty$ as the noise magnitude ϵ is sent to 0; however, it turns out that we can more closely quantify this growth rate.

In addition to quantifying the escape times, it is also possible to identify the most *likely* path that X_t^ϵ will take to exit the bounded domain D. In view of the structure of the large deviations theory in the previous sections, we expect that the most likely exit path will satisfy a least action principle as before, and this is indeed the case. Since we now must consider paths of arbitrary time length T to minimize over, it is useful to define the cost function

$$U(x,y) := \inf\{S_T(\varphi) : T > 0, \quad \varphi \in C([0,T];\mathbb{R}^d), \quad \varphi(0) = x, \quad \varphi(T) = y\}, \tag{2.3.1}$$

where S_T denotes the action integral

$$S_T(\varphi) := \begin{cases} \frac{1}{2}\int_0^T \|\dot{\varphi}_t - b(\varphi_t)\|^2 \, dt, & \text{if } \varphi \text{ is absolutely continuous,} \\ \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.3.2}$$

The quantity $U(x, y)$ thus represents the cost, in terms of the action, of the cheapest path from x to y. However, it is important to note that in this exit problem, *any* path starting in D will first get sucked in very close to the attractor $\mathcal{O}$ with overwhelming probability, regardless of where it starts. In other words, the initial condition of the path is effectively transient, so that the only cost that will matter when determining exit times will be the cost of moving from the attractor $\mathcal{O}$ to the boundary ∂D. In view of this, we define the function $U : \mathbb{R}^d \to [0, +\infty]$, called the *quasi-potential* with respect to the attractor $\mathcal{O}$,

$$U(x) := U(\mathcal{O}, x).$$

It turns out that the quasi-potential controls much of the asymptotics of the escape time and most likely exit trajectory. Indeed, if there is a unique minimizer z^* of U on the boundary ∂D, then with overwhelming probability as $\epsilon \to 0$, the escape trajectory will exit in an arbitrarily small neighborhood of x^*. Moreover, the portion of the escape trajectory away from $\mathcal{O}$ will fall within an arbitrary delta tube of the minimum action path ϕ^* that terminates at x^*. These results concerning the manner of exit are proven by Freidlin and Wentzell [37], and are, respectively, Theorems 2.1 and 2.3 of [37]. Since the latter is somewhat complicated, we state only the former here.

Theorem 2.3.1 (Theorem 2.1 of [37]). *Suppose there exists $x^* \in \partial D$ which is the unique minimizer of U on ∂D. Then for any $\delta > 0$.*

$$\lim_{\epsilon \to 0} \mathbb{P}(\|X^{\epsilon}_{\tau^{\epsilon}} - x^*\| < \delta) = 1.$$

Remark 2.3.1. If there is not a unique minimizer of U on the boundary, the distribution of exit locations on ∂D will converge to a measure on the set of global minimizers of U on ∂D.

□

Perhaps more surprising is that the leading order term in the asymptotics of the first escape time τ^{ϵ} are also entirely determined by the minimum of the quasi-potential along the boundary.

Theorem 2.3.2 (Theorem 4.1 of [37]). *Assume that the domain D is attracted to $\mathcal{O}$ and that the boundary ∂D is smooth. Then for any initial condition x in the interior of D,*

$$\lim_{\epsilon \to 0} \epsilon \log \mathbb{E}\tau^{\epsilon}_x = \min_{y \in \partial D} U(y), \tag{2.3.3}$$

where τ^{ϵ}_x is the first exit time corresponding to the process X^{ϵ}_t with initial condition x.

Outline of proof of Theorem 2.3.2. (Upper bound) The critical step in the upper bound is the claim that for any $\eta > 0$ there exists time T_0 large enough that

$$\liminf_{\epsilon \to 0} \epsilon \log \inf_{x \in D} \mathbb{P}(\tau^{\epsilon}_x \le T_0) > - \min_{y \in \partial D} U(y) - \eta. \tag{2.3.4}$$

Equation (2.3.4) immediately provides an exponential lower bound on the probability of τ_x^ϵ landing in a fixed ϵ-independent time interval. From there, the upper bound of (2.3.3) can be established by applying (2.3.4) on each of the intervals $[0, T_0], [T_0, 2T_0], \ldots$ etc. Indeed,

$$\begin{aligned}
\mathbb{E}\tau_x^\epsilon = \int_0^\infty \mathbb{P}(\tau_x^\epsilon > t)dt \ &\leq \ T_0 \sum_{k=0}^{\infty} \sup_{x \in D} \mathbb{P}(\tau_x^\epsilon > kT_0) \\
&\leq T_0 \sum_{k=0}^{\infty} (1 - \inf_{x \in D} \mathbb{P}(\tau_x^\epsilon \leq T_0))^k \ = \ \frac{T_0}{\inf_{x \in D} \mathbb{P}(\tau_x^\epsilon \leq T_0)} \\
&\leq T_0 \exp\Big(\frac{\min_{y \in \partial D} U(y) + \eta}{\epsilon}\Big),
\end{aligned}$$

and the upper bound follows. The critical component is then the justification of (2.3.4). This statement can be deduced from the large deviations statement for the paths. In particular, for any initial point $x \in D$, one can construct the three-stage path $\varphi_x \in C([0, T_0]; \mathbb{R}^d)$ for $T_0 = T_1 + 1 + T_2$ by the recipe below where T_1 and T_2 are to be determined but picked independently of x. The construction is illustrated in Figure 2.3.

- Stage (i): Follow the deterministic trajectory (2.0.2) on $[0, T_1]$ for fixed large time T_1.

- Stage (ii): Follow a straight line path on $[T_1, T_1 + 1]$ with uniform speed chosen such that $\varphi(T_1 + 1) = \mathcal{O}$.

- Stage (iii): Follow an "almost" optimal escape path $\tilde{\phi}$ from $\mathcal{O}$ to just past the boundary that satisfies $S_{T_2}(\tilde{\phi}) \leq \min_{y \in \partial D} U(y) + \eta/2$. Let $\delta > 0$ be the final

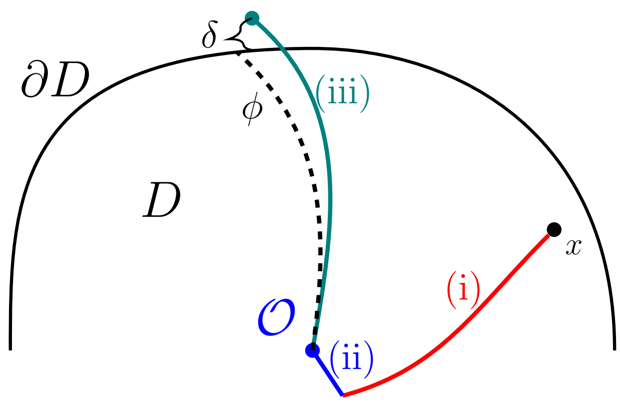


Fig. 2.3: Illustration of the three stage path $\varphi_x \in C([0, T_1 + 1 + T_2]; \mathbb{R}^d)$. The dashed path ϕ is the optimal escape trajectory from $\mathcal{O}$ to the boundary ∂D.

distance past the boundary.

Note that Stage (i) contributes nothing to the action $S_{T_0}(\varphi_x)$ and that T_1 can be picked large enough that the action contribution from Stage (ii) is less than $\eta/2$ for any x. Therefore, $S_{T_0}(\varphi_x) \leq \min_{y \in \partial D} U(y) + \eta$ for each $x \in D$. Then, equation (2.3.4) follows by applying the uniform large deviations principle lower bound to the set

$$\Psi := \bigcup_{x \in \bar{D}} \{\psi \in C([0, T_0]; \mathbb{R}^d) : \sup_{0 \leq t \leq T_0} \|\psi_t - \varphi_x(t)\| < \delta/2\},$$

and noting that $\mathbb{P}(X_t^\epsilon \in \Psi) \leq \mathbb{P}(\tau_x^\epsilon \leq T_0)$ since every path in Ψ exits the domain. Actually, since the set Ψ is open in $C([0, T_0]; \mathbb{R}^d)$, either the FWULDP or the DZULDP lower bound is sufficient.

□

Remark 2.3.2. We emphasize that it is crucial that the large deviations principle for the paths is uniform on the domain D so that (2.3.4) holds with the infimum within the limit. This entails the need for a large deviations principle that is uniform with respect to initial conditions in bounded sets. This does not present an issue in the finite dimensional setting where bounded sets are pre-compact; however, when one tries to prove the analog of Theorem 2.3.2 for an infinite dimensional problem, one typically encounters the problem of only being able to prove a Laplace principle that is uniform on compact sets.

□

2.3.2 Gradient systems

In order to understand the structure of the quasi-potential, it is helpful to consider a simpler class of SDEs: namely, those given by a gradient system. We assume for this section that $b = -\nabla F$ for some smooth potential $F : \mathbb{R}^d \to \mathbb{R}$. We assume also that F is bounded below with a global minimum at $\mathcal{O}$ and that $\mathcal{O}$ is globally attracting.

In this case, the quasi-potential actually coincides with the potential function, up to a multiplicative factor of 2. Indeed, for any path $\varphi \in C([0,T];\mathbb{R}^d)$

$$\begin{aligned} S_T(\varphi) &= \frac{1}{2}\int_0^T \|\dot{\varphi}_t + \nabla F(\varphi_t)\|^2\, dt \\ &= \frac{1}{2}\int_0^T \|\dot{\varphi}_t - \nabla F(\varphi_t)\|^2\, dt + 2\int_0^T \dot{\varphi}_t \cdot \nabla F(\varphi_t) dt \\ &= \frac{1}{2}\int_0^T \|\dot{\varphi}_t - \nabla F(\varphi_t)\|^2\, dt + 2F(\varphi_T) - 2F(\varphi_0). \end{aligned} \tag{2.3.5}$$

In view of the definition of the quasi-potential (2.3.1) and assuming without loss of generality that $F(\mathcal{O}) = 0$, we then have

$$U(x) = 2F(x) + \inf_{T>0} \left\{ \frac{1}{2} \int_0^T \|\dot{\varphi}_t - \nabla F(\varphi_t)\|^2 \, dt : \right.$$
$$\left. \varphi \in C([0,T]; \mathbb{R}^d, \ \varphi(0) = \mathcal{O}, \ \varphi(T) = x \right\}. \tag{2.3.6}$$

However, the infimum on the right hand side can be made arbitrarily small by considering the time reversal of the solution to

$$\dot{\phi}_t = -\nabla F(\phi_t), \qquad \phi(0) = x. \tag{2.3.7}$$

Indeed, if T is large, then the path $\varphi_t = \phi_{T-t}$ starts near $\mathcal{O}$, contribute nothing to the integral in (2.3.6) and ends at x. Moreover, the integral contribution of adding a small segment from $\mathcal{O}$ to ϕ_T at the beginning of φ can be made arbitrarily small by taking T arbitrarily large. Thus for the gradient system case, we simply have

$$U(x) = 2F(x). \tag{2.3.8}$$

Remark 2.3.3. The characterization (2.3.8) also holds true in the event where the drift has the structure

$$b(x) = -\nabla F(x) + \ell(x), \tag{2.3.9}$$

where F is continuously differentiable and $\ell(x) \cdot \nabla F(x)$ for each $x \in \mathbb{R}^d$. Indeed, the computation (2.3.5) holds true in the same manner.

□

We next consider the invariant probability $\mu^\epsilon \in \mathcal{P}(\mathbb{R}^d)$ measure of system (2.0.1) in the case where $b(x) = -\nabla F(x)$. There are a number of ways to identify μ^ϵ. One is to consider the stationary Fokker-Planck equations for the density $p^\epsilon(x)$ of μ^ϵ:

$$0 = -\nabla \cdot \Big[b(x)p^\epsilon(x)\Big] + \frac{\epsilon}{2}\Delta p^\epsilon(x), \qquad \int_{\mathbb{R}^d} p^\epsilon(x)dx = 1. \tag{2.3.10}$$

Indeed by inspection, one can see that the Gibbs measures on $\mathbb{R}^d$ with density

$$p^\epsilon(x) = C \exp\Big(-\frac{2F(x)}{\epsilon}\Big) = C \exp\Big(-\frac{U(x)}{\epsilon}\Big), \tag{2.3.11}$$

where C is a normalization factor, gives a solution to (2.3.10). In particular, the Gibbsian nature of the invariant measure implies that for a (sufficiently regular) set $D \subset \mathbb{R}^d$

$$\lim_{\epsilon \to 0} \epsilon \log \mu^\epsilon(D) = -\inf_{y \in D} U(y). \tag{2.3.12}$$

Remark 2.3.4. In the case of $b(x) = -\nabla F(x) + \ell(x)$, we see that the time invariant measure μ^ϵ is the same Gibbs measure if in addition ℓ is divergence free.

2.3.3 Stationary measures

In the case of a general drift field b admitting a globally attracting equilibrium $\mathcal{O}$, explicit formulas for the quasi-potential and invariant measure are unavailable. Indeed, the stationary measure will not in general be a Gibbs measure. However, it turns out that the asymptotic relationship (2.3.12) between the quasi-potential and stationary measure μ^ϵ will still hold. In fact, (2.3.12) is really a statement that the invariant measures satisfy a large deviations principle in $\mathbb{R}^d$ with good rate function U.

Theorem 2.3.3 (Theorem 4.3 of [37])**.** *Assume the whole space is attracting to $\mathcal{O}$ and assume b is such that equation (2.0.1) admits a unique invariant measure μ^ϵ. Then the family μ^ϵ satisfies a large deviations principle in $\mathbb{R}^d$ with good rate function U.*

Outline of proof. The structure of the proof to this is similar to the proof of 2.3.2.

(Lower Bound) The lower bound follows by using the invariance property of the measures μ^ϵ. Let $\eta > 0$. For any $x \in \mathbb{R}^d$ and $\delta > 0$, we have for any $t \geq 0$

$$\mu^\epsilon(B(x,\delta)) = \int_{\mathbb{R}^d} \mathbb{P}(\|X_t^\epsilon - x\| < \delta) d\mu(y).$$

If we then restrict to any ball $B(0,R)$ for $R > 0$, we can construct a collection of paths $\{\varphi^y\}_{y \in B(0,R)} \subset C([0,T_0];\mathbb{R}^d)$, constructed in exactly the same way and with same T_0 as in the proof of 2.3.2, such that $\varphi^y(0) = y$ and $S_{T_0}(\varphi^y) < U(x) + \eta/2$. Here, $B(0,R)$ is taking the place of the bounded set D. Then, taking $t = T_0$ we have

$$\begin{aligned}\mu^\epsilon(B(x,\delta)) &\geq \int_{\mathbb{R}^d} \mathbb{P}(\sup_{0 \leq t \leq T_0} \|X_t^\epsilon - \varphi_t^y\| < \delta/2) d\mu(y) \\ &\geq \mu^\epsilon(B(0,R)) \inf_{y \in B(0,R)} \mathbb{P}(\|X^\epsilon - \varphi^y\|_{C([0,T_0];\mathbb{R}^d)} < \delta/2)\end{aligned}$$

Then, since $\mu^\epsilon(B(0,R))$ goes to 1 as $\epsilon \to 0$, by the uniform large deviations principle, we have for sufficiently small ϵ

$$\mu^\epsilon(B(x,\delta)) \geq \frac{1}{2} \exp\Big(- \frac{1}{\epsilon}(U(x) + \eta)\Big).$$

$\square$

Theorem 2.3.3 tells us that the primary term in the Gibbsian asymptotics of the invariant measure is the quasi-potential. We can investigate the invariant measure further by considering again the Fokker-Planck equations with the WKB ansatz that the density takes the form of Gibbs density multiplied by a subexponential pre-factor

$$p^\epsilon(x) = C(x)\exp\Big(-\frac{U(x)}{\epsilon}\Big).$$

Justification for the WKB approximation is given in e.g. [62, 44]. By plugging the ansatz into the Fokker-Planck equations and grouping by order in ϵ, we obtain

$$\begin{aligned}
0 = &-\frac{p^\epsilon(x)}{\epsilon}\Big[b(x)\cdot\nabla U(x) + \frac{1}{2}\left\|\nabla U(x)\right\|^2\Big] \\
&- \exp\Big(-\frac{U(x)}{\epsilon}\Big)\Big[(\nabla\cdot b + \frac{1}{2}\Delta U)(x)C(x) + (b+\nabla U)(x)\cdot\nabla C(x)\Big] \\
&+ \frac{\epsilon}{2}(\nabla\cdot C)(x)\exp\Big(-\frac{U(x)}{\epsilon}\Big).
\end{aligned}$$

Taking the lowest order expansion then gives the Hamilton-Jacobi-Bellman equation

$$b\cdot\nabla U + \frac{1}{2}\left\|\nabla U\right\|^2 = 0, \tag{2.3.13}$$

while taking the next order in the perturbation series provides a transport equation for the leading order term in the exponential prefactor

$$(\nabla\cdot b + \frac{1}{2}\Delta U)C + (b+\nabla U)\cdot\nabla C = 0. \tag{2.3.14}$$

Remark 2.3.5. We remark that in general the quasi-potential is not continuously differentiable. However, it can be shown that it is at worst Lipschitz continuous [37] and the Lebesgue measure of the set of points for which it is not differentiable is 0. Moreover, it can be shown that when U is continuously differentiable, it is a classical solution to (2.3.13) endowed with the boundary condition $U(\mathcal{O}) = 0$, and a viscosity solution otherwise.

□

2.3.4 The quasi-potential

It is important to note that in general the infimum in T in the definition of the quasi-potential is never achieved. We observed this directly in the case of a gradient system where one cannot have the integral term in (2.3.6) be 0 for *finite* T, but instead must be taken as a limit as $T \to \infty$. Physically, this occurs because the optimal escape path leaves the attractor $\mathcal{O}$ with infinitesimal initial speed.

On the other hand, a single minimizing *path* in $\mathbb{R}^d$ in general will exist if a different parametrization, such as an arclength parametrization, is considered. Toward the purpose of identifying this minmizing path, it is convenient to introduce an alternate action functional, called the geometric action $\tilde{S}_L : C([0,T];\mathbb{R}^d) \to [0,+\infty]$ given by

$$\tilde{S}_L(\varphi) := \begin{cases} \int_0^L \Big[\|b(\varphi_r)\| \, \|\dot{\varphi}_r\| - b(\varphi_r) \cdot \dot{\varphi}_r \Big] dr, & \text{if } \varphi \text{ is absolutely continuous,} \\ +\infty, & \text{otherwise.} \end{cases} \tag{2.3.15}$$

Lemma 2.3.1. *For any x in the well of attraction of $\mathcal{O}$, it holds that*

$$U(x) = \inf\{\tilde{S}_L(\phi) : L > 0, \quad \phi \in C([0,L];\mathbb{R}^d), \quad \phi_0 = \mathcal{O}, \quad \phi_L = x\}. \tag{2.3.16}$$

Proof. Observe that

$$
\begin{aligned}
S_T(\varphi) &= \frac{1}{2}\int_0^T \|\dot{\varphi}_t - b(\varphi_t)\|^2\, dt \\
&= \frac{1}{2}\int_0^T \Big[\, \|\dot{\varphi}_t\|^2 + \|b(\varphi_t)\|^2 - 2b(\varphi_t)\cdot\dot{\varphi}_t\Big] dt \\
&\geq \int_0^T \Big[\, \|b(\varphi_t)\|\,\|\dot{\varphi}_t\| - b(\varphi_t)\cdot\dot{\varphi}_t\Big] dt
\end{aligned}
\tag{2.3.17}
$$

Hence $U(x)$ is greater than the right hand side of (2.3.16). Next, suppose that $\phi \in C([0,T];\mathbb{R}^d)$ is such that $\phi(0) = \mathcal{O}$ and $\phi(T) = x$. Let $\tilde{\phi}$ be a re-parametrization of ϕ such that $\left\|\dot{\phi}_r\right\| = \|b(\phi_r)\|$ for all r in the domain of the new variable $[0,L]$. Then the inequality in (2.3.17) is an equality and $S_T(\phi) = \tilde{S}_L(\tilde{\phi})$. Hence we get (2.3.16). □

The Hamilton-Jacobi equation (2.3.13) for the quasi-potential we obtained from the WKB ansatz can be derived rigorously from the minimum action definition of U using standard techniques in Hamiltonian mechanics (see [1], for example). We present a third route to this equation by viewing the quasi-potential definition as an instantaneous optimal control problem. This route will provide useful information that we will use in Chapter 3. Consider U defined via the geometric action (2.3.16). We can re-write the quasi-potential as

$$
\begin{aligned}
U(x) &= \inf_{y\in\mathbb{R}^d}\{U(x-y) + U(x-y,x)\} \\
&= \inf_{L>0} \inf_{\substack{\varphi\in C([0,L];\mathbb{R}^d):\\ \varphi_L = x}} \Big\{U(\varphi(0)) + \tilde{S}_L(\varphi)\Big\} \\
&= \inf_{L>0} \inf_{\substack{\varphi\in C([0,L];\mathbb{R}^d):\\ \varphi_L = x}} \Big\{U(x - \int_0^L \dot{\varphi}_r dr) + \tilde{S}_L(\varphi)\Big\}.
\end{aligned}
\tag{2.3.18}
$$

In addition, if we restrict to an arclength parameterization of the paths in the geometric action so that L indicates the length of curve φ, then the equality in (2.3.18) actually holds for each L individually less than the arclength of the minimum action path (MAP) passing through x. Thus for any small $\delta > 0$, we have

$$\begin{aligned} U(x) &= \inf_{\substack{\varphi \in C([0,\delta];\mathbb{R}^d): \\ \varphi_\delta = x,\ \|\dot{\varphi}_r\|=1}} \Big\{ U(x - \int_0^\delta \dot{\varphi}_r dr) + \int_0^\delta [\|b(\varphi_r)\| - b(\varphi_r)\cdot\dot{\varphi}_r] dr \Big\} \\ &= \inf_{\phi\in\mathbb{R}^d:\|\phi\|=1} \Big\{ U(x) - \delta\nabla U(x)\cdot\phi + \delta(\|b(x)\| - b(x)\cdot\phi) + \mathcal{O}(\delta^2) \Big\}, \end{aligned}$$

where in the last line, we assume that the quasi-potential is continuously differentiable at x. By taking $\delta \to 0$, we obtain the instantaneous cost minimization problem

$$0 = \inf_{\phi\in\mathbb{R}^d:\|\phi\|=1} \Big[\|b(x)\| - (b+\nabla U)(x)\cdot\phi \Big].$$

In this formulation, the minimizing direction ϕ corresponds to the direction at x of the minimum action path from $\mathcal{O}$ to x. We see also that this minimizing direction ϕ is in the $b + \nabla U$ direction. Indeed, taking $\phi = \dfrac{b+\nabla U}{\|b+\nabla U\|}$, we obtain

$$0 = \|b(x)\| - \|b(x) + \nabla U(x)\|\,, \tag{2.3.19}$$

which can be rewritten as equation (2.3.13). The utility of this approach, however, is that we see directly the geometric relationship between the quasi-potential at a point and the MAP. Namely, for each x where the quasi-potential is differentiable, we have

$$\nabla U(x) = \|b(x)\|\,\phi_x - b(x), \tag{2.3.20}$$

where ϕ_x is the unit velocity vector of the MAP from $\mathcal{O}$ to x at the point x. This will be a fundamental piece of information in our development of our Jet solver in Chapter 3.

Chapter 3: Efficient Jet Marcher for computing the quasi-potential

3.1 Introduction

We return again to the finite-dimensional setting. Consider the stochastic differential equations in $\mathbb{R}^d$,

$$dX_t = b(X_t)dt + \sqrt{\epsilon}dW_t, \qquad X_0 = x_0 \in \mathbb{R}^d, \tag{3.1.1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, W_t is a standard Brownian motion in $\mathbb{R}^d$ and $\epsilon > 0$ is a small parameter. As in Section 2.3.1, we assume that b admits a stable, attracting equilibrium at $\mathcal{O} \in \mathbb{R}^d$. As we have seen, the quasi-potential $U(x)$ (defined by (2.3.1)) with respect to $\mathcal{O}$ controls much of the small-noise and long-time asymptotic behavior of equation (3.1.1). In this chapter, we discuss existing techniques for numerically computing the quasi-potential, and then describe a novel algorithm for computing the quasi-potential on a mesh in the 2-dimensional case.

As discussed in Chapter 2, the quasi-potential $U(x)$ can be seen as the cost of the cheapest possible path that begins at the attractor $\mathcal{O}$ and ends at the point x, where the cost is measured by either the Freidlin-Wentzell action functional (2.3.2) or the geometric action functional (2.3.15). Hence, a logical way to compute $U(x)$

for a given x is to try to conduct this minimization numerically. Indeed, one can construct a numerical version of the action function via quadrature and then perform a high-dimensional minimization over a suitably rich path space. Since methods of this form invariably return the minimum action paths (MAP) themselves, we refer to them as *path-based* methods.

In practice, techniques conducting the minimizations of both types of action functionals are used. The Minimum Action Method (MAM) [31] and Adaptive Minimum Action Method (AMAM) [67] find MAPs by minimizing the Freidlin-Wentzell action, while the Geometric Minimum Action Method (GMAM) [38] does so by minimizing the geometric action. We describe the key details of these approaches in Section 3.2. By design, these techniques are applicable to both finite and infinite dimensional problems. For example, the GMAM was used to find the MAP between two stationary solutions of a 2D reaction-diffusion partial differential equation in [38] and to find the MAP between two solitary waves of different amplitudes of nonlinear Schrodinger equations in [53].

An important advantage of path-based methods is that they are computationally cheap and suitable for any dimension. However, a number of key drawbacks. First, they allow for computation of the quasi-potential along a *single* MAP, but do not provide a mechanism for efficient computation of the quasi-potential over an entire region of space. In particular, they are ill suited for the task of identifying quasi-potential minima. As we have seen, locating quasi-potential minima is an important step in quantifying exit locations and escape times for the exit problem (Section 2.3.1). Second, path-based techniques work well if the MAP is relatively simple, but their convergence tends to stall if the MAP exhibits spiraling or other complicated behavior. Finally, the MAPs obtained by these methods are *biased* by the initial guess, typically taken to be straight line segments, and may only be *local* minimizers of the action that lead to incorrect estimates of the quasi-potential. While various attempts have been undertaken to address these issues [63, 42], none of the proposed solutions have become commonly used due to complexity and lack of robustness.

With the application of determining quasi-potential minima in mind, the majority of this chapter focuses on *mesh-based* quasi-potential solvers, whose objective is the efficient computation of $U(x)$ on a mesh. As with all numerical PDE solvers, these methods suffer from the curse of dimensionality, so that solving for $U(x)$ on a mesh in a high-dimensional space is infeasible. Current mesh-based quasi-potential solvers are implemented only in 2 [14, 21, 22] and 3 dimensions [66]. Our algorithm, the Efficient Jet Marching (EJM) method, which is described in full detail in Section 3.4, addresses only the $d = 2$ case with additive noise. The task of extending this algorithm to 3-dimensions remains as future work.

A full description of mesh-based quasi-potential solvers is provided in Section 3.3, although we make a few preliminary remarks here. The discussed methods are all descendants of Dijkstra's algorithm for identifying the shortest path in a network [29], and, more specifically, of Sethian's Fast Marching algorithm for solving the eikonal equation [57]. Indeed, recall from Section 2.3.4 that the quasi-potential is a viscosity solution to the Hamilton-Jacobi-Bellman equation

$$\|\nabla U(x)\|^2 + 2b(x) \cdot \nabla U(x) = 0, \qquad U(\mathcal{O}) = 0, \tag{3.1.2}$$

which can be interpreted as an anisotropic eikonal equation for the evolution of a wavefront so that the ideas of Sethian's fast marching algorithm and its related descendants [58, 59] can be applied.

At their core, these types of methods treats the quasi-potential $U(x)$ as the first hitting time at location x of a fictitious wavefront propagating outwards from the point attractor $\mathcal{O}$, as illustrated in Figure 3.1. In this analogy, the level set $\{x : U(x) = T\}$ represents the state of the wavefront at "time" T. The MAP passing through a point x can be seen as representing the liftetime trajectory of the particular particle that was the first to reach x. Fast-marching algorithms seek to identify these "leading" particles and make local approximations of their paths between nearby mesh points, from which estimates of the wavefront arrival times are made. There are, however, some key difficulties in extending the fast marching algorithm to the quasi-potential problem; most notably the potential unboundedness of the particle velocity function. The first such algorithm for solving the quasi-potential was introduced in [14], in which a thorough discussion of the the obstacles and necessary modifications is given.

In Section 3.4, we describe in detail our EJM algorithm, which is a descendant of the fast-marching based methods of [14, 21]. We also briefly mention here the two key features that distinguish it from previous quasi-potential solvers.

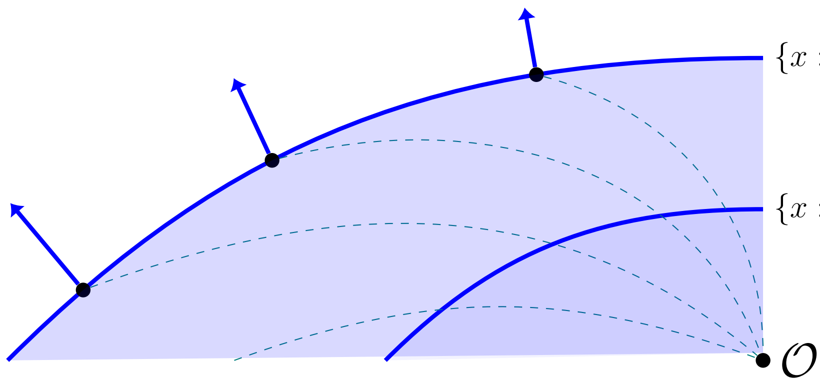


Fig. 3.1: The quasi-potential problem as a wavefront propagation problem.

The *first* key feature is the use of higher-order approximations of (a) the quasipotential between mesh points and (b) local segments of MAPs. Concerning (a), our solver propagates the quasipotential *and* its gradient by taking advantage of the geometric relationship (see Section 2.3.4)

$$\nabla U(x) = \|b(x)\| \, \dot{\phi} - b(x), \tag{3.1.3}$$

between the quasi-potential $U(x)$ and MAP ϕ passing through x. This enables us to use Hermite cubic interpolation of U between mesh points rather than linear interpolation as done in [14, 21, 22, 66]. The idea of using Hermite interpolation originally comes from the jet scheme for solving an advection equation [48]. Concerning (b), our solver approximates MAP segments using cubic curves, while all previous quasi-potential solvers have approximated them with straight line segments. We remark that a similar family of jet marching methods for isotropic eikonal equations was recently introduced in [54].

Although rigorous numerical analysis remains as future work, we expected the improved approximation scheme to provide our method with an $\mathcal{O}(h^2)$ accuracy convergence rate with respect to the mesh spacing h. As we show in Section 3.5, this is precisely what we have observed empirically in all but the most extreme scenarios. To our knowledge, this is the first such quasi-potential solver with second order in h accuracy.

The aforementioned improvement in accuracy does come at the cost of an increase in computing time due to the more computation-intensive MAP approximations. However, this increase in runtimes is offset by the algorithm's *second* key feature: implementation of pre-computed anisotropic stencils, inspired by the Anisotropic Stencil Refinement (ASR) algorithm of Mirebeau [47]. Roughly speaking, the presence of potentially unbounded particle speeds in the quasi-potential problem requires the searching of very large "neighborhoods" of mesh points in the search for MAPs. Mirebeau's ASR algorithm pre-computes smaller, more targeted neighborhoods, which results in a significantly smaller number of total MAP searches at a comparable overall accuracy. Adoption of a modified version of these ideas provided EJM with a significant reduction in computation time. A description of Mirebeau's ASR algorithm is given in Section 3.3.

A full analysis of the accuracy-speed trade-off is provided in Section 3.5. There, we compare the performance of the EJM algorithm with other mesh-based quasi-potential solvers on a variety of different drift fields b.

3.2 Path-based approaches

In this section, we briefly describe two commonly used path-based techniques for identifying the MAP from an attractor $\mathcal{O}$ to a point x. Both techniques are suitable for finite and infinite dimensional problems.

AMAM. The adaptive minimum-action method (AMAM) [67] solves for the MAP from $\mathcal{O}$ to point x by numerically minimizing the Freidlin-Wentzell action functional S_T (2.3.2) for some large value of T. To do so, the action S_T is approximated using a midpoint quadrature rule on a partition $\{t_k\}_{k=0}^m$ of $[0,T]$. More precisely, if the path $\varphi \in C([0,T];\mathbb{R}^d)$ is approximated by its values at its values $\Phi_k := \varphi_{t_k}$, then the numerical action is given by

$$S_{t_0,\dots,t_m} = \frac{1}{2}\sum_{k=0}^{m-1} \left\| \frac{\Phi_{k+1}-\Phi_k}{\Delta t_k} - b(\varphi_{k+1/2}) \right\|^2 \Delta t_k, \tag{3.2.1}$$

where $\Delta t_k = t_{k+1} - t_k$ and $\Phi_{k+1/2} = (\Phi_{k+1}+\Phi_k)/2$. The key difficulty with minimizing this discretized action is the question of how to select the mesh $\{t_k\}_{k=1}^m$. Indeed, the true MAP spends much more time near the attractor $\mathcal{O}$, whence the integrand of S_T is very small. For this reason, minimization of (3.2.1) along a simple uniformly spaced mesh on $[0,T]$ will provide poor results, and a more strategically refined mesh is needed.

The AMAM solves this issue by implementing a moving mesh technique, in which the t-mesh for discretizing $S_T(\varphi)$ changes along with the optimal path $\{\Phi_k\}_{k=0}^m$.

Indeed, an satisfactory time mesh $\{t_k\}_{k=0}^m$ is one for which each term in the sum (3.2.1) is of the same order of magnitude. To achieve this AMAM *weights* the time steps by the monitor function

$$w(t) := \frac{1}{\sqrt{1 + C\left\|\dot{\varphi}_t\right\|^2}},$$

for some large $C > 0$ and choose $\{t_k\}_{k=0}^m$ such that the *product* $w(t_k)\Delta t_k$ is approximately constant for each $k = 0, ..., m-1$. These can be found by solving the Euler-Lagrange equation

$$\frac{d}{dt}\Big(\frac{1}{w(t)}\frac{d\alpha}{dt}\Big) = 0, \quad \alpha(0) = 0, \quad \alpha(T) = T, \tag{3.2.2}$$

by finite difference for the re-weighted mesh variable α over the uniform partition of $[0, T]$. The optimal mesh is then given by the solution α discretized over the uniform partition of $[0, T]$, i.e. $t_k := \alpha(\frac{kT}{m})$.

In view of this re-weighting procedure, the AMAM algorithm attempts to solve for the MAP terminating at x by proceeding as follows.

- *(Step 1).* Set mesh $\{t_k\}_{k=0}^m$ to the uniform partition of $[0, T]$, and start with the initial guess $\{\Phi_k\}_{k=0}^m$ corresponding to a straight line path φ from $\mathcal{O}$ to x.

- *(Step 2).* Solve the Euler-Lagrange equation (3.2.2) numerically to construct a new mesh $\{t_k\}_{k=0}^m$.

- *(Step 3).* Run R iterations of a high-dimensional optimization routine (such as L-BFGS) on the right hand side of (3.2.1) for the partition $\{t_k\}$.

- *(Step 4).* Terminate if the optimization termination condition is met *and* if the quantity $\{w(t_k)\Delta t_k\}$ are sufficiently uniform in k to meet a mesh termination condition. Otherwise, return to step 2 to re-compute the mesh on the current iterate of $\{\Phi_k\}_{k=0}^m$.

GMAM. The geometric minimum action method (GMAM) [38] solves for the MAP from $\mathcal{O}$ to x by instead minimizing the geometric action $\tilde{S}_L$ (2.3.15). By a standard calculus of variations computation, it can be shown that a minimizing path ϕ of $\tilde{S}_L$ satisfies the Euler-Lagrange equations

$$0 = \frac{d\tilde{S}}{d\varphi}(\phi) = -\lambda^2\ddot{\phi}_r + \lambda\Big[\nabla b(\phi_r) - (\nabla b(\phi_r))^T\Big]\dot{\phi}_r + (\nabla b(\phi_r))^T b(\phi_r) - \lambda\frac{d\lambda}{d\varphi}\dot{\phi}_r, \quad (3.2.3)$$

where $\lambda = \|b(\phi_r)\| \,/\, \left\|\dot{\phi}_r\right\|$ and $\frac{d}{d\varphi}$ denotes functional derivative. GMAM will solve a discretized version of (3.2.3) with $L = 1$, subject to the constraints $\left\|\dot{\phi}_r\right\| = \textit{constant}$, $\phi_0 = \mathcal{O}$ and $\phi_1 = x$. The constant speed condition will ensure that the sampled points along the path ϕ are equi-spaced to avoid the issue of under-sampling important regions discussed in the previous section.

GMAM solves for a discretized version $\{\Phi_k\}_{k=0}^m$ of the MAP by following an implicit gradient descent scheme. The scheme begins with an initial guess, usually taken to be an uniformly-spaced sampling of the straight line path from $\mathcal{O}$ to x. The next iteration of $\{\Phi_k\}_{k=0}^m$ is computed via a gradient descent step using a semi-implicit discretization of (3.2.3). The exact discretization of (3.2.3), which we omit to preserve simplicity, contains *implicit* centered finite differences for the second derivatives $\ddot{\phi}_r$ and *explicit* centered finite differences for the first derivatives $\dot{\phi}_r$. Each update step thus requires the solving of a tri-diagonal matrix system. After every update step, the solution is re-normalized such that points Φ_k are uniformly spaced apart. More precisely, by interpolating the discretized values of $\{\Phi_k\}_{k=0}^m$ into a continuous path, one can set $\Phi_k := \Psi_k$ where the $\{\Psi_k\}_{k=0}^m$ lie along the interpolated path and satisfy $\|\Psi_{k+1} - \Psi_k\| = \|\Psi_k - \Psi_{k-1}\|$ for each $k = 1, ..., m-1$.

3.3 Mesh-based label-setting algorithms

In this section, we describe the structure of Dijkstra-like solvers of anisotropic eikonal equations. We also provide a description of Mirebeau's design for the anisotropic (ASR) algorithm [47] as well as the Ordered Line Integral Methods (OLIM), the most recent mesh-based quasi-potential solvers [21].

3.3.1 Dijkstra-like eikonal solvers

Consider the geometric action (2.3.15) expressed in the line integral form

$$\tilde{S}_L(\varphi) = \int_0^L s(\varphi_r, \dot{\varphi}_r)\,\|\dot{\varphi}_r\|\,dr = \int_0^L \frac{\|\dot{\varphi}_r\|}{f(\varphi_r, \dot{\varphi}_r)}dr, \tag{3.3.1}$$

where $s(x, v) = \|b(x)\| - b(x) \cdot \dfrac{v}{\|v\|} =: \dfrac{1}{f(x, v)}$. Written in this way, $\tilde{S}_L(\varphi)$ gives the cost of trajectory φ using the instantaneous anisotropic cost function $s(x, v)$. If we are to proceed with the wavefront analogy introduced in Section 3.1, we can interpret $f(x, v)$ as the instantaneous speed of particles moving in the v direction at point x. At each point along the wavefront, one can envision more particles being spawned and sent out in all directions with speed $f(x, v)$. For this reason, we call the function

$f(x,v)$ the *speed* function and $s(x,v)$ the *slowness* function. The quasi-potential $U(x)$ will then measure the quickest time for one of these particle to hit the point x, while the MAP passing through x will trace that particles path back to its origin at $\mathcal{O}$.

An alternative perspective can be taken by expressing the HJB equation (3.1.2) for the quasi-potential in the form

$$F(x,\hat{n})\,\|\nabla U(x)\| = 1, \qquad U(\mathcal{O}) = 0, \tag{3.3.2}$$

where $F(x,\hat{n}) = \dfrac{1}{-2b(x)\cdot\hat{n}(x)}$ and $\hat{n}(x) = \dfrac{\nabla U(x)}{\|\nabla U(x)\|}$ is the outward pointing normal vector to the level set of the quasi-potential. Here, the quantity $F(x,\hat{n}) \geq 0$ can be interpreted as the speed at which the wavefront expands outward in the normal direction.

For a general F, equation (3.3.2) is referred to as an anisotropic eikonal equation or static Hamilton-Jacobi equation. When the front speed F does not depend on direction v, we refer to it just as an eikonal equation. In general, the front speed F can be reconstructed from the particle speed function f by taking the weighted dual norm

$$F(x,u) = \max_{v\neq 0} \frac{u\cdot v}{\|v\|} f(x,v). \tag{3.3.3}$$

Moreover, equation (3.3.2) can be derived from the least cost problem with action (3.3.1) for a general f in a manner similar to as done in Section 2.3.4.

Note that in the isotropic case, it is clear from (3.3.3) that $F(x) = f(x)$. Namely, particle speeds are equal in all directions. In particular, this implies that the MAPs will always be normal to the wavefront. This is not the case for the quasi-potential problem, where instead the direction of travel is given by (3.1.3). As we will see, this lack of orthogonality will pose some computational difficulties.

We now introduce the main ideas of Sethian's fast marching algorithm for computing the viscosity solution $U(x)$ to equation (3.3.2) in 2-dimensions. Let $\mathcal{X}$ be a discretization of the domain D into a uniform rectangular mesh with common horizontal and vertical spacing $h > 0$. We assume for simplicity that the attractor $\mathcal{O}$ is a member of the mesh $\mathcal{X}$. Dijkstra-like algorithms rely on the partition of $\mathcal{X}$ into the following three disjoint groups:

- Unknown: *Mesh points for which no value of U has been computed. The value of U defaults to $+\infty$.*
- Considered: *Mesh points for which a tentative value of U has been computed.*
- Accepted: *Mesh points for which a final, immutable value of U has been computed.*

Initially, all mesh points will begin in the Unknown category. As the algorithm proceeds, points will be gradually moved from Unknown to Considered as tentative values of U are computed. For each mesh point, one of those tentative values will eventually be finalized, at which time that point will be moved into the Accepted category. Figure 3.2 displays a snapshot in time of what this setup may look like, while Algorithm 1 provides a template for fast marching Dijkstra-like eikonal solvers.

Algorithm 1 Fast marching method template for solving eikonal equations

Initialization

Start with all mesh points in $\mathcal{X}$ in Unknown and set them to $U = +\infty$.

Set $U(\mathcal{O}) = 0$ and add $\mathcal{O}$ to Considered.

Main Body

```
while (Considered is non-empty) do
    Set x := argmin_{z∈X} { U(z) :  z ∈ Considered }.
    Switch x from Considered to Accepted and finalize its current value of U(x).
    for each neighbor y of x such that y ∉ Accepted do
        Compute a value U_new(y) using the values of U(x) and possibly U(z) for
  other z ∈ Accepted.
        Set U(y) := min(U(y), U_new(y)).
        Switch y to Considered if it was previously in Unknown.
```

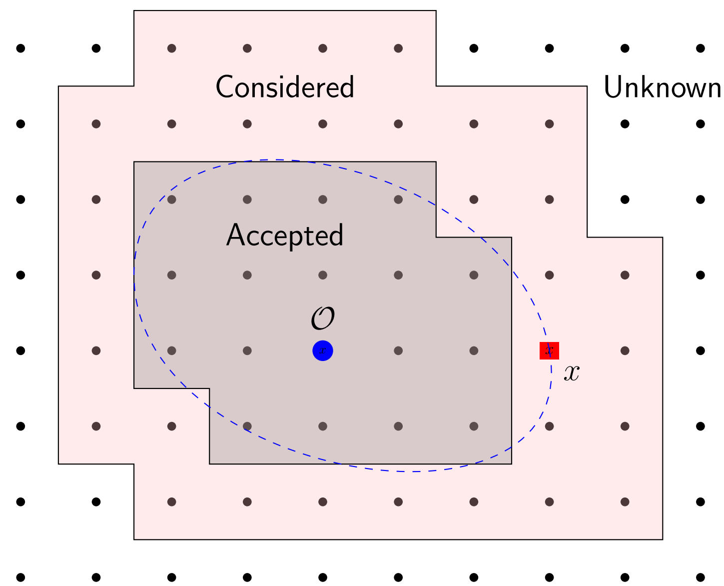


Fig. 3.2: Algorithm snapshot as mesh point x is added to the Accepted list. The dashed blue line represents a possible implied current state of the wavefront.

The procedure described in Algorithm 1 can be visualized quite handily in terms of the propagating wavefront analogy, in which the objective is to determine the first arrival time $U(x)$ of the front at $x \in \mathcal{X}$. A mesh point $x \in \mathcal{X}$ begins in the Unknown category when the wavefront is far away from x. Once the front gets sufficiently near, an observer at x can construct an estimated time of arrival (eta) of the wavefront at x, based on the arrival times reported by observers at nearby Accepted mesh points. Once x's first eta is calculated, the mesh point is switched from Unknown to Considered. Then, once the front actually hits x, it is switched to the Accepted category, and one of the previously computed eta's becomes the locked-in value of $U(x)$. This now finalized value of $U(x)$ is then subsequently used in the eta computations of other Considered and Unknown neighbors.

For the purposes of this analogy, we assume the observers do not have clocks or other means of actually telling the passage of time. Instead, they have to estimate the wave arrival time by using only the reported values of U for Accepted points and their knowledge of the velocity field f. In order to make such an estimate, the observer at x must determine (a) where the particle that will hit x first is currently located, (b) the remaining path this particle will take before it gets to x, and (c) the time it will take to traverse this path. Indeed, these are also three of the main decisions that a numerical algorithm must make, and largely will be the key factors in determining the accuracy of the method.

Let us remark on some of the key steps of Algorithm 1. First, in line 2, the mesh point x with the smallest tentative value of U among all Considered points is selected. This point is effectively the next point to be "hit" by the expanding wavefront, and one can envision the level set $\{r \in \mathbb{R}^d : U(r) = U(x)\}$ as actually representing the current state of the front, as illustrated in Figure 3.2. Since the minimizer will need to be extracted from the Considered list at each iteration, the Considered list is often given a heap-sort structure in practice so that the argmin can be computed in $\mathcal{O}(\log N)$ operations.

Next, we note that the definition of *neighbors* (line 4) and the process for computing estimates $U_{\text{new}}(y)$ (line 5) are left unspecified in Algorithm 1. These are the main areas in which Dijkstra-like eikonal solvers will differ, and we describe below some of the possible choices for these procedures in full detail.

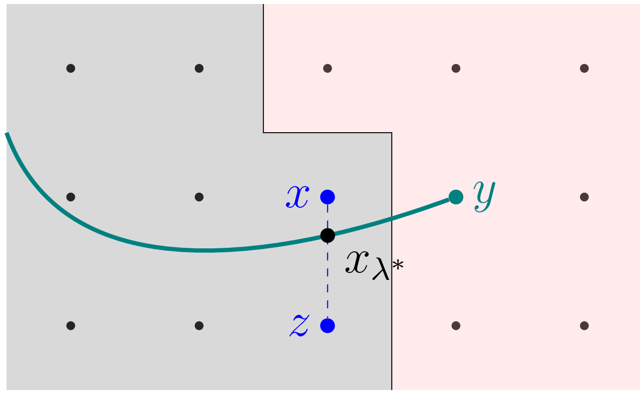


Fig. 3.3: The MAP of Considered or Unknown point y passes between nearby Accepted points x and z.

Finally, we note that after computing a new tentative value $U_{\text{new}}(y)$, this value replaces the previous tentative value only if it is smaller. This particular choice of Accept/Reject rule tries to filter out those estimates of $U(y)$ that come from mesh points x which are not necessarily near the MAP passing through y.

Computation of $\mathbf{U}_{\text{new}}$: We discuss techniques for computing the update value $U_{\text{new}}(y)$ in line 5 of Algorithm 1. Suppose, as in Algorithm 1 that x has just been switched to Accepted and we are now interested in using $U(x)$ to compute a tentative value of $U(y)$ for a nearby "neighbor" y. Appropriate definitions of "neighbor" are made precise under the next subheading.

To compute $U(y)$, we seek to identify the MAP that passes through y (Figure 3.3). In general, this MAP will not pass directly through x; however, it may pass between x and another neighboring mesh point z that also lies in the Accepted group. In this case, the MAP will pass through the point $x_{\lambda^*} = (1-\lambda^*)x + \lambda^* z$ for some

$\lambda^* \in [0,1]$. By noting that $U(y)$ is the minimum cost (geometric action in the quasi-potential case) over all paths terminating at y, we can write

$$U(y) = \min_{\lambda\in[0,1]} \Big[U(x_\lambda) \tag{3.3.4}$$

$$+ \inf_{L>0,\varphi\in C([0,L];\mathbb{R}^d)} \{\int_0^L \frac{\|\dot\varphi(r)\|\,dr}{f(\varphi(r),\dot\varphi(r))} : \varphi(0) = x_{\lambda^*}, \varphi(L) = y\}\Big]. \tag{3.3.5}$$

The key decisions to be made concern how best to approximate the right hand side of equation (3.3.4). This requires determining (a) how to interpolate $U(x_\lambda)$, since x_λ in general lies between mesh points, (b) what path-space to take the inner minimization over, and (c) what quadrature rule to use to approximate the action integral.

The simplest answers to these three questions are to (a) linearly interpolate $U(x_\lambda)$ between the finalized values of $U(x)$ and $U(z)$, (b) use a linear path connecting x_λ and y, so that no inner minimization need occur at all, and (c) use right endpoint quadrature (at y) of the action integral. This route is taken, for instance, by the Ordered Upwind Method (OUM) of [59]. With these approximations, (3.3.4) becomes

$$U_{\text{new}}(y) = \min_{\lambda\in[0,1]} \Big[(1-\lambda)U(x) + \lambda U(z) + \frac{\|y-x_\lambda\|}{f(y, \frac{y-x_\lambda}{\|y-x_\lambda\|})}\Big]. \tag{3.3.6}$$

This univariate minimization can then easily be carried out numerically. For the isotropic case where $f(y,v) = f(y)$, the minimizer λ^* can be found analytically by solving a quadratic equation.

The ordered line integral methods (OLIMs) [21, 22] for solving for the quasi-potential also use linear interpolation for $U(x_\lambda)$ and linear paths, but they employ higher order quadrature rules instead of the right-hand rule. The most efficient choice turned out to be the midpoint rule, which reduced the error constants by two-to-three orders of magnitude compared to right-hand rule quadrature.

Update formula (3.3.6) in general leads to a first order $\mathcal{O}(h)$ error convergence rate. It is perfectly valid for the quasi-potential problem, however in EJM we will opt for higher order approximations of (3.3.4) in pursuit of a $\mathcal{O}(h^2)$ convergence rate.

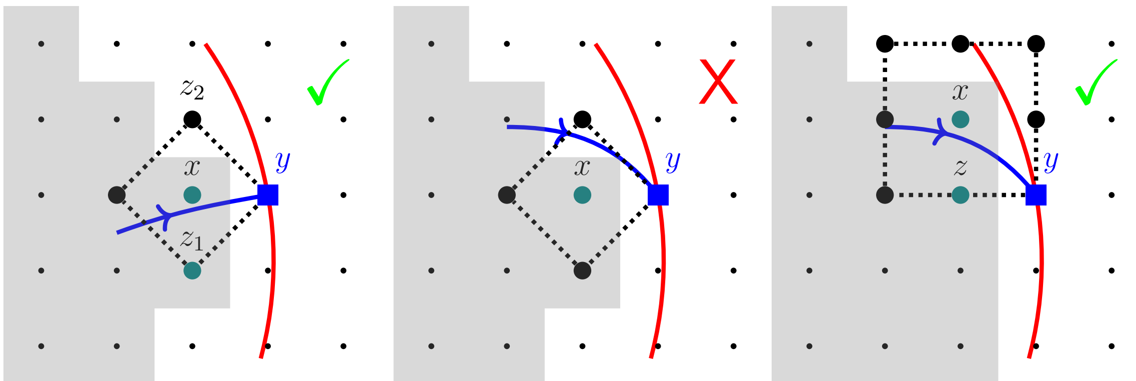
Neighborhoods The remaining undiscussed components of Algorithm 1 are (i) how to choose the neighbors y to update from x (line 4) and (ii) how to choose points z to pair with a given x and y for the computation of (3.3.6) (line 5).

The goal of *neighborhood* design is to assure that every mesh point y experiences at least 1 of the (3.3.4) updates with an x and z that straddle its MAP as in Figure 3.3. Since we do not a priori know where the MAPs will be coming from, many updates will necessarily have to be performed where x and z do *not* surround the MAP at all. These minimizations should result in boundary ($\lambda = 0$ and $\lambda = 1$) minimizers of equation (3.3.4). Moreover, these solutions will typically result in proposed values of $U_{\text{new}}(y)$ that are *larger* than the true value. Therefore, they should eventually be discarded by the Accept/Reject rule of line 6 if a proper interior solution is found. However, for this to happen, the neighborhoods must be large enough to include a valid $\triangle xzy$ triangle where the MAP passes between x and z. We refer to such a triangle $\triangle xzy$ of mesh points, where $U(x) < U(y)$, $U(z) < U(y)$, and y's MAP passes between x and z, as a *causal* triangle.

Due to this filtering system, it is in general harmless to increase the size of the neighborhoods, since any extraneous updates will automatically be filtered out by the Accept/Reject rule in the end. The drawback of using very large neighborhoods is simply the additional computation time present in performing the additional minimizations (3.3.4). The main goal of design is then to create neighborhoods that are just large enough to ensure that at least one causal triangle is checked for each y.

To discuss minimum neighborhood sizes, we assume first the isotropic case where $f(y, v) = f(y)$. As we have seen, the MAPs are guaranteed to be orthogonal to the level sets of U wherever U is differentiable, which is the case on a set of full Lebesgue measure. Provided the MAP and level set of U are sufficiently flat (which can be achieved for small enough h), a sufficient "neighborhood" of x in line 4 of Algorithm 1 is simply the 4-point diamond nearest neighborhood of x (Figure 3.4a). In such a case, one could conduct update (3.3.4) using two different values of z: the two neighboring members of y in the 4-point neighborhood of x (z_1 and z_2 in Figure 3.4a). It is easy to see geometrically that because of the orthogonality relation between MAP and level set, this procedure guarantees that a causal triangle is checked for every y. In fact, the such a triangle will also necessarily be a right triangle with hypotenuse $\sqrt{2}h$.

Unfortunately, in the anisotropic case, where the orthogonality relation between MAP and level set of U no longer holds, it is clear that the 4-point diamond neighborhood is no longer sufficient, as shown by Figure 3.4b. The neighborhoods must be larger in order to guarantee that update (3.3.4) is checked on a causal triangle. For



(a) Isotropic *(b)* Anisotropic - 4 point stencil *(c)* Anisotropic - 8 point stencil

Fig. 3.4: An illustration for what is a sufficiently large neighborhood. Red curves depict level sets of the solution, blue curves represent MAPs, and gray-shaded areas represent Accepted mesh points. (a) *Isotropic eikonal equation*: 4-point neighborhoods always guarantee a causal triangle since MAPs and level sets are orthogonal. (b,c) *Anisotropic eikonal equation*: Larger neighborhoods are needed. In this example, a 4-point neighborhood is insufficient, but an 8-point neighborhood is sufficient to guarantee a causal triangle.

instance, in Figure 3.4b, an 8-point square nearest neighborhood would be sufficiently large, as illustrated by Figure 3.4c.

It is intuitively clear that the necessary size of neighborhood will increase as the angle between MAP and level set of U decreases. Since this angle is related to the particle velociy speed anisotropy, a minimum possible neighborhood size can be determined by the anisotropy ratio

$$\rho := \sup_x \frac{\sup_v f(x,v)}{\inf_v f(x,v)}.$$

For full descriptions of how neighborhood size can be determined from the anisotropy ratio, see [59] and [21].

3.3.2 Anisotropic stencil refinement

In this section we describe the Anisotropic Stencil Refinement (ASR) algorithm of Mirebeau [47] for solving anisotropic eikonal equations. The ASR algorithm follows the prescription of Algorithm 1 and uses the approximation scheme (3.3.6), but with pre-computed stencils serving as the "neighborhoods". These stencils are significantly smaller than the neighborhoods used in [59], so that the ASR algorithm is significantly faster than OUM and only slightly less accurate.

Suppose we wish to solve the minimum cost problem associate with an action of the form (3.3.1) where f has a finite anisotropy ratio ρ. For each mesh point y, we first construct the *stencil* $\mathcal{N}(y)$, representing the candidates for x and z in a potentially causal triangle $\triangle xyz$. The neighborhood referred to in line 4 of Algorithm 1 will then be the *reversed stencil*

$$\mathcal{N}^{-1}(x) := \{y \in \mathcal{X} : x \in \mathcal{N}(y)\},$$

which represents the points y of which x might be a member of a causal triangle.

Next, we define the function

$$\mathcal{F}(x,v) := s(x,v)\,\|v\|\,,$$

representing the integrand in the action functional (3.3.1). We assume that for each x, the mapping $\mathcal{F}(x, \cdot) : \mathbb{R}^2 \to [0, +\infty)$ is an *asymmetric* norm, that is, it is subadditive and positive definite, but only satisfies positive homogeneity: $\mathcal{F}(x, \lambda v) = \lambda \mathcal{F}(x, v)$ for all $\lambda \geq 0$.

Definition 3.3.1. The vectors u and v are said to form an acute angle with respect to an asymmetric norm $G : \mathbb{R}^2 \to [0, +\infty)$ (or G-acute angle, for short) provided that

$$G(u + \delta v) \geq G(u), \quad \text{and} \quad G(v + \delta u) \geq G(v), \tag{3.3.7}$$

for all $\delta \geq 0$.

Remark 3.3.1. We note that if the asymmetric norm G is differentiable, it is easy to see that acuteness conditions (3.3.7) are equivalent to

$$u \cdot \nabla_v G(v) \geq 0, \quad \text{and} \quad v \cdot \nabla_v G(u) \geq 0.$$

□

The objective will then be to construct a stencil $\mathcal{N}(y)$ for any y that is a collection of directions such that neighboring line segments terminating at y form $\mathcal{F}(y, \cdot)$-acute angles. That is, for each mesh point y, we seek to construct a finite (and rotationally ordered) collection of mesh points $\mathcal{N}(y) = \{y_k\}_{k=1}^{n_y}$, such that $y - y_k$ and $y - y_{k+1}$ form a $\mathcal{F}(y, \cdot)$-acute angle for each $k = 1, ..., n_y$.

Note that when $s(x, v) = s(x)$, the norm $\mathcal{F}(x, \cdot)$ is a multiple of the standard Euclidean norm so that $\mathcal{F}(x, \cdot)$-acute angles become ordinary Euclidean acute angles.

Thus, the standard 4-point diamond neighborhood discussed in the previous section is an admissible stencil, since neighboring directions are separated by right angles.

The guarantee that such a stencil creates causal triangles is provided by the following causality property.

Proposition 3.3.1 (Proposition 1.3 of [47]: Causality Property)**.** *Let G be an asymmetric norm on $\mathbb{R}^2$. Let $u, v \in \mathbb{R}^2$ be linearly independent and let $d_u, d_v \in \mathbb{R}$. Assume that u and v form an G-acute angle. Define*

$$d_w := \min_{t\in[0,1]} td_u + (1-t)d_v + G(tu + (1-t)v), \tag{3.3.8}$$

and assume that this minimum is not attained for $t \in \{0, 1\}$. Then $d_u < d_w$ and $d_v < d_w$.

We do not provide the proof here, but rather defer it to [47]. It can be done rather succinctly using Lagrange multipliers. We do, however, provide a geometric proof in Section 3.4 of a similar statement in the quasi-potential case that sheds more light on why this acuteness condition is relevant.

Let us remark on Proposition 3.3.1. The right hand side of (3.3.8) is precisely the right hand side of the update formula (3.3.6) for $U_{\text{new}}(y)$, if $u = x$, $v = z$, $d_u = U(x)$ and $d_v = U(z)$. Thus Proposition 3.3.1 can be interpreted as saying that if the xy and zy legs of the xyz triangle form a $\mathcal{F}(y, \cdot)$-acute angle and an interior minima is found, then necessarily $U(x) < U_{\text{new}}(y)$ and $U(z) < U_{\text{new}}(y)$. This means, assuming $U(x) > U(z)$ without loss of generality, that when x is switched to Accepted, z would already be an Accepted point so that the triangle $\triangle xyz$ would yield a valid interior update, as desired.

The final step is then to define a method for creating causal stencils for a given problem. In [47], Mirebeau defines stencils via Algorithm 2, in which the base stencil is a 4-point diamond that will be refined as necessary, until all neighboring angle are $\mathcal{F}(y,\cdot)$-acute. For completeness, we also provide the full description of the FM-ASR in Algorithm 3.

Algorithm 2 FM-ASR Stencil Design [47]

for $y \in \mathcal{X}$ **do**

Set $L := [(1,0)]$ and $M := [(1,0),(0,-1),(-1,0),(0,1)]$.

while M is non-empty. **do**

Set u and v to the last elements of L and M, respectively.

if u and v are $\mathcal{F}(y,\cdot)$-acute **then**

Remove v from M and append v to L.

else

Append $u+v$ to M.

Set $\mathcal{N}(y) := y - L$.

Output $\{\mathcal{N}(y)\}_{y\in\mathcal{X}}$.

Algorithm 3 Fast March Anisotropis Stencil Refinement (FM-ASR) [47]

Initialization and Pre-processing

Start with all mesh points in $\mathcal{X}$ in $\mathsf{Unknown}$ and set them to $U = +\infty$.

Set $U(\mathcal{O}) = 0$ and add $\mathcal{O}$ to $\mathsf{Considered}$.

Compute the stencils $\{\mathcal{N}(y)\}_{y\in\mathcal{X}}$ via Algorithm 2.

Compute the reversed stencil $\{\mathcal{N}^{-1}(x)\}_{x\in\mathcal{X}}$ by inverting the stencils $\{\mathcal{N}(y)\}_{y\in\mathcal{X}}$.

Main Body

while ($\mathsf{Considered}$ is non-empty) **do**

  Set $x := \arg\min_{z\in\mathcal{X}} \left\{U(z) : \ z \in \mathsf{Considered}\right\}$.

  Switch x from $\mathsf{Considered}$ to $\mathsf{Accepted}$ and lock in its current value of $U(x)$.

  for each $y \in \mathcal{N}^{-1}(x)$ such that $y \notin \mathsf{Accepted}$ **do**

    Letting $\mathcal{N}(y) = \{y_i\}_{i=1}^{n_y}$ and supposing $x = y_k$, define $z_1 := y_{k-1}$ and $z_2 := y_{k+1}$.

    Set $U_{\text{new}} = \min(U_{\text{new},1}, U_{\text{new},2})$, where $U_{\text{new},i}$ is computed via (3.3.6) over the triangle $\triangle x z_i y$.

    Set $U(y) := \min(U(y), U_{\text{new}}(y))$.

    Switch y to $\mathsf{Considered}$ if it was previously in $\mathsf{Unknown}$.

3.4 Our algorithm: Efficient Jet Marcher

Like the previously discussed solvers, our Efficient Jet Marcher (EJM) algorithm follows the general template of Algorithm 1. Unlike the previously discussed solvers, EJM will instead use a higher order MAP approximation scheme as compared to (3.3.6), as well as a modified version of Mirebeau's pre-computed stencils. The full structure of the solver is shown in Algorithm 4.

We remark briefly on key observations about Algorithm 4; detailed descriptions are provided in the subsequent sections. The first observation is that the gradient ∇U, is now treated as part of the solution and is included in all of the update computations. This will be necessary to perform the higher order MAP interpolations in the update step of line 7 (Section 3.4.2).

As in the OLIM methods [21, 22], we apply a slightly a more involved initialization process. This is typically only needed when the MAPs exhibit significant curvature near the origin. The most common technique is to simply initialize U and ∇U with the exact solution corresponding to a linearized version of b around the attractor $\mathcal{O}$. We discuss initializations in more detail in Section 3.4.5.

Concerning the construction of anisotropic stencils, we follow the ideas of Mirebeau [47], discussed in Section 3.3.2. However, we also use the fact that stencils only depend on the angle of $b(x)$ and not its magnitude, to save the stencils only for a binned collection of possible angles of b. This drastically reduces the memory requirement and the runtime of pre-processing phase. Moreover, this allows us to skip the stencils entirely and compute the reversed stencils directly, since these are ultimately the objects used in the body of Algorithm 4. We also use a slightly different stencil construction algorithm. All of which is described in Section 3.4.1.

The if-statement and subsequent fail-safe method, mentioned in lines 3 and 4, are used to prevent values of $U(x)$ computed from one-point updates (boundary solutions to the update minimization problem (3.3.4)) from becoming finalized values. As we will see, due to practical reasons the stencils are not perfect and will occasionally fail to find causal triangles. If these failures are not caught and corrected, the higher order accuracy may not be achieved. This fail-safe is called only very rarely. When called, it searches a much larger area than the stencil until it finds a successful triangle update (interior solution of (3.3.4)).

Finally, the most important difference lies in the structure of the update procedure (line 7) and prescription for computing the right hand side of (3.3.4). This procedures is responsible for the $\mathcal{O}(h^2)$ error convergence rate. It is discussed in detail in Section 3.4.2.

3.4.1 Anisotropic stencils

We return to the anisotropic stencil ideas introduced in Section 3.3.2. For the quasi-potential problem, we have

$$\mathcal{F}(y, v) = \|b(x)\| \, \|v\| - b(x) \cdot v.$$

Algorithm 4 Efficient Jet Marching Algorithm (EJM)

Initialization and Pre-processing

Start with all mesh points in $\mathcal{X}$ in $\mathsf{Unknown}$ and set them to $U = +\infty$ and $\nabla U = (\infty, \infty)$.

Initialize the U and ∇U values of an 8-point rectangular neighborhood of the attractor $\mathcal{O}$ (Section 3.4.5).

Switch each of these points into $\mathsf{Considered}$.

For each $\theta_k = \frac{2\pi k}{N_{\text{bins}}}$, $k = 0, ..., N_{\text{bins}} - 1$, construct the reversed stencils $\mathcal{N}^{-1}(\theta)$ (Section 3.4.1).

For each $x \in \mathcal{X} \sim \{\mathcal{O}\}$, assign $\mathcal{N}^{-1}(x) := \mathcal{N}^{-1}(\theta_k)$ where k is such

$$-\angle b \in \left(\theta_k - \frac{\pi}{N_{\text{bins}}}, \theta_k + \frac{\pi}{N_{\text{bins}}}\right].$$

Main Body

while ($\mathsf{Considered}$ is non-empty) **do**

  Set $x := \arg\min_{z \in \mathcal{X}} \{U(z) : \ z \in \mathsf{Considered}\}$.

  if x's last update is a one-point update **then**

    Run the fail-safe on x (Section 3.4.4).

  Switch x from $\mathsf{Considered}$ to $\mathsf{Accepted}$.

  for each $y \in \mathcal{N}^{-1}(x)$ such that $y \notin \mathsf{Accepted}$ **do**

    Update $U(y)$ and $\nabla U(y)$ from x and possibly other $z \in \mathsf{Accepted}$ (Section 3.4.2).

    Switch y to $\mathsf{Considered}$ if it was previously in $\mathsf{Unknown}$.

This is subadditive and positive homogeneous, but not positive definite, since

$$\mathcal{F}(y, \lambda b(x)) = 0,$$

for any y and any $\lambda > 0$. Nonetheless, the machinery of the ASR method can still be applied, for instance, by considering instead a modified function

$$\mathcal{F}^{\alpha}(y, v) = \|b(y)\| \, \|v\| - \alpha b(y) \cdot v,$$

where $\alpha = 1 - \delta$ for some $\delta > 0$ very small. Such an $\mathcal{F}^{\alpha}(y, \cdot)$ is an asymmetric norm for any x such that $b(x) \neq 0$.

For each $y \in \mathcal{X}$, we seek to construct a stencil $\mathcal{N}(y)$ such that neighboring points $y_k, y_{k+1} \in \mathcal{N}(y)$ form line segments $u = \frac{y - y_k}{\|y - y_k\|}$ and $v = \frac{y - y_{k+1}}{\|y - y_{k+1}\|}$ that satisfy the acuteness conditions (3.3.7), which, in the quasi-potential case, reduces to the conditions

$$u \cdot v \geq u \cdot (-b(y)), \quad \text{and} \quad u \cdot v \geq v \cdot (-b(y)).$$

It is not intuitively clear where these acuteness conditions come from. To motivate them, we provide a simple geometric proof of the following proposition, which is essentially a corollary of Proposition 3.3.1 for the specific case of the quasi-potential problem.

Proposition 3.4.1. *Suppose the drift field b is smooth, fix $y \in \mathbb{R}^2$ and let ϕ denote the MAP passing through y. Suppose that D is a rotationally ordered collection of unit directions in $\mathbb{R}^2$ such that any neighbors $\hat{u}, \hat{v} \in D$ satisfy*

$$\hat{u} \cdot \hat{v} \geq \max\Big(- \hat{u} \cdot \hat{b}, -\hat{v} \cdot \hat{b}, 0 \Big), \tag{3.4.1}$$

where $\hat{b} = b(y)/|b(y)|$. Then there exists neighbors $\hat{u}, \hat{v} \in D$ and $h > 0$ small enough that ϕ passes between $A := y + h\hat{u}$ and $B := y + h\hat{v}$ while $U(A) < U(z)$ and $U(B) < U(z)$.

Proof. Suppose first that U is continuously differentiable at y. Let $\hat{\tau}$ be the tangent vector of the level set $\{y : U(y) = U(z)\}$ at z. Picking scale $h > 0$ such that U and ϕ are locally flat, the setup looks like Figure 3.5. By taking the dot product of both sides of equation (3.1.3) with $\hat{\tau}$, we immediately have that

$$\left(\frac{\phi'}{\|\phi'\|}\right) \cdot \hat{\tau} = \hat{b} \cdot \hat{\tau} =: \cos(\theta).$$

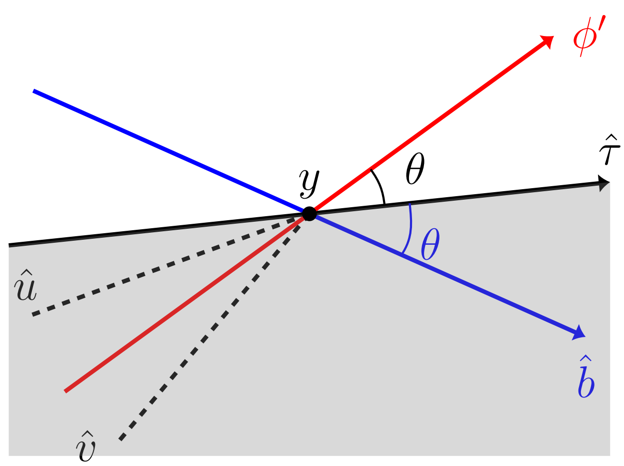


Fig. 3.5: If (3.4.1) is satisfied, there must exist neighbors $\hat{u}$, $\hat{v} \in D$ that sandwich the incoming MAP φ and both lie in the shaded region $\{x : U(x) < U(z)\}$.

The result then follows by noting that there *must* exist a direction $\hat{u} \in D$ lying between $-\phi'$ and $-\hat{\tau}$, as in Figure 3.5. We can see this by realizing that there *cannot* exist neighbors $\hat{u}, \hat{v} \in D$ that straddle the wedge created by $-\hat{\tau}$ and $-\phi'$. If there did exist such neighbors, $\hat{u}$ between $-\hat{b}$ and $-\hat{\tau}$ and $\hat{v}$ between $\hat{b}$ and $-\phi'$, then we would immediately contradict the acuteness requirement, since

$$-\hat{b} \cdot \hat{u} > -\hat{b} \cdot -\hat{\tau} = \cos\theta = -\Big(\frac{\phi'}{\|\phi'\|}\Big) \cdot -\hat{\tau} > \hat{u} \cdot \hat{v}.$$

Hence nearest neighbors cannot straddle the wedge, and there must be neighbors $\hat{u}, \hat{v} \in D$ surrounding ϕ' that both lie in the half-plane of directions opposite ∇U. $\square$

We now seek to construct stencils such that neighboring directions satisfy acuteness condition (3.4.1). To do this, we make the following two approximations to reduce computation time and memory requirements.

(1) We take the reversed stencils to be

$$\mathcal{N}^{-1}(y) = -\mathcal{N}(y), \tag{3.4.2}$$

so that the *reversed* stencils are readily obtained by choosing a set of directions satisfying

$$\hat{u} \cdot \hat{v} \geq \max(\hat{u} \cdot \hat{b}, \hat{v} \cdot \hat{b}, 0).$$

and the original stencils $\mathcal{N}(y)$ need not be computed at all.

(2) We save reversed stencils only for N_{bin} different uniformly spaced values of the angle of $b(x)$.

Approximation (1) is taken to speed up the computation time of the preprocessing phase of Algorithm 4. The process of inverting the stencil can be time consuming for fine meshes, since its complexity is $\mathcal{O}(K(\mathcal{N})h^{-2})$ where $K(\mathcal{N})$ is the average stencil cardinality. Moreover, the approximation (3.4.2) is very close to exact on a scale $h > 0$ such that the MAPs are approximately flat.

Approximation (2) is taken to reduce memory requirements of the algorithm. Here, we take use of the fact that the metric $\mathcal{F}(x, \cdot)$ only depends on the angle of $b(x)$. Therefore, the reversed stencils can be saved for a common set of $\theta = \angle b$ values rather than one for every mesh point $x \in \mathcal{X}$. Angles are binned as mentioned in the pre-processing phase of Algorithm 4. We do note however, that this shortcut is only possible because we are using a mesh that is translation invariant, so that we need only to store the shifts in the reversed stencil. This would not be possible if instead we were using a non-uniform mesh.

To construct the reversed stencils, one option is to follow the prescription of Algorithm 2. However, since $\mathcal{F}(x, \cdot)$ is not positive definite, one should modify the process or else it will never terminate, as it will attempt to refine the reversed stencil in the direction of $-b(x)$ indefinitely. One solution, is to instead use the modified version of $\mathcal{F}^\alpha$, mentioned above, for some α very slightly below 1. The value of α will control how refined the reversed stencil is in the $-b(y)$ direction: our typical values is 0.9999. An alternative options is simply to cut off the production algorithm once the leg in the direction of $-b(y)$ reaches some maximum threshold value.

We opted to construct the reversed stencils via another route. Instead, we noted that for a given value of $\angle b$, a minimal set of directions satisfying relations (3.4.1) can

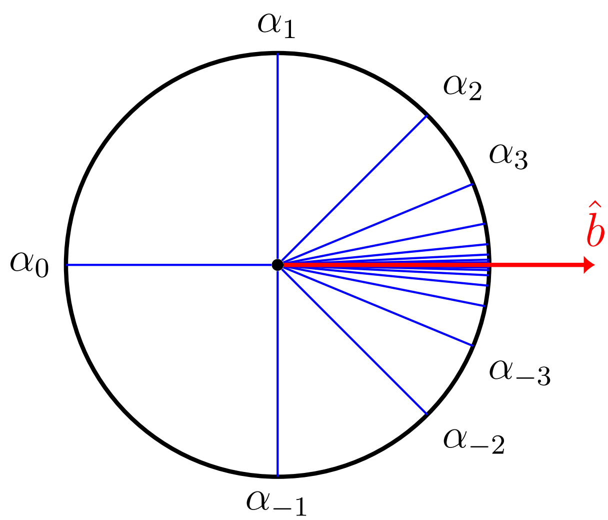


Fig. 3.6: The infinite collection of angles $\{\alpha_k\}_{k=0}^{\infty}$ with α_k given by (3.4.3) satisfy (3.3.1) with equality except right at the direction $\hat{b}$.

be easily written down. In fact, consider the infinite collection of angles $\{\alpha_k\}_{k=-\infty}^{\infty}$ where

$$\alpha_k = \angle\hat{b} + \operatorname{sgn}(k)\frac{\pi}{2^{|k|}}, \tag{3.4.3}$$

which is displayed in Figure 3.6. Here, it is clear that $\alpha_{k+1} - \alpha_k = \alpha_k - \angle b$ so that relation (3.4.1) holds with equality. Such a simple characterization of an "ideal" stencil is certainly not possible for the general case of an anisotropic eikonal equation. Thus, rather than following Algorithm 2 which is designed to work for the general case, we seek to construct stencils by directly discretizing rotated versions of the collection shown in Figure 3.6.

The construction of our discretization of Figure 3.6 is given in Algorithm 5, with the supporting Figure 3.7. The reversed stencil consists of a discretization of $|k| - 1$

equi-spaced points along each of the α_k rays in Figure 3.6, up until some maximum k value. Since these are all non-integer points, their nearest neighbor in $\mathbb{Z}^2$ is added instead. In addition to these points, the standard 8-point neighborhood is included.

Algorithm 5 is a rather simple method for discretizing Figure 3.6 in a way that is both non-hollow and elongated in the direction of $\hat{b}$. Example stencils created via the two methods are displayed in Figure 3.8. The stencil created via Algorithm 5 is slightly less concentrated in the direction of anisotropy and has denser interior: properties that have shown to perform slightly better in practice.

Algorithm 5 Construction of dense oblong reversed stencils

Let θ be the binned value of $-\angle b$.

Set k_{cutoff} = largest k to include in discretization of α_k (Figure 3.6). Value $k_{\text{cutoff}} = 4$ is typical.

Set d_{gap} = spacing between sampled points on each ray (in multiples of h). Value $d_{\text{gap}} = 3$ is typical.

1: Add $\{(1,0),(1,1),(0,1),(-1,1),(-1,0),(-1,-1),(0,-1),(1,-1)\}$ to $\mathcal{N}^{-1}(\theta)$.
2: Let M be an empty list of points.
3: **for** $k = 2 : k_{\text{cutoff}}$ **do**
4:     **for** $n = 1 : |k| - 1$ **do**
5:         Add $n \cdot d_{\text{gap}} \cdot (\cos(\theta + \frac{\pi}{2^k}), \sin(\theta + \frac{\pi}{2^k})$ to M
6: **for** each point $(x, y) \in M$ **do**
7:     Add the closest integer pair $(round(x), round(y))$ to $\mathcal{N}^{-1}(\theta)$.

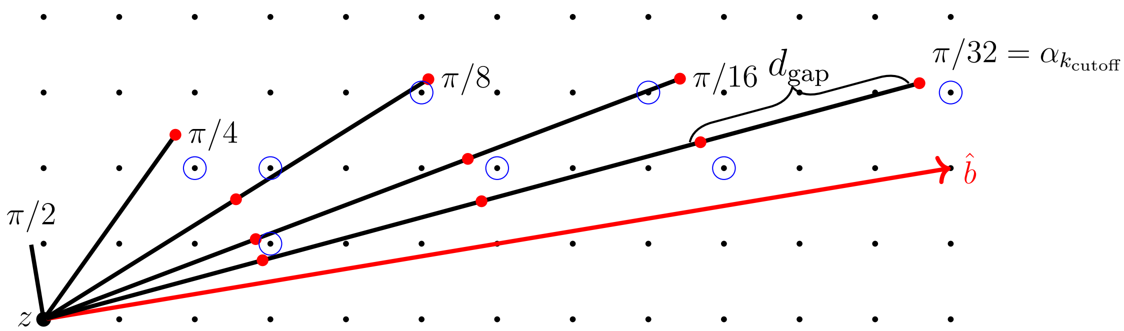


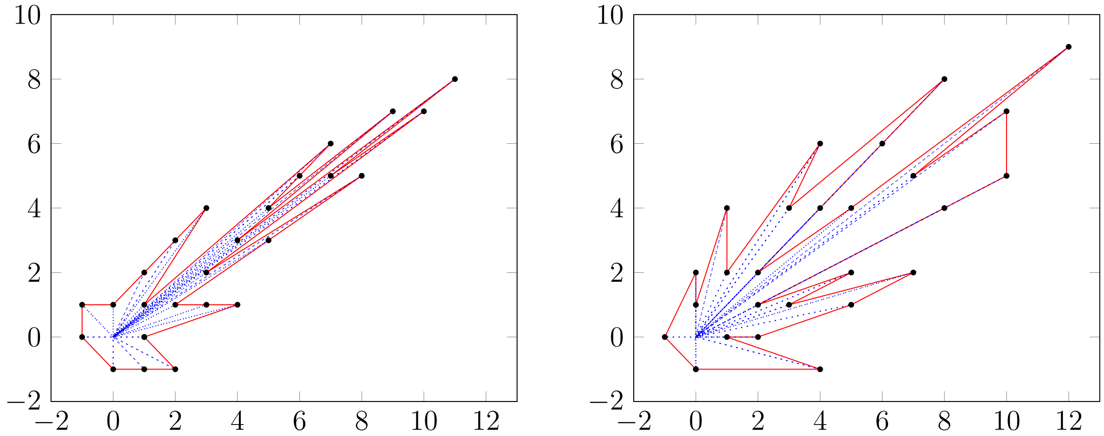
Fig. 3.7: The stencil consists of all of the blue circled mesh points, along with the analog procedure done on the other side of $\hat{b}$ and overlayed with a standard 8-point neighborhood of z.

3.4.2 Update procedure

In this section, we discuss how the update procedure (line 7 of Algorithm 4) is conducted. In EJM, this procedure is considerably more involved than the OUM where a quadratic equation is solved, and the OLIMs where only a 1D minimization problem is solved. We provide a full description of the update process in Algorithm 6 at the end of this section.

As mentioned in Section 3.4, we henceforth distinguish between the process of checking for interior and exterior solutions to (3.3.4), which we refer to as one-point updates and triangle updates, respectively. That is, one-point updates compute values of $U(y)$ and $\nabla U(y)$ from a single mesh point x, while triangle updates, when successful, compute values of $U(y)$ and $\nabla U(y)$ from two mesh points x and z.

One-point updates: As in Algorithm 4, we suppose that mesh point x has just been switched to Accepted, and the mesh point $y \in \mathcal{N}^{-1}(x)$ is to be updated from x.



(a) FM-ASR Reversed Stencil

(b) Dense Oblong Reversed Stencil

Fig. 3.8: Sample Reversed Stencils created using Algorithms 2 and 5, respectively. Both of these stencils have the same cardinality.

Let ϕ be the MAP that passes through y. We assume throughout the remainder that the drift field $b(x)$ is twice continuously differentiable. In view of equation (3.3.4), it follows that

$$U(y) \geq U(x) + \inf_{L>0, \varphi \in C([0,L];\mathbb{R}^d)} \{ \int_0^L \frac{\|\dot{\varphi}(r)\| \, dr}{f(\varphi(r), \dot{\varphi}(r))} : \varphi(0) = x, \varphi(L) = y \}, \quad (3.4.4)$$

with equality if and only if x lies on the MAP ϕ. The one-point update will be a numerical approximation of this right-hand side. The resulting proposed value of $U(y)$, will in general only be a good estimate of the true value if x lies very near ϕ. It will be a significant overestimate otherwise.

In approximating the right hand side of (3.4.4), we opt to conduct the minimization over the two-parameter family of paths $\tilde{\varphi}_{\alpha,\beta} : [0,1] \to \mathbb{R}^2$ consisting of cubic curves with fixed endpoints at x and y. This family is parametrized by the entry and exit angles α and β (Figure 3.9a), while each individual path $\tilde{\varphi}_{\alpha,\beta}$ is parametrized by its normalized coordinate in the $y - x$ direction. Complete formulas can be found in the Appendix.

For a given path $\tilde{\varphi}_{\alpha,\beta}$, we approximate the integral in (3.4.4) by using a Simpson's quadrature rule. We also adopt the notation $U^x_{\text{new}}(y)$ and $\nabla U^x_{\text{new}}(y)$ to denote the proposed values of $U(y)$ and $\nabla U(y)$, respectively, from the one-point update from mesh point x. Hence, we set

$$U^x_{\text{new}}(y) := U(x) + \min_{\alpha,\beta\in[0,2\pi]} \frac{\|y-x\|}{6}\Big[\frac{\|\tilde{\varphi}'_{\alpha,\beta}(0)\|}{f(x,\tilde{\varphi}'_{\alpha,\beta}(0))} \tag{3.4.5}$$

$$+ 4\frac{\|\tilde{\varphi}'_{\alpha,\beta}(1/2)\|}{f(\tilde{\varphi}_{\alpha,\beta}(1/2),\tilde{\varphi}'_{\alpha,\beta}(1/2))} + \frac{\|\tilde{\varphi}'_{\alpha,\beta}(1)\|}{f(y,\tilde{\varphi}'_{\alpha,\beta}(1))}\Big]. \tag{3.4.6}$$

We conduct this two-dimensional minimization over α and β by using Newton's method. The cumbersome derivatives are implemented by hand, but we note that this could potentially be improved by implementing automatic differentiation techniques.

In view of relation (3.1.3), the proposed one-point update value of the gradient U is

$$\nabla U^x_{\text{new}}(y) = \|b(y)\|\,\tilde{\varphi}'_{\alpha^*,\beta^*}(1) - b(y),$$

where α^* and β^* are the minimizing angles of (3.4.5).

We remark that if the numerical solver fails to find a minimizer of (3.4.5), then the values $\alpha = 0$ and $\beta = 0$ are used to compute $U^x_{\text{new}}(y)$ and $\nabla U^x_{\text{new}}(y)$. The proposed update value then represents an approximation of using a linear MAP and Simpson's rule quadrature. From our empirical observations, such a failure is extremely rare

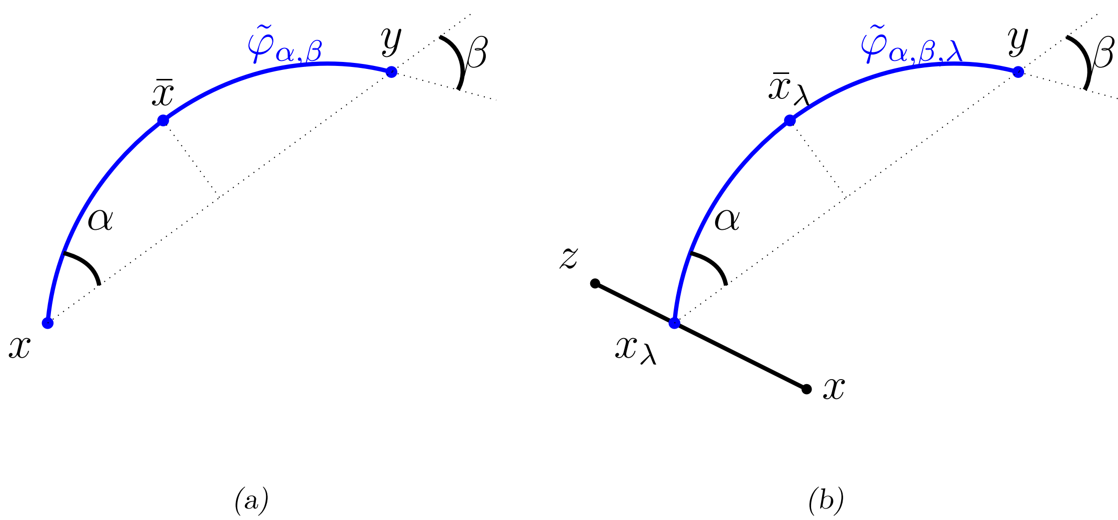


Fig. 3.9: Diagrams for the (a) One-point updates and (b) triangle updates. The MAP is approximated by minimizing the geometric action over the families $\tilde{\varphi}_{\alpha,\beta}$ and $\tilde{\varphi}_{\alpha,\beta,\lambda}$, respectively.

unless there is a significant amount of curvature of the MAP over the one-point update in question. In scenarios where the MAP is very flat, minimizing values of α and β are very close to 0 and the Newton solver has little difficulty finding them.

Triangle Updates: Conversely, suppose now that points x and z are Accepted and that we seek to update the values of $U(y)$ and $\nabla U(y)$ using x and z. The triangle update will search for interior (in λ) minimizers of (3.3.4) corresponding to a situation described by Figure 3.3.

As in the case of the one-point update, we approximate the MAP with a cubic polynomial terminating at y. However, here the starting point $x_\lambda = (1-\lambda)x + \lambda z$ is allowed to vary on along the xz line segment as part of the minimization.

In approximating the right hand side of (3.3.4), we opt to conduct the minimization over the three-parameter family of paths $\tilde{\varphi}_{\alpha,\beta,\lambda} : [0,1] \to \mathbb{R}^2$ consisting of cubic curves with fixed endpoints at x_λ and y (Figure 3.9b). As before, this family is parametrized by the entry and exit angles α and β, in addition to the value of $\lambda \in [0,1]$. Each path $\tilde{\varphi}_{\alpha,\beta,\lambda}$ is parametrized by its normalized coordinate in the $y - x_\lambda$ direction. In particular, the coordinate systems vary with the value of parameter λ.

Notice that in the one-point update, we did not use the value of $\nabla U(x)$ in order to perform the update. We use it here, however, to interpolate the value of $U(x_\lambda)$. In particular, we interpolate $U(x_\lambda)$ with the unique cubic Hermite polynomial $p : [0,1] \to \mathbb{R}$ satisfying boundary conditions

$$\begin{cases} p(0) = U(x), \qquad p'(0) = \nabla U(x) \cdot (z - x), \\ p(1) = U(z), \qquad p'(1) = \nabla U(z) \cdot (z - x). \end{cases}$$

Then, similar to the case of the one-point update, we take a Simpson's rule approximation of the integral in (3.3.4). As before, we adopt the notation $U^{x,z}_{\text{new}}(y)$ and $\nabla U^{x,z}_{\text{new}}(y)$ to denote the proposed values of $U(y)$ and $\nabla U(y)$, respectively, from the triangle update of y from x and z. Thus, we set

$$\begin{aligned} U^{x,z}_{\text{new}}(y) := \min_{\lambda\in[0,1],\alpha,\beta\in\mathbb{R}} \Big(p(\lambda) + \frac{\|y - x_\lambda\|}{6}\Big[\frac{\|\tilde{\varphi}'_{\alpha,\beta,\lambda}(0)\|}{f(x_\lambda, \tilde{\varphi}'_{\alpha,\beta,\lambda}(0))} \\ + 4\frac{\|\tilde{\varphi}'_{\alpha,\beta,\lambda}(1/2)\|}{f(\tilde{\varphi}_{\alpha,\beta,\lambda}(1/2), \tilde{\varphi}'_{\alpha,\beta,\lambda}(1/2))} + \frac{\|\tilde{\varphi}'_{\alpha,\beta,\lambda}(1)\|}{f(y, \tilde{\varphi}'_{\alpha,\beta,\lambda}(1))}\Big]\Big). \end{aligned} \tag{3.4.7}$$

As before, we perform the three-dimensional minimization over α, β and λ using Newton's method with exact derivatives. Similarly, the proposed update value of the gradient is

$$\nabla U^{x,z}_{\text{new}}(y) = \|b(y)\| \, \tilde{\varphi}'_{\alpha^*,\beta^*,\lambda^*}(1) - b(y),$$

where α^*, β^*, λ^* are the minimizing parameter values i of (3.4.7).

Unlike in the linear approximation minimization problems, (3.3.6), here we are only interested in interior solutions since exterior solutions are handled by the separate one-point updates. As such, we terminate the numerical solver once the working value of λ leaves the interval $[0, 1]$ and return *failure* of the triangle update. Since the triangle update should only succeed if the MAP passes through the xz line segment, the vast majority of attempted triangle updates should result in failure.

We now state the full update procedure (Algorithm 6). Justification for the Accept/Reject rule (lines 2 and 7, and Condition 3.4.1) and the choice of points z for the triangle updates (line 4) are given in the next section.

Condition 3.4.1. Let $U^{x,z}_{\text{new}}(y)$ and $\nabla U^{x,z}_{\text{new}}(y)$ be the values proposed from a successful triangle update with minimizing λ value λ^*. The proposed values are Accepted if and only if

(A) $U^{x,z}_{\text{new}}(y)$ is smaller than *all* previous proposed one-point update values $U^{\cdot}_{\text{new}}(y)$ of y.

(B) If the current tentative value of $U(y)$ came from a *triangle update* $\triangle x_{\text{old}} z_{\text{old}} y$, then we must have

$$\|x_{\lambda^*} - y\| < \left\| x_{\lambda^*_{\text{old}}} - y \right\|,$$

Algorithm 6 Update neighbors y of x (line 7 of Algorithm 4)

Let x be newly $\mathsf{Accepted}$ and $y \in \mathcal{N}^{-1}(x)$ be $\mathsf{Unknown}$ or $\mathsf{Considered}$.

Compute one-point update values $U^x_{\text{new}}(y)$ and $\nabla U^x_{\text{new}}(y)$ from (3.4.5).
if $U^x_{\text{new}}(y) < U(y)$ **then**
    Set $U(y) := U^x_{\text{new}}(y)$ and $\nabla U(y) := \nabla U^x_{\text{new}}(y)$.
for z in 8 point nearest neighborhood of x (Figure 3.10a) **do**
    if z is $\mathsf{Accepted}$ **then**
        Compute triangle update values $U^{x,z}_{\text{new}}(y)$ and $\nabla U^{x,z}_{\text{new}}(y)$ from (3.4.7).
        if triangle update is successful and Condition 3.4.1 is met. **then**
            Set $U(y) := U^{xz}_{\text{new}}(y)$ and $\nabla U(y) := \nabla U^{xz}_{\text{new}}(y)$.

where $x_{\lambda^*_{\text{old}}} = (1 - \lambda^*_{\text{old}})x_{\text{old}} + \lambda^*_{\text{old}} z_{\text{old}}$ and λ^*_{old} is the minimizing λ from the $x_{\text{old}} z_{\text{old}} y$ update.

3.4.3 Practical difficulties

There are several key algorithmic difficulties that arise in our higher order $\mathcal{O}(h^2)$ solver but are not present in the previous $\mathcal{O}(h)$ solvers. We describe these in detail in this section, in order to justify the decisions made in Algorithm 6 and the fail-safe method, discussed in the next section.

Large triangles. First, we note that with cubic interpolation, the error in the approximation of $U(x_\lambda)$ in the triangle updates grows more quickly with respect to the length of the xz line segment. If we suppose that $U(x)$, $U(z)$, $\nabla U(x)$ and $\nabla U(z)$

are equal to the correct values, then the interpolated polynomial $p(\lambda)$ should satisfy

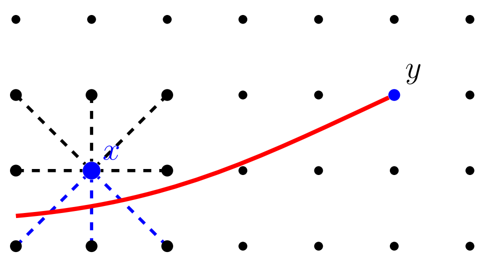
$$\sup_{\lambda \in [0,1]} |p(\lambda) - U_{\text{correct}}(x_\lambda)| \approx \mathcal{O}(|x - z|^4),$$

whereas a linear interpolation has worst case error that grows with $\mathcal{O}(|x - z|^2)$.

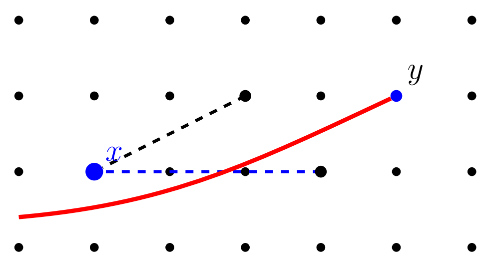
As such, it is critical to reduce the size of the xz leg of triangle updates as much as possible. To this purpose, we deviate from the procedure used in the ASR algorithm. Instead of using x and its two neighboring elements in the stencil $\mathcal{N}(y)$ (line 5 of Algorithm 3) as candidates for the point z, we opt to run a triangle update with x and *each* of its 8-point nearest neighbors as z. This limits the xz leg to a length of either h or $\sqrt{2}h$, compared to a potentially much larger xz length when neighboring elements of the stencil are used (Figure 3.10). The cost of this improvement in accuracy, however, is the requirement of between 2 and 3 times as many total triangle updates to be performed.

A more difficult but related problem to handle is the reduction of the size of the update length $|x_\lambda - y|$ in both the triangle updates and the one-point updates. Here, due to the Simpsons rule quadrature we expect the local error of a (3.4.7) update to scale with update length on the order of $\mathcal{O}(|x_\lambda - y|^5)$.

This issue unfortunately is much more difficult to handle. There are cases where it is unavoidable to have a large update length $|x_\lambda - y|$ when the angle between the MAP and level set of U is small (see Figure 3.12 for example). In fact, we see very clearly that in those situations the accuracy benefits of the EJM algorithm are significantly reduced (see Section 3.5).



(a) Triangle update is performed for each z in the 8-point neighborhood of x, as in Algorithm 6.



(b) Triangle update is performed with z being the two mesh points next to x in $\mathcal{N}(y)$, as in Algorithm 3.

Fig. 3.10: Possible choices of z for triangle updates. Dashed lines indicate bases used for triangle updates, while blue dashed lines indicate bases that should result in successful triangle updates.

Our solution to minimizing this triangle leg is the enforcement of Condition 3.4.1 (B). Instead of using the $U(y) := \min(U(y), U_{\text{new}}(y))$ Accept/Reject rule of Algorithm 1, we instead accept a new triangle update over a previous triangle update, only if the new update length $|x_\lambda - y|$ leg is smaller than the previous update length. We do not, however, require that the new triangle update propose a smaller value of $U(y)$ than the previous triangle update value because the value with a longer update length may be artificially too small.

In addition, to reduce the effect of the increased quadrature error of formulas (3.4.5) and (3.4.7) for large triangles, we increase the refinement of the Simpsons rule quadrature. Specifically, once the minimizing path $\tilde{\varphi}_{\alpha^*,\beta^*,\lambda^*}$ of (3.4.7) is found, the *actual* proposed update value $U^{x,z}_{\text{new}}(y)$ is computed with a Simpson's rule approximation of (3.3.4) along $\tilde{\varphi}_{\alpha^*,\beta^*,\lambda^*}$, but with a *larger* number of nodes sampled along $\tilde{\varphi}_{\alpha^*,\beta^*,\lambda^*}$. For the number of nodes, we typically use the smallest odd number greater than $1+|x_{\lambda^*}-y|/h$. Thus, for update lengths of h and $\sqrt{2}h$, the number of nodes used in the Simpsons rule evaluation is the usual 3, but for larger triangles, the number of nodes is proportional to $|x_{\lambda^*}-y|$. This refinement significantly reduce quadrature error when there is a large number of long-distance updates.

Missed triangle updates. Second, the vast reduction in the error from triangle updates (3.4.7), causes the error from one-point updates to be unacceptably large by comparison. The error discrepancy between the two types of updates is still present in earlier quasi-potential solvers, but is of far smaller magnitude. In fact, with the EJM method, a small handful of mesh points whose final updates are one-point updates will destroy the expected $\mathcal{O}(h^2)$ error convergence rate. It is therefore essential that the final $\mathsf{Accepted}$ value of $U(y)$ for each y comes from a triangle update rather than a one-point update. This is precisely the purpose of the fail-safe discussed in Section 3.4.4.

Undesirable local minima. Finally, there is a third more subtle problem introduced by adopting (B) of Condition 3.4.1 as part of the Accept/Reject rule: namely, that the right hand side of (3.4.7) may have local minima that have nothing to do with the desired MAP. An example of such a local minimum that does not correspond to a MAP is shown in Figure 3.11a. Under the traditional Accept/Reject rule $U(y) := \min(U(y), U_{\text{new}}(y))$, the presence of these local minima is not an issue

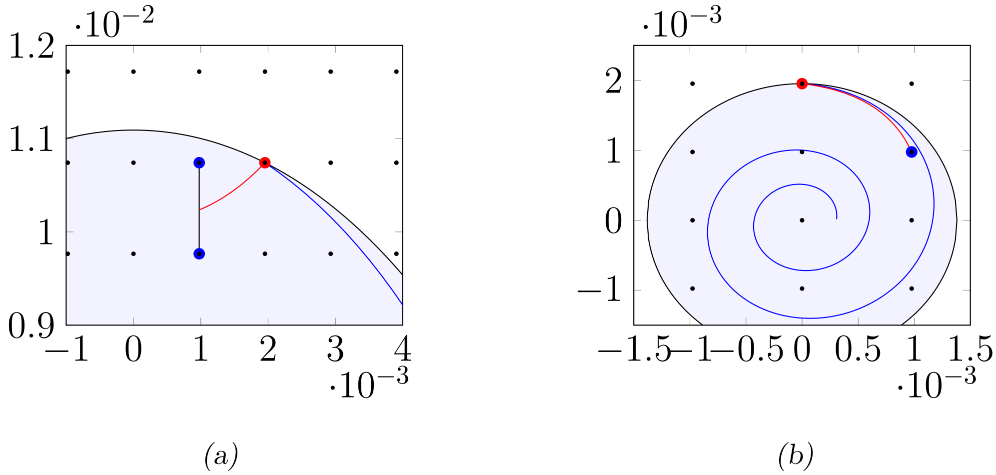


Fig. 3.11: Pitfalls: (a) A situation where a local minimizer (red) of (3.4.7) exists which does not correspond to the true MAP (blue) and (b) a situation where no successful triangle update occurs for a point y near the attractor $\mathcal{O}$ because the curvature of the MAP (blue) is too great .

since these "fake" paths necessarily correspond to *overestimates* of $U(y)$ and are hence filtered out. This is not the case when we just use Condition (B) as the accept rule, since we are filtering based on update length.

To mitigate this problem, we introduction (A) of Condition 3.4.1. With (A), a valid triangle update must propose a value of $U(y)$ smaller than any previously proposed one-point update values of $U(y)$. This filters out any "fake" paths that lead to answers worse than the *best* of the one-point updates. This is certainly not a guarantee that *all* of the other possible local minima will be prevented, but rather, a guarantee that the ones that do slip by, will provide update values at least as good as the best one-point update.

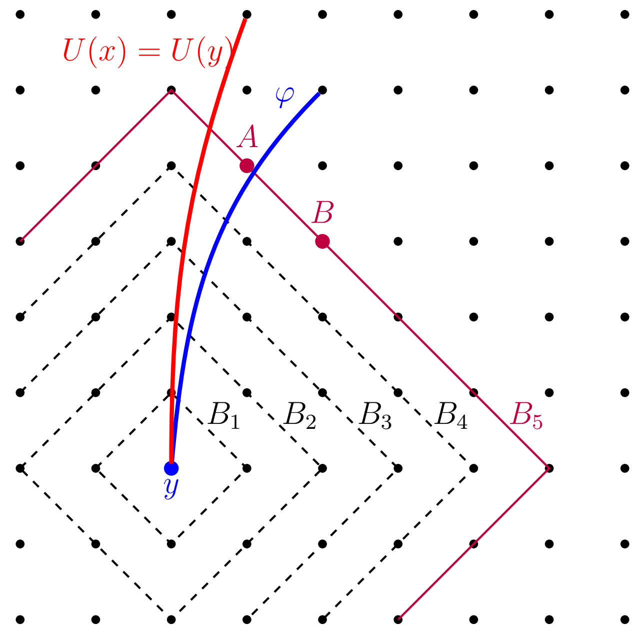


Fig. 3.12: The fail-safe will check B_k for each k until a successful two-point update is found. Here the first time a two-point update should be successful is on iteration $k = 5$ with $x = A$ and $z = B$.

3.4.4 Fail-safe

As mentioned above, it is critical that Accepted mesh points have a final value of U coming from a triangle update, rather than from a one-point update. To ensure this is the case we run the following fail-safe procedure on any mesh points whose final update is a one-point update.

Let x be the Considered mesh point with the smallest value of U in Considered as in Algorithm 4. Suppose that x's most recent update is a one-point update. As shown in Figure 3.12, let B_k be the set of mesh points a distance kh in ℓ^1 norm away from y. We simply search each B_k, starting from $k = 1$, for a pair of Accepted neighbors A

and B in B_k that provide a successful triangle update of x. If we find such a pair, we use the associated triangle update values of $U(x)$ and $\nabla U(x)$. If the procedure does not succeed through some large k threshold (we typically use $k = 20$), we revert to the prior one-point update value of $U(x)$.

It is important to note that since the value of $U(x)$ is changed (and typically lowered) *after* x was selected from the Considered list, we may lose the monotonic ordering of points in Accepted. That is, a mesh point may have its value of U lowered during the fail-safe procedure such that it is now smaller than some points already in the Accepted list. This would have some significant adverse consequences if the fail-safe is called too frequently. However, provided that the fail-safe is called only a small percentage of mesh points, these potential deviations from causality have a negligible impact.

We note also that the presence of the fail-safe allows some flexibility in the creation of the stencils. As long as the stencils are *reasonably* well refined in the direction of anisotropy, the fail-safe will correct for the few cases where the MAP might be able to sneak by the stencils.

3.4.5 Initialization

When the attractor $\mathcal{O}$ is a point attractor, the importance of the initialization largely depends on the linearization of the drift field around $\mathcal{O}$. For highly rotational fields (see Figure 3.11b) where the MAP undergoes infinite curvature at $\mathcal{O}$, the everimportant triangle updates are not going provide satisfactory approximations of the MAP. Since the performance fundamentally depends on the triangle updates, the standard initialization procedure of Algorithm 1 is insufficient here.

In most cases, it is sufficient to initialize the MAP on the 8-point nearest neighborhood of the attractor $\mathcal{O}$. Here, we approximate the solution U and ∇U on this neighborhood by the linearized solution. That is, suppose that b is continuously differentiable in a neighborhood of $\mathcal{O}$, and let

$$b(x) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \mathcal{O}(|x|^2),$$

in a neighborhood of $\mathcal{O}$ (assuming without loss of generality that $\mathcal{O} = (0,0)$). The quasi-potential corresponding to equation (3.1.1) with instead the linear drift is explicitly known (Section 4.1.2 of [14]) and we use this known formula to initialize the values of U and ∇U.

Due to the possibility of the situation shown in Figure 3.11b, this may not be sufficient if the MAPs still display heavy curvature in the 8-point neighborhood of $\mathcal{O}$. For these situations, it may be necessary to initialize by running run a separate, more heavily refined quasi-potential solving routine on the neighborhood where curvature is strong.

3.5 Results

In this section, we compare the performance of the EJM with several other methods on a pair of test problems. In particular, we are interested in the error convergence rates with respect to the mesh spacing h displayed by the EJM.

3.5.1 Nonlinear drift with varying rotational components

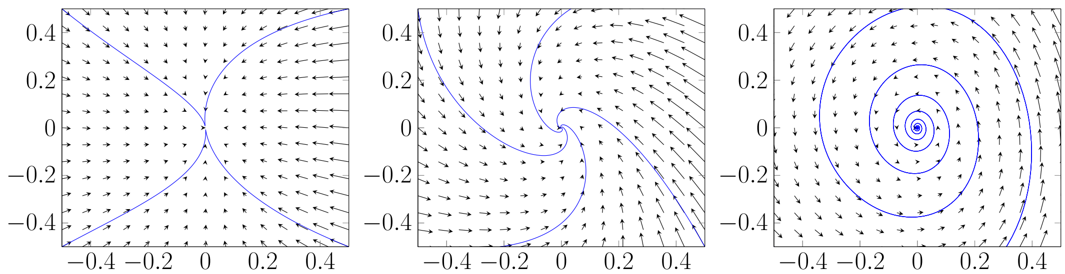
As our first example, we consider the nonlinear drift field

$$b(x,y) = -\frac{1}{2}\begin{bmatrix} 4x+3x^2 \\ 2y \end{bmatrix} + \frac{a}{2}\begin{bmatrix} -2y \\ 4x+3x^2 \end{bmatrix}. \tag{3.5.1}$$

Written this way, we immediately have the orthogonal decomposition $b = -\frac{1}{2}\nabla U + \ell$ with $\ell \perp \nabla U$, as in Remark 2.3.3. Here, the potential is given by $U(x,y) = 2x^2 + x^3 + y^2$. The parameter a controls the rotational component of the drift field. The flow lines of (3.5.1) (sample trajectories of the deterministic system $\dot{X} = b(X)$) are shown in Figures 3.13 and 3.15 for different values of a.

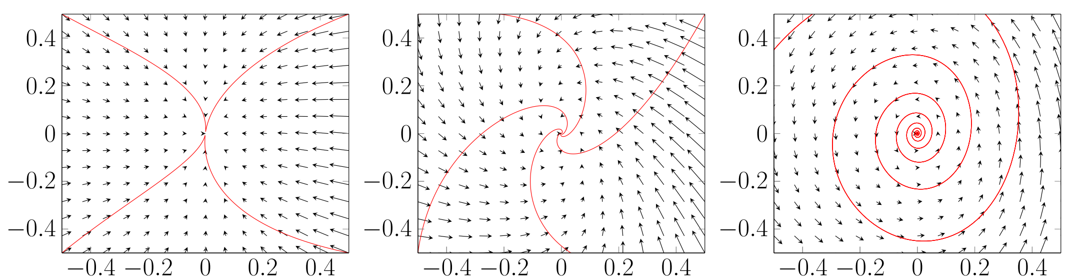
The drift field b admits a stable attracting equilibrium at $\mathcal{O} = (0,0)$ and a saddle at the point $x_0 = (-4/3, 0)$, for any value of a. As we see in Figure 3.13, for $a = 0.1$ the eigenvalues of the linearization of b around $\mathcal{O}$ are both real, and there is no spiraling of the flow lines. The cases $a = 1$ and $a = 10$ correspond to flow lines displaying a moderate and large amount, respectively, of spiraling around $\mathcal{O}$.

Moreover, as can be seen by Figure 3.5, the MAPs behave in the same way as the flow lines, but with a *flipped* rotational component. Thus, they exhibit the same degree of spiraling. Some sample MAPs are shown for (3.5.1) in Figure 3.14.



(a) $a = 0.1$ *(b)* $a = 1$ *(c)* $a = 10$

Fig. 3.13: Flow lines for drift field (3.5.1) for three values of a.



(a) $a = 0.1$ *(b)* $a = 1$ *(c)* $a = 10$

Fig. 3.14: Some MAPs for drift field (3.5.1) for three values of a.

For this problem, we compute the quasi-potential on the box $D = [-1, 1] \times [-1, 1]$, discretized into an $N \times N$ square mesh with common horizontal and vertical mesh spacing $h = 2/N$. We run EJM for the values $N = 2^k$ for $k = 7, ..., 12$. The solver is terminated once the first *boundary* mesh point is added to Accepted, corresponding to the first time the "wave" hits the edge of the box. Since this domain

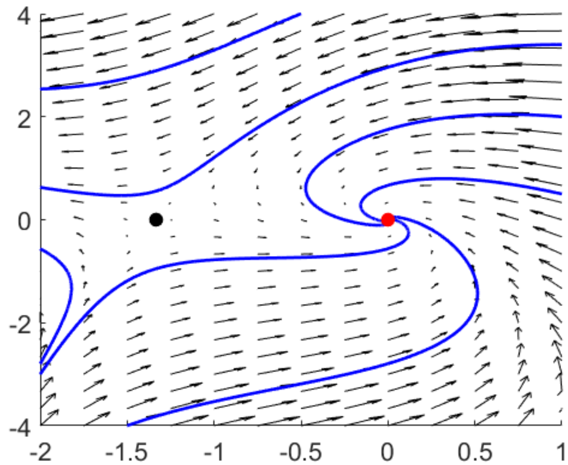


Fig. 3.15: Flow lines for drift field (3.5.1) with $a = 1$. The red marker indicates the attractor $\mathcal{O} = (0, 0)$ while the black marker indicates the unstable saddle at $(-4/3, 0)$.

D is within the well of attraction of $\mathcal{O}$, the solution $U(x) = 2x^2 + x^3 + y^2$ is the quasi-potential. We remark that outside the well of attraction, this is not the case.

We compare the performance of EJM to three other quasi-potential solvers. The primary comparison benchmark is the *OLIM with midpoint quadrature* [21, 22], the highest prior performing mesh-based quasi-potential solver. In addition, we compare EJM to a routine that uses the same cubic Hermite update formulas developed for EJM, but with the *neighborhood* selection strategy used in the OLIMs, rather than the anisotropic stencil strategy used in EJM. This benchmark is used to gauge the improvement in computing time obtained by the use of anisotropic stencils. Finally, we compare the EJM to Mirebeau's ASR algorithm directly [47], using linear update formulas with endpoint quadrature and anisotropic stencils constructed from the asymmetric norms $\mathcal{F}^\alpha$ (see Section 3.4.1). Each of these three benchmark methods we run on the same set of N values for the three different values of a.

Errors in U. In the left column of Figure 3.16, we plot the maximum error of U over all Accepted points versus N for each of the 4 methods and each of the three values of a. Best polynomial fits were computed and are shown in Table 3.1.

The most important takeaway is the clear 2nd order convergence rate in the errors for EJM for the cases of *mild* and *moderate* rotational components of the drift (left panel of Figures 3.16a and 3.16b). Best fit rates are 1.96 and 1.94 for $a = 0.1$ and $a = 1$, respectively. On the other hand, OLIM-midpoint, the prime metric of comparison, displays rates of 1.24 and 1.41, respectively, consistent with the super-linear, but not 2nd order convergence rates seen in [21]. For the largest grid size $N = 4096$ that we sampled, the performance difference between EJM and OLIM-midpoint is over 4 orders of magnitude for $a = 0.1$ and $a = 1$.

The accuracy improvement when the rotational component is large ($a = 10$, left panel of Figure 3.16c), is much smaller. Part of this is due to the significantly *improved* performance of OLIM-midpoint on drift fields with high-rotational components (discussed in [21]), while part of this due to the poorer performance of EJM on high-rotational drift fields. Nonetheless, EJM still performs better but not quite at a second order rate – the best convergence rates is 1.73. The decreased performance of EJM for this case is explained by the presence of more long-distance triangle updates, which is a consequence of the small angles between MAP and level set of U (see Section 3.4.3 and Figure 3.12).

Runtime. To analyze the utility of the anisotropic stencils, we compare the EJM with the OLIM-cubic method. As we see in the left panel of Figure 3.16, in the $a = 0.1$ and $a = 1$ case the errors of these two methods are indistinguishable, while in the $a = 10$ case they are very close. This is expected since these two methods differ only in the choice of neighborhoods and not in the update procedures.

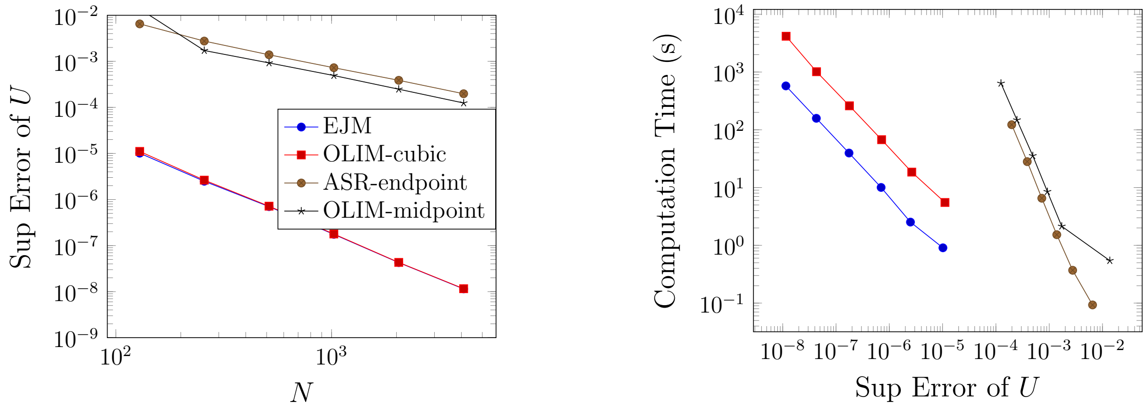
The important piece of information is the difference in runtime between these two methods, shown in the right panel of Figure 3.16. For the $a = 0.1$ and $a = 1$ cases, EJM is consistently between 7 and 8 times faster. In the $a = 10$ case, the improvement factor is only about 5, likely due to a larger number of fail-safe calls due to stencil sparsity.

Tab. 3.1: Best fits for the supremum error of U as a function of N.

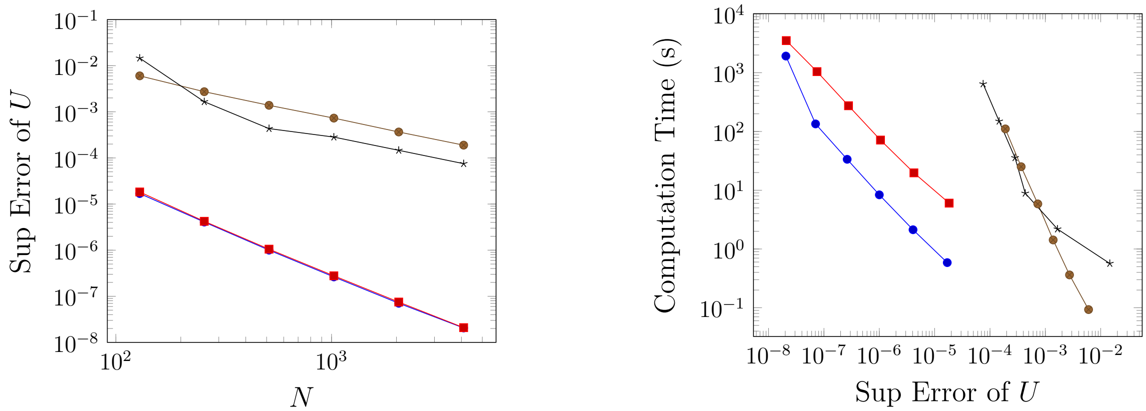
	$a = 0.1$	$a = 1$	$a = 10$
EJM	$0.14 * N^{-1.96}$	$0.20 * N^{-1.94}$	$3.54 * N^{-1.73}$
OLIM-cubic	$0.17 * N^{-1.98}$	$0.22 * N^{-1.95}$	$4.13 * N^{-1.85}$
ASR-endpoint	$0.72 * N^{-0.99}$	$0.70 * N^{-0.99}$	$0.52 * N^{-1.02}$
OLIM-midpoint	$2.86 * N^{-1.24}$	$5.95 * N^{-1.41}$	$5737.55 * N^{-2.32}$

Errors in ∇U. Next we consider the error of ∇U. Although rigorous numerical analysis remains as future work, we expect that the gradient ∇U should demonstrate 2nd order in h convergence.

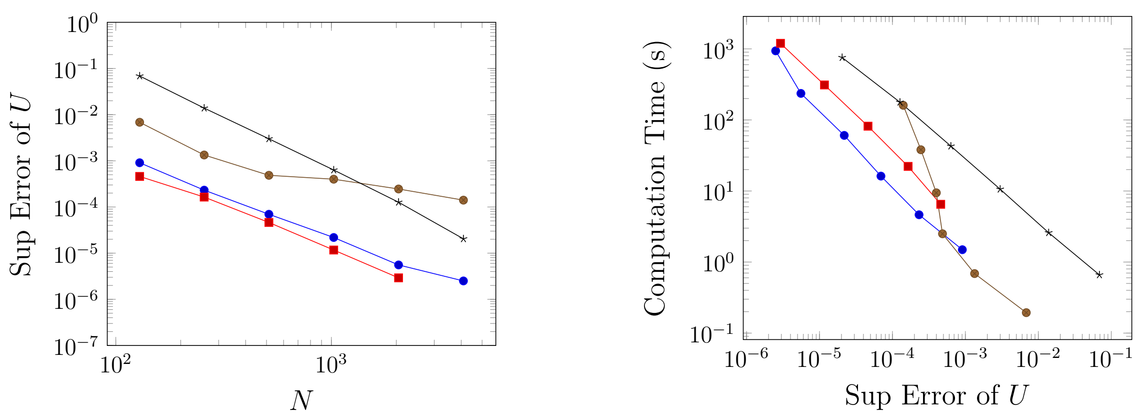
In Figure 3.17 we plot both the sup error (left panels) and root mean square



(a) $a = 0.1$



(b) $a = 1$



(c) $a = 10$

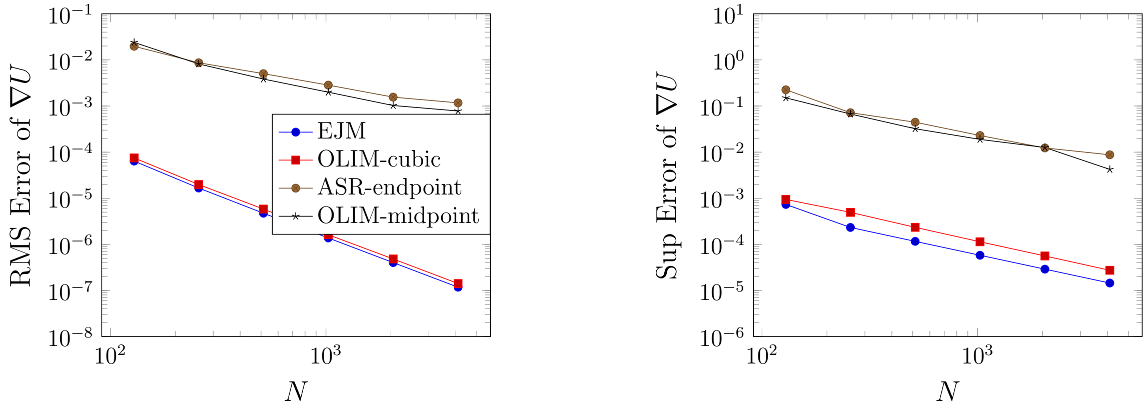
Fig. 3.16: Error plots of U for drift field (3.5.1) for the three values of a. Sup error of U over all Accepted mesh points is plotted against N and computation time.

(RMS) error (right panels) of ∇U against N. Best polynomial fits for these plots are included in Tables 3.2 and 3.3. The behavior of sup error and RMS error are markedly different. For the sup error (right panel), all 4 methods, including EJM, display linear in h convergence rates. For the EJM, this is simply due to a few one-point updates making it into the final cut, and is not overly significant.

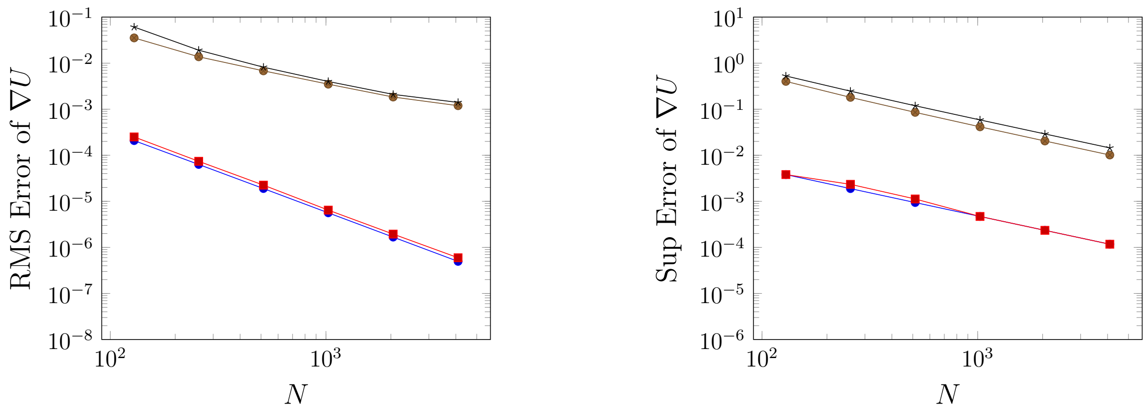
The RMS errors for EJM, on the other hand, display higher order convergence. Indeed, the best fit convergence rates (Table 3.2) are 1.81, 1.75 and 1.73 for $a = 0.1$, 1 and 10, respectively. These are superlinear, but not quite 2nd order. Presently, the reason for this is unclear. On the other hand, OLIM-midpoint displays the expected 1st order in h convergence of ∇U associated with linear methods, and the super-convergence in U displayed by OLIM-midpoint does not carry over to ∇U.

Tab. 3.2: Best fits for the RMS error of ∇U as a function of N.

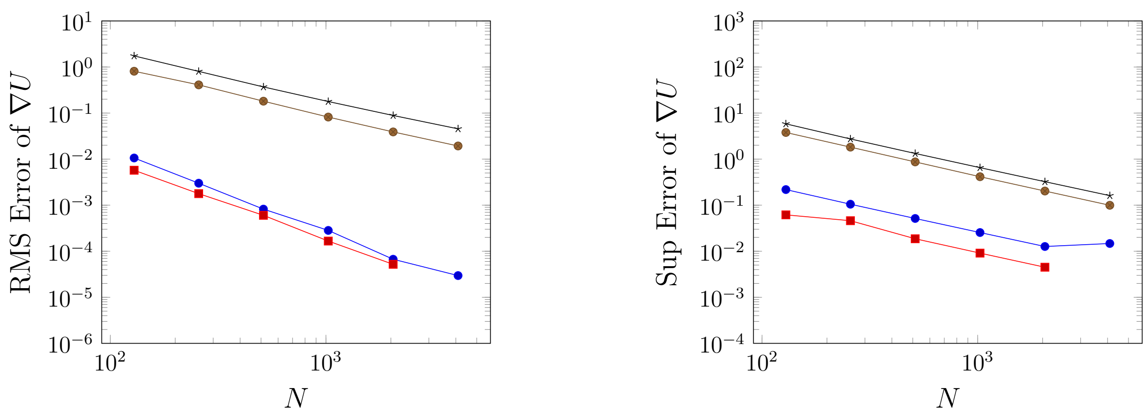
	$a = 0.1$	$a = 1$	$a = 10$
EJM	$0.40 * N^{-1.81}$	$1.02 * N^{-1.75}$	$43.74 * N^{-1.73}$
OLIM-cubic	$0.47 * N^{-1.81}$	$1.22 * N^{-1.75}$	$23.00 * N^{-1.70}$
ASR-endpoint	$0.90 * N^{-0.82}$	$3.38 * N^{-0.98}$	$169.58 * N^{-1.10}$
OLIM-midpoint	$2.24 * N^{-0.99}$	$8.63 * N^{-1.08}$	$281.61 * N^{-1.06}$



(a) $a = 0.1$



(b) $a = 1$



(c) $a = 10$

Fig. 3.17: Error plots of ∇U for drift field (3.5.1) for the three values of a. Sup error and RMS error of ∇U are plotted against N.

Tab. 3.3: Best fits for the supremum error of ∇U as a function of N.

	$a = 0.1$	$a = 1$	$a = 10$
EJM	$0.12 * N^{-1.10}$	$0.50 * N^{-1.01}$	$11.21 * N^{-0.85}$
OLIM-cubic	$0.14 * N^{-1.03}$	$0.67 * N^{-1.04}$	$9.02 * N^{-0.99}$
ASR-endpoint	$14.48 * N^{-0.92}$	$66.24 * N^{-1.06}$	$630.60 * N^{-1.05}$
OLIM-midpoint	$15.31 * N^{-0.97}$	$78.83 * N^{-1.04}$	$866.65 * N^{-1.04}$

3.5.2 Maier-Stein model

The problems of determining exit locations and exit trajectories from arbitrary domains are particularly suited for mesh-based solvers. In order to apply *path-based* methods, one has to know a-priori the exit location, which in general corresponds to a quasi-potential minima. Mesh-based solvers provide an easy way to determine this minima by simply interpolating the quasi-potential along the boundary of the domain of interest. Once the exit location is discovered, a path-based method can be used to determine the exit trajectory. *However*, one can also construct the exit trajectory by using the already calculatd mesh of ∇U values. In fact, this can be done with a negligible runtime, without the need to use a path-based method at all. In this section, we investigate the accuracy of reconstructing MAPs from the ∇U field obtained by EJM.

The test example we consider is the Maier-Stein system [44]

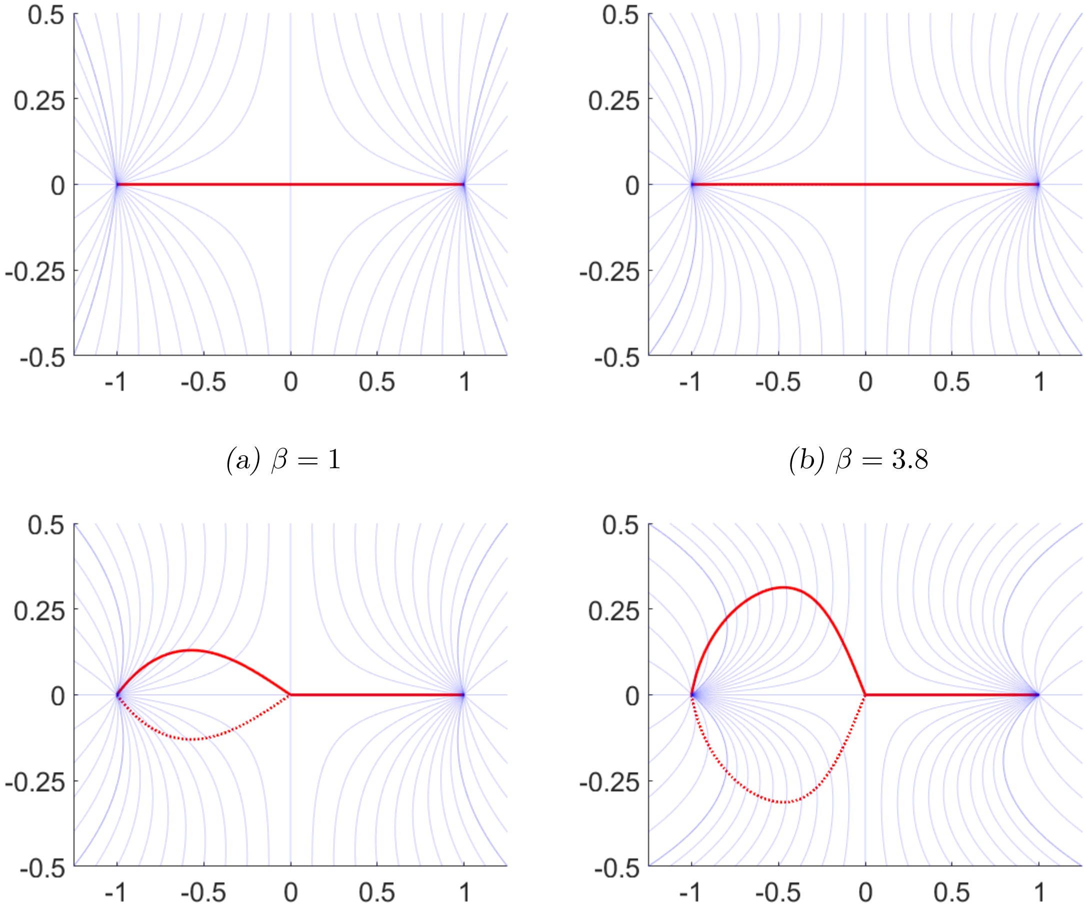
$$dX_t = b_{\mathrm{MS}}(X_t)dt + \sqrt{\epsilon}dW_t, \qquad X_0 = x_0 \in \mathbb{R}^2,$$

where

$$b_{\mathrm{MS}}(x,y) := \begin{bmatrix} x - x^3 - \beta x y^2 \\ -(1+x^2)y \end{bmatrix}. \tag{3.5.2}$$

Here, β is a positive parameter. For any $\beta > 0$, the drift (3.5.2) admits stable attracting equilibria at $\mathcal{O}_- := (-1, 0)$ and $\mathcal{O}_+ := (1, 0)$ and an unstable saddle at $\mathcal{O}_* := (0, 0)$. In particular, the solution X_t displays *metastability* in the $\epsilon \to 0$ limit. Namely, the solution will spend exponentially long periods of time near attractor $\mathcal{O}_-$, before experiencing a noise-induced transition to attractor $\mathcal{O}_+$, followed by another exponentially long time period near $\mathcal{O}_+$, and so forth. We are interested in exploring the maximum likelihood switching path between the two attractors. For this purpose, we assume $x_0 = O_-$ and consider the MAP φ^* that starts at $\mathcal{O}_-$ and ends at $\mathcal{O}_+$. As we saw in Chapter 2, as $\epsilon \to 0$, the actual transition path becomes overwhelmingly likely to occur within an arbitrarily narrow tube around this path φ^*. The flow lines of b_{MS} and maximum likelihood transition paths for a variety of β values are shown in Figure 3.18.

The first thing to notice about the Maier-Stein drift b_{MS} is that for any value of $\beta > 0$, the half-plane $\{x < 0\}$ lies in the well of attraction of $\mathcal{O}_-$ while the half-plane $\{x > 0\}$ lies in the well of attraction of $\mathcal{O}_+$. This can be seen by symmetry or by viewing the plots of the drift field in Figure 3.18. Since the drift field is symmetric



(a) $\beta = 1$

(b) $\beta = 3.8$

(c) $\beta = 4.2$

(d) $\beta = 10$

Fig. 3.18: Flow lines for the Maier Stein system as well as the maximum likelihood transition path (red) from $\mathcal{O}_- = (-1, 0)$ to $\mathcal{O}_+ = (1, 0)$ for four values of β. For (c) and (d) where $\beta > 4$, the transition path is not unique, as both the positive and negative arcs in the left half-plane constitute minimum action paths.

about $y = 0$, the maximum likelihood transition path φ^* will necessarily pass through the saddle $\mathcal{O}_*$. Moreover, since the MAP from $\mathcal{O}^*$ to $\mathcal{O}_+$ is just the flow line from $\mathcal{O}_*$ to $\mathcal{O}_+$, the portion of φ^* in the half-plane $\{x > 0\}$ will simply be the straight line from $\mathcal{O}_*$ to $\mathcal{O}_+$. The behavior of φ^* in the left half-plane $\{x < 0\}$ on the other hand, will depend on the value of β. It turns out that the MAP will exhibit different behaviors for $\beta < 4$ and $\beta > 4$.

The simplest case to study is the $\beta = 1$ case, where b_{MS} is the gradient of a potential function. In particular,

$$b_{\mathrm{MS}}(x, y) = \nabla\Big[\frac{1}{2}x^2 - \frac{1}{2}y^2 - \frac{1}{4}x^4 - \frac{1}{2}x^2y^2\Big].$$

As we saw in Chapter 2, for gradient systems, MAPs are given by time-reversals of the deterministic flows. Therefore, the portion of the MAP φ_* in the left half-plane is simply the straight line connecting $\mathcal{O}_-$ to $\mathcal{O}_*$ (Figure 3.18a). It turns out that this is also the case for any $\beta < 4$. One can interpret these values of β as being the "near" gradient system cases.

At the critical value of $\beta = 4$, however, two non-trivial MAPs begin to emerge, as shown in Figures 3.18c and 3.18d. Indeed, for $\beta > 4$, the MAP is no longer unique and there exist symmetric positive and negative arcs which both minimize the geometric action in the left half-plane. An interesting consequence of this non-uniqueness is that the quasi-potential at their point of intersection, the saddle $\mathcal{O}_*$, is not differentiable.

The task in this section is to reconstruct the MAP φ^* for the $\beta = 10$ case from the EJM mesh of ∇U values by using the following procedure. Assume that we have run EJM on a uniform discretization of $[-2, 0] \times [-1, 1]$. In view of (3.1.3), we can construct a discrete approximation to the path φ_* by defining iteratively the sequence

$\varphi_k = (x_k, y_k)$

$$\varphi_{k+1} = \varphi_k - \bar{h}\Big(\frac{\nabla U(\varphi_k) + b(\varphi_k)}{\|b(\varphi_k)\|}\Big), \qquad \varphi_0 = \mathcal{O}_*, \tag{3.5.3}$$

for some step size $\bar{h}$. We took $\bar{h} = h/10$ where h is the mesh spacing used in the run of the EJM. The iteration is terminated once φ_k is sufficiently close to the attractor $\mathcal{O}_-$. To compute the values $\nabla U(\varphi_k)$ in (3.5.3), when $\varphi_k = (x_k, y_k)$ is not a mesh point, we simply use bi-linear interpolation of ∇U between the nearest mesh points.

To evaluate the error of our computed paths $\{\varphi_k\}_{k=0}^K$ by this method, we construct a master solution φ_{GMAM} by running GMAM (Section 3.2) with a large number of degrees of freedom. In particular, φ_{GMAM} is a piecewise-linear path with 10,000 nodes. To compute the error of a discretized path $\varphi = \{\varphi_k\}_{k=0}^K$ reconstructed from EJM, we simply compute

$$\mathrm{Err}(\varphi) = \sup_{k=0,\ldots,K} \mathrm{dist}(\varphi_k, \varphi_{\mathrm{GMAM}}).$$

The distances can be computed analytically since φ_{GMAM} is piecewise-linear. These errors are plotted in Figure 3.19 as a function of the number of steps $N = 2/h$ used in the EJM algorithm. We compare this to the errors of the MAP reconstructed in the same way from the OLIM with midpoint quadrature.

In both cases, the convergence is 1st order in h, although the error constant is significantly better for the EJM method. However, in view of the nearly 2nd order convergence of ∇U displayed by EJM (see Figure 3.17), we would expect near 2nd order convergence in MAP accuracy. This is not achieved because we are using only

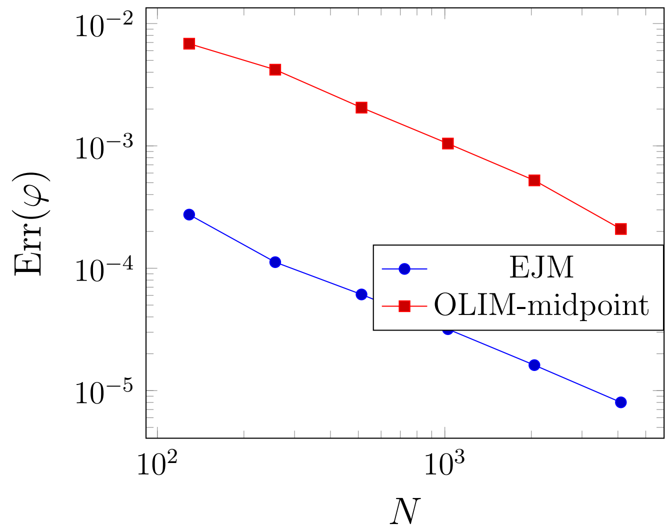


Fig. 3.19: Errors of the reconstructed maximum likelihood transition path from $\mathcal{O}_-$ to $\mathcal{O}_*$ for the Maier-Stein model.

bi-linear interpolation of ∇U between mesh points. Indeed, if we change this to bi-cubic interpolation, we expect to get near second order convergence in Err(φ).

3.5.3 Future applications

When the quasi-potential is twice continuously differentiable, a key benefit of a 2nd order convergence rate for the gradient ∇U, is the implied convergence of 2nd derivative matrix $\nabla^2 U$. Since, the sub-exponential pre-factors of stationary measures and expected transition times in general depend on *second* derivative information about U (as in equation (2.3.14), for instance), one may hope to numerically compute pre-factors to reasonable accuracy.

Indeed, suppose for instance that, as in the Maier-Stein model, a drift field b admits two stable attractors at x_- and x_+. Let $\tau^{x_- \to x_+}$ be the expected transition time from x_- to x_+. If the quasi-potential is twice continuously differentiable at the saddle point x_* between x_- and x_+, Bouchet and Reygner [6] showed that

$$\mathbb{E}\tau^{x_- \to x_+} \asymp \frac{2\pi}{\lambda_+^*}\sqrt{\frac{|\mathrm{det}H(x^*)|}{\mathrm{det}H(x_-)}}\exp\Big(\int_0^L F(\varphi_s^*)ds\Big)\exp\Big(\frac{U(x^*)}{\epsilon}\Big). \tag{3.5.4}$$

where H is the Jacobian matrix of the quasi-potential, λ_+^* is the unique positive eigenvalue of the Jacobian matrix at the saddle x^*, φ^* is the MAP from x_- to x_+ and

$$F(x) := \nabla \cdot \Big(b(x) + \frac{1}{2}\nabla U(x)\Big).$$

This can be seen as a generalization of the Eyring-Kramer formula to non-gradient systems. For a nice survey of results and techniques in this direction, see the article [2].

Convergence of the 2nd derivative matrix $\nabla^2 U$ obtained by the EJM would suggest the possibility of computing the exponential prefactor numerically. Indeed, using the MAP solver discussed above (or a path-basd method) along with values of ΔU computed using finite differences, one could expect to obtain convergence of a numerical computation of the right hand side of (3.5.4).

We note however, that the requirement of U being twice continuously differentiable at the saddle is often not met in practice. For instance, the Maier Stein example with $\beta = 10$, does not satisfy this condition. This can be seen as a consequence of the fact that the two MAPs (up and down arcs) intersect at the saddle point $\mathcal{O}_*$. On the other hand, for $\beta < 4$, we would expect that the quasi-potential is twice continuously differentiable at $\mathcal{O}_*$. In fact, for the $\beta = 1$ case, where the solution U is explicit, one could compute the pre-factor analytically and compare this to an approximate solution obtained by computing the right hand side of (3.5.4) using the techniques discussed. This remains as future work.

3.6 Conclusions

Thus far, we have tested the efficacy of EJM on nonlinear drift fields that have known, *smooth* solutions. In such cases, it has proven *far* more accurate than previous techniques when the drift fields contain at most moderate rotational components. In particular, it displays 2nd order convergence in the error of U and close to 2nd order convergence in the gradient ∇U. Nonetheless, it remains to test our algorithm in more diverse settings, including in situations where the quasi-potential may be rougher.

Admittedly, our solver suffers from an issue of over-complexity in a couple of areas. First, the higher order updates require minimizations (3.4.5) and (3.4.7) that are performed via Newton's method with first and second order derivatives computed and implemented by hand. These formulas are extremely lengthy and complicated, which make EJM a very difficult algorithm to implement from scratch. We have yet to explore the efficacy of using finite differences for the 2nd order derivatives and of implementing automatic differentiation techniques.

Second, the Accept/Reject procedure described in Algorithm 6 and Condition 3.4.1 are more complicated than in traditional Fast Marching methods, which accept a new proposed update U_{new} value if it is smaller than any previously proposed values.

Instead, EJM requires a more complicated rule. The two key issues cause this are the (a) unacceptably large errors coming from triangle updates computed over large distances and (b) the presence of *undesired* local minimizers of (3.4.7). The procedure presented in this chapter is our best attempt to simultaneously circumvent both of these issues, but it is possible there exist more concise solutions.

Chapter 4: Large deviations for fast transport reaction-diffusion equations

4.1 Introduction

In the remaining two chapters, we shift our focus towards the application of Freidlin-Wentzell theory in the infinite-dimensional setting. Using the flexibility of the weak convergence approach (Section 2.2) as well as many of the ideas detailed in Chapter 2, we study the small-noise asymptotics of a pair of important multi-scale stochastic partial differential equations.

In this chapter we study reaction-diffusion equations, a class of partial differential equations most commonly used to describe the evolution of chemical concentrations in reactions in which the concentrations vary in space. We consider the case where *random* changes in time and space of the rates of reaction occur, and we allow for the possibility of different sources of noise acting on the interior and exterior of the spatial domain, respectively. Equations containing boundary noise are typically used when there exists some random mechanism that acts only along a particular interface; for example, in the modeling of heat transfer within a solid that is in a contact with a fluid [40] or in the modeling of interactions between air and water on the surface of oceans [50]. In applications, it is often the case that the rates of chemical reactions and the diffusion coefficients have different orders of magnitude. Here, we address the particular regime where the relative size of the diffusion is much larger than the rates of reaction and, in addition, the deterministic rate of reaction is much larger than the stochastic rate of reaction.

More precisely, we consider the following class of stochastic reaction-diffusion equations,

$$\begin{cases} \dfrac{\partial u_\epsilon}{\partial t}(t,\xi) = \epsilon^{-1}\,\mathcal{A}u_\epsilon(t,\xi) + f(t,\xi,u_\epsilon(t,\xi)) + \alpha(\epsilon)\,g(t,\xi,u_\epsilon(t,\xi))\dfrac{\partial W^Q}{\partial t}(t,\xi), \quad \xi \in \mathcal{D}, \\ \dfrac{\partial u_\epsilon}{\partial \nu}(t,\xi) = \epsilon\,\beta(\epsilon)\,\sigma(t,\xi)\dfrac{\partial W^B}{\partial t}(t,\xi), \quad \xi \in \partial\mathcal{D}, \qquad u_\epsilon(0,\xi) = x(\xi), \quad \xi \in \mathcal{D}, \end{cases} \tag{4.1.1}$$

for $0 < \epsilon \ll 1$ and for some positive functions $\alpha(\epsilon)$ and $\beta(\epsilon)$, both converging to zero, as $\epsilon \to 0$. Here, $\mathcal{D}$ is a bounded domain in $\mathbb{R}^d$, $d \geq 1$, with a smooth boundary, $\mathcal{A}$ is a uniformly elliptic second order differential operator, and $\partial/\partial\nu$ is the associated co-normal derivative acting at $\partial\mathcal{D}$. The coefficients $f, g : [0,\infty) \times \mathcal{D} \times \mathbb{R}$ satisfy a Lipschitz condition with respect to the third variable. The noises W^Q and W^B are cylindrical Wiener processes valued in $H = L^2(\mathcal{D})$ and $Z = L^2(\partial\mathcal{D})$, respectively, with covariances $Q \in \mathcal{L}^+(H)$ and $B \in \mathcal{L}^+(Z)$. The full description of our assumptions is given in the next section.

We assume here that the diffusion X_t associated with the operator $\mathcal{A}$, endowed with the co-normal boundary condition, admits a unique invariant measure μ and a spectral gap occurs. Namely, there exists some $\gamma > 0$ such that for any $h \in L^2(\mathcal{D}, \mu)$,

$$\int_{\mathcal{D}} |\mathbb{E}_\xi h(X_t) - \langle h, \mu \rangle|^2 \, d\mu(\xi) \leq c \, e^{-2\gamma t} \, \|h\|^2_{L^2(\mathcal{D},\mu)}, \qquad t \geq 0, \tag{4.1.2}$$

where $\langle h, \mu \rangle = \int_{\mathcal{D}} h(\xi) d\mu(\xi)$. Physically, the measure μ and the spectral gap γ represent the *averaged* distribution across $\mathcal{D}$ resulting from the diffusion $\mathcal{A}$ and the corresponding exponential *rate* of averaging, respectively. In the prototypical case of uniform diffusion in which $\mathcal{A}$ is the Laplacian operator Δ, the invariant measure μ is simply the scaled Lebesgue measure, so that the system tends towards a concentration that is uniform in space.

In fact, if the deterministic and stochastic rates of reaction in (4.1.1) were of order one (i.e. $\alpha(\epsilon) = \mathcal{O}(1)$), then the diffusion term would disappear in the $\epsilon \to 0$ limit, and the effective dynamic would be described by an "averaged" ordinary stochastic differential equation. Indeed, in [16] (see also [5]) it was shown that, for every $0 < \delta < T$ and $p \geq 1$, the solutions u_ϵ^x to (4.1.1), corresponding to $\alpha(\epsilon) = \beta(\epsilon) = 1$, converge in $L^p(\Omega; C([\delta, T]; L^2(\mathcal{D}, \mu)))$ to the solution of the averaged one-dimensional stochastic differential equation

$$du(t) = \bar{F}(t, u(t))dt + \bar{G}(t, u(t))dW^Q(t) + \bar{\Sigma}(t)dW^B(t), \qquad u(0) = \langle x, \mu \rangle. \tag{4.1.3}$$

Here, $\bar{F}$, $\bar{G}$, and $\bar{\Sigma}$ are all obtained by taking suitable spatial averages of their counterparts, f, g and σ, with respect to the invariant measure μ. Since the averaging still takes time, convergence in $C([0, T]; L^2(\mathcal{D}, \mu))$ only occurs if the initial condition x is already constant in space.

We are interested in studying the fast transport approximation described above in the small noise regime (i.e. $\alpha(\epsilon) \to 0$ and $\beta(\epsilon) \to 0$). In this case, the noisy terms vanish entirely from the limit and the solution to (4.1.1) converges in $L^p(\Omega; C([\delta, T]; L^2(\mathcal{D}, \mu)))$ to the solution of the ODE

$$\frac{du}{dt} = \bar{F}(t, u(t)), \qquad u(0) = \langle x, \mu \rangle. \tag{4.1.4}$$

Thus, we believe it is of interest to study the validity of a large deviation principle for the family $\{u_\epsilon^x\}_{\epsilon>0}$ in the space $C([\delta, T]; L^2(\mathcal{D}, \mu))$, and, in particular, to understand its interplay with the fast transport limit. It turns out that, depending on the following different scalings between $\alpha(\epsilon)$ and $\beta(\epsilon)$

$$\lim_{\epsilon \to 0} \frac{\beta(\epsilon)}{\alpha(\epsilon)} = \bar{\rho} \in [0, +\infty],$$

the action functional and the speed governing the large deviation principle for equation (4.1.1) are precisely the same as those governing the large deviation principle for the stochastic *ordinary* differential equation

$$du(t) = \bar{F}(u(t))\, dt + (\alpha(\epsilon) + \beta(\epsilon)) \sqrt{\mathcal{H}_{\bar{\rho}}(t, u(t))}\, d\beta_t, \qquad u(0) = \langle x, \mu \rangle,$$

where

$$\mathcal{H}_{\bar{\rho}}(t, u) = \frac{1}{(1+\bar{\rho})^2} \left[\left\| \sqrt{Q}\, [G(t, u)m] \right\|_H^2 + \bar{\rho}^2 \left\| \delta_0 \sqrt{B} \left[\Sigma(t) N_{\delta_0}^* m \right] \right\|_Z^2 \right], \tag{4.1.5}$$

(here m is the density of the invariant measure μ). This means in particular that the fast transport asymptotics for equation (4.1.1) is consistent with the small noise limit.

After obtaining the large deviations principle for the paths, we then study the problem of the exit of the solutions u_ϵ^x to (4.1.1) from a bounded domain D in the functional space $L^2(\mathcal{D},\mu)$. As in Section 2.3.1, we consider the case where the limiting equation (4.1.4) has an attractive equilibrium at 0, and we prove Freidlin-Wentzell type exit time estimates of the form (2.3.3). More precisely, if we define

$$\tau_\epsilon^x := \inf\{t \geq 0 : u_\epsilon^x(t) \in \partial D\},$$

then we show that for any initial condition $x \in D \subset L^2(\mathcal{D},\mu)$,

$$\lim_{\epsilon\to 0} \left(\alpha(\epsilon) + \beta(\epsilon)\right)^2 \log \mathbb{E}\tau_\epsilon^x = \inf_{y\in\partial D} V(y), \tag{4.1.6}$$

where $V : L^2(\mathcal{D},\mu) \to \mathbb{R}^+$ is the quasi-potential corresponding to the action functional governing the large deviation principle, defined as in (2.3.1). If the interior noise is additive, i.e. $g \equiv 1$ in (4.1.1), then the quasi-potential can be written explicitly. Namely,

$$V(y) = -2\,\mathcal{H}_{\bar\rho}^{-1} \int_0^y \bar F(r)dr.$$

where $\mathcal{H}_{\bar\rho}$ is obtained from (4.1.5) by setting $G(t,u) = \mathrm{Id}$ and by assuming Σ constant in time. For example, when $\mathcal{A}$ is a divergence type operator, we have $m = |\mathcal{D}|^{-1}$ and hence

$$V(y) = -\frac{(1+\bar\rho)^2}{c_1 + c_2\,\bar\rho^2} \int_0^y \int_{\mathcal{D}} f(\xi,\sigma)\,d\xi\,d\sigma,$$

for some non-negative constants c_1 and c_2, depending on Q and B, and not simultaneously zero. In the general case of multiplicative noise, we do not have such an explicit representation of the quasi-potential; however, the result (4.1.6) still holds.

As we saw in Section 2.3.1, in order to obtain results of the form (4.1.6), a large deviations principle for the paths that is *uniform* with respect to initial conditions in bounded sets of $L^2(\mathcal{D},\mu)$ is needed. Here, we are able to prove the validity of a large deviations principle for the family $\{u_\epsilon^x\}_{\epsilon>0}$ in the space $C([\delta,T];L^2(\mathcal{D},\mu))$ by using the weak convergence approach described in Section 2.2. In particular, we let the space of initial conditions ($\mathcal{E}_0$ in Section 2.2) be the compact Polish space of $L^2(\mathcal{D},\mu)$ endowed with the topology of *weak* convergence, so that the Laplace principle proven via the weak convergence approach will be uniform with respect to initial conditions in bounded sets of $L^2(\mathcal{D},\mu)$. This can then be extended into a uniform large deviations principle by verifying the conditions of Propositions 2.2.3 and 2.2.4. The task of proving the uniform large deviations principle for the paths then reduces to the problem of proving the validity of Hypothesis 2.2.1.

Once we have established a large deviation principle that is uniform with respect to initial conditions in a bounded set, we prove (4.1.6) by adapting the method used in finite dimension (see Chapter 4, Section 2 of [37] and Chapter 5.7 of [28]) to our infinite dimensional setting (see [17], [9] and [20] for some previous results in this direction). In our model, several complications arise in obtaining the lower bound of $\mathbb{E}\tau_\epsilon^x$. Actually, when ϵ is small, equation (4.1.1) behaves like the linear heat equation for t on the order of ϵ. However, for times on the order of 1, the averaging has already taken place so that the solution is essentially constant in space and evolves according to (4.1.4). So to establish any kind of lower bound on the exit time, we require a domain that is both invariant with respect to the semigroup e^{tA} and invariant with respect to trajectories of equation (4.1.4).

4.2 Notations and preliminaries

4.2.1 Assumptions on the semigroup

We assume that $\mathcal{D}$ is a bounded domain in $\mathbb{R}^d$, $d \geq 1$, with a smooth boundary, satisfying the extension and exterior cone properties. We denote $H := L^2(\mathcal{D})$ and $Z := L^2(\partial\mathcal{D})$, and, for any $\alpha > 0$, we denote $H^\alpha := H^\alpha(\mathcal{D})$ and $Z^\alpha := H^\alpha(\partial\mathcal{D})$.

We assume that $\mathcal{A}$ is a second order differential operator of the form

$$\mathcal{A} = \sum_{i,j=1}^{d} \frac{\partial}{\partial \xi_i}\left(a_{ij}(\xi)\frac{\partial}{\partial \xi_j}\right) + \sum_{i=1}^{d} b_i(\xi)\frac{\partial}{\partial \xi_i}, \qquad \xi \in \mathcal{D}.$$

The matrix $a(\xi) = [a_{ij}(\xi)]_{i,j}$ is symmetric and all entries a_{ij} are differentiable, with continuous derivatives in $\bar{\mathcal{D}}$. Moreover, there exists some $a_0 > 0$ such that

$$\inf_{\xi \in \bar{\mathcal{D}}} \langle a(\xi)\eta, \eta\rangle \geq a_0|\eta|^2, \qquad \eta \in \mathbb{R}^d. \tag{4.2.1}$$

Finally, the coefficients b_i are continuous on $\bar{\mathcal{D}}$.

In what follows, we shall denote by A the realization in H of the differential operator $\mathcal{A}$, endowed with the conormal boundary condition

$$\frac{\partial h}{\partial \nu}(\xi) := \langle a(\xi)\nu(\xi), \nabla h(\xi)\rangle = 0, \qquad \xi \in \partial\mathcal{D}.$$

The operator A generates a strongly continuous analytic semigroup in H, which we will denote by e^{tA}. Moreover (see [41] for a proof)

$$D(A^\alpha) \subseteq H^{2\alpha}, \quad \text{for } \alpha \geq 0, \qquad D(A^\alpha) = H^{2\alpha}, \quad \text{for } 0 \leq \alpha < \frac{3}{4}, \tag{4.2.2}$$

In general, the realization of $\mathcal{A}$ in L^p spaces under the same boundary conditions will also generate a strongly continuous, analytic semigroup, for $p > 1$. It is proved in [27] that under the above conditions on $\mathcal{A}$ and $\mathcal{D}$, the semigroup admits an integral kernel $k_t(\xi, \eta)$ that satisfies

$$0 \leq k_t(\xi,\eta) \leq c\,(t^{-\frac{d}{2}} + 1), \qquad t > 0, \tag{4.2.3}$$

In what follows, we shall assume that e^{tA} satisfies the following condition.

Hypothesis 4.2.1. The semigroup e^{tA} admits a unique invariant measure μ, and there exist $\gamma > 0$ and $c > 0$ such that, for any $h \in L^2(\mathcal{D}, \mu)$ and $t \geq 0$,

$$\left\| e^{tA} h - \int_{\mathcal{D}} h(\xi) d\mu(\xi) \right\|_{L^2(\mathcal{D},\mu)} \leq c\, e^{-\gamma t} \, \|h\|_{L^2(\mathcal{D},\mu)} \,. \tag{4.2.4}$$

In what follows, we shall denote

$$H_\mu := L^2(\mathcal{D}, \mu), \qquad \langle h, \mu \rangle := \int_{\mathcal{D}} h(\xi) d\mu(\xi), \quad h \in H_\mu.$$

Remark 4.2.1. Hypothesis 4.2.1 is satisfied for example if $\mathcal{A}$ is a divergence-type operator. Actually, in this case the Lebesgue measure is invariant under the semigroup e^{tA}, so that we can define

$$\mu = |\mathcal{D}|^{-1} \lambda_d,$$

where λ_d is the Lebesgue measure on $\mathbb{R}^d$. Since A is self-adjoint, we can find a complete orthonormal system $\{e_k\}_{k\geq 0}$ in H, and an increasing nonnegative sequence $\{\alpha_k\}_{k \geq 0}$

such that $Ae_k = -\alpha_k e_k$. Clearly, $\alpha_0 = 0$ and $e_0 = |\mathcal{D}|^{-1/2}$, so that $\langle x, \mu \rangle = \langle x, e_0 \rangle_H e_0$, for any $x \in H$. This implies that

$$\left\| e^{tA} x - \langle x, \mu \rangle \right\|_{H_\mu}^2 = |\mathcal{D}|^{-1} \sum_{i=1}^{\infty} e^{-2t\alpha_i} \langle x, e_i \rangle_H^2 \leq e^{-2t\alpha_1} \left\| x \right\|_{H_\mu}^2,$$

so that (4.2.4) holds for $\gamma = \alpha_1$.

$\square$

Remark 4.2.2. We have the continuous embedding $H \hookrightarrow H_\mu$. This follows from the invariance of μ with respect to e^{tA}, and from the boundedness of the integral kernel (4.2.3). Actually, for $h \in H$, we have

$$\left\| h \right\|_{H_\mu}^2 = \int_{\mathcal{D}} e^{1A} |h|^2(\xi) d\mu(\xi) = \int_{\mathcal{D}} \int_{\mathcal{D}} k_1(\xi, \eta) |h(\eta)|^2 d\eta \, d\mu(\xi) \leq c \left\| h \right\|_H^2.$$

We also note that, due to the invariance of μ, e^{tA} acts as a contraction in H_μ

$$\left\| e^{tA} h \right\|_{H_\mu}^2 \leq \int_{\mathcal{D}} e^{tA} \left| h(\xi) \right|^2 d\mu(\xi) = \left\| h \right\|_{H_\mu}^2.$$

$\square$

Remark 4.2.3. In fact, one can show that the invariant measure μ is absolutely continuous with respect to the Lebesgue measure on $\mathcal{D}$ and has a nonnegative density $m \in L^\infty(\mathcal{D})$ (for a proof, see [16]).

$\square$

4.2.2 Assumptions on the coefficients and noise

Concerning the coefficients f, g, and σ, we make the following assumptions.

Hypothesis 4.2.2.

(i) The mappings $f, g : [0,\infty)\times\mathcal{D}\times\mathbb{R} \to \mathbb{R}$ are measurable and Lipschitz continuous in the third variable, uniformly with respect to $(t,\xi) \in [0,T]\times\mathcal{D}$, for any fixed $T > 0$. In addition, for any $T > 0$, f and g satisfy

$$\sup_{0\leq t\leq T} \|f(t,\cdot,0)\|_{L^\infty(\mathcal{D})} < +\infty, \qquad \sup_{0\leq t\leq T} \|g(t,\cdot,0)\|_{L^\infty(\mathcal{D})} < +\infty,$$

(ii) The mapping $\sigma : [0,\infty)\times\partial\mathcal{D} \to \mathbb{R}$ is measurable and satisfies for any $T > 0$

$$\sup_{0\leq t\leq T} \|\sigma(t,\cdot)\|_{L^\infty(\partial\mathcal{D})} < +\infty,$$

In what follows, for $h_1, h_2 \in H$ and $\xi \in \mathcal{D}$, we define

$$F(t,h_1)(\xi) := f(t,\xi,h_1(\xi)),$$

and

$$[G(t,h_1)h_2](\xi) := g(t,\xi,h_1(\xi))h_2(\xi).$$

The uniform Lipschitz assumptions on f and g in Hypothesis 4.2.2 imply that the mappings $F(t,\cdot) : H \to H$, $G(t,\cdot) : H \to \mathcal{L}(H, L^1(\mathcal{D}))$, and $G(t,\cdot) : H \to \mathcal{L}(L^\infty(\mathcal{D}), H)$ are all well-defined and Lipschitz continuous, uniformly with respect to $t \in [0,T]$, for any $T > 0$.

Next, for $z \in Z$ and $\xi \in \partial\mathcal{D}$, we set

$$[\Sigma(t)z](\xi) := \sigma(t,\xi)z(\xi).$$

Hypothesis 4.2.2 implies that $\Sigma(t) \in \mathcal{L}(Z)$ and $\sup_{0\leq t\leq T} \|\Sigma(t)\|_{\mathcal{L}(Z)} < \infty$.

Concerning the noisy terms, we assume that $W^Q(t)$ and $W^B(t)$ are cylindrical Wiener processes in H and Z, with covariances $Q \in \mathcal{L}^+(H)$ and $B \in \mathcal{L}^+(Z)$, respectively. That is,

$$W^Q(t) = \sum_{k=0}^{\infty} \sqrt{Q} e_k \beta_k(t), \qquad W^B(t) = \sum_{k=0}^{\infty} \sqrt{B} f_k \tilde{\beta}_k(t),$$

where $\{e_k\}_{k\geq 0}$ is an orthonormal basis of H, $\{f_k\}_{k\geq 0}$ is an orthonormal basis of Z and $\{\beta_k(t)\}_{k\geq 0}$ and $\{\tilde{\beta}_k(t)\}_{k\geq 0}$ are sequences of independent real-valued Brownian motions defined on a common stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$.

We assume for simplicity that $\{e_k\}_{k\geq 0}$ diagonalizes $\sqrt{Q}$ with eigenvalues $\{\lambda_k\}_{k\geq 0}$, and $\{f_k\}_{k\geq 0}$ diagonalizes $\sqrt{B}$ with eigenvalues $\{\theta_k\}_{k\geq 0}$. We do not assume that the operators Q and B are trace class, so the sums above do not necessarily converge in H and Z, respectively. However, both of the sums converge in larger Hilbert spaces containing H and Z, respectively, with Hilbert-Schmidt embeddings.

We make the following assumption regarding the eigenvalues of Q and B.

Hypothesis 4.2.3. If $d \geq 2$, then there exist $\rho < 2d/(d-2)$ and $\beta < 2d/(d-1)$ such that

$$\sum_{k\in\mathbb{N}} \lambda_k^{\rho} \|e_k\|_{\infty}^2 =: \kappa_Q < \infty, \qquad \sum_{k\in\mathbb{N}} \theta_k^{\beta} =: \kappa_B < \infty. \tag{4.2.5}$$

4.2.3 Mild solutions

In the present paper, we are dealing with the following class of equations

$$\begin{cases} \dfrac{\partial u_\epsilon}{\partial t}(t,\xi) = \epsilon^{-1}\,\mathcal{A}u_\epsilon(t,\xi) + f(t,\xi,u_\epsilon(t,\xi)) + \alpha(\epsilon)\,g(t,\xi,u_\epsilon(t,\xi))\dfrac{\partial W^Q}{\partial t}(t,\xi), \quad \xi \in \mathcal{D}, \\[2em] \dfrac{\partial u_\epsilon}{\partial \nu}(t,\xi) = \epsilon\,\beta(\epsilon)\,\sigma(t,\xi)\dfrac{\partial W^B}{\partial t}(t,\xi), \quad \xi \in \partial\mathcal{D}, \qquad u_\epsilon(0,\xi) = x(\xi), \quad \xi \in \mathcal{D}. \end{cases} \tag{4.2.6}$$

Under the above assumptions on the differential operator $\mathcal{A}$ and the domain $\mathcal{D}$, it can be shown (see [43]), that there exists $\delta_0 \in \mathbb{R}$ such that for any $\delta \geq \delta_0$ and $h \in Z$, the elliptic boundary value problem

$$(\delta - \mathcal{A})u(\xi) = 0, \quad \xi \in \mathcal{D}, \qquad \frac{\partial u}{\partial \nu} = h(\xi), \quad \xi \in \partial\mathcal{D},$$

admits a unique solution $u \in H$. We define the Neumann map, $N_\delta : Z \to H$, to be the solution map of this equation, i.e. $N_\delta h := u$. One can show that

$$N_\delta \in \mathcal{L}(Z^\alpha, H^{\alpha+3/2}). \tag{4.2.7}$$

Next, we consider the deterministic parabolic problem

$$\begin{cases} \dfrac{\partial y}{\partial t}(t,\xi) = \mathcal{A}y(t,\xi), \quad \xi \in \mathcal{D} \\[2em] \dfrac{\partial y}{\partial \nu} = v(t,\xi), \quad \xi \in \partial\mathcal{D}, \qquad y(0,\xi) = 0, \quad \xi \in \mathcal{D}. \end{cases}$$

One can show that for smooth v and large enough δ, the solution to this equation is given explicitly by

$$y(t) = (\delta - A) \int_0^t e^{(t-s)A} N_\delta v(s) ds.$$

This formula can be extended by continuity to provide a notion of mild solution for less regular v. In our case, we are interested in the boundary value problem

$$\begin{cases} \dfrac{\partial y}{\partial t}(t,\xi) = \dfrac{1}{\epsilon}\mathcal{A} y(t,\xi), \quad \xi \in \mathcal{D}, \\ \\ \dfrac{\partial y}{\partial \nu} = \epsilon\, \beta(\epsilon)\, \sigma(t,\xi) \dfrac{\partial W^B}{\partial t}(t,\xi), \quad \xi \in \partial\mathcal{D}, \qquad y(0,\xi) = 0, \quad \xi \in \mathcal{D}, \end{cases} \tag{4.2.8}$$

So, upon taking $\delta = \delta_0/\epsilon$, we say that the process

$$\beta(\epsilon)\, w^\epsilon_{A,B}(t) := \beta(\epsilon)\,(\delta_0 - A) \int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0}[\Sigma(s)\, dW^B(s)]$$

is a mild solution to problem (4.2.8). (see [25] for details, and see [36], [56] and [61] for other papers where the same type of equations has been studied). This motivates the following.

Definition 4.2.1. Let $p \geq 1$ and $T > 0$. An adapted process $u_\epsilon \in L^p(\Omega; C([0,T]; H))$ is called a mild solution to problem (4.2.6) if, for any $t \in [0,T]$,

$$u_\epsilon(t) = e^{t\frac{A}{\epsilon}} x + \int_0^t e^{(t-s)\frac{A}{\epsilon}} F(s, u_\epsilon(s)) ds + \alpha(\epsilon)\, w^\epsilon_{A,Q}(u_\epsilon)(t) + \beta(\epsilon)\, w^\epsilon_{A,B}(t),$$

where, for any $u \in L^p(\Omega; C([0,T]; H))$, we define

$$w^\epsilon_{A,Q}(u)(t) := \int_0^t e^{(t-s)\frac{A}{\epsilon}} G(s, u(s)) dW^Q(s).$$

4.2.4 Well-posedness and averaging results

In this section, we recall some important preliminary results from [16].

Lemma (**3.1 of** [**16**]). *Assume Hypotheses 4.2.2 and 4.2.3 hold. Then, for any* $\epsilon > 0$*,* $p \geq 1$ *and* $T > 0$*, the process* $w^{\epsilon}_{A,B}$ *belongs to* $L^p(\Omega; C([0,T]; H))$ *and satisfies*

$$\sup_{\epsilon \in (0,1]} \mathbb{E} \left\| w^{\epsilon}_{A,B} \right\|^p_{C([0,T];H)} < +\infty. \tag{4.2.9}$$

Lemma (**3.3 of** [**16**]). *Assume Hypotheses 4.2.2 and 4.2.3 hold. Then for any* $\epsilon > 0$*,* $p \geq 1$ *and* $T > 0$*, the mapping* $w^{\epsilon}_{A,Q}(\cdot)$ *maps* $L^p(\Omega; C([0,T]; H)$ *into itself and satisfies*

$$\sup_{\epsilon \in (0,1]} \mathbb{E} \left\| w^{\epsilon}_{A,Q}(u) \right\|^p_{C([0,T];H)} \leq c_{T,p} \left(1 + \mathbb{E} \int_0^T \|u(s)\|^p_H \, ds \right). \tag{4.2.10}$$

Moreover, it is Lipschitz continuous and

$$\sup_{\epsilon \in (0,1]} \left\| w^{\epsilon}_{A,Q}(u) - w^{\epsilon}_{A,Q}(v) \right\|_{L^p(\Omega;C([0,T];H))} \leq L_T \|u - v\|_{L^p(\Omega;C([0,T];H))}, \tag{4.2.11}$$

for some constant $L_T > 0$*, independent of* $\epsilon \in (0,1]$*, such that* $L_T \to 0$*, as* $T \to 0$*.*

Theorem (**3.4 of** [**16**]). *Assume Hypotheses 4.2.2 and 4.2.3 hold. Then for any* $\epsilon > 0$*,* $p \geq 1$ *and* $T > 0$ *and for any initial condition* $x \in H$*, equation* (4.2.6) *has a unique adapted mild solution* $u^x_\epsilon \in L^p(\Omega; C([0,T]; H))$*, which satisfies*

$$\sup_{\epsilon \in (0,1]} \mathbb{E} \|u^x_\epsilon\|^p_{C([0,T];H)} \leq c_{T,p} (1 + x^p_H). \tag{4.2.12}$$

Next, for any $t \geq 0$ and $h \in H_\mu$, we define

$$\bar{F}(t,h) := \langle F(t,h), \mu \rangle = \int_{\mathcal{D}} f(t, \xi, h(\xi)) d\mu(\xi).$$

Moreover, for any $t \geq 0$ and $h_1, h_2 \in H_\mu$, we define

$$\bar{G}(t, h_1)h_2 := \langle G(t, h_1)h_2, \mu \rangle = \int_{\mathcal{D}} g(t, \xi, h_1(\xi))h_2(\xi)\, d\mu(\xi), \tag{4.2.13}$$

and for any $t \geq 0$ and $z \in Z$, we define

$$\bar{\Sigma}(t)z = \delta_0 \langle N_{\delta_0} \Sigma(t)z, \mu \rangle = \delta_0 \int_{\mathcal{D}} N_{\delta_0}[\sigma(t, \cdot)z](\xi)\, d\mu(\xi).$$

Hypothesis 4.2.2 implies that $\bar{F}(t, \cdot) : H_\mu \to \mathbb{R}$ is Lipschitz continuous, uniformly with respect to $t \in [0, T]$. Concerning $\bar{G}$, we observe that for any $h \in H_\mu$ and $T > 0$

$$|\bar{G}(t, h_1)h - \bar{G}(t, h_2)h|^2 \leq \|h\|^2_{H_\mu} \int_{\mathcal{D}} |g(t, \xi, h_1(\xi)) - g(t, \xi, h_2(\xi))|^2 d\mu(\xi) \tag{4.2.14}$$

$$\leq c \, \|h\|^2_{H_\mu} \|h_1 - h_2\|^2_{H_\mu}, \qquad h_1, h_2 \in H_\mu, \quad t \in [0, T].$$

Therefore, $\bar{G}(t, \cdot)h : H_\mu \to \mathbb{R}$ is Lipschitz continuous, uniformly with respect to $t \in [0, T]$ and h in a bounded set of H_μ (and hence H). Finally, the linear functional $\bar{\Sigma}(t) : Z \to \mathbb{R}$ is bounded due to (4.2.7).

With these notations, we introduce the equation

$$dv^x(t) = \bar{F}(t, v^x(t))dt + \bar{G}(t, v^x(t))dW^Q(t) + \bar{\Sigma}(t)dW^B(t), \qquad v^x(0) = \langle x, \mu \rangle. \tag{4.2.15}$$

Theorem (4.1 of [16]). *Assume that Hypotheses 4.2.1, 4.2.2 and 4.2.3 hold, and let $\alpha(\epsilon) = \beta(\epsilon) \equiv 1$. Then, for any $x \in H$, $p \geq 1$, and $0 < \delta < T$, we have*

$$\lim_{\epsilon \to 0} \mathbb{E} \sup_{\delta \leq t \leq T} \|v^x_\epsilon(t) - v^x(t)\|^p_{H_\mu} = 0, \tag{4.2.16}$$

where v^x_ϵ is the mild solution to (4.2.6) with $\alpha(\epsilon) = \beta(\epsilon) \equiv 1$ and v^x is the solution of equation (4.2.15).

4.3 Main results and description of the methods

We are here interested in the validity of a large deviation principle for the family $\{\mathcal{L}(u^x_\epsilon)\}_{\epsilon\in(0,1]}$, as $\epsilon \to 0$, where u^x_ϵ is the solution to the equation (4.2.6) with initial condition $x \in H$.

In [16], equation (4.2.6) was studied with $\alpha(\epsilon) = \beta(\epsilon) \equiv 1$, and it was shown that for every $\delta > 0$, the solutions converge in $L^p(\Omega; C([\delta, T]; H_\mu))$ to the solution of the one-dimensional stochastic differential equation (4.2.15). Therefore, if

$$\lim_{\epsilon\to 0} \alpha(\epsilon) = \lim_{\epsilon\to 0} \beta(\epsilon) = 0,$$

thanks to the bounds (4.2.9), (4.2.10) and (4.2.12), the solution u^x_ϵ will converge in the space $L^p(\Omega; C([\delta, T]; H_\mu))$ to the solution of the deterministic one-dimensional differential equation,

$$\frac{du}{dt} = \bar{F}(t, u(t)), \qquad u(0) = \langle x, \mu\rangle. \tag{4.3.1}$$

In what follows, we shall assume that the following conditions are satisfied

Hypothesis 4.3.1. (i) We have

$$\lim_{\epsilon\to 0} \alpha(\epsilon) = \lim_{\epsilon\to 0} \beta(\epsilon) = 0, \qquad \lim_{\epsilon\to 0} \frac{\beta(\epsilon)}{\alpha(\epsilon)} = \bar{\rho} \in [0, +\infty]. \tag{4.3.2}$$

(ii) For every $t \geq 0$, $w \in \mathbb{R}$ and $\rho \in [0, +\infty]$, we define

$$\mathcal{H}_\rho(t, w) = \frac{1}{(1+\rho)^2} \left[\left\| \sqrt{Q}\, [G(t, w)m] \right\|^2_H + \rho^2 \left\| \delta_0 \sqrt{B} \left[\Sigma(t) N^*_{\delta_0} m \right] \right\|^2_Z \right], \tag{4.3.3}$$

where m is the density of the invariant measure μ. Then, if $\bar{\rho}$ is the constant introduced in (4.3.2), we have

$$\inf_{(t,w)\in[0,\infty)\times\mathbb{R}} \mathcal{H}_{\bar{\rho}}(t,w) > 0. \tag{4.3.4}$$

Now, for every $0 \leq \delta < T$, we denote by $\Psi_{\delta,T}$ the subset of $C([\delta,T];H_\mu)$ containing all functions $u \in C([\delta,T];H_\mu)$ that are absolutely continuous in t and are constant in the spatial variable ξ. Then, if $u \in \Psi_{\delta,T}$, we define

$$I^x_{\delta,T}(u) = \inf_{\substack{w\in C([0,T];\mathbb{R}) \\ w(0)=\langle x,\mu\rangle,\ w|_{[\delta,T]}=u}} \frac{1}{2}\int_0^T \frac{|w'(t) - \bar{F}(t,w(t))|^2}{\mathcal{H}_{\bar{\rho}}(t,w(t))}\, dt. \tag{4.3.5}$$

For any other $u \in C([\delta,T];H_\mu)$, we set $I^x_{\delta,T}(u) = +\infty$.

We will show that, in fact, the laws of the family $\{u^x_\epsilon(t)\}_{\epsilon\in(0,1]}$ satisfy a large deviation principle in the space $C([\delta,T];H_\mu)$, with respect to the action functional $I^x_{\delta,T}$.

Theorem 4.3.1. *Assume that all Hypotheses 4.2.1 to 4.3.1 are satisfied. Fix any $T > 0$ and $0 < \delta < T$ and let u^x_ϵ denote the solution to equation (4.2.6), with initial condition $x \in H$. Moreover, let us define*

$$\gamma(\epsilon) =: (\alpha(\epsilon) + \beta(\epsilon))^2, \qquad \epsilon > 0. \tag{4.3.6}$$

Then, the following facts hold.

(i) The family $\{\mathcal{L}(u^x_\epsilon)\}_{\epsilon\in(0,1]}$ satisfies both a Freidlin-Wentzell large deviations principle and Dembo-Zeitouni large deviations principle in $C([\delta,T];H_\mu)$ with speed

$\gamma(\epsilon)$ and good rate functions functional $I^x_{\delta,T}$, uniformly for x in any closed, bounded subset of H.

(ii) If in addition x is constant, then $\{\mathcal{L}(u^x_\epsilon)\}_{\epsilon\in(0,1]}$ satisfies both a Freidlin-Wentzell large deviations principle and Dembo-Zeitouni large deviations principle in $C([0,T];H_\mu)$ with speed $\gamma(\epsilon)$ and good rate functions $I^x_{0,T}$, uniformly for x in any closed, bounded subset of H.

To prove Theorem 4.3.1, we follow the weak convergence approach, and instead prove the validity of a uniform Laplace principle, as described in Theorem 4.3.2. This, combined with Propositions 2.2.3 and 2.2.4 will yield both versions of the uniform large deviations principle, with same rate and same action functional.

In view of the formalisms introduced in Section 2.2.4, we first introduce some notation. We denote by V the product Hilbert space $H \times Z$, endowed with the inner product

$$\langle v_1, v_2\rangle_V := \langle h_1, h_2\rangle_H + \langle z_1, z_2\rangle_Z,$$

for every $v_1 = (h_1, z_1),\ v_2 = (h_2, z_2) \in V$. Next, we define the linear operator,

$$Sv = (Qh, Bz), \qquad v = (h, z) \in V.$$

Notice that $S \in \mathcal{L}^+(V)$ and the process $w^S(t) := (W^Q(t), W^B(t)),\ t \geq 0$ is an S-Wiener process. Next, we let $\mathcal{P}(V)$ be the set of predictable processes in $L^2(\Omega \times [0,T];V)$. For every fixed $M > 0$, we define

$$S^M(V) := \left\{ u \in L^2(0,T;V) : \int_0^T \|u(s)\|_V^2 \, ds \leq M \right\},$$

and

$$\mathcal{P}^M(V) := \left\{\varphi \in \mathcal{P}(V) : \varphi \in S^M(V),\ \mathbb{P}-a.s.\right\}$$

For any $\varphi(t) = (\varphi_H(t), \varphi_Z(t)) \in \mathcal{P}^M(V)$, we denote by $u_\epsilon^{x,\varphi}$ the unique mild solution of the controlled stochastic PDE

$$\begin{cases} \dfrac{\partial u}{\partial t}(t,\xi) = \epsilon^{-1}\,\mathcal{A}u(t,\xi) + f(t,\xi,u(t,\xi)) \\ \qquad\qquad + \dfrac{\alpha(\epsilon)}{\sqrt{\gamma(\epsilon)}}\, g(t,\xi,u(t,\xi)) \left[\sqrt{Q}\varphi_H(t,\xi) + \sqrt{\gamma(\epsilon)}\,\dfrac{\partial W^Q}{\partial t}(t,\xi)\right], & \xi \in \mathcal{D}, \\ \dfrac{\partial u}{\partial \nu}(t,\xi) = \epsilon\,\beta(\epsilon)\sqrt{\gamma(\epsilon)}\,\sigma(t,\xi)\left[\sqrt{B}\varphi_Z(t,\xi) + \sqrt{\gamma(\epsilon)}\,\dfrac{\partial W^B}{\partial t}(t,\xi)\right], & \xi \in \partial\mathcal{D}, \\ u(0,\xi) = x(\xi), & \xi \in \mathcal{D}, \end{cases} \tag{4.3.7}$$

where $\gamma(\epsilon)$ is the function defined in (4.3.6). Moreover, we denote by $u^{x,\varphi}$ the unique solution of the random ODE

$$\frac{du}{dt} = \bar{F}(t,u(t)) + \frac{1}{1+\bar{\rho}}\,\bar{G}(t,u(t))\left[\sqrt{Q}\varphi_H(t)\right] + \frac{\bar{\rho}}{1+\bar{\rho}}\,\bar{\Sigma}(t)\left[\sqrt{B}\varphi_Z(t)\right], \quad u(0) = \langle x, \mu\rangle. \tag{4.3.8}$$

We will prove the well-posedness of both of these equations in the next section. In what follows, we denote

$$\mathcal{G}_\delta(x,\varphi) := u^{x,\varphi}|_{[\delta,T]}.$$

Theorem 4.3.2. *For any $x \in H$, $0 < \delta < T$ and $u \in C([\delta,T];H_\mu)$, let*

$$\hat{I}^x_{\delta,T}(u) := \inf_{\substack{\varphi \in L^2(0,T;V) \\ \mathcal{G}_\delta(x,\varphi) = u}} \frac{1}{2}\int_0^T \|\varphi(s)\|_V^2\, ds. \tag{4.3.9}$$

Suppose that the following conditions hold.

(i) If $B \subset H$ is a closed and bounded set, then for every $M < \infty$ the set

$$F_{M,B} := \left\{u \in C([\delta, T]; \mathbb{R}) : u = \mathcal{G}_\delta(x, \varphi),\ \varphi \in S^M(V),\ x \in B\right\},$$

is compact in $C([\delta, T]; H_\mu)$.

(ii) If $\{\varphi_\epsilon\}_{\epsilon>0} \subset \mathcal{P}^M(V)$ is any family that converges in distribution, as $\epsilon \to 0$, to some $\varphi \in \mathcal{P}^M(V)$ with respect to the weak topology of $L^2(0, T; V)$, and if $\{x_\epsilon\}_{\epsilon>0} \subset H$ is any family that converges weakly in H, as $\epsilon \to 0$, to some $x \in H$, then the family $\{u_\epsilon^{x_\epsilon, \varphi_\epsilon}\}_{\epsilon>0}$ converges in distribution, as $\epsilon \to 0$, to $u^{x,\varphi}$ in the space $C([\delta, T]; H_\mu)$, endowed with the strong topology.

(iii) For every $u \in C([\delta, T; H_\mu])$, the mapping $x \mapsto \hat{I}^x_{\delta,T}(u)$ is weakly lower semicontinuous from H into $[0, +\infty]$.

Then the family $\{\mathcal{L}(u_\epsilon^x)\}_{\epsilon>0}$ satisfies a Laplace principle in $C([\delta, T]; H_\mu)$, with speed $\gamma(\epsilon)$ and action functional $\hat{I}^x_{\delta,T}$, uniformly in x on any closed, bounded subset of H. Moreover, for any closed bounded set $B \subset H$ and any $s \geq 0$, the set

$$\Lambda_{s,B} := \bigcup_{x\in B} \left\{u \in C([\delta, T]; H_\mu) : \hat{I}^x_{\delta,T}(u) \leq s\right\}$$

is compact in $C([\delta, T]; H_\mu)$.

Remark 4.3.1. In the theorem above we allow the uniformity of the Laplace principle for initial conditions x in closed and bounded sets $B \subset H$ (rather than compact sets). As noted in Remark 2.2.4, this is possible by simply changing the topology of H to the weak topology. Specifically, we require that the mapping $x \mapsto \hat{I}_{\delta,T}(u)$ is weakly lower semicontinuous and that (2.2.13) must hold for any sequence $\{x_\epsilon\}_{\epsilon>0}$ converging weakly to x. The reason *why* we will be able to prove this stronger form of condition is because the limiting equation (4.3.1) is finite dimensional. If, for example, the averaging were only to occur in some but not all of the coordinates, then we would not have this property and the stronger version of Hausdorff continuity would not be met.

$\square$

4.4 Proof of Theorem 4.3.1

4.4.1 Well-Posedness of the skeleton equations

Proposition 4.4.1. *Assume that Hypotheses 4.2.2 and 4.2.3 hold, and fix any $\epsilon, M > 0$ and $p \geq 1$. Then for any $\varphi = (\varphi_H, \varphi_Z) \in \mathcal{P}^M(V)$ and $x \in H$, equation (4.3.7) has a unique adapted mild solution, $u_\epsilon^{x,\varphi} \in L^p(\Omega; C([0,T]; H))$. Furthermore, if (4.3.2) holds, we have*

$$\sup_{\epsilon\in(0,1]} \mathbb{E} \sup_{0\leq t\leq T} \|u_\epsilon^{x,\varphi}(t)\|_H^p \leq c_{p,T,M} \left(1 + \|x\|_H^p\right). \tag{4.4.1}$$

Proof. The well-posedness of equation (4.3.7) follows from a fixed point argument in the space of adapted processes in $L^p(\Omega; C([0,T]; H))$. For $u \in L^p(\Omega; C([0,T]; H))$, we define

$$\mathcal{K}_\epsilon u(t) := e^{t\frac{A}{\epsilon}}x + \int_0^t e^{(t-s)\frac{A}{\epsilon}} F(s,u(s))ds + \alpha(\epsilon)\, w^\epsilon_{A,Q}(u)(t) + \beta(\epsilon)\, w^\epsilon_{A,B}(t)$$

$$+\frac{\alpha(\epsilon)}{\sqrt{\gamma(\epsilon)}} \int_0^t e^{(t-s)\frac{A}{\epsilon}} G(s,u(s)) \left[\sqrt{Q}\varphi_H(s)\right] ds$$

$$+\frac{\beta(\epsilon)}{\sqrt{\gamma(\epsilon)}} (\delta_0 - A) \int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0} \left[\Sigma(s)\sqrt{B}\varphi_Z(s)\right] ds.$$

We show that $\mathcal{K}_\epsilon$ is Lipschitz continuous from $L^p(\Omega; C([0,T];H))$ into itself, with Lipschitz constant going to 0 as $T \to 0$. This clearly implies the well-posedness of equation (4.3.7) in $L^p(\Omega; C([0,T];H))$.

Thanks to (4.2.9), (4.2.10) and (4.2.11), since $F(t,\cdot) : H \to H$ is Lipschitz-continuous, it suffices to show that the mapping Γ_ϵ, defined by

$$\Gamma_\epsilon(u)(t) = \frac{\alpha(\epsilon)}{\sqrt{\gamma(\epsilon)}} \int_0^t e^{(t-s)\frac{A}{\epsilon}} G(s,u(s))\sqrt{Q}\varphi_H(s)ds$$

$$+\frac{\beta(\epsilon)}{\sqrt{\gamma(\epsilon)}} (\delta_0 - A) \int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0} \left[\Sigma(s)\sqrt{B}\varphi_Z(s)\right] ds,$$

maps $L^p(\Omega; C([0,T];H))$ into itself and is Lipschitz continuous with Lipschitz constant going to 0 as $T \to 0$. To this purpose, we define $\zeta = \frac{2\rho}{\rho-2}$, where $\rho < \frac{2d}{d-2}$ satisfies

(4.2.5). Since $\zeta < d$ thanks to (4.6.1), for any $u, v \in L^p(\Omega; C([0,T]; H))$, we have

$$
\begin{aligned}
&\sup_{0\le t\le T} \left\| \int_0^t e^{(t-s)\frac{A}{\epsilon}} \left(G(s,u(s)) - G(s,v(s))\right) \left[\sqrt{Q}\varphi_H(s)\right] ds \right\|_H^p \\
&\le c \sup_{0\le t\le T} \left(\int_0^t \left(((t-s)/\epsilon)^{-\frac{d}{2\zeta}} + 1 \right) \|u(s) - v(s)\|_H \, \|\varphi_H(s)\|_H \, ds \right)^p \\
&\le c_p \left(T^{\frac{p}{2}} + T^{(1-\frac{d}{\zeta})\frac{p}{2}}\right) \left[\left(\int_0^T \|\varphi_H(s)\|_H^2 \, ds \right)^{\frac{p}{2}} \sup_{0\le t\le T} \|u(s) - v(s)\|_H^p \right] \\
&\le c_{p,M,T} \sup_{0\le t\le T} \|u(s) - v(s)\|_H^p ,
\end{aligned}
\tag{4.4.2}
$$

where, in the last step, we used the fact that $\|\varphi_H\|^2_{L^2(0,T;H)} \le \|\varphi\|^2_{L^2(0,T;V)} \le M$, $\mathbb{P}$-a.s..

To conclude our proof of the well-posedness, we show that the second term in Γ_ϵ is in $L^p(\Omega; C([0,T]; H))$. Due to (4.2.7) and (4.2.2), the operator

$$
S_\rho := (\delta_0 - A)^{\frac{3-\rho}{4}} N_{\delta_0} \tag{4.4.3}
$$

belongs to $\mathcal{L}(Z, H)$, for any $\rho > 0$. Therefore, for any $t > 0$, we have

$$
(\delta_0 - A) e^{tA} N_{\delta_0} = e^{\frac{t}{2}A} (\delta_0 - A)^{\frac{1+\rho}{4}} e^{\frac{t}{2}A} S_\rho, \tag{4.4.4}
$$

and

$$\left\|(\delta_0 - A)\int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi_Z(s)]ds\right\|_H$$

$$\leq \int_0^t \left\|e^{(t-s)\frac{A}{2\epsilon}}(\delta_0 - A)^{\frac{1+\rho}{4}} e^{(t-s)\frac{A}{2\epsilon}} S_\rho\left[\Sigma(s)\sqrt{B}\varphi_Z(s)\right]\right\|_H ds$$

$$\leq c\int_0^t \left[1 + (t-s)^{-\frac{1+\rho}{4}}\right]\|\varphi_Z(s)\|_Z\, ds.$$

Thus, by taking the p-th moment and choosing $\rho < 1$, we get

$$\sup_{0\leq t\leq T}\left\|(\delta_0 - A)\int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi_Z(s)]ds\right\|_H^p$$

$$\leq c\sup_{0\leq t\leq T}\left(\int_0^t [1 + s^{-\frac{1+\rho}{2}}]ds\right)^{\frac{p}{2}}\left(\int_0^t \|\varphi_Z(s)\|_Z^2\, ds\right)^{\frac{p}{2}} \leq c_p\, M^{\frac{p}{2}}(T^{\frac{p}{2}} + T^{\frac{p(1-\rho)}{4}}). \tag{4.4.5}$$

Next, we prove that estimate (4.4.1) holds. To this purpose, we first remark that due to (4.3.2)

$$\lim_{\epsilon\to 0}\frac{\alpha(\epsilon)}{\sqrt{\gamma(\epsilon)}} = \frac{1}{1+\bar{\rho}} \in [0,1], \quad \lim_{\epsilon\to 0}\frac{\beta(\epsilon)}{\sqrt{\gamma(\epsilon)}} = \frac{\bar{\rho}}{1+\bar{\rho}} \in [0,1]. \tag{4.4.6}$$

In particular, both $\alpha(\epsilon)/\sqrt{\gamma(\epsilon)}$ and $\beta/\sqrt{\gamma(\epsilon)}$ remain uniformly bounded, with respect to $\epsilon \in (0,1]$.

Thus, thanks to (4.4.6), by proceeding as in the proofs of (4.4.2) and (4.4.5), due to (4.2.9) and (4.2.10), we have

$$\mathbb{E}\sup_{0\leq t\leq T}\|u_\epsilon^{x,\varphi}(t)\|_H^p \leq c_p\,(1 + \|x\|_H^p) + c_{p,M,T}\left(1 + \mathbb{E}\sup_{0\leq t\leq T}\|u_\epsilon^{x,\varphi}(t)\|_H^p\right), \quad \epsilon \in (0,1],$$

for some constant $c_{p,M,T} > 0$ such that $c_{p,M,T} \to 0$, as $T \to 0$. This means that there exists $T_0 > 0$ such that

$$\sup_{\epsilon \in (0,1]} \mathbb{E} \sup_{0 \le t \le T_0} \|u_\epsilon^{x,\varphi}(t)\|_H^p \le 2\, c_p\, (1 + \|x\|_H^p)\,.$$

By a bootstrap argument, this yields (4.4.1).

□

Proposition 4.4.2. *Assume that Hypothesis 4.2.2 hold, and fix any $M > 0$. Then, for any $\varphi = (\varphi_H, \varphi_Z) \in \mathcal{P}^M(V)$, the random differential equation (4.3.8) has a unique adapted solution, $u^{x,\varphi} \in L^p(\Omega; C([0,T];\mathbb{R}))$, for any $T > 0$ and $p \ge 1$.*

Proof. As before, existence and uniqueness follows from the Lipschitz continuity of the mapping

$$\mathcal{K} : L^p(\Omega; C([0,T];\mathbb{R})) \to L^p(\Omega; C([0,T];\mathbb{R})),$$

defined by

$$\mathcal{K}u(t) = \langle x, \mu \rangle + \int_0^t \bar{F}(s, u(s))ds$$

$$+\frac{1}{1+\bar{\rho}} \int_0^t \bar{G}(s, u(s)) \left[\sqrt{Q}\varphi_H(s)\right] ds + \frac{\bar{\rho}}{1+\bar{\rho}} \int_0^t \bar{\Sigma}(t) \left[\sqrt{B}\varphi_Z(s)\right] ds.$$

Let $u, v \in L^p(\Omega; C([0,T];\mathbb{R}))$. Due to (4.2.14) and the fact that $\varphi \in \mathcal{P}^M(V)$, we have

$$\mathbb{E} \sup_{0 \le t \le T} \left| \int_0^t \left(\bar{G}(s, u(s)) - \bar{G}(s, v(s))\right) \left[\sqrt{Q}\varphi_H(s)\right] ds \right|^p$$

$$\le c\,\mathbb{E} \sup_{0 \le t \le T} \left(\int_0^t \|\varphi_H(s)\|_H \,|u(s) - v(s)|\, ds \right)^p \le c_p\, T^{\frac{p}{2}} M^{\frac{p}{2}} \mathbb{E} \sup_{0 \le t \le T} |u(t) - v(t)|^p.$$

Moreover, due to (4.2.7), we easily have

$$\mathbb{E}\sup_{0\leq t\leq T}\left|\int_0^t \bar{\Sigma}(t)\left[\sqrt{B}\varphi_Z(s)\right]ds\right|^p = \mathbb{E}\sup_{0\leq t\leq T}\left|\delta_0\int_0^t \langle N_{\delta_0}\left(\Sigma(s)[\sqrt{B}\varphi_Z(s)]\right), \mu\rangle ds\right|^p \leq c\,T^{\frac{p}{2}}M^{\frac{p}{2}}.$$

Since $\bar{F}(t,\cdot):\mathbb{R}\to\mathbb{R}$ is Lipschitz continuous, we conclude that $\mathcal{K}$ is Lipschitz continuous from $L^p(\Omega; C([0,T];\mathbb{R})$ into itself, and the well-posedness of equation (4.3.8) follows. □

4.4.2 Convergence

Clearly, in Theorem 4.3.2 Condition (i) follows immediately from Condition (ii). On the other hand, due to Skorokhod's theorem, Condition (ii) in Theorem 4.3.2 follows from the following convergence result.

Proposition 4.4.3. *Assume that Hypotheses 4.2.1, 4.2.2 and 4.2.3 hold. Moreover, assume that* (4.3.2) *holds. Suppose that* $\{x_\epsilon\}_{\epsilon>0}\subset H$ *converges weakly to* $x\in H$*, as* $\epsilon\to 0$*, and suppose that* $\{\varphi_\epsilon\}_{\epsilon>0}\subset\mathcal{P}^M(V)$ *converges weakly in* $L^2(0,T;V)$ *to* φ*, as* $\epsilon\to 0$*,* $\mathbb{P}$*-a.s. Then for any* $\delta>0$ *and* $p\geq 1$*,*

$$\lim_{\epsilon\to 0}\mathbb{E}\sup_{\delta\leq t\leq T}\|u_\epsilon^{x_\epsilon,\varphi_\epsilon}-u^{x,\varphi}\|^p_{H_\mu}=0. \tag{4.4.7}$$

Proof. We denote $\varphi_\epsilon = (\varphi^\epsilon_H, \varphi^\epsilon_Z)$ and $\varphi = (\varphi_H, \varphi_Z)$. We first write

$$u^{x_\epsilon,\varphi_\epsilon}_\epsilon(t) - u^{x,\varphi}(t) = \left(e^{t\frac{A}{\epsilon}} x_\epsilon - \langle x, \mu\rangle\right) + \alpha(\epsilon)\, w^\epsilon_{A,Q}(u^{x_\epsilon,\varphi_\epsilon}_\epsilon)(t) + \beta(\epsilon)\, w^\epsilon_{A,B}(t)$$

$$+ \left(\int_0^t e^{(t-s)\frac{A}{\epsilon}} F(s, u^{x_\epsilon,\varphi_\epsilon}_\epsilon(s))ds - \int_0^t \bar{F}(s, u^{x,\varphi}(s))ds\right)$$

$$+\frac{\alpha(\epsilon)}{\sqrt{\gamma(\epsilon)}} \int_0^t e^{(t-s)\frac{A}{\epsilon}} G(s, u^{x_\epsilon,\varphi_\epsilon}_\epsilon(s)) \left[\sqrt{Q}\varphi^\epsilon_H(s)\right] ds - \frac{1}{1+\bar{\rho}} \int_0^t \bar{G}(s, u^{x,\varphi}(s)) \left[\sqrt{Q}\varphi_H(s)\right] ds$$

$$+\frac{\beta(\epsilon)}{\sqrt{\gamma(\epsilon)}}(\delta_0 - A) \int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0} \left[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)\right] ds - \frac{\bar{\rho}}{1+\bar{\rho}} \int_0^t \bar{\Sigma}(s) \left[\sqrt{B}\varphi_Z(s)\right] ds$$

$$=: \left(e^{t\frac{A}{\epsilon}} x_\epsilon - \langle x, \mu\rangle\right) + \alpha(\epsilon)\, w^\epsilon_{A,Q}(u^{x_\epsilon,\varphi_\epsilon}_\epsilon)(t) + \beta(\epsilon)\, w^\epsilon_{A,B}(t) + \sum_{i=1}^3 I^i_\epsilon(t).$$

Thanks to estimates (4.2.9), (4.2.10) and (4.4.1), as well as Lemmas 4.6.2, 4.6.3, and 4.6.4 (where a-priori bounds for the terms $I^i_\epsilon(t)$ are proven), there exists some non-negative function $r_{T,p}(\epsilon)$ going to 0, as $\epsilon \to 0$, such that

$$\mathbb{E} \sup_{\delta\le t\le T} \|u^{x_\epsilon,\varphi_\epsilon}_\epsilon(t) - u^{x,\varphi}(t)\|^2_{H_\mu}$$

$$\le \sup_{\delta\le t\le T} \left\|e^{t\frac{A}{\epsilon}} x_\epsilon - \langle x, \mu\rangle\right\|^2_{H_\mu} + r_{T,p}(\epsilon) + c_T \int_\delta^T \mathbb{E} \sup_{\delta\le s\le t} \|u^{x_\epsilon,\varphi_\epsilon}_\epsilon(s) - u^{x,\varphi}(s)\|^2_{H_\mu}. \tag{4.4.8}$$

We have

$$\sup_{\delta\le t\le T}\left\|e^{t\frac{A}{\epsilon}}x_\epsilon-\langle x,\mu\rangle\right\|_{H_\mu}^2\le 2\sup_{\delta\le t\le T}\left\|e^{t\frac{A}{\epsilon}}x_\epsilon-\langle x_\epsilon,\mu\rangle\right\|_{H_\mu}^2+2\big|\langle x_\epsilon-x,\mu\rangle\big|^2$$

$$\le 2\,c\,e^{-\frac{2\gamma\delta}{\epsilon}}\,\|x_\epsilon\|_{H_\mu}+2\big|\langle x_\epsilon-x,\mu\rangle\big|^2.$$

Therefore, since the sequence $\{x_\epsilon\}_{\epsilon>0}$ converges weakly to x in H_μ, we have

$$\lim_{\epsilon\to 0}\sup_{\delta\le t\le T}\left\|e^{t\frac{A}{\epsilon}}x_\epsilon-\langle x,\mu\rangle\right\|_{H_\mu}^2=0.$$

This fact, together with (4.4.8) and Gronwall's Lemma, allows us to conclude that (4.4.7) holds for $p\ge 2$. To obtain the result for $p>2$, we use estimate (4.4.1) and the dominated convergence theorem. □

In the next section we show that for every $u\in C([\delta,T];H_\mu)$ the mapping $x\in H\mapsto \hat{I}^x_{\delta,T}(u)\in[0,+\infty]$ is weakly lower semicontinuous. Due to the convergence result proved in Proposition 4.4.3 and to Theorem 4.3.2, this implies that the family $\{u^x_\epsilon\}_{\epsilon>0}$ satisfies a uniform Laplace principle in $C([\delta,T];H_\mu)$, with speed $\gamma(\epsilon)$ and action functional $\hat{I}^x_{\delta,T}$.

4.4.3 Conclusion

In this section, we first show that under Hypothesis 4.3.1, we have $\hat{I}^x_{\delta,T}=I^x_{\delta,T}$ where $I^x_{\delta,T}$ is the action functional defined in (4.3.1). Then, we show that the action functionals $I^x_{\delta,T}$ satisfy the properties required to extend the uniform Laplace principle into a uniform large deviation principle (see Proposition 2.2.4). In particular, the mapping

$x \mapsto \hat{I}^x_{\delta,T}(u) \in [0,+\infty]$ is weakly lower semicontinuous, for every $u \in C([0,T];H_\mu)$. This will conclude the proof of Theorem 4.3.1.

Lemma 4.4.1. *For every $0 \leq a < b$, let us define*

$$I_{a,b}(w) := \frac{1}{2}\int_a^b \frac{\left|w'(t) - \bar{F}(t,w(t))\right|^2}{\mathcal{H}_{\bar{\rho}}(t,w(t))}\,dt.$$

Then, we have

$$\hat{I}^x_{\delta,T}(u) = I^x_{\delta,T}(u) = \inf_{\substack{w\in C([0,T];\mathbb{R})\\ w(0)=\langle x,\mu\rangle,\ w|_{[\delta,T]}=u}} I_{0,T}(w). \tag{4.4.9}$$

Proof. First, we observe that $\hat{I}^x_{\delta,T}(u) = \infty$, if $u(t,\xi)$ is any function depending on the spatial variable ξ. Next, we notice that $\hat{I}^x_{\delta,T}(u)$ can be rewritten as

$$\hat{I}^x_{\delta,T}(u) = \inf_{\substack{w\in C([0,T];\mathbb{R})\\ w|_{[\delta,T]}=u}}\ \inf_{\substack{\varphi\in L^2(0,T;V)\\ u^{x,\varphi}=w}} \frac{1}{2}\int_0^T \|\varphi(s)\|_V^2\,ds,$$

because the condition $\mathcal{G}_\delta(x,\varphi) = u$ does not constrain the values of φ on the interval $(0,\delta)$. We suppose now that $w = u^{x,\varphi}$, for some $x \in H$ and $\varphi = (\varphi_H,\varphi_Z) \in$

$L^2(0,T;V)$. Then, recalling that μ has a density $m \in L^\infty(\mathcal{D})$, we have

$$|w'(t) - \bar{F}(t,w(t))| = \frac{1}{1+\bar\rho}\left|\bar{G}(t,w(t))\left[\sqrt{Q}\varphi_H(t)\right] + \bar\rho\,\bar\Sigma(t)\left[\sqrt{B}\varphi_Z(t)\right]\right|$$

$$= \frac{1}{1+\bar\rho}\left|\langle\varphi_H(t), \sqrt{Q}\left[G(t,w(t))m\right]\rangle_H + \bar\rho\,\delta_0\langle\varphi_Z(t), \sqrt{B}\left[\Sigma(t)N^*_{\delta_0}m\right]\rangle_Z\right|$$

$$\leq \frac{1}{1+\bar\rho}\,\|\varphi_H(t)\|_H\left(\left\|\sqrt{Q}\left[G(t,w(t))m\right]\right\|_H + \bar\rho\,\delta_0\,\|\varphi_Z(t)\|_Z\left\|\sqrt{B}\left[\Sigma(t)N^*_{\delta_0}m\right]\right\|_Z\right)$$

$$\leq \frac{1}{1+\bar\rho}\,\|\varphi(t)\|_V\left(\left\|\sqrt{Q}\left[G(t,w(t))m\right]\right\|_H^2 + \bar\rho^2\,\delta_0^2\left\|\sqrt{B}\left[\Sigma(t)N^*_{\delta_0}m\right]\right\|_Z^2\right)^{1/2}$$

$$= \|\varphi(t)\|_V\,\sqrt{\mathcal{H}_{\bar\rho}(t,w(t))}.$$

On the other hand, equality is achieved with the choice,

$$\hat\varphi(t) := \frac{1}{1+\bar\rho}\,\frac{w'(t)-\bar{F}(w(t))}{\mathcal{H}_{\bar\rho}(t,w(t))}\left(\sqrt{Q}\left[G(t,w(t))m\right], \bar\rho\,\delta_0\sqrt{B}\left[\Sigma(t)N^*_{\delta_0}m\right]\right).$$

Notice that $\hat\varphi$ is well defined due to the non-degeneracy condition in Hypothesis 4.3.1. Moreover, it is easy to see that w solves equation (4.3.8) with the control $\hat\varphi$, so that $u^{x,\hat\varphi} = w$. This minimizing choice of $\varphi = \hat\varphi$ gives rise to the action functional $I^x_{\delta,T}$. $\square$

Alternatively, we can write the action functional as

$$I^x_{\delta,T}(u) = \inf_{\substack{w\in C([0,\delta];\mathbb{R})\\ w(0)=\langle x,\mu\rangle,\ w(\delta)=u(\delta)}} I_{0,\delta}(w) + I_{\delta,T}(u) =: J_\delta(x,u) + I_{\delta,T}(u) \tag{4.4.10}$$

$J_\delta(x,u)$ depends only on the initial condition $x \in H$ and the value of the path u at $t=\delta$, while $I_{\delta,T}(u)$ only depends on the path u.

Lemma 4.4.2. *Assume H is endowed with the weak topology. Then, for every $0 < \delta < T$, the mapping $J_\delta : H \times C([0,T];\mathbb{R}) \to [0,+\infty)$ is continuous.*

Proof. For every $x \in H$, $u \in C([0,T];\mathbb{R})$ and $\eta > 0$, we denote by $w_\eta(x,u)$ a path in $C([0,T];\mathbb{R})$ such that

$$w_\eta(x,u)(0) = \langle x, \mu\rangle, \qquad w_\eta(x,u)(\delta) = u(\delta), \qquad J_\delta(x,u) \geq I_{0,\delta}(w_\eta(x,u)) - \frac{\eta}{4}.$$

Moreover, for every $y \in H$, $v, w \in C([0,T];\mathbb{R})$ and $\delta' \in (0,\delta)$, we denote by $\rho_{\delta'}(y,v,w)$ the path in $C([0,T];\mathbb{R})$ defined by

$$\rho_{\delta'}(y,v,w)(t) = \begin{cases} (\delta' - t)/\delta' \,\langle y, \mu\rangle + t/\delta' w(\delta'), & t \in [0,\delta'], \\ w(t), & t \in [\delta', \delta - \delta'] \\ (\delta - t)/\delta' \, w(\delta - \delta') + (t - (\delta - \delta'))/\delta' \, v(\delta), & t \in [\delta - \delta', \delta]. \end{cases}$$

Since $\rho_{\delta'}(y,v,w)$ and w coincide in the interval $[\delta', \delta - \delta']$, we have

$$\begin{aligned} &\left| I_{0,\delta}(\rho_{\delta'}(y,v,w)) - I_{0,\delta}(w)\right| \\ &\leq \left|I_{0,\delta'}(\rho_{\delta'}(y,v,w))\right| + \left|I_{\delta-\delta',\delta'}(\rho_{\delta'}(y,v,w))\right| + |I_{0,\delta'}(w)| + |I_{\delta-\delta',\delta}(w)|. \end{aligned} \tag{4.4.11}$$

Now, let us fix $x \in H$, $u \in C([0,T];\mathbb{R})$ and $\eta > 0$. Let $\{x_n\}_{n\geq 1} \subset H$ be a sequence weakly convergent to x and let $\{u_n\} \subset C([0.T];\mathbb{R})$ be a sequence convergent to u. For every $n \in \mathbb{N}$ and $\delta' \in (0,\delta)$, we have

$$\begin{aligned} &J_\delta(x_n, u_n) \leq I_{0,\delta}(\rho_{\delta'}(x_n, u_n, w_\eta(x,u))) \\ &\leq I_{0,\delta}(\rho_{\delta'}(x_n, u_n, w_\eta(x,u))) - I_{0,\delta}(w_\eta(x,u)) + J_\delta(x,u) + \eta/4. \end{aligned} \tag{4.4.12}$$

Since the sequences $\{x_n\}_{n\geq 1}$ and $\{u_n\}_{n\geq 1}$ are bounded and Hypothesis 4.3.1 holds true, we have

$$|I_{0,\delta'}(\rho_{\delta'}(x_n,u_n,w_\eta(x,u)))| \leq c\int_0^{\delta'}\left(\frac{|w_\eta(x,u)(\delta')-\langle x_n,\mu\rangle_H|}{\delta'}+1\right)^2 dt$$

$$\leq c\left(\frac{|w_\eta(x,u)(\delta')-w_\eta(x,u)(0)|^2}{\delta'}+\frac{|\langle x_n-x,\mu\rangle|^2}{\delta'}+\delta'\right).$$

Analogously,

$$|I_{\delta-\delta',\delta}(\rho_{\delta'}(x_n,u_n,w_\eta(x,u)))| \leq c\int_{\delta-\delta'}^{\delta}\left(\frac{|w_\eta(x,u)(\delta')-u_n(\delta)|}{\delta'}+1\right)^2 dt$$

$$\leq c\left(\frac{|w_\eta(x,u)(\delta-\delta')-w_\eta(x,u)(\delta)|^2}{\delta'}+\frac{|u_n(\delta)-u(\delta)|^2}{\delta'}+\delta'\right).$$

Therefore, as $w_\eta(x,u)\in W^{1,2}(0,\delta)$, we can find $\delta_1'>0$ such that

$$|I_{0,\delta'}(\rho_{\delta'}(x_n,u_n,w_\eta(x,u)))|+|I_{\delta-\delta',\delta}(\rho_{\delta'}(x_n,u_n,w_\eta(x,u)))|\leq \eta/4, \qquad \delta'\leq\delta_1'.$$

Moreover, as $I_{0,\delta}(w_\eta(x,u))<\infty$, we can find $\delta_2'>0$ such that

$$|I_{0,\delta'}(w_\eta(x,u))|+|I_{\delta-\delta',\delta}(w_\eta(x,u))|\leq\eta/4, \qquad \delta'\leq\delta_2'.$$

Thus, if we pick $\bar{\delta}'=\min(\delta_1',\delta_2')$, thanks to (4.4.11) we conclude that

$$\limsup_{n\to\infty}|I_{0,\delta}(\rho_{\delta'}(x_n,u_n,w_\eta(x,u)))-I_{0,\delta}(w_\eta(x,u))|<\frac{3}{4}\eta.$$

Thanks to (4.4.12), this implies that there exists $n_\eta^1\in\mathbb{N}$ such that

$$J_\delta(x_n,u_n)\leq J_\delta(x,u)+\eta, \qquad n\geq n_\eta^1. \tag{4.4.13}$$

Next, we want to prove that there exists $n_\eta^2\in\mathbb{N}$ such that

$$J_\delta(x_n,u_n)\geq J_\delta(x,u)-\eta, \qquad n\geq n_\eta^2. \tag{4.4.14}$$

The proof of the inequality above follows the same line of the proof of inequality (4.4.13). Actually, as in (4.4.12) we have

$$J_\delta(x,u) \leq |I_{0,\delta}(\rho_{\delta'}(x,u,w_\eta(x_n,u_n))) - I_{0,\delta}(w_\eta(x_n,u_n))| - J_{0,\delta}(x_n,u_n) + \eta/4.$$

Then, by using the same arguments used above, we can find a sequence $\{\delta'_n\}_{n\geq 1} \subset (0,\delta)$ such that

$$\limsup_{n\to\infty} \left| I_{0,\delta}(\rho_{\delta'_n}(x,u,w_\eta(x_n,u_n))) - I_{0,\delta}(w_\eta(x_n,u_n)) \right| < \frac{3}{4}\eta,$$

and (4.4.14) follows.

□

The continuity above is strictly due to the fact that $\delta > 0$. If $\delta = 0$ then certainly the mapping $x \mapsto I^x(u)$ is not continuous, since $I^x(u)$ is finite only if $u(0) = x$. However, the lemma above easily implies the following weaker condition, which is also true in the $\delta = 0$ case.

Lemma 4.4.3. *For every sequence $\{x_n\}_{n\geq 1} \subset H$, weakly convergent to some x, and for every $u \in \{\varphi \in C([\delta,T];\mathbb{R}) : I^x_{\delta,T}(\varphi) \leq s\}$, there exists a sequence $\{u_n\}_{n\geq 1}$ such that $u_n \to u$ in $C([\delta,T];H)$ and*

$$\limsup_{n\to\infty} I^{x_n}_{\delta,T}(u_n) \leq s. \tag{4.4.15}$$

When $\delta > 0$, this is trivially satisfied by the sequence $u_n = u$ by the previous lemma. This condition can be used along with the following lemma to prove that the conditions of Proposition 2.2.4 is satisfied in the model we are studying.

Lemma 4.4.4. *For any $\delta > 0, T > 0, s \geq 0$ and $x \in H$, we define the set*

$$\Phi^x_{\delta,T}(s) := \{\varphi \in C([\delta, T]; \mathbb{R}) : I^x_{\delta,T}(\varphi) \leq s\}.$$

Then, for any bounded set $B \subset H$, we have

$$\lim_{r \to 0^+} \sup_{x \in B} \lambda(\Phi^x_{\delta,T}(s), \Phi^x_{\delta,T}(s + r)) = 0.$$

Proof. Fix an $\epsilon > 0$ and $s > 0$. We will show that there exist $r > 0$ small enough that for any $x \in B$ and $u \in \Phi^x_{\delta,T}(s + r)$, there exists $z_u \in \Phi^x_{\delta,T}(s)$ such that $\|u - z_u\|_{C([\delta,T];\mathbb{R})} < \epsilon$.

Fix an $r > 0$. First we consider the case of $x \in B$ and $u \in \Phi^x_{\delta,T}(s + r)$ such that $I_{\delta,T}(u) > r$. For such a path u, we may consider the continuous path $z_u \in C([\delta, T]; \mathbb{R})$ defined by

$$z_u(t) = \begin{cases} u(t), & \text{if } t \in [\delta, T^*], \\ \\ u^{u(T^*)}(t - T^*), & \text{if } t \in [T^*, T], \end{cases}$$

where

$$T^* = T^*(u, r) := \inf\{t \in [\delta, T] : I_{t,T}(u) \leq r\}.$$

Hence, $z_u \in \Phi^x_{\delta,T}(s)$. Moreover, since $W^{1,2}([\delta, T]) \hookrightarrow C([\delta, T]; \mathbb{R})$, it is easy to see that

$$\sup_{x \in B} \sup_{u \in \Phi^x_{\delta,T}(s+r)} \|u\|_{C([\delta,T];\mathbb{R})} < \infty.$$

Thanks to the Lipschitz condition on g, this implies that

$$\sup_{x \in B} \sup_{u \in \Phi^x_{\delta,T}(s+r)} \|\mathcal{H}_{\bar{\rho}}(u)\|_{C([\delta,T];\mathbb{R})} < \infty.$$

Next, for any $t \in [T^*, T]$ we have that

$$|u(t) - z_u(t)|^2 \leq \Big(\int_{T_*}^t |u'(s) - \bar{F}(s, u(s))| ds + \int_{T_*}^t |\bar{F}(s, u(s)) - \bar{F}(s, z_u(s))| ds \Big)^2$$

$$\leq c_B \, (t - T^*) \left(I_{T^*,t}(u) + \int_{T_*}^t |u(s) - z_u(s)|^2 \, ds \right),$$

so that, thanks to the Gronwall lemma,

$$|u(t) - z_u(t)|^2 \leq c_{T,B} \, I_{T_\star,T}(u).$$

Now, if we fix $r < (\epsilon \, c_{T,B}^{-1})^{1/2}$, then we have $\|u - z_u\|_{C([\delta,T];\mathbb{R})} < \epsilon$. Since the constant $c_{T,B}$ is independent of x, this proves the result.

Next, we consider the case where $u \in \Phi^x_{\delta,T}(s + r)$, but $I_{\delta,T}(u) \leq r$. Let $w \in C([0, \delta]; \mathbb{R})$ be a path such that $w(0) = \langle x, \mu \rangle$, $w(\delta) = u(\delta)$ and $I_{0,\delta}(w) \leq J_{0,\delta}(x, u) + r$. Then similar to before we may define the path $z_u \in C([\delta, T]; \mathbb{R})$ by

$$z_u(t) = u^{w(T^*)}(t - T^*), \qquad t \in [\delta, T],$$

where

$$T^* = T^*(w, r) := \inf\{t \in [0, \delta] : I_{t,\delta}(w) \leq 2r\}.$$

This implies that

$$I^x_{\delta,T}(z_u) = J_{0,\delta}(x, z_u) \leq I_{0,T^*}(w) = I_{0,\delta}(w) - I_{T^*,\delta}(w) \leq J_{0,\delta}(x, u) - r \leq I^x_{\delta,T}(u) - r.$$

Therefore, $z_u \in \Phi^x_{\delta,T}(s)$. Finally, if we consider the path $\tilde{u}(t) := w(t)\mathbb{I}_{[T^*,\delta]}(t) + u(t)\mathbb{I}_{[\delta,T]}(t)$, then $I_{T^*,T}(\tilde{u}) \leq 3r$. Thus by the same calculation as before we obtain

$$\|u - z_u\|_{C([\delta,T];\mathbb{R})} \leq \left\|\tilde{u} - u^{w(T^*)}(\cdot - T^*)\right\|_{C([T^*,T];\mathbb{R})} \leq c_{T,B} \, I_{T^*,T}(\tilde{u}),$$

which completes the proof upon taking r small enough. $\square$

Lemma 4.4.5. *Suppose $x_n \rightharpoonup x$ in H. Then for any $\delta, T > 0$ and $s \geq 0$, we have*

$$\lim_{n\to\infty} \sup_{u\in\Phi^x_{\delta,T}(s)} \operatorname{dist}(u, \Phi^{x_n}_{\delta,T}(s)) = 0,$$

and

$$\lim_{n\to\infty} \sup_{u\in\Phi^{x_n}_{\delta,T}(s)} \operatorname{dist}(u, \Phi^{x}_{\delta,T}(s)) = 0.$$

In particular, the requirements of Proposition 2.2.4 are satisfied.

Proof. For fixed $u \in C([0,T];\mathbb{R})$, the mapping $x \mapsto I^x_{\delta,T}(u)$ is lower semi-continuous. Condition (i) of Proposition 2.2.4 then follows from Condition (i) in Theorem 4.3.2 (see the proof of Theorem 5 in [13]).

To show the first limit, it suffices to prove that for any $\{u_n\}_{n=1}^\infty \subset \Phi^x_{\delta,T}(s)$ we have

$$\liminf_{n\to\infty} \operatorname{dist}(u_n, \Phi^{x_n}_{\delta,T}(s)) = 0. \tag{4.4.16}$$

Since $I^x_{\delta,T}$ is a good rate function, we may assume by taking a subsequence, if necessary, that $u_n \to u \in \Phi^x_{\delta,T}(s)$. By (4.4.15), we may also find a sequence $\{z_n\}_{n=1}^\infty$ such that $z_n \to u$ and

$$\limsup_{n\to\infty} I^{x_n}_{\delta,T}(z_n) \leq s.$$

Then, for any $r > 0$ we have that

$$\begin{aligned}\operatorname{dist}(u_n, \Phi^{x_n}_{\delta,T}(s)) &\leq \|u_n - z_n\|_{C([\delta,T];\mathbb{R})} + \operatorname{dist}(z_n, \Phi^{x_n}_{\delta,T}(s+r)) + \operatorname{dist}(\Phi^{x_n}_{\delta,T}(s+r), \Phi^{x_n}_{\delta,T}(s))\\ &= \|u_n - z_n\|_{C([\delta,T];\mathbb{R})} + \operatorname{dist}(\Phi^{x_n}_{\delta,T}(s+r), \Phi^{x_n}_{\delta,T}(s)).\end{aligned}$$

Therefore, due to the previous lemma, for every $\epsilon > 0$ we can find $r_\epsilon > 0$ such that

$$\operatorname{dist}(u_n, \Phi^{x_n}_{\delta,T}(s)) \leq \|u_n - z_n\|_{C([\delta,T];\mathbb{R})} + \epsilon, \qquad n \geq 0,$$

and this implies (4.4.16).

To show the second limit, it suffices to prove that for any $\{u_n\}_{n=1}^\infty \subset C([\delta,T];\mathbb{R})$ such that $u_n \in \Phi^{x_n}_{\delta,T}(s)$, we have

$$\liminf_{n\to\infty} \operatorname{dist}(u_n, \Phi^{x}_{\delta,T}(s)) = 0.$$

By condition (i), we may assume, by taking a subsequence if necessary, that $u_n \to u$. Then, thanks to Lemma 4.4.2, we obtain

$$\liminf_{n\to\infty} I^{x_n}_{\delta,T}(u_n) \geq \liminf_{n\to\infty} J_{0,\delta}(x_n,u_n) + \liminf_{n\to\infty} I_{\delta,T}(u_n) \geq J_{0,\delta}(x,u) + I_{\delta,T}(u) = I^x_{\delta,T}(u).$$

In particular, this implies that $I^x_{\delta,T}(u) \leq s$ so that $u \in \Phi^x_{\delta,T}(s))$. Therefore,

$$\operatorname{dist}_{C([\delta,T];\mathbb{R})}(u_n, \Phi^x_{\delta,T}(s)) \leq \|u - u_n\|_{C([\delta,T];\mathbb{R})},$$

which concludes the proof. □

Remark 4.4.1. In Proposition 4.4.3, we have proven that $u_\epsilon^{x_\epsilon,\varphi_\epsilon}$ converges to $u^{x,\varphi}$ in $C([\delta,T];H_\mu)$, $\mathbb{P}$-a.s., for every $0 < \delta < T$. The reason we do not have convergence (and hence a large deviation principle) in $C([0,T];H_\mu)$ is because $e^{t\frac{A}{\epsilon}}x$ does not converge to $\langle x,\mu\rangle$ uniformly on $t \in [0,T]$ as $\epsilon \to 0$. On the other hand, for any $k \geq 1$, due to (4.2.4) we have

$$\int_0^T \left\| e^{t\frac{A}{\epsilon}}x - \langle x,\mu\rangle \right\|^k_{H_\mu} dt \leq c\int_0^T e^{-t\gamma k/\epsilon}\|x\|^k_{H_\mu}\,dt \leq c\,\epsilon\,\|x\|^k_{H_\mu}.$$

This implies that $u_\epsilon^{x_\epsilon,\varphi_\epsilon}$ converges to $u^{x,\varphi}$, as $\epsilon \to 0$, in the space $L^p(\Omega; L^k(0,T;H_\mu))$ for any $p, k \geq 1$. Consequently, the family $\{\mathcal{L}(u_\epsilon^x)\}_{\epsilon\in(0,1]}$ satisfies a large deviation principle in $L^k(0,T;H_\mu)$.

$\square$

Remark 4.4.2. The action functional $I_{\delta,T}^x$ is the same action functional that governs the large deviation principle in $C([\delta,T];\mathbb{R})$ satisfied by the family $\{\mathcal{L}(v_\epsilon^x)\}_{\epsilon>0}$, where v_ϵ^x is the solution to the one-dimensional SDE,

$$dv(t) = \bar{F}(v(t))dt + \sqrt{\gamma(\epsilon)\,\mathcal{H}(t,v(t))}\,d\beta(t), \qquad v(0) = \langle x, \mu\rangle. \tag{4.4.17}$$

The law of the solutions to (4.4.17) is equal to the law of the solutions to the SDE,

$$du(t) = \bar{F}(t,u(t))dt + \frac{\sqrt{\gamma(\epsilon)}}{1+\bar{\rho}}\left(\bar{G}(t,u(t))dW^Q(t) + \bar{\rho}\,\bar{\Sigma}(t)dW^B(t)\right), \quad u(0) = \langle x, \mu\rangle. \tag{4.4.18}$$

Now, in view of equation (4.2.16), we see that (4.4.18) is precisely the limiting equation of (4.2.6) if the coefficients $\alpha(\epsilon)$ and $\beta(\epsilon)$ are held fixed, while only the ϵ terms with the diffusion $\mathcal{A}$ are taken to 0. Therefore, the large deviation principle would not be affected if we were to take the spatial averaging limit to completion before allowing the noises to decay.

$\square$

4.5 Applications to the exit problem

In this section we consider the problem of the exit of the process u_ϵ^x, the solution of equation (4.2.6), from a bounded domain $D \subset H_\mu$. With this in mind, we make the following assumptions on the domain D and the coefficients f, g and σ.

Hypothesis 4.5.1. (i) The coefficients f, g and σ are all independent of t. In addition,

$$\sup_{(\xi,r)\in\mathcal{D}\times\mathbb{R}} |g(\xi,r)| < \infty.$$

(ii) For any $x \in \bar{D}$, the unique solution u^x of the one-dimensional ODE

$$\frac{du}{dt} = \bar{F}(u(t)), \qquad u(0) = \langle x, \mu\rangle,$$

satisfies $u^x(t) \in \bar{D}$, for any $t \geq 0$. Moreover, for every $c_1, c_2 > 0$ there exists $T = T(c_1, c_2) > 0$ such that

$$\|x\|_{H_\mu} \leq c_2 \Longrightarrow \|u^x(t)\|_{H_\mu} \leq c_1, \qquad t \geq T.$$

(iii) The domain $D \subset H_\mu$ is an open, bounded, connected set that contains $x = 0$. In addition, D is invariant under the semigroup e^{tA} and $\langle x, \mu\rangle \in D$, for each $x \in D$.

Remark 4.5.1. The invariance of D under e^{tA} will be necessary in order to prove a lower bound on the exit time of the process u_ϵ^x from the domain D. This is because when ϵ is small, equation (4.2.6) behaves likes the heat equation,

$$\frac{\partial u}{\partial t} = \frac{1}{\epsilon}\mathcal{A}u,$$

for t on the order of ϵ. In fact, if D is not invariant under the semigroup e^{tA}, then for some $x \in D$ the process u_ϵ^x will immediately exit the domain, as $\epsilon \to 0$.

□

Lemma 4.5.1. *Assume that $\mathcal{A}$ is a divergence type operator and pick any function $g : \mathbb{R} \to \mathbb{R}$ that is of class C^2 and convex and has quadratic growth at infinity. For every $r \in \mathbb{R}$, we define*

$$\mathcal{D}_g(r) := \{\, x \in H \,:\, G(x) < r \,\},$$

where

$$G(h) = \int_{\mathcal{D}} g(h(\xi))d\xi, \qquad h \in H.$$

Then, there exists $\bar{r} \in \mathbb{R}$ such that the domain $\mathcal{D}_g(r)$ satisfies Condition (iii) in Hypothesis 4.5.1, for every $r > \bar{r}$.

Proof. First of all, since g has no more that quadratic growth at infinity, the mapping $G : H \to \mathbb{R}$ is well defined. It is differentiable and $G'(h) = g' \circ h$, for every $h \in H$. Moreover, since $\mathcal{A}$ is a divergence type operator, $H = H_\mu$.

The convexity and the quadratic growth at infinity of g imply, respectively, that $\mathcal{D}_g(r)$ is convex and bounded, for every $r \in \mathbb{R}$. Moreover, $0 \in \mathcal{D}_g(r)$, for every $r > g(0)\,|\mathcal{D}| =: \bar{r}$.

Now, we show that $\mathcal{D}_g(r)$ is invariant under the semigroup e^{tA}. Actually, if $x \in H$ and $u(t) := e^{tA}x$, by differentiating and integrating by parts we have

$$\frac{d}{dt}\,G(u(t)) = \langle G'(u(t)), \partial_t u(t)\rangle_H = \langle g'(u(t)), \mathcal{A}u(t)\rangle_H$$

$$= -\int_{\mathcal{D}} g''(u(t,\xi))\langle a(\xi)\nabla u(t,\xi), \nabla u(t,\xi)\rangle\, d\xi \leq 0,$$

last inequality following from the fact that g is convex and from (4.2.1). This means that the mapping $t \mapsto G(u(t))$ is non-increasing, so that

$$x \in \mathcal{D}_g(r) \Longrightarrow G(e^{tA}x) \leq G(x) < r, \quad t \geq 0.$$

Finally, we show that if $x \in \mathcal{D}_g(r)$, then $\langle x, \mu\rangle \in \mathcal{D}_g(r)$. We have

$$G(\langle x, \mu\rangle) = \int_{\mathcal{D}} g(\langle x, \mu\rangle)\, d\xi = |\mathcal{D}|\, g(\langle x, \mu\rangle) \leq \int_{\mathcal{D}} g(x(\xi))\, d\xi = G(x) < r.$$

□

We have seen that for every $x \in H$ and $\delta > 0$, the family $\{\mathcal{L}(u_\epsilon^x)\}_{\epsilon>0}$ satisfies a uniform large deviation principle in $C([\delta, T]; H_\mu)$ with action functional $I_{\delta,T}^x$. Moreover, if x is constant then $\{\mathcal{L}(u_\epsilon^x)\}_{\epsilon>0}$ satisfies a large deviation principle in $C([0,T]; H_\mu)$ with action functional $I_{0,T}^x$. On the basis of this, we define the quasipotential $V : H_\mu \to [0, +\infty]$, by

$$V(y) := \inf\{\, I_{0,T}^0(u) \;:\; u \in C([0,T]; H_\mu),\; u(T) = y,\; T > 0\,\}.$$

Recalling that $I_{0,T}^x$ is finite only if $u \in C([0,T]; \mathbb{R})$, it follows that

$$V(y) < +\infty \implies y \text{ is constant.}$$

Moreover, since we assume that D contains a ball around 0, it follows that both D and ∂D will contain some constant $y \in H_\mu$. In particular, there will exist paths starting at 0 and ending at $z \in \partial D$ that only travel along the subspace $\{y \in H_\mu : y \text{ is constant}\}$. These paths will have finite values of the action functional, so that

$$\bar{V}(D) := \inf_{y \in \partial D} V(y) < +\infty. \tag{4.5.1}$$

In addition, due to Condition (ii) of Hypothesis 4.5.1, the intersection of D and the subspace $\mathbb{R} \subset H_\mu$ is precisely an open interval containing 0. Therefore, if we denote the endpoints of the interval $\mathbb{R} \cap D$ by y_1 and y_2, then $\bar{V}(D) = \min(V(y_1), V(y_2))$.

Remark 4.5.2. Suppose $g \equiv 1$, so that the noise is additive. As discussed in Remark 4.4.2, $I^x_{0,T}$ is the action functional for the large deviation principle satisfied by the family $\{\mathcal{L}(v^x_\epsilon)\}$, where v^x_ϵ is the solution of

$$dv(t) = \bar{F}(v(t))dt + \sqrt{\gamma(\epsilon)\,\mathcal{H}_{\bar{\rho}}}\,d\beta(t), \qquad v(0) = \langle x, \mu \rangle,$$

with

$$\mathcal{H}_{\bar{\rho}} = \frac{1}{(1+\bar{\rho})^2}\left(\left\|\sqrt{Q}m\right\|_H^2 + \bar{\rho}^2\,\delta_0^2\left\|\sqrt{B}[\Sigma N^*_{\delta_0} m]\right\|_Z^2\right).$$

Therefore, due to classical results (see [37]), we will have the explicit formula,

$$V(y) = -\frac{2}{\mathcal{H}_{\bar{\rho}}}\int_0^y \bar{F}(\sigma)d\sigma.$$

In the case that the noise is multiplicative, there is no such explicit representation of the quasipotential, but the exit results we discuss below will still hold.

□

Our goal in this section is to a prove Freidlin-Wentzell type estimates on the exit time of u_ϵ^x from the domain D. With this in mind, we define the stopping times,

$$\tau_\epsilon^x := \inf\{t \geq 0 : u_\epsilon^x(t) \in \partial D\}.$$

The main result is the following.

Theorem 4.5.1. *Assume that all Hypotheses 4.2.1 to 4.5.1 are satisfied. Then for any $x \in D$, we have*

$$\lim_{\epsilon \to 0} \epsilon \log \mathbb{E}\tau_\epsilon^x = \bar{V}(D).$$

The proof of Theorem 4.5.1 is a consequence of the following series of lemmas. Once these lemmas are established, the proof of Theorem 4.5.1 proceeds as in the finite dimensional case (see Theorem 5.7.11 of [28]). We list the lemmas below, and postpone their proofs until Appendix 4.7.

In what follows, we set

$$B_\rho := \{y \in H_\mu : \|y\|_{H_\mu} \leq \rho\},$$

and, for every $\rho > 0$ such that $B_\rho \subset D$, we define the stopping times

$$\sigma_\epsilon^x(\rho) := \inf\{t \geq 0 : u_\epsilon^x(t) \in B_\rho \cup \partial D\}.$$

Lemma 4.5.2. *For any $\eta > 0$, there exists a $T < \infty$ such that*

$$\liminf_{\epsilon \to 0} \gamma(\epsilon) \log \inf_{x \in D} \mathbb{P}_x(\tau_\epsilon^x \leq T) > -(\bar{V}(D) + \eta).$$

Lemma 4.5.3. *Let $\rho > 0$ be such that $B_\rho \subset D$. Then*

$$\lim_{t\to\infty} \limsup_{\epsilon\to 0} \gamma(\epsilon) \log \sup_{x\in D} \mathbb{P}(\sigma_\epsilon^x(\rho) > t) = -\infty.$$

Lemma 4.5.4. *Let $\rho > 0$ be such that $B_\rho \subset D$. Then, for any $x \in D$,*

$$\lim_{\epsilon\to 0} \mathbb{P}\left(u_\epsilon^x(\sigma_\epsilon^x(\rho)) \in B_\rho\right) = 1.$$

Lemma 4.5.5. *Let $\rho > 0$ be such that $B_\rho \subset D$. Then for any $\eta > 0$, there exists $T < \infty$ such that*

$$\limsup_{\epsilon\to 0} \gamma(\epsilon) \log \sup_{x\in B_\rho} \mathbb{P}(\sup_{0\le t\le T} \|u_\epsilon^x(t) - x\|_{H_\mu} \geq 3\rho) < -\eta.$$

Lemma 4.5.6. *Let $\rho > 0$ be such that $B_{2\rho} \subset D$. Then, for any closed set $N \subset \partial D$, we have*

$$\lim_{\rho\to 0} \limsup_{\epsilon\to 0} \gamma(\epsilon) \log \sup_{x\in\partial B_{2\rho}} \mathbb{P}\left(u_\epsilon^x(\sigma_\epsilon^x(\rho)) \in N\right) \leq -\inf_{z\in N} V(z).$$

4.6 Appendix A: Some Lemmas used in Section 4.4

We start with a first preliminary result.

Lemma 4.6.1. *For every $\varphi, \psi \in H$ and $t > 0$ we have*

$$\left\| e^{tA}\left(\varphi\sqrt{Q}\psi\right)\right\|_{H_\mu} \leq c\left(t^{-\frac{d}{2\zeta}} + 1\right) \|\varphi\|_{H_\mu} \|\psi\|_H . \tag{4.6.1}$$

Proof. If we set $\zeta = \frac{2\rho}{\rho-2}$, we have $\frac{1}{\zeta} + \frac{1}{\rho} + \frac{1}{2} = 1$. Thus, for any $t > 0$ and $\psi \in H$, due to condition (4.2.5) we have

$$\left\|e^{tA}(\varphi\sqrt{Q}\psi)\right\|_{H_\mu}=\left\|\sum_k \lambda_k\langle\psi,e_k\rangle e^{tA}(\varphi e_k)\right\|_{H_\mu}$$

$$\leq\left(\sum_k \lambda_k^\rho\,\|e_k\|_\infty^2\right)^{\frac{1}{\rho}}\left(\sum_k \|e_k\|_\infty^{-\frac{2\zeta}{\rho}}\left\|e^{tA}(\varphi e_k)\right\|_{H_\mu}^{\zeta}\right)^{\frac{1}{\zeta}}|\varphi|_H \tag{4.6.2}$$

$$\leq\kappa_Q^{\frac{1}{\rho}}\,\|\psi\|_H\left(\sum_k \left\|e^{tA}(\varphi e_k)\right\|_{H_\mu}^2\right)^{\frac{1}{\zeta}}\sup_k\left(\|e_k\|_\infty^{-\frac{2}{\rho}}\left\|e^{tA}(\varphi e_k)\right\|_{H_\mu}^{\frac{\zeta-2}{\zeta}}\right).$$

By Remark 4.2.2, the semigroup is a contraction on H_μ. Then, since $(\zeta-2)/\zeta=2/\rho$,

$$\sup_{k\geq 0}\|e_k\|_\infty^{-\frac{2}{\rho}}\left\|e^{tA}(\varphi e_k)\right\|_{H_\mu}^{\frac{\zeta-2}{\zeta}}\leq\sup_{k\geq 0}\|e_k\|_\infty^{-\frac{2}{\rho}}\|\varphi e_k\|_{H_\mu}^{\frac{2}{\rho}}\leq\|\varphi\|_{H_\mu}^{\frac{2}{\rho}}. \tag{4.6.3}$$

Moreover, thanks to (4.2.3) and the invariance of the semigroup with respect to the measure μ, we obtain

$$\sum_k\left\|e^{tA}(\varphi e_k)\right\|_{H_\mu}^2=\int_{\mathcal{D}}\sum_k|\langle k_t(\xi,\cdot)\varphi(\cdot),e_k(\cdot)\rangle|^2d\mu(\xi)=\int_{\mathcal{D}}\|k_t(\xi,\cdot)\varphi(\cdot)\|_H^2\,d\mu(\xi)$$

$$\leq c\,(t^{-\frac{d}{2}}+1)\int_{\mathcal{D}}e^{tA}|\varphi(\xi)|^2d\mu(\xi)=c\,(t^{-\frac{d}{2}}+1)\,\|\varphi\|_{H_\mu}^2.$$

Due to (4.6.2) and (4.6.3), this implies that (4.6.1) holds. □

Now, we are ready to state and prove all lemmas used in Section 4.4.

Lemma 4.6.2. *For every $\epsilon>0$, let us define*

$$I_\epsilon^1(t)=\int_0^t e^{(t-s)\frac{A}{\epsilon}}F(s,u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s))ds-\int_0^t\bar{F}(s,u^{x,\varphi}(s))ds,$$

as in Proposition 4.4.3. Then for any $p \geq 1$ and $T > 0$, we have

$$\mathbb{E} \sup_{0 \leq t \leq T} \left\| I^1_\epsilon(t) \right\|^p_{H_\mu} \leq r_{T,p}(\epsilon) + c_{T,p} \int_0^T \mathbb{E} \sup_{0 \leq s \leq t} \| u^{x_\epsilon, \varphi_\epsilon}_\epsilon(s) - u^{x,\varphi}(s) \|^p_{H_\mu} \, ds, \quad (4.6.4)$$

where $r_{T,p}(\epsilon)$ is some non-negative function such that $r_{T,p}(\epsilon) \to 0$, as $\epsilon \to 0$.

Proof. We can rewrite $I^1_\epsilon(t)$ as follows.

$$I^1_\epsilon(t) = J^\epsilon_1(t) + J^\epsilon_2(t) := \left(\int_0^t e^{(t-s)\frac{A}{\epsilon}} F(s, u^{x_\epsilon, \varphi_\epsilon}_\epsilon(s)) ds - \int_0^t \bar{F}(s, u^{x_\epsilon, \varphi_\epsilon}_\epsilon(s)) ds \right)$$

$$+ \left(\int_0^t \left[\bar{F}(s, u^{x_\epsilon, \varphi_\epsilon}_\epsilon(s)) - \bar{F}(s, u^{x,\varphi}(s)) \right] ds \right).$$

Concerning J^ϵ_1, thanks to (4.2.4) and the a priori estimate (4.4.1), we obtain

$$\mathbb{E} \sup_{0 \leq t \leq T} \| J^\epsilon_1(t) \|^p_{H_\mu} \leq c \, \mathbb{E} \sup_{0 \leq t \leq T} \left(\int_0^t e^{-\frac{\gamma(t-s)}{\epsilon}} |F(s, u^{x_\epsilon, \varphi_\epsilon}_\epsilon(s))|_{H_\mu} ds \right)^p$$

$$\leq c_p \left(1 + \mathbb{E} \sup_{0 \leq t \leq T} \| u^{x_\epsilon, \varphi_\epsilon}_\epsilon(t) \|^p_{H_\mu} \right) \left(\int_0^T e^{-\frac{\gamma s}{\epsilon}} ds \right)^p \leq \epsilon^p c_{T,p,M} (1 + \| x_\epsilon \|^p_H) \leq \epsilon^p c_{T,p,M}, \quad (4.6.5)$$

where the last inequality follows from the fact that the sequence $\{x_\epsilon\}_{\epsilon > 0}$ is weakly convergent and hence resides in a bounded set of H. Next, concerning $J^\epsilon_2(t)$, we have

$$\mathbb{E} \sup_{0 \leq t \leq T} |J^\epsilon_2(t)|^p \leq T^{p-1} \mathbb{E} \int_0^T |\bar{F}(s, u^{x_\epsilon, \varphi_\epsilon}_\epsilon(s)) - \bar{F}(s, u^{x,\varphi}(s))|^p ds$$

$$\leq c \, T^{p-1} \int_0^T \mathbb{E} \sup_{0 \leq s \leq t} |u^{x_\epsilon, \varphi_\epsilon}_\epsilon(s) - u^{x,\varphi}(s)|^p_{H_\mu} dt.$$

This inequality, together with (4.6.5), implies (4.6.4). □

Lemma 4.6.3. *For every $\epsilon > 0$, let us define*

$$I_\epsilon^2(t) = \int_0^t e^{(t-s)\frac{A}{\epsilon}} G(s, u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s)) \left[\sqrt{Q}\varphi_H^\epsilon(s)\right] ds - \int_0^t \bar{G}(s, u^{x,\varphi}(s)) \left[\sqrt{Q}\varphi_H(s)\right] ds,$$

as in Proposition 4.4.3. Then, for every $p \geq 1$ and $T > 0$, the following estimate holds.

$$\mathbb{E} \sup_{0\leq t\leq T} \left\|I_\epsilon^2(t)\right\|_{H_\mu}^2 \leq r_{T,p}(\epsilon) + c_{T,p} \int_0^T \mathbb{E} \sup_{0\leq s\leq t} \|u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s) - u^{x,\varphi}(s)\|_{H_\mu}^2 \, dt,$$

where $r_{T,p}(\epsilon)$ is some non-negative function such that $r_{T,p}(\epsilon) \to 0$, as $\epsilon \to 0$.

Proof. We can rewrite $I_\epsilon^2(t)$ as follows.

$$I_\epsilon^2(t) = \left(\int_0^t e^{(t-s)\frac{A}{\epsilon}} G(s, u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s)) \left[\sqrt{Q}\varphi_H^\epsilon(s)\right] \, ds - \int_0^t \bar{G}(s, u^{x_\epsilon,\varphi_\epsilon}(s)) \left[\sqrt{Q}\varphi_H^\epsilon(s)\right] \, ds\right)$$

$$+ \left(\int_0^t \bar{G}(s, u^{x,\varphi}(s)) \left[\sqrt{Q}(\varphi_H^\epsilon(s) - \varphi_H(s))\right] \, ds\right)$$

$$+ \left(\int_0^t \left(\bar{G}(s, u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s)) - \bar{G}(s, u^{x,\varphi}(s))\right]\right) \left[\sqrt{Q}\varphi_H^\epsilon(s)\right] \, ds\right) =: \sum_{i=1}^3 J_i^\epsilon(t).$$

Step 1. We first show that for any $p \geq 1$,

$$\lim_{\epsilon\to 0} \mathbb{E} \sup_{0\leq t\leq T} \|J_1^\epsilon(t)\|_{H_\mu}^p = 0. \tag{4.6.6}$$

Due to the invariance of the semigroup with respect to μ and (4.2.4), we have

$$\|J_1^\epsilon(t)\|_{H_\mu} \leq \Big\| \int_0^t e^{(t-s)\frac{A}{\epsilon}} G(s, u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s)) \left[\sqrt{Q}\varphi_H^\epsilon\right] ds$$

$$- \int_0^t \langle e^{(t-s)\frac{A}{2\epsilon}} G(s, u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s)) \left[\sqrt{Q}\varphi_H^\epsilon(s)\right], \mu\rangle ds \Big\|_{H_\mu}$$

$$\leq \int_0^t e^{-\frac{\gamma(t-s)}{2\epsilon}} \left\| e^{(t-s)\frac{A}{2\epsilon}} G(u_\epsilon^{x_\epsilon,\varphi_\epsilon}) \left[\sqrt{Q}\varphi_H^\epsilon(s)\right] \right\|_{H_\mu} ds.$$

Note that $\frac{d}{\zeta} < 1$ since $\rho < \frac{2d}{d-2}$. Then, by applying inequality (4.6.1) with $\theta = g(s, \cdot, u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s, \cdot))$ we conclude that

$$\|J_1^\epsilon(t)\|_{H_\mu} \leq c \int_0^t e^{-\frac{\gamma(t-s)}{2\epsilon}} \left[((t-s)/\epsilon)^{-\frac{d}{2\zeta}} + 1\right] \|g(s, \cdot, u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s))\|_{H_\mu} \|\varphi_H^\epsilon(s)\|_H \, ds$$

$$\leq c \left(\int_0^T e^{-\frac{\gamma t}{\epsilon}} \left[(t/\epsilon)^{-\frac{d}{\zeta}} + 1\right] ds \right)^{1/2} \left(\int_0^T |\varphi_H^\epsilon(s)|_H^2 ds \right)^{1/2} \left(1 + \sup_{0 \leq s \leq t} \|u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s)\|_{H_\mu} \right)$$

$$\leq c_M \epsilon^{\frac{1}{2}} \left(1 + \sup_{0 \leq s \leq t} \|u_\epsilon^{x_\epsilon,\varphi_\epsilon}(s)\|_{H_\mu} \right).$$

In view of estimate (4.4.1), since $\sup_{\epsilon \in (0,1]} |x_\epsilon|_{H_\mu} < \infty$, we obtain (4.6.6) upon taking the pth moment.

Step 2. We show that for any $p \geq 1$

$$\lim_{\epsilon \to 0} \mathbb{E} \sup_{0 \leq t \leq T} |J_2^\epsilon(t)|^p = 0. \tag{4.6.7}$$

For every $\psi \in L^2(0, T; H)$, we define

$$\Lambda_\psi(t) := \int_0^t \bar{G}(s, u^{x,\varphi}(s)) \left[\sqrt{Q}\psi(s)\right] ds, \qquad t \in [0, T].$$

First, we show that the family $\{\Lambda_{\varphi^\epsilon_H}\}_{\epsilon\in(0,1]}$ is equi-continuous and equi-bounded in $[0,T]$, $\mathbb{P}$-a.s. Actually, we have

$$|\Lambda_{\varphi^\epsilon_H}(t+h)-\Lambda_{\varphi^\epsilon_H}(t)| = \left|\int_t^{t+h} \bar{G}(s,u^{x,\varphi}(s))\left[\sqrt{Q}\varphi^\epsilon_H(s)\right]\,ds\right|$$

$$\leq c\left(\int_t^{t+h}\int_{\mathcal{D}}|g(s,\xi,u^{x,\varphi}(s))|^2 d\mu(\xi)ds\right)^{1/2}\left(\int_t^{t+h}\int_{\mathcal{D}}|\varphi^\epsilon_H(s,\xi)|^2 d\mu(\xi)ds\right)^{1/2}$$

$$\leq c\left(\int_t^{t+h}(1+|u^{x,\varphi}(s)|^2)ds\right)^{1/2}\|\varphi^\epsilon_H\|_{L^2(0,T;H)}\,.$$

Then, since $u^{x,\varphi}\in C([0,T];\mathbb{R})$, $\mathbb{P}$-a.s., we have that

$$\sup_{\epsilon\in(0,1]}\sup_{0\leq t\leq T}|\Lambda_{\varphi^\epsilon_H}(t+h)-\Lambda_{\varphi^\epsilon_H}(t)|\leq c_M\sqrt{h},\qquad \mathbb{P}-a.s., \tag{4.6.8}$$

for some random variable $c_M\in L^2(\Omega)$. Next, we observe that for each fixed $t\in[0,T]$ the linear functional $\psi\in L^2(0,T;H)\mapsto\Lambda_\psi(t)\in\mathbb{R}$ is bounded. Therefore, by the weak convergence of the sequence $\{\varphi^\epsilon_H\}$ to φ_H, we may conclude that

$$\lim_{\epsilon\to 0}\Lambda_{\varphi^\epsilon_H}(t)=\Lambda_{\varphi_H}(t)=\int_0^t\bar{G}(s,u^{x,\varphi}(s))\left[\sqrt{Q}\varphi_H(s)\right]ds,\qquad \mathbb{P}-a.s.,$$

and estimate (4.6.8) implies that this convergence is uniform with respect to $t\in[0,T]$. Finally, noting that $J^\epsilon_2(t)=\Lambda_{\varphi^\epsilon_H}(t)-\Lambda_{\varphi_H}(t)$, we conclude that (4.6.7) holds from the dominated convergence theorem.

Step 3. Using the Lipschitz continuity of g, we have

$$|J^\epsilon_3(t)|^2\leq\left(\int_0^t\int_{\mathcal{D}}\left|(G(s,u^{x_\epsilon,\varphi_\epsilon}_\epsilon)-G(s,u^{x,\varphi}))\left[\sqrt{Q}\varphi^\epsilon_H(s)\right]\right|d\mu(\xi)ds\right)^2$$

$$\leq c\,\|u^{x_\epsilon,\varphi_\epsilon}_\epsilon-u^{x,\varphi}\|^2_{L^2(0,T;H_\mu)}\|\varphi^\epsilon_H\|^2_{L^2(0,T;H_\mu)}\leq c_M\int_0^T\sup_{0\leq s\leq t}\|u^{x_\epsilon,\varphi_\epsilon}_\epsilon(s)-u^{x,\varphi}(s)\|^2_{H_\mu}\,dt.$$

This, together with (4.6.6) and (4.6.7), concludes the proof. □

Lemma 4.6.4. *For every $\epsilon > 0$, let us define*

$$I^3_\epsilon(t) = (\delta_0 - A)\int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0}\left[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)\right] ds - \delta_0 \int_0^t \langle N_{\delta_0}\left[\Sigma(s)\sqrt{B}\varphi_Z(s)\right], \mu\rangle ds,$$

as in Proposition 4.4.3. Then for any $p \geq 1$,

$$\lim_{\epsilon\to 0} \mathbb{E} \sup_{0\leq t\leq T} \left\|I^3_\epsilon(t)\right\|^p_{H_\mu} = 0.$$

Proof. We can rewrite I^3_ϵ as follows.

$$I^3_\epsilon(t) = \left((\delta_0 - A)\int_0^t e^{(t-s)\frac{A}{\epsilon}} N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)]ds - \delta_0\int_0^t \langle N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)], \mu\rangle ds\right)$$

$$+\delta_0 \int_0^t \langle N_{\delta_0}[\Sigma(s)\sqrt{B}(\varphi^\epsilon_Z(s) - \varphi_Z(s))], \mu\rangle ds =: J^\epsilon_1(t) + J^\epsilon_2(t).$$

Concerning J^ϵ_1, the invariance of μ gives

$$\|J^\epsilon_1(t)\|_{H_\mu} = \|\int_0^t e^{(t-s)\frac{A}{2\epsilon}}(\delta_0 - A)e^{(t-s)\frac{A}{2\epsilon}} N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)]ds$$

$$-\int_0^t \langle(\delta_0 - A)e^{(t-s)\frac{A}{2\epsilon}} N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)], \mu\rangle ds\|_{H_\mu}$$

$$\leq c\int_0^t e^{-\frac{\gamma(t-s)}{2\epsilon}} \left\|(\delta_0 - A)e^{(t-s)\frac{A}{2\epsilon}} N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)]\right\|_{H_\mu} ds.$$

Then, thanks to (4.4.4) and the boundedness of the operator S_ρ defined in (4.4.3), for any $\rho > 0$, we have

$$\left\|(\delta_0 - A)e^{(t-s)\frac{A}{2\epsilon}} N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi^\epsilon_Z(s)]\right\|_{H_\mu} \leq c\left[1 + ((t-s)/\epsilon)^{-\frac{1+\rho}{4}}\right] \|\varphi^\epsilon_Z(s)\|_Z.$$

Hence, if $\rho < 1$, we obtain

$$\mathbb{E} \sup_{0\leq t\leq T} \|J_1^\epsilon(t)\|_{H_\mu}^p \leq c \left(\int_0^T e^{-\frac{\gamma t}{\epsilon}} \left[(t/\epsilon)^{-\frac{1+\rho}{2}} + 1 \right] ds \right)^{\frac{p}{2}} \mathbb{E}\, \|\varphi_Z^\epsilon\|_{L^2(0,T;Z)}^{\frac{p}{2}} \leq c_M\, \epsilon^{\frac{p}{2}}. \tag{4.6.9}$$

To estimate $J_2^\epsilon(t)$, we proceed as in Lemma 4.6.3 and for every $\psi \in L^2(0,T;Z)$ we define

$$\Lambda_\psi(t) := \delta_0 \int_0^t \langle N_{\delta_0}[\Sigma(s)\sqrt{B}\psi(s)], \mu\rangle ds.$$

Then the family $\{\Lambda_{\varphi_Z^\epsilon}\}_{\epsilon\in(0,1]}$ is uniformly equi-continuous in $[0,T]$, since

$$|\Lambda_{\varphi_Z^\epsilon}(t+h) - \Lambda_{\varphi_Z^\epsilon}(t)| = \left| \delta_0 \int_t^{t+h} \langle N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi_Z^\epsilon(s)], \mu\rangle ds \right|$$

$$\leq \delta_0 \sqrt{h} \left(\int_t^{t+h} \int_{\mathcal{D}} |N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi_Z^\epsilon(s)](\xi)|^2 d\mu(\xi)\ ds \right)^{1/2}$$

$$\leq c\, \delta_0 \sqrt{h}\, \|N_{\delta_0}[\Sigma(\cdot)\varphi_Z^\epsilon]\|_{L^2(0,T;H_\mu)} \leq c_M \sqrt{h},$$

where the last inequality holds $\mathbb{P}$-a.s., for some random variable $c_M \in L^1(\Omega)$. In addition, for fixed $t \in [0,T]$, the linear functional $\psi \in L^2(0,T;Z) \mapsto \Lambda\psi(t) \in \mathbb{R}$ is bounded. Hence by the weak convergence of the sequence $\{\varphi_Z^\epsilon\}$ to φ_Z, we have

$$\lim_{\epsilon\to 0} \Lambda_{\varphi_Z^\epsilon}(t) = \Lambda_{\varphi_Z}(t) = \delta_0 \int_0^t \langle N_{\delta_0}[\Sigma(s)\sqrt{B}\varphi_Z(s)], \mu\rangle ds.$$

Moreover, this convergence is uniform in $t \in [0,T]$, so that

$$\lim_{\epsilon\to 0} \mathbb{E} \sup_{0\leq t\leq T} |J_2^\epsilon(t)|^p = \lim_{\epsilon\to 0} \sup_{0\leq t\leq T} |\Lambda_{\varphi_Z^\epsilon}(t) - \Lambda_{\varphi_Z}(t)|^p = 0$$

from the dominated convergence theorem. This, together with (4.6.9), concludes the proof. □

4.7 Appendix B: Proofs of Lemmas in Section 4.5

Proof of Lemma 4.5.2. Fix $\eta > 0$. We first construct a collection of paths $\{z^x\}_{x\in D} \subset C([0,T];\mathbb{R})$ that leave the domain with a close to minimal energy.

Let $\rho > 0$ such that $B_\rho \subset D$. Due to Condition (ii) in Hypothesis 4.5.1, we can fix T_1 large enough that $u^x(T_1) \in B_\rho$, for any $x \in \bar{D}$, where u^x is the solution of (4.3.1). Thus, we set $z^x(t) = u^x(t)$ on the interval $[0, T_1]$. Next, we set

$$z^x(t) = z^x(T_1)(T_1 + 1 - t), \qquad \text{if } t \in [T_1, T_1 + 1],$$

so that $z^x(T_1 + 1) = 0$. Now, due to (4.5.1), there exists some $T_2 > 0$ and some path $v(t) \in C([0, T_2], \mathbb{R})$ such that $v(0) = 0$, $v(T_2) \notin \bar{D}$ and $I^0_{0,T_2}(v) < \bar{V}(D) + \eta/4$. We then set $z^x(T_1 + 1 + t) = v(t)$ for $t \in [0, T_2]$. Hence, upon defining $T^* = T_1 + T_2 + 1$, we have

$$I^x_{0,T^*}(z^x) = I^x_{0,T_1}(z^x) + I_{T_1,T_1+1}(z^x) + I^0_{T_1+1,T^*}(z^x)$$

$$\leq c \int_{T_1}^{T_1+1} |z^{x\prime}(t) - \bar{F}(z^x(t))|^2 dt + (\bar{V}(D) + \eta/4) \leq c\,\rho^2 + \big(\bar{V}(D) + \eta/4\big)\,.$$

Thus, taking ρ small enough, we obtain $I^x_{0,T}(z^x) < \bar{V}(D) + \eta/2$. We note that all of these paths $\{z^x\}_{x\in D}$ agree on the interval $[T_1 + 1, T^*]$ and exit the domain on this time interval. Let us now denote

$$h := \sup_{T_1+1\leq t\leq T^*} \text{dist}_{H_\mu}(z^x(t), \bar{D}) > 0.$$

To prove the lemma, we pick any $0 < \delta < T_1 + 1$ and define the open set

$$\Psi = \bigcup_{x \in D} \left\{ u \in C([\delta, T^*]; H_\mu) : \sup_{\delta \le t \le T^*} |u(t) - z^x(t)|_{H_\mu} < h \right\}.$$

Then, thanks to Theorem 4.3.1 and the FWULDP lower bound (Definition 2.2.5), there exists $\epsilon_0 > 0$ such that for any $\epsilon < \epsilon_0$,

$$\inf_{x \in D} \mathbb{P}(\tau_x^\epsilon < T^*) \ge \inf_{x \in D} \mathbb{P}(u_\epsilon^x \in \Psi) \ge \exp\left(-\frac{1}{\gamma(\epsilon)} \left[\sup_{x \in D} \inf_{\varphi \in \Psi} I^x_{\delta, T^*}(\varphi) + \frac{\eta}{2} \right] \right)$$

$$\ge \exp\left(-\frac{1}{\gamma(\epsilon)} \left[\sup_{x \in D} I^x_{\delta, T^*}(z^x|_{[\delta, T^*]}) + \frac{\eta}{2} \right] \right\} \ge \exp\left\{ -\frac{1}{\gamma(\epsilon)} (\bar{V}(D) + \eta) \right).$$

□

Proof of Lemma 4.5.3. In this lemma, the behavior of the process near $t = 0$ is not a concern and so the same proof as in Lemma 5.7.19 in [28] holds. □

Proof of Lemma 4.5.4. Fix some $x \in D$ and let $\rho > 0$ be such that $B_\rho \subset D$. If $x \in B_\rho$, there nothing to prove. Thus, we can assume that $x \notin B_\rho$.

We denote $T_x := \inf\{t \ge 0 : u^x(t) \in B_{\rho/2}\}$ and $\Delta_x := \inf_{t \ge 0} \operatorname{dist}_{H_\mu}(u^x(t), \partial D)$. We clearly have $T_x > 0$ and, due to Condition (iii) of Hypothesis 4.5.1, we have $\Delta_x > 0$. Moreover, again thanks to Condition (iii) of Hypothesis 4.5.1, we have

$$d_x := \inf_{t \ge 0} \operatorname{dist}_{H_\mu}(e^{tA} x, \partial D) > 0.$$

This implies that for every $0 < \delta < T_x$

$$\mathbb{P}\left(u_\epsilon^x(\sigma_\epsilon^x(\rho)) \in \partial D\right)$$

$$\le \mathbb{P}\left(\sup_{0 \le t \le \delta} \left\| u_\epsilon^x(t) - e^{t\frac{A}{\epsilon}} x \right\|_{H_\mu} > d_x \right) + \mathbb{P}\left(\sup_{\delta \le t \le T_x} \|u_\epsilon^x(t) - u^x(t)\|_{H_\mu} > \Delta_x \wedge \rho/2 \right).$$

$$(4.7.1)$$

Now, thanks to (4.2.9) and (4.2.10) and Lemma 4.6.2, for every $T > 0$ there exists some function $r_T(\epsilon)$ going to 0, as $\epsilon \to 0$, such that

$$\mathbb{E} \sup_{\delta \leq t \leq T} \|u_\epsilon^x(t) - u^x(t)\|_{H_\mu} \leq c\, e^{-\frac{\gamma\delta}{\epsilon}} \|x\|_{H_\mu} + r_T(\epsilon) + c_T \int_\delta^T \mathbb{E} \sup_{\delta \leq s \leq t} \|u_\epsilon^x(s) - u^x(s)\|_{H_\mu} dt.$$

Then, using Gronwall's Lemma, we have

$$\mathbb{E} \sup_{\delta \leq t \leq T} \|u_\epsilon^x - u^x\|_{H_\mu} \leq c \left(e^{-\frac{\gamma\delta}{\epsilon}} \|x\|_{H_\mu} + r_T(\epsilon) \right) e^{c_T T}. \tag{4.7.2}$$

Meanwhile, we can estimate the second term in (4.7.1) by using the bounds (4.2.9), (4.2.10) and (4.4.1) to obtain

$$\mathbb{E} \sup_{0 \leq t \leq \delta} \left\| u_\epsilon^x(t) - e^{t\frac{A}{\epsilon}} x \right\|_{H_\mu} \leq c \sqrt{\gamma(\epsilon)} + \mathbb{E} \sup_{0 \leq t \leq \delta} \left\| \int_0^t e^{(t-s)\frac{A}{\epsilon}} F(s, u_\epsilon^x(s)) ds \right\|_{H_\mu} \tag{4.7.3}$$

$$\leq c \left(\sqrt{\gamma(\epsilon)} + \delta \left(1 + \mathbb{E} \|u_\epsilon^x\|_{C([0,\delta];H_\mu)} \right) \right) \leq c\,(\sqrt{\gamma(\epsilon)} + \delta).$$

This, together with (4.7.1) and (4.7.2), implies that for every $\delta \in (0, T_x)$

$$\mathbb{P}\,(u_\epsilon^x(\sigma_\epsilon^x(\rho)) \in \partial D) \leq c_T \left(\sqrt{\gamma(\epsilon)} + \delta + r_T(\epsilon) + e^{-\frac{\gamma\delta}{\epsilon}} \|x\|_{H_\mu} \right).$$

Thus, by taking $\delta = \epsilon^r$ for some $0 < r < 1$, we get

$$\mathbb{P}\,(u_\epsilon^x(\sigma_\epsilon^x(\rho)) \in \partial D) = 0.$$

$\square$

Proof of Lemma 4.5.5. We have

$$u_\epsilon^x(t) - x = e^{t\frac{A}{\epsilon}} x - x + \int_0^t e^{(t-s)\frac{A}{\epsilon}} F(s, u_\epsilon^x(s)) ds + \alpha(\epsilon)\, w_{A,Q}^\epsilon(u_\epsilon^x)(t) + \beta(\epsilon)\, w_{A,B}^\epsilon(t).$$

Since the semigroup e^{tA} acts as a contraction on H_μ, we have that $\left\| e^{t\frac{A}{\epsilon}}x - x \right\|_{H_\mu} \leq 2\,\|x\|_{H_\mu}$. Next we observe that, for $t \in [0, T]$,

$$\left\| \int_0^t e^{(t-s)\frac{A}{\epsilon}} F(s, u_\epsilon^x(s)) ds \right\|_{H_\mu} \leq \int_0^t \|F(s, u_\epsilon^x(s))\|_{H_\mu} ds$$

$$\leq c\,T \left(1 + \sup_{0\leq s\leq t} \|u_\epsilon^x(s)\|_{H_\mu}\right) \leq c\,T \left(1 + \|x\|_{H_\mu} + \sup_{0\leq s\leq T} \|u_\epsilon^x(s) - x\|_{H_\mu}\right).$$

Therefore, if $x \in B_\rho$, we can find a $T_\rho > 0$ small enough that

$$\sup_{0\leq t\leq T} \|u_\epsilon^x(t) - x\|_{H_\mu} \leq \frac{7\rho}{3} + \alpha(\epsilon) \sup_{0\leq t\leq T_\rho} \left\|w_{A,Q}^\epsilon(u_\epsilon^x)(t)\right\|_{H_\mu} + \beta(\epsilon) \sup_{0\leq t\leq T_\rho} \left\|w_{A,B}^\epsilon(t)\right\|_{H_\mu}.$$

Hence,

$$\mathbb{P}\left(\|u_\epsilon^x - x\|_{C([0,T];H_\mu)} \geq 3\rho\right)$$

$$\leq \mathbb{P}\left(\alpha(\epsilon)\, \left\|w_{A,Q}^\epsilon(u_\epsilon^x)\right\|_{C([0,T];H_\mu)} \geq \rho/3\right) + \mathbb{P}\left(\beta(\epsilon)\, \left\|w_{A,B}^\epsilon\right\|_{C([0,T];H_\mu)} \geq \rho/3\right).$$

Thanks to Condition (i) of Hypothesis 4.5.1, the integrand of $w^{A,Q}(u_\epsilon^x)$ is bounded, so that we can use the exponential estimates for the stochastic convolution (see [51]). In particular, for every $T > 0$ we have

$$\mathbb{P}\left(\sup_{0\leq t\leq T} \left\|w_{A,Q}^\epsilon(u_\epsilon^x)(t)\right\|_{H_\mu} \geq \frac{\rho}{3\,\alpha(\epsilon)}\right) \leq c \exp\left(-\frac{\rho^2}{3c_T\,\alpha(\epsilon)}\right) \leq c\, \exp\left(-\frac{\rho^2}{3\,c_t\,\gamma(\epsilon)}\right),$$

where c_T is a constant going to 0, as $T \to 0$. We obtain a similar estimate for $w_{A,B}^\epsilon$, with $\alpha(\epsilon)$ replaced by $\beta(\epsilon)$. All together, for every $T \leq T_\rho$ we have

$$\gamma(\epsilon) \log \sup_{x\in B_\rho} \mathbb{P}(\sup_{0\leq t\leq T} \|u_\epsilon^x(t) - x\|_{H_\mu} \geq 3\rho) \leq c\,\gamma(\epsilon) - \frac{\rho^2}{c_T}.$$

Upon taking T small enough, this gives us the desired result. □

Proof of Lemma 4.5.6. We modify the proof of Lemma 5.7.21 in [28] to account for the behavior of $u_\epsilon^x(t)$ near $t = 0$. Let $N \subset \partial D$ be a closed set. Define the closed set

$$\Psi_{\delta,T}(N) := \{u \in C([0,T]; H_\mu) : \exists\, t \in [\delta, T] \text{ such that } u(t) \in N\}.$$

Then, for any $T > 0$ and $\delta < T$,

$$\mathbb{P}(u_\epsilon^x(\sigma_\epsilon^x(\rho)) \in N) \leq \mathbb{P}(\tau_\epsilon^x < \delta) + \mathbb{P}(\sigma_\epsilon^x(\rho) > T) + \mathbb{P}(u_\epsilon^x \in \Psi_{\delta,T}(N)). \tag{4.7.4}$$

To bound the first term from above, we notice that

$$\sup_{x \in \partial B_{2\rho}} \mathbb{P}(\tau_x^\epsilon < \delta) \leq \sup_{x \in \partial B_{2\rho}} \mathbb{P}\left(\sup_{0 \leq t \leq \delta} \|u_\epsilon^x(t) - x\|_{H_\mu} \geq \operatorname{dist}_{H_\mu}(x, \partial D)\right).$$

Now, let $\rho > 0$ be small enough that $\inf_{x \in \partial B_{2\rho}} \operatorname{dist}(x, \partial D) \geq 6\rho$. Then, by Lemma 4.5.5, the inequality above implies that for any $\eta > 0$ there exists $\delta > 0$ small enough that

$$\limsup_{\epsilon \to 0} \gamma(\epsilon) \log \sup_{x \in \partial B_{2\rho}} \mathbb{P}(\tau_x^\epsilon < \delta) \leq -\eta. \tag{4.7.5}$$

Next, thanks to Lemma 4.5.3, we can find $T > 0$ large enough that

$$\limsup_{\epsilon \to 0} \gamma(\epsilon) \log \sup_{x \in \partial B_{2\rho}} \mathbb{P}(\sigma_\epsilon^x(\rho) > T) < -\eta. \tag{4.7.6}$$

Since the set $\Psi_{\delta,T}(N)$ is closed, we can use the upper bound in the DZULDP (Definition 2.2.4) to obtain that

$$\limsup_{\epsilon \to 0} \gamma(\epsilon) \log \sup_{x \in \partial B_{2\rho}} \mathbb{P}(u_\epsilon^x \in \Psi_{\delta,T}(N)) \leq - \inf_{x \in \partial B_{2\rho}} I_{\delta,T}^x(\Psi_{\delta,T}(N)). \tag{4.7.7}$$

On the other hand, for fixed x, we have that

$$I_{\delta,T}^x(\Psi_{\delta,T}(N)) = \inf_{\varphi \in \Psi_{\delta,T}(N)} I_{\delta,T}^x(\varphi) \geq \inf_{\varphi \in \Psi_{0,T}(N)} I_{0,T}^x(\varphi),$$

because every path hitting N in the interval $[\delta, T]$ has an extension to a path on $[0, T]$ starting at x.

Next, we notice that for any $x \in \partial B_{2\rho}$,

$$V(x) + \inf_{\varphi \in \Psi_{0,T}(N)} I^x_{0,T}(\varphi) \geq \inf_{z \in N} V(z),$$

since any path on the left hand side is also Considered in the infima on the right hand side. Now, due to Hypotheses 4.5.1, it is clear that $\lim_{x \to 0} V(x) = 0$. Hence, for any $\gamma > 0$, if we choose $\rho > 0$ small enough then, thanks to (4.7.7), we have

$$\limsup_{\epsilon \to 0} \gamma(\epsilon) \log \sup_{x \in \partial B_{2\rho}} \mathbb{P}(u^x_\epsilon \in \Psi_{\delta,T}(N)) \leq \gamma - \inf_{z \in N} V(z). \tag{4.7.8}$$

Due to (4.7.4), (4.7.5), (4.7.6), (4.7.8) and the arbitrariness of γ, the result then follows by picking $\eta > \inf_{z \in N} V(z)$. □

Chapter 5: Large deviations for the invariant measures of the stochastic Navier-Stokes equations

5.1 Introduction

In this section, we continue our study of small-noise limits for multi-scale stochastic partial differential equation. We consider the following two-dimensional incompressible Navier-Stokes equations on the torus $\mathbb{T}^2 = [0, 2\pi]^2$, perturbed by a small additive noise:

$$\begin{cases} \partial_t u(t,x) + (u(t,x)\cdot\nabla)u(t,x) = \Delta u(t,x) + \nabla p(t,x) + \sqrt{\epsilon\, Q_\epsilon}\; \partial_t W(t,x), \\ \\ \text{div } u(t,x) = 0, \qquad u(0,x) = u_0(x), \qquad u \text{ is periodic.} \end{cases} \tag{5.1.1}$$

Here, the functions $u(t,x) \in \mathbb{R}^2$ and $p(t,x) \in \mathbb{R}$ denote the velocity and the pressure of the fluid at any $(t,x) \in \mathbb{R}^+ \times \mathbb{T}^2$. The random forcing $\partial_t W(t,x)$ is a space-time white noise, while the operator $\sqrt{Q_\epsilon}$ provides spatial correlation to the noise on the scale of size $\delta(\epsilon)$. We are interested in the behavior of equation (5.1.1) as the noise magnitude $\sqrt{\epsilon}$ and the correlation scale $\delta(\epsilon)$ are simultaneously sent to 0.

In two dimensions, the incompressible Navier-Stokes equation driven by space-time white noise is well-posed only in spaces of negative regularity (see [23]). The driving noise must have more regularity in the spatial variable in order to have function-valued solutions. In our case, we consider a smoothing operator $\sqrt{Q_\epsilon}$ that provides sufficient regularity to interpret equation (5.1.1) in the space $C([0,T];[L^2(\mathbb{T}^2)]^2)$ for any fixed $\epsilon > 0$. In fact, the regularization $\sqrt{Q_\epsilon}$ can be chosen to decay to the identity operator slowly enough for the $\sqrt{\epsilon}$ factor to compensate and produce a function-valued limit.

Under present assumptions, the $\epsilon \downarrow 0$ limit of equation (5.1.1) in $C([0,T];[L^2(\mathbb{T}^2)]^2)$ is unsurprisingly the corresponding unforced Navier-Stokes equation. A more interesting problem is the quantification of the convergence rate via large deviations theory. In [15], it was shown that the solutions to the Leray-projected version of equation (5.1.1) satisfy a large deviations principle in $C([0,T];[L^2(\mathbb{T}^2)]^2)$ with rate function

$$I(u) = \frac{1}{2}\int_0^T \|u'(t) + Au(t) + B(u(t))\|^2_{[L^2(\mathbb{T}^2)]^2}\,dt,$$

where A is the Stokes operator and B is the Navier-Stokes nonlinearity. This result was proven using the weak convergence approach. The weak convergence method was also used in [3] and [4] to prove large deviations principles for the stochastic Navier-Stokes with a fixed noise regularization and viscosity vanishing at a rate proportional to the strength of the noise, which is believed to be a relevant problem in the study of turbulent fluid dynamics.

If the operator $\sqrt{Q_\epsilon}$ is simultaneously smoothing enough but not too degenerate, then equation (5.1.1) will possess a unique ergodic invariant probability measure (see [32]). In the $\epsilon \downarrow 0$ limit, it can be shown that these measures converge weakly to the

Dirac measure at 0. For fixed correlation strength $\delta(\epsilon) = \delta > 0$, it was proven in [8] that the invariant measures also satisfy a large deviations principle in $[L^2(\mathbb{T}^2)]^2$ with rate function given by the quasipotential

$$U_\delta(x) = \inf \left\{ I_T^\delta(u) \,:\, T > 0,\ u \in C([0,T]; [L^2(\mathbb{T}^2)]^2),\ u(0) = 0,\ u(T) = x \right\},$$

where $I_T : C([0,T]; [L^2(\mathbb{T}^2)]^2) \to [0, +\infty]$ is the action functional for the paths defined by

$$I_T^\delta(u) := \frac{1}{2} \int_0^T \left\| Q_\delta^{-1} \Big[u'(t) + Au(t) + B(u(t)) \Big] \right\|_{[L^2(\mathbb{T}^2)]^2}^2 dt.$$

This result was generalized in [46] to the case of the Navier-Stokes equations posed on a bounded domain with Dirichlet boundary conditions. In [46] they also considered the case where the equation has a deterministic, time-independent forcing so that the limiting dynamics may be nontrivial point attractors or sets of attractors. Both papers established their results by following the general strategy of [60] for proving large deviations principles for families of invariant measures.

In [9], it was also proven that the quasipotential $U_\delta(x)$, corresponding to the problem on the torus with fixed correlation strength δ, converges pointwise to

$$U(x) = \|x\|_{[H^1(\mathbb{T}^2)]^2}^2, \tag{5.1.2}$$

as $\delta \downarrow 0$. This is a consequence of the orthogonality of Au and $B(u)$ in $[L^2(\mathbb{T}^2)]^2$, which in general does not hold for the problem posed on a bounded domain. In some sense, $U(x)$ is what one would expect the quasi-potential for the space-time white noise case to be if the time-stationary problem were well-posed.

The purpose of this chapter is to bridge the results of [8] and [9] with the result of [15]. Rather than first taking $\epsilon \downarrow 0$ and then studying what happens as the regularization is removed, we take ϵ and δ to 0 *simultaneously*. In analogy with the result Theorem 2.3.3 in Chapter 2, we prove that the invariant measures of equation (5.1.1) satisfy a large deviations principle with rate function given by the quasi-potential (5.1.2), under suitable conditions on the regularization $\sqrt{Q_\epsilon}$.

To prove this result, we first prove a large deviations principle for the solutions of equation (5.1.1) in $C([0,T];[L^2(\mathbb{T}^2)]^2)$ that is uniform with respect to initial conditions in appropriate sets of initial conditions. This is done by proving a large deviations principle for the linearized problem using the weak convergence approach and then transferring this to the nonlinear problem via the contraction principle. We note that this method allows for slower decay of the correlation scale $\delta(\epsilon)$ than the methods used in [15]. The proof of the large deviations principle for the invariant measures then follows in a similar manner as in [8], but requires crucial modifications to account for the decaying regularity of the driving noise.

Remark 5.1.1. Let us briefly put the results of this section into the context of the finite dimensional results discussed in Section 2.3. Suppose that the covariance Q_ϵ in (5.1.1) is instead the identity operator so that (5.1.1) can be formally re-written as the evolution equation

$$du_t = b(u_t)dt + \sqrt{\epsilon}dW_t, \tag{5.1.3}$$

in $[L^2(\mathbb{T}^2)]^2$. Here, $b(u) = -Au - B(u)$ and W_t is a cylindrical Wiener process in

$[L^2(\mathbb{T}^2)]^2$, the infinite-dimensional analog of a standard Brownian motion. Let us pretend as if this equation were well-posed so that the solution $u(t,x)$ could be interpreted as a continuous $[L^2(\mathbb{T}^2)]^2$-valued random process. As we saw in Chapter 2, one does not expect to be able to write down the quasi-potential for (5.1.3) explicitly, except in the case of a gradient system. *But*, in this particular case where the problem is studied on the torus in two-dimensions, $b(u)$ is already in the form of the orthogonal decomposition $-\nabla U + \ell$ mentioned in Remark 2.3.3. In fact, the orthogonality $\langle Au, B(u)\rangle_{[L^2(\mathbb{T}^2)]^2} = 0$ implies that

$$-\nabla U(u) = Au, \qquad \ell(u) = B(u),$$

where $\nabla U(u)$ denotes the functional derivative of U in $[L^2(\mathbb{T}^2)]^2$. This implies the identification of (5.1.2) as the quasi-potential.

It should also be noted that the quasi-potential will still exist for the problem studied on domains other than the torus, but it will not have the simple representation given by (5.1.2). Nonetheless, one may still hope that a large deviations principle for the invariant measures ν_ϵ as $\epsilon \to 0$ and $Q_\epsilon \to \mathbb{1}$ will hold. While we suspect this is likely true, presently, we are unable to prove this for a general domain. Our proof of the upper bound of the large deviations principle for ν^ϵ requires exponential bounds of the solution on an H^1 ball that we are presently only able to obtain in the case of the problem on the torus by exploiting the orthogonality of Au and $B(u)$. The lower bound, on the other hand, remains valid for a general bounded domain.

□

5.2 Preliminaries

We consider equation (5.1.1) posed on the space of square-integrable, mean zero, space-periodic functions. For an introduction to the 2D Navier-Stokes equations on the torus, see the book [64] by Temam. We follow the notations and conventions used there. Denoting $\mathbb{T}^2 := [0, 2\pi]^2$, we let

$$H := \left\{ f \in [L^2(\mathbb{T}^2)]^2 : \int_{\mathbb{T}^2} f(x)dx = 0, \quad \operatorname{div} f = 0, \quad f \text{ is periodic in } \mathbb{T}^2 \right\},$$

where the periodic boundary conditions are interpreted in the sense of trace. It can be shown that H is a Hilbert space when endowed with the standard $L^2(\mathbb{T}^2)$ inner product. We denote the norm and inner product on H by $\|\cdot\|_0$ and $\langle \cdot, \cdot \rangle_0$, respectively. Moreover, in what follows, for every $p \geq 1$ we shall write L^p instead of $L^p(\mathbb{T}^2)$.

We denote by $H_{\mathbb{C}}$, the complexification of H, and by $\mathbb{Z}_0^2$ the set $\mathbb{Z}^2 \setminus \{(0,0)\}$. The family $\{e_k\}_{k \in \mathbb{Z}_0^2} \subset H_{\mathbb{C}}$ defined by

$$e_k(x) = \frac{1}{2\pi} \frac{(k_2, -k_1)}{\sqrt{k_1^2 + k_2^2}} \, e^{ix \cdot k}, \qquad x \in \mathbb{T}^2, \quad k = (k_1, k_2) \in \mathbb{Z}_0^2,$$

form a complete orthonormal system in $H_{\mathbb{C}}$. Similarly, the family $\{\mathrm{Re}(e_k)\}_{k \in \mathbb{Z}_0^2} \subset H$ form a complete orthonormal system in H. In the remainder, we use the basis $\{e_k\}_{k \in \mathbb{Z}_0^2}$ with the implicit assumption that we are only considering the real components.

Next, we let P be the orthogonal projection from $[L^2(\mathbb{T}^2)]^2$ onto H, known as the Leray projection. We define the Stokes operator by setting

$$Au := -P\Delta u, \quad u \in D(A) := H \cap [W^{2,2}(\mathbb{T}^2)]^2.$$

It is easy to see that A is a diagonal operator on H with respect to the basis $\{e_k\}_{k\in\mathbb{Z}_0^2}$. In particular, for any $k \in \mathbb{Z}_0^2$ we have

$$Ae_k = |k|^2 e_k.$$

Since A is a positive, self-adjoint operator, for any $r \in \mathbb{R}$ we can define the fractional power A^r with domain $D(A^r)$. In fact, it can be shown that $D(A^r)$ is the closure of $\text{span}(\{e_k\}_{k\in\mathbb{Z}_0^2})$ with respect to the $[W^{2r,2}(\mathbb{T}^2)]^2$ Sobolev norm. To simplify our notations, we will denote $V^r := D(A^{r/2})$ with norm given by the $[W^{2r,2}(\mathbb{T}^2)]^2$ Sobolev semi-norm

$$\|u\|_r^2 := \|u\|_{D(A^{r/2})}^2 = \|u\|_{[H^r(\mathbb{T}^2)]^2}^2 = \sum_{k\in\mathbb{Z}_0^2} |k|^{2r} \langle u, e_k\rangle_H^2.$$

In particular, we have that $V^2 = D(A)$ and $V := V^1 = D(A^{1/2})$. For any $r \geq 0$, we denote by V^{-r} the dual space of V^r. In addition, for any $p \geq 1$, we will use the shorthands

$$L^p := [L^p(\mathbb{T}^2)]^2, \qquad W^{k,p} := [W^{k,p}(\mathbb{T}^2)]^2.$$

Next, we define the tri-linear form, $b : V \times V \times V \to \mathbb{R}$, by

$$b(u, v, w) := \int_{\mathbb{T}^2} (u(x) \cdot \nabla) v(x) \cdot w(x) dx, \qquad u, v, w \in V.$$

From standard interpolation inequalities and Sobolev embeddings, it follows that

$$|b(u,v,w)| \le c \begin{cases} \|u\|_H^{1/2} \|u\|_1^{1/2} \|v\|_1 \|w\|_H^{1/2} \|w\|_1^{1/2}, \\ \\ \|u\|_H^{1/2} \|u\|_2^{1/2} \|v\|_1 \|w\|_H, \\ \\ \|u\|_H \|v\|_1 \|w\|_H^{1/2} \|w\|_2^{1/2}, \\ \\ \|u\|_H^{1/2} \|u\|_1^{1/2} \|v\|_1 \|w\|_H^{1/2} \|w\|_1^{1/2}, \end{cases} \tag{5.2.1}$$

for smooth u, v, w. These inequalities can then be extended to the appropriate Sobolev spaces by continuity. We note that the first inequality in (5.2.1) implies that b is indeed well-defined and continuous on $V \times V \times V$. The tri-linear form b also induces the continuous mappings $B : V \times V \to V'$ and $B : V \to V'$ defined by

$$\langle B(u,v), w\rangle := b(u,v,w),$$

$$B(u) := B(u,u),$$

for $u, v, w \in V$. Moreover, it can be shown that for any $u, v \in D(A)$

$$B(u,v) = P[(u \cdot \nabla)v].$$

It can be proven that

$$\langle B(u,v), w\rangle_H = -\langle B(u,w), v\rangle_H, \qquad u, v, w \in V. \tag{5.2.2}$$

Moreover

$$\langle B(u), Au\rangle_H = 0, \qquad u \in D(A), \tag{5.2.3}$$

which implies that

$$\langle B(u,v), v\rangle_H = 0, \qquad u, v \in V.$$

Equation (5.2.2) is still true when considering the problem posed on a bounded domains with Dirichlet boundary conditions. Equation (5.2.3), on the other hand, only holds for the problem posed on the torus with periodic boundary conditions. For a proof of equation (5.2.3), see for example [39]. We note that the proof of our main result relies on equation (5.2.3) in several places, and hence will not immediately generalize to the case of the Navier-Stokes equation on a bounded domain.

As for the random forcing in equation (5.1.1), we assume that $W(t,x)$ is a cylindrical Wiener process on the Hilbert space of mean-zero functions in $[L^2(\mathbb{T}^2)]^2$. We then set $w(t) := PW(t)$ so that w has the formal expansion

$$w(t,x) = \sum_{k\in\mathbb{Z}_0^2} e_k(x)\beta_k(t), \qquad t \geq 0, \quad x \in \mathbb{T}^2,$$

where $\{\beta_k\}_{k\in\mathbb{Z}_0^2}$ are a collection of independent, real-valued Brownian motions on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We assume that the covariance operator $Q_\epsilon \in \mathcal{L}(H;H)$ takes the following form:

$$Q_\epsilon := (I + \delta(\epsilon)A^\beta)^{-1}, \tag{5.2.4}$$

for some $\beta > 0$ and $\delta(\epsilon) > 0$. Since we are concerned with the singular noise limit, $\delta(\epsilon)$ will be taken to be a strictly decreasing function of ϵ such that

$$\lim_{\epsilon\to 0}\delta(\epsilon) = 0.$$

Definition (5.2.4) implies that Q_ϵ is diagonal with respect to the basis $\{e_k\}_{k\in\mathbb{Z}_0^2}$. It is worth remarking that we only take this particular form of the covariance operator in order to simplify the presentation. The results below can easily be adapted to more general covariance operators with the same smoothing and ergodic properties.

The driving noise, $\sqrt{Q_\epsilon}\, w(t)$, can thus formally be written as the infinite series

$$\sqrt{Q_\epsilon}\, w(t) = \sum_{k\in\mathbb{Z}_0^2} \sigma_{\epsilon,k}\, e_k\, \beta_k(t) := \sum_{k\in\mathbb{Z}_0^2} (1+\delta(\epsilon)|k|^{2\beta})^{-1/2} e_k\, \beta_k(t).$$

Since $\delta(\epsilon)$ converges to zero, as $\epsilon \downarrow 0$, the covariance operator Q_ϵ convergences pointwise to the identity operator as $\epsilon \downarrow 0$. For each fixed $\epsilon > 0$, it is immediate to check that $\sqrt{Q_\epsilon} \in \mathcal{L}(V^r, V^{r+\beta})$. In fact, one can show that

$$\left\| \sqrt{Q_\epsilon} f \right\|_{r+q} \leq \frac{1}{\sqrt{\delta(\epsilon)}} \|f\|_r\,, \tag{5.2.5}$$

for any $r \in \mathbb{R}$, $q \leq \beta$ and $f \in V^r$. Moreover, Q_ϵ is a trace class operator in H if and only if $\beta > 1$. This means that the Wiener process $\sqrt{Q_\epsilon}\, w$ is H-valued only when $\beta > 1$.

By taking the Leray projection of equation (5.1.1), we obtain the following stochastic evolution problem:

$$\begin{cases} du(t) + (Au(t) + B(u(t)))dt = \sqrt{\epsilon\, Q_\epsilon}\, dw(t), \\ \\ u(0) = x. \end{cases} \tag{5.2.6}$$

We assume the initial condition x is an element of H. As is well-known, (see [10] or Chapter 15 of [25]), under the assumption that $\beta > 0$, equation (5.2.6) admits a unique generalized solution, $u_\epsilon^x \in C([0,T];H)$. That is, there exists a progressively measurable process, u_ϵ^x taking values in $C([0,T];H)$, $\mathbb{P}$-a.s. for any $T>0$, such that

$$
\begin{aligned}
\langle u_\epsilon^x(t), h\rangle_H = \langle x, h\rangle_H - \int_0^t \langle u_\epsilon^x(s), Ah\rangle_H \\
+ \int_0^t \langle B(u_\epsilon^x(s), h), u_\epsilon(s)\rangle_H + \langle \sqrt{\epsilon\, Q_\epsilon}\, w(t), h\rangle_H, \quad \mathbb{P}-a.s.,
\end{aligned}
$$

for any $h \in D(A)$ and $t \in [0,T]$.

The condition $\beta > 0$ is not enough to ensure the existence and uniqueness of an invariant measure for equation (5.2.6). In the last twenty five years there has been an extremely intense activity aimed to the study of the ergodic properties of randomly perturbed PDEs in fluid dynamics and in particular of equation (5.1.1). As shown for instance in the monograph [39], a sufficient condition for this is that Q_ϵ be trace-class in H and $\sigma_{\delta(\epsilon),k} \neq 0$ for all k. Notice that if $\beta > 1$, then Q_ϵ is a trace-class operator and by applying Itô's formula we get

$$
\mathbb{E}\,\|u_\epsilon^x(t)\|_H^2 + 2\int_0^t \mathbb{E}\|u_\epsilon^x(s)\|_V^2\, ds = \|x\|_H^2 + t\,\epsilon\,\mathrm{Tr}\, Q_\epsilon \leq \|x\|_H^2 + t\,\epsilon\,\delta_\epsilon^{-1/\beta}. \tag{5.2.7}
$$

This means that $u_\epsilon^x \in L^2(\Omega; C([0,T];H) \cap L^2(0,T;V))$ and, in particular, for every $\epsilon > 0$ there exists an invariant measure.

Now, let $\{\nu_\epsilon\}_{\epsilon>0}$ be this family of invariant measures. Each ν_ϵ is ergodic in the sense that

$$
\lim_{T\to\infty} \frac{1}{T}\int_0^T f(u_\epsilon^x(t))dt = \int_H f(h)d\nu_\epsilon(x),
$$

for all $x \in H$ and Borel-measurable $f : H \to \mathbb{R}$. If

$$
\sup_{\epsilon\in(0,1)} \epsilon\,\delta_\epsilon^{-1/\beta} < \infty, \tag{5.2.8}
$$

we have that the family $\{\nu_\epsilon\}_{\epsilon>0}$ is tight in H. Actually, due to (5.2.7) and the invariance of ν_ϵ, for every $T>0$ we have

$$\int_H \|x\|_V^2\, d\nu_\epsilon(x) = \frac{1}{T}\int_0^T \int_H \mathbb{E}\,\|u_\epsilon^x(t)\|_V^2\, d\nu_\epsilon(x)\, dt = \frac{1}{T}\int_H \int_0^T \mathbb{E}\,\|u_\epsilon^x(t)\|_V^2\, dt\, d\nu_\epsilon(x)$$

$$\leq \frac{1}{2\,T}\int_H \|x\|_H^2\, d\nu_\epsilon(x) + \frac{1}{2}\,\epsilon\,\delta_\epsilon^{-1/\beta} \leq \frac{1}{2\,T}\int_H \|x\|_V^2\, d\nu_\epsilon(x) + \frac{1}{2}\,\epsilon\,\delta_\epsilon^{-1/\beta}.$$

Then, thanks to (5.2.8), if we choose $T>1$ we get

$$\sup_{\epsilon\in(0,1)} \int_H \|x\|_V^2\, d\nu_\epsilon(x) < \infty,$$

and this implies the tightness of $\{\nu_\epsilon\}_{\epsilon\in(0,1)}$ in H. In fact, provided that

$$\lim_{\epsilon\to 0} \epsilon\,\delta_\epsilon^{-1/\beta} = 0,$$

we have that

$$\nu_\epsilon \rightharpoonup \delta_0, \qquad \text{as } \epsilon \downarrow 0.$$

The purpose of this paper is to quantify the rate of this convergence through a large deviations principle. The main result of this paper is the following.

Theorem 5.2.1. *Assume that Q_ϵ has the form given in (5.2.4), for some $\beta>2$. Moreover, suppose that*

$$\lim_{\epsilon\to 0} \delta(\epsilon) = 0, \qquad \lim_{\epsilon\to 0} \epsilon\,\delta(\epsilon)^{-2/\beta} = 0.$$

Then the family of invariant measures $\{\nu_\epsilon\}_{\epsilon>0}$ of equation (5.2.6) satisfies a large *deviations principle in H with rate function given by*

$$U(x) = \begin{cases} \|x\|_1^2, & x \in V, \\ \\ +\infty, & x \in V \setminus H. \end{cases} \tag{5.2.9}$$

We remark here that the rate function, $U(x)$, is really the quasipotential corresponding to equation (5.2.6), whose definition is given in equation (5.4.1). The quasi-potential has the explicit representation given in (5.2.9) in the case of the problem posed on a torus. That formula does not hold in general for the problem posed on a bounded domain with Dirichlet boundary conditions.

5.3 Large deviation principle for the paths

The proof of Theorem 5.2.1 requires a large deviations principle for the solutions to equation (5.2.6). One such large deviations principle is proven in [15], but here we have to proceed differently in order to obtain a result that is uniform with respect to initial conditions in bounded subsets of H. Unlike in [15], we first prove a large deviation principle for the linearized Ornstein-Uhlenbeck process in the space $C([0,T]; L^4)$, and then transfer it back to the appropriate Navier-Stokes process by means of the contraction principle.

5.3.1 LDP for the Ornstein-Uhlenbeck process

Assume that Q_ϵ has the form given in (5.2.4) for $\beta > 0$. For every $\epsilon > 0$ let z_ϵ denote the mild solution to the equation

$$\begin{cases} dz_\epsilon + Az_\epsilon dt = \sqrt{\epsilon\, Q_\epsilon}\, dw(t), \\ \\ z_\epsilon(0) = 0. \end{cases} \tag{5.3.1}$$

It is well-known that z_ϵ is given by the stochastic convolution

$$z_\epsilon(t) = \int_0^t S(t-s)\sqrt{\epsilon\, Q_\epsilon}\, dw(s), \qquad t \geq 0,$$

where $\{S(t)\}_{t\geq 0}$ is the analytic semigroup generated by the operator $-A$ on H. Moreover, it can be shown that $z_\epsilon \in L^p(\Omega; C([0,T]; V^r))$ for any $r < \beta$ and $p \geq 1$ (e.g. see [26]). In this subsection, we prove that the family $\{z_\epsilon\}_{\epsilon>0}$ satisfies a large deviations principle in $C([0,T]; L^4)$. To do so, we first prove that the stochastic convolution z_ϵ converges to 0 in $C([0,T]; L^4)$.

Lemma 5.3.1. *For any $\epsilon > 0$, the solution z_ϵ has trajectories in $C([0,T]; L^p)$, $\mathbb{P}$-a.s. for any $p \in [1, \infty)$. Moreover,*

$$\lim_{\epsilon\to 0} \epsilon \log \delta(\epsilon)^{-1} = 0 \implies \lim_{\epsilon\to 0} \mathbb{E} \sup_{t\in[0,T]} \|z_\epsilon(t)\|^p_{L^p} = 0. \tag{5.3.2}$$

Proof. Fix any $p < \infty$. Thanks to the Burkholder-Davis-Gundy inequality and the uniform boundedness of the basis $\{e_k\}_{k\in\mathbb{Z}_0^2}$, we have

$$
\begin{aligned}
\mathbb{E}\sup_{0\le t\le T}\|z_\epsilon(t)\|_{L^p}^p &= \epsilon^{p/2}\mathbb{E}\sup_{0\le t\le T}\left\|\int_0^t S(t-s)\sqrt{Q_\epsilon}dW(s)\right\|_{L^p}^p \\
&\le \epsilon^{p/2}\int_{\mathbb{T}^2}\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t\sum_{k\in\mathbb{Z}_0^2}e^{-|k|^2(t-s)}\sigma_{\delta(\epsilon),k}e_k(x)d\beta_k(s)\right|^p dx \\
&\le c_p\,\epsilon^{p/2}\int_{\mathbb{T}^2}\Big(\sum_{k\in\mathbb{Z}_0^2}\sigma^2_{\delta(\epsilon),k}|e_k(x)|^2\int_0^T e^{-2|k|^2 s}ds\Big)^{p/2}dx \\
&\le c_p\,\epsilon^{p/2}\Big(\sum_{k\in\mathbb{Z}_0^2}\frac{1}{|k|^2(1+\delta(\epsilon)|k|^{2\beta})}\Big)^{p/2}<\infty.
\end{aligned}
$$

Moreover, (5.3.2) follows by noting that

$$
\begin{aligned}
\sum_{k\in\mathbb{Z}_0^2}\frac{1}{|k|^2(1+\delta(\epsilon)|k|^{2\beta})} &\le \int_1^\infty\frac{1}{r(1+\delta(\epsilon)r^{2\beta})}dr=\int_{\delta(\epsilon)^{1/(2\beta)}}^\infty\frac{1}{r(1+r^{2\beta})}dr \\
&\le \int_{\delta(\epsilon)^{1/(2\beta)}}^1\frac{dr}{r}+\int_1^\infty\frac{dr}{r^{2\beta+1}}\le\frac{1}{2\beta}\log\frac{1}{\delta(\epsilon)}+\frac{1}{2\beta}.
\end{aligned}
$$

□

To prove that the family $\{z_\epsilon\}_{\epsilon>0}$ satisfies a large deviations principle in $C([0,T];L^4)$ we follow the weak convergence approach, outlined in Section 2.2.Recall that this approach involves proving convergence of the solutions to a sequence of controlled versions of the equations. For $\phi\in L^2(\Omega;L^2(0,T;H))$, we denote by $z_{\epsilon,\phi}$ the solution to the equation

$$
dz_{\epsilon,\phi}(t)+Az_{\epsilon,\phi}(t)\,dt=\sqrt{\epsilon\,Q_\epsilon}\,dw(t)+\sqrt{Q_\epsilon}\,\phi(t)\,dt,\qquad z_{\epsilon,\phi}(0)=0,
$$

and we denote by z_ϕ the solution to the so-called skeleton equation

$$
\frac{dz_\phi}{dt}(t)+Az_\phi(t)=\phi(t),\qquad z_\phi(0)=0. \tag{5.3.3}
$$

In the present case, we are proving the non-uniform version of the Laplace principle, so that the set $\mathcal{E}_0$ of initial conditions in Hypothesis 2.2.1 is simply $\mathcal{E}_0 = \{0\}$. In particular, the non-uniform Laplace principle implies the validity of the large deviations principle. Therefore, Theorem 2.2.3 implies the following.

Theorem 5.3.1. *Assume the following hold for any $M \in [0, \infty)$.*

(i) The set

$$\Phi(M) := \Big\{ z \in C([0,T]; L^4) : z = z_\phi, \quad \phi \in L^2(0,T;H),\ \frac{1}{2}\int_0^T \|\phi(t)\|_0^2 \, dt \leq M, \Big\},$$

is a compact subset of $C([0,T]; L^4)$.

(ii) For every $\{\varphi_\epsilon\}_{\epsilon \geq 0} \subset L^2(\Omega; L^2(0,T;H))$, such that

$$\sup_{\epsilon \in (0,1)} \frac{1}{2}\int_0^T \|\varphi_\epsilon(t)\|_0^2 \, dt \leq M,\ \mathbb{P} - a.s., \tag{5.3.4}$$

if φ_ϵ converges to φ_0 in distribution with respect to the weak topology of $L^2(0,T;H)$, as $\epsilon \downarrow 0$, then $z_{\epsilon,\varphi_\epsilon}$ converges to z_{φ_0} in distribution in $C([0,T]; L^4)$, as $\epsilon \downarrow 0$.

Then the family $\{\mathcal{L}(z_\epsilon)\}_{\epsilon > 0}$ satisfies a large deviations principle in $C([0,T]; L^4)$ with rate function

$$J_T(z) = \frac{1}{2} \inf \Big\{ \int_0^T \|\phi(t)\|_0^2 \, dt : \phi \in L^2(0,T;H), z = z_\phi \Big\}. \tag{5.3.5}$$

Thus, to prove the large deviations principle it remains to prove hypotheses (i) and (ii) of the above theorem. Note that hypothesis (i) is precisely the statement that J_T is a good rate function.

Theorem 5.3.2. *Assume that*

$$\lim_{\epsilon\to 0}\delta(\epsilon)=0, \qquad \lim_{\epsilon\to 0}\epsilon\log\frac{1}{\delta(\epsilon)}=0.$$

Then the family $\{\mathcal{L}(z_\epsilon)\}_{\epsilon>0}$ *of solutions to equation* (5.3.1) *satisfies a large deviations principle in* $C([0,T];L^4)$ *with rate function*

$$J_T(z)=\begin{cases}\frac{1}{2}\int_0^T \|z'(t)+Az(t)\|_0^2\,dt & \text{if } z\in W^{1,2}(0,T;H)\cap L^2(0,T;D(A)),\\ +\infty & \text{otherwise.}\end{cases} \tag{5.3.6}$$

Proof. In view of Theorem 5.3.1, it suffices to show that conditions (i) and (ii) in Theorem 5.3.1 hold true. Equality of the rate functions defined in equations (5.3.5) and (5.3.6) follows immediately from the fact that $z_\phi=z_\varphi$ implies that $\phi=\varphi$.

Step 1. We first verify condition (i). Suppose that $z\in\Phi(M)$, so that $z=z_\phi$ for some $\phi\in L^2(0,T;H)$ satisfying

$$\frac{1}{2}\int_0^T\|\phi(t)\|_0^2\,dt\le M. \tag{5.3.7}$$

For any $\zeta\in(0,1)$, the function $z_\phi(t)=\int_0^t S(t-s)\phi(s)ds$ can be rewritten as $z_\phi=\Gamma_\zeta(Y_\zeta(\phi))$ where

$$\Gamma_\zeta(Y)(t):=c_\zeta\int_0^t(t-s)^{\zeta-1}S(t-s)Y(s)ds,$$

and

$$Y_\zeta(\phi)(s):=\int_0^s(s-r)^{-\zeta}S(s-r)\phi(r)dr.$$

It is possible to show that for any $\zeta \in (0, \frac{1}{2})$, $p \geq 2$, $\rho \in (0, 1)$ and $\delta \in (0, \frac{1}{2})$ such that $\delta + \dfrac{\rho}{2} < \zeta - \dfrac{1}{p}$,

$$\Gamma_\zeta : L^p(0, T; H) \to C^\delta([0, T]; V^\rho), \tag{5.3.8}$$

is a continuous linear mapping (see Appendix A of [25]). Moreover, we have by Young's inequality that

$$\begin{aligned} \|Y_\zeta(\phi)\|^p_{L^p(0,T;H)} &= \int_0^T \left\| \int_0^s (s-r)^{-\zeta} S(s-r)\phi(r)dr \right\|_0^p dt \\ &\leq \int_0^T \Big(\int_0^s (s-r)^{-\zeta} \|\phi(r)\|_0 \, dr \Big)^p dt \\ &\leq \Big(\int_0^T t^{-\frac{2\zeta p}{p+2}} dt \Big)^{\frac{p+2}{2}} \|\phi\|^p_{L^2(0,T;H)}, \end{aligned} \tag{5.3.9}$$

which is finite provided that $\zeta < \frac{1}{2} + \frac{1}{p}$. Hence $z \in C^\delta(0, T; V^\rho)$ for any δ, ρ satisfying $\delta + \frac{\rho}{2} < \frac{1}{2}$. Moreover, thanks to (5.3.7),

$$\|z\|_{C^\delta(0,T;V^\rho)} \leq c_{p,\delta,\rho}\sqrt{M},$$

so that $\Phi(M)$ is a bounded subset of $C^\delta(0, T; V^\rho)$ and thus a compact subset of $C([0, T]; L^4)$.

Step 2. Next, we verify condition (ii) in Theorem 5.3.1. Let $M > 0$ and let $\{\varphi_\epsilon\}_{\epsilon \geq 0}$ be a sequence in $L^2(\Omega; L^2(0, T; H))$ satisfying (5.3.4). Thanks to the Skorohod theorem, there exists a probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \{\bar{\mathcal{F}}_t\}_{t \geq 0}, \bar{\mathbb{P}})$, a cylindrical Wiener process $\bar{W}(t)$, and collection $\{\bar{\varphi}_\epsilon\}_{\epsilon \geq 0}$ in $L^2(\bar{\Omega}; L^2(0, T; H))$ such that φ_ϵ and $\bar{\varphi}_\epsilon$ have the same distributions and

$$\lim_{\epsilon \to 0} \bar{\varphi}_\epsilon = \bar{\varphi}_0, \ \mathbb{P} - \text{a.s.},$$

with respect to the weak topology of $L^2(0,T;H)$. If we show that $z_{\epsilon,\bar{\varphi}_\epsilon}$ converges to $z_{\bar{\varphi}_0}$ in $C([0,T];L^4)$, $\mathbb{P}$-a.s., then condition (ii) will follow.

To simplify our notation, we dispense with the bars. Now, for any $t \geq 0$, we have

$$z_{\epsilon,\varphi_\epsilon}(t) - z_\varphi(t) = \sqrt{\epsilon}\int_0^t S(t-s)\sqrt{Q_\epsilon}dW(s) + \int_0^t S(t-s)\Big[\sqrt{Q_\epsilon}\varphi_\epsilon(s) - \varphi(s)\Big]ds$$

$$=: J_1^\epsilon(t) + J_2^\epsilon(t).$$

Thanks to Lemma 5.3.1, we have

$$\lim_{\epsilon\to 0}\mathbb{E}\|J_1^\epsilon\|^4_{C([0,T];L^4)} = 0.$$

To handle the control terms, we observe that $\sqrt{Q_\epsilon}\,\varphi_\epsilon$ converges to φ weakly in $L^2(0,T;H)$. Indeed, for any $h \in L^2(0,T;H)$, it follows that

$$\Big|\langle\sqrt{Q_\epsilon}\varphi_\epsilon - \varphi, h\rangle_{L^2(0,T;H)}\Big| = \Big|\langle\varphi_\epsilon, (\sqrt{Q_\epsilon} - I)h\rangle_{L^2(0,T;H)} + \langle\varphi_\epsilon - \varphi, h\rangle_{L^2(0,T;H)}\Big|$$

$$\leq \sqrt{2M}\,\Big\|(\sqrt{Q_\epsilon} - I)h\Big\|_{L^2(0,T;H)} + \Big|\langle\varphi_\epsilon - \varphi, h\rangle_{L^2(0,T;H)}\Big|,$$

which converges to 0, $\mathbb{P}-$a.s. as $\epsilon \to 0$ since Q_ϵ converges to $\mathbb{1}$ pointwise in H and φ_ϵ converges to φ weakly. Moreover, we already showed in Step 1 that the solution map $\Gamma : L^2(0,T;H) \to C([0,T];L^4)$ given by

$$\Gamma(\phi)(t) = \int_0^t S(t-s)\phi(s)ds,$$

is a compact operator. Since compact operators map weakly convergent sequences to strongly convergent sequences, it follows

$$\lim_{\epsilon\to 0}\|J_2\|_{C([0,T];L^4)} = 0, \qquad \mathbb{P} - \text{a.s.}$$

$\square$

5.3.2 Uniform LDP for the Navier-Stokes process

To obtain a uniform large deviations principle for the solutions to equation (5.2.6), we will apply the contraction principle to the large deviations principle for the solutions to equation (5.3.1).

For every $x \in H$, let $\mathcal{F}_x : L^4(0,T;L^4) \to C([0,T];H)$ be the family of mappings that associate to any $z \in L^4(0,T;L^4)$ the solution to the equation

$$\begin{cases} du(t) + Au(t)dt + B\Big(u(t) + z(t)\Big)dt = 0, \\ \\ u(0) = x \in H. \end{cases}$$

In particular, we see that $I + \mathcal{F}_x$ maps a trajectory of z_ϵ to a trajectory of u_ϵ^x. Throughout the remainder, we will use the shorthand

$$B_Y(y,r) := \{h \in Y : \|h - y\|_Y < r\}, \qquad y \in Y, \quad r > 0,$$

for the open ball in the Banach space Y, centered at y and of radius r. When $y = 0$ we will just write $B_Y(r)$.

The proof of the following result is from [10]. Here we give a brief sketch of it to emphasize the right dependence on the initial conditions.

Lemma 5.3.2. *For every fixed $T > 0$ the mappings $\mathcal{F}_x : L^4(0,T;H) \to C([0,T];L^4)$ are locally Lipschitz continuous, uniformly over x in bounded sets of H. That is, for* *any $r>0$ and $R>0$, there exists constant $L_{r,R}>0$ such that*

$$\sup_{x\in B_H(r)} \|\mathcal{F}_x(f)-\mathcal{F}_x(g)\|_{C([0,T];H)} \leq L_{r,R}\,\|f-g\|_{L^4(0,T;L^4)}\,, \qquad f,g\in B_{L^4(0,T;L^4)}(R). \tag{5.3.10}$$

Proof. For every $f,g\in B_{L^4(0,T;L^4)}(R)$, we define $u:=\mathcal{F}_x(f)-\mathcal{F}_x(g)$. By proceeding as in [10], we have

$$\begin{aligned}\|u(t)\|_H^2+\int_0^t \|u(s)\|_1^2\,ds &\leq c\,\|g-f\|^2_{L^4(0,t;L^4)}\Big[\,\|\mathcal{F}_x(f)\|_{L^\infty(0,T;H)}\,\|\mathcal{F}_x(f)\|_{L^2(0,T;V)}\\
&\quad+\|\mathcal{F}_x(g)\|_{L^\infty(0,T;H)}\,\|\mathcal{F}_x(g)\|_{L^2(0,T;V)}+\|f\|^2_{L^4(0,T;L^4)}+\|g\|^2_{L^4(0,T;L^4)}\Big]\\
&\quad+c\int_0^t\Big[\,\|\mathcal{F}_x(g)(s)\|_1^2+\|g(s)\|^4_{L^4}\Big]\,\|u(s)\|_H^2\,ds.\end{aligned}$$

Now, for an arbitrary $f\in\ L^4(0,T;L^4)$

$$\begin{aligned}&\frac{1}{2}\,\|\mathcal{F}_x(f)(t)\|_H^2+\int_0^t\|\mathcal{F}_x(f)(t)\|_1^2\,ds\\
&\leq\frac{1}{2}\,\|x\|_H^2+\int_0^t\Big[|b(\mathcal{F}_x(f)(s),f(s),\mathcal{F}_x(f)(s))|+|b(f(s),f(s),\mathcal{F}_x(f)(s))|\Big]ds\\
&\leq\frac{1}{2}\,\|x\|_H^2+\int_0^t\Big[\,\|\mathcal{F}_x(f)(s)\|_H^{1/2}\,\|\mathcal{F}_x(f)(s)\|_1^{3/2}\,\|f(s)\|_{L^4}+\|f(s)\|^2_{L^4}\,\|\mathcal{F}_x(f)(s)\|_1\Big]ds\\
&\leq\frac{1}{2}\,\|x\|_H^2+\frac{1}{2}\int_0^t\|\mathcal{F}_x(f)(s)\|_1^2\,ds+c\int_0^t\|f(s)\|^4_{L^4}\,\|\mathcal{F}_x(f)(s)\|_H^2\,ds+c\,\|f\|^4_{L^4(0,t;L^4)}\,,\end{aligned} \tag{5.3.11}$$

which implies that

$$\|\mathcal{F}_x(f)(t)\|_H^2+\int_0^t\|\mathcal{F}_x(f)(s)\|_1^2\,ds\leq(\|x\|_H^2+\|f\|^4_{L^4(0,t;L^4)})\exp\Big(\,\|f\|^4_{L^4(0,t;L^4)}\Big).$$

This implies that if $f, g \in B_{L^4(0,T;L^4)}(R)$, and $x \in B_H(R)$, there exists $L_{r,R} > 0$ such that

$$\begin{aligned} &\|u(t)\|_H^2 + \int_0^t \|u(s)\|_V^2\, ds \\ &\leq L_{r,R}\, \|g - f\|^2_{L^4(0,t;L^4)} \exp\Big(c \int_0^t \Big[\, \|\mathcal{F}_x(g)(s)\|_V^2 + \|g(s)\|_{L^4}^4 \Big] ds \Big). \end{aligned} \tag{5.3.12}$$

By using again (5.3.11) to estimate $\mathcal{F}_x(g)$, we obtain (5.3.10) □

In the proof of the main result, Theorem 5.2.1, both versions of the uniform large deviations principle (see Section 2.2.5) will be required. As we saw in the sketch of the proof of the lower bound of Theorem 2.3.3, we needed the validity of a FWULDP that is uniform over initial conditions in *bounded* sets. Moreover, it turns out that in the proof of the upper bound, we need the validity of a DZULDP that is uniform over initial conditions in *compact* sets. The FWULDP over bounded sets will be obtained via the uniform contraction principle Theorem 2.2.4 along with Lemma 5.3.2 and Theorem 5.3.2. The DZULDP over compact sets will be obtained from the FWULDP using Proposition 2.2.4.

To define the Navier-Stokes rate function, we first define the Hamiltonian

$$\mathcal{H}(u) := u' + Au + B(u), \qquad u \in D(\mathcal{H}) := W^{1,2}(0,T;V^{-1}) \cap L^2(0,T;V)).$$

For $u \in D(\mathcal{H})$ in this space, the nonlinearity $B(u)$ is a well-defined element of $L^2(0,T;V^{-1})$. For any $x \in H$ and $u \in C([0,T];H)$, we define

$$I^x(u) = \begin{cases} \dfrac{1}{2} \displaystyle\int_0^T \|\mathcal{H}(u)(t)\|_H^2\, dt & \text{if } \mathcal{H}(u) \in L^2(0,T;H), \text{ and } u(0) = x, \\ \\ +\infty & \text{otherwise.} \end{cases}$$

Theorem 5.3.3. *Assume that*

$$\lim_{\epsilon\to 0}\delta(\epsilon)=,\qquad \lim_{\epsilon\to 0}\epsilon\log\frac{1}{\delta(\epsilon)}=0.$$

If u_ϵ^x is the solution to equation (5.2.6), then for any $R>0$, the family $\{\mathcal{L}(u_\epsilon^x)\}_{\epsilon>0}$ satisfies a Freidlin-Wentzell uniform large deviations principle in $C([0,T];H)$ with rate functions I^x uniformly with respect to $x\in B_H(R)$.

Proof. First of all, notice that $u_\epsilon^x=(I+\mathcal{F}_x)(z_\epsilon)$. Lemma 5.3.2 implies that the mapping

$$I+\mathcal{F}_x : C([0,T];L^4)\to C([0,T];H),$$

is locally Lipschitz, uniformly on bounded sets. Therefore, thanks to the contraction principle, Theorem 2.2.4, and Theorem 5.3.2 the family $\{(I+\mathcal{F}_x)(z_\epsilon)\}=\{u_\epsilon^x\}$ satisfies a Freidlin-Wentzell uniform large deviations principle with rate function

$$I_T^x(u)=\inf\Big\{J_T(z): u=z+\mathcal{F}_x(z),\ z\in W^{1,2}(0,T;H)\cap L^2(0,T;D(A))\Big\}.$$

If $u\in D(\mathcal{H})$ and $u(0)=x$, then $\mathcal{H}(u)\in L^2(0,T;V^{-1})$ and u is a weak solution to

$$\begin{cases} du(t)+[Au(t)+B(u(t))]dt=\mathcal{H}(u)(t)dt, \\ u(0)=x. \end{cases}\tag{5.3.13}$$

Note that $u\in D(\mathcal{H})$ implies that equation (5.3.3) with forcing $\phi=\mathcal{H}(u)$ has a unique weak solution $z_\phi\in X$. In particular this also implies that $\mathcal{F}_x(z_\phi)\in D(\mathcal{H})$ and $u=z_\phi+\mathcal{F}_x(z_\phi)$. This decomposition is unique. Indeed, if $u=z+\mathcal{F}_x(z)$ for some other $z \in D(\mathcal{H})$, then $u - \mathcal{F}_x(z)$ would again be a weak solution to equation (5.3.3) with forcing $\phi = \mathcal{H}(u)$ so that $z = z_\phi$. This implies that

$$J_T(z_\phi) = \frac{1}{2}\int_0^T \|\mathcal{H}(u)(t)\|_H^2 \, dt,$$

whenever $\mathcal{H}(u) \in L^2(0,T;H)$. □

Remark 5.3.1. *Notice that both the proof of Theorem 5.3.2 and the proof of Theorem 5.3.3 do not require periodic boundary conditions.*

□

Corollary 5.3.1. *Assume that*

$$\lim_{\epsilon\to 0}\delta(\epsilon) = 0, \qquad \lim_{\epsilon\to 0}\epsilon\log\frac{1}{\delta(\epsilon)} = 0.$$

Let $K \subset H$ be a compact set. Then the family $\{\mathcal{L}(u_\epsilon^x)\}_{\epsilon>0}$ of solutions to equation (5.2.6) satisfies a Dembo-Zeitouni uniform large deviations principle in $C([0,T];H)$ with rate functions $\{I^x\}_{x\in K}$ uniformly with respect to $x \in K$.

Proof. In view of Proposition 2.2.4 (Theorem 2.7 of [55]), to prove equivalence of the two uniform large deviation principles over a compact subset of H, it suffices to show that for every fixed $s \geq 0$ the mapping

$$x \in H \mapsto \Phi^x(s) := \{u \in C([0,T];H) : I^x(u) \leq s\},$$

is continuous with respect to the Hausdorff metric. That is, we must show that for any $\{x_n\}_{n=1}^\infty \subset H$ such that $x_n \to x \in H$,

$$\lim_{n\to\infty}\max\left(\sup_{u\in\Phi^{x_n}(s)}\mathrm{dist}_{C([0,T];H)}(u,\Phi^x(s)), \sup_{u\in\Phi^x(s)}\mathrm{dist}_{C([0,T];H)}(u,\Phi^{x_n}(s))\right) = 0.$$

This is immediately implied by the continuity of the Navier-Stokes equations with respect to initial conditions. Indeed, suppose that u_ϕ^x is a solution to the equation,

$$\begin{cases} du(t) + (Au(t) + B(u(t)))dt = \phi(t)dt, \\ \\ u(0) = x, \end{cases} \tag{5.3.14}$$

with driving force $\phi \in L^2(0,T;H)$. Then by standard energy estimates (see for instance, [39]), we have

$$\begin{aligned} \sup_{0\le t\le T} \left\| u_\phi^x(t) - u_\phi^y(t) \right\|_H^2 + \int_0^T \left\| u_\phi^x(t) - u_\phi^y(t) \right\|_V^2 dt \\ \le \|x-y\|_H^2 \exp\left(c \int_0^T \left\| u_\phi^y(t) \right\|_V^2 dt \right) \\ \le \|x-y\|_H^2 \exp\left(c\big[\, \|y\|_H^2 + \|\phi\|_{L^2(0,T;H)}^2 \big] \right). \end{aligned}$$

Now, if $u \in \Phi^x(s)$, then $\varphi := \mathcal{H}(u) \in L^2(0,T;H)$, $\frac{1}{2}\|\varphi\|_{L^2(0,T;H)}^2 \le s$ and u solves equation (5.3.13). But then, the weak solution $v \in W^{1,2}(0,T;V^{-1}) \cap L^2(0,T;V)$ to

$$\begin{cases} dv(t) + [Av(t) + B(v(t))]dt = \varphi(t)dt, \\ \\ v(0) = y, \end{cases}$$

belongs to $\Phi^y(s)$. Therefore,

$$\mathrm{dist}_{C([0,T];H)}(u, \Phi^y(s)) \le \|u - v\|_{C([0,T];H)} \le c_s(\|y\|_H)\ \|x-y\|_H\,,$$

for some continuous increasing function $c_s : [0,+\infty) \to [0,+\infty)$. Since this is true for arbitrary $u \in \Phi^x(s)$, it follows that

$$\sup_{u\in\Phi^x(s)} \mathrm{dist}_{C([0,T];H)}(u, \Phi^y(s)) \le c_s(\|y\|_H)\ \|x-y\|_H\,,$$

which implies the result, since $\sup_{n\in\mathbb{N}} \|x_n\|_H < \infty$. $\square$

5.4 Proof of Theorem 5.2.1

We start this section with the description of the quasi-potential associated with equation (5.2.6). To simplify notation, for any $T > 0$ we will denote

$$I_T(u) := \frac{1}{2}\int_0^T \|\mathcal{H}(u)(t)\|_H^2 \, dt,$$

whenever $\mathcal{H}(u) \in L^2(0,T;H)$. In addition, we set

$$I_T^y(u) := \begin{cases} I_T(u), & \text{if } u(0) = y, \\ \\ +\infty, & \text{otherwise.} \end{cases}$$

The quasi-potential, $U : H \to [0,+\infty]$ is then defined as

$$U(x) := \inf\{I_T(u) : T > 0, u \in C([0,T];H), u(0) = 0, u(T) = x\}. \tag{5.4.1}$$

For any $x \in H$, the quasipotential $U(x)$ gives the minimum action of all paths that start at 0 and end at x. Since 0 is an asymptotically attracting equilibria for the Navier-Stokes equations, $U(x)$ will govern the long-time dynamics and asymptotic behavior of the invariant measures.

In the particular case of the Navier-Stokes equations on the torus, the orthogonality of $B(u)$ and Au can be taken advantage of to provide an explicit formula for the quasipotential. In fact, as proven in [9, Theorem 7.1] we have that for any $x \in H$

$$U(x) = \begin{cases} \|x\|_V^2, & x \in V, \\ \\ +\infty, & x \in H \setminus V. \end{cases} \tag{5.4.2}$$

Now, we proceed with the proof of Theorem 5.2.1. Some of the steps of the proof are analogous to those used in [8, Theorem 4.5], where a large deviation principle for the invariant measures of the $2D$ stochastic Navier-stokes equation is studied, under the assumption that the covariance of the noise does not depend on ϵ. In those steps our arguments will be less detailed and we refer the reader to [8]. On the other hand, our arguments will be fully detailed in those steps of the proof that deviate from [8].

5.4.1 Lower bound

Proposition 5.4.1. *Under the assumptions of Theorem 5.2.1, the family of invariant measures $\{\nu_\epsilon\}_{\epsilon>0}$ of equation (5.2.6) satisfies the large deviations principle lower bound in H with rate function $U(x)$. That is, for any $x \in H$, $\delta > 0$ and $\gamma > 0$, there exists $\epsilon_0 > 0$ such that*

$$\nu_\epsilon(B_H(x,\delta)) \geq \exp\Big(-\frac{U(x)+\gamma}{\epsilon}\Big), \qquad \epsilon \leq \epsilon_0.$$

Proof. Fix $x \in H$, and any $\delta > 0$, $\gamma > 0$ and $T > 0$. We assume that $U(x) < \infty$ or else there is nothing to prove. Suppose that $\{v^y\}_{y\in H} \subset C([0,T];H)$ is a family of paths satisfying

$$\sup_{y\in H} \|v^y(T) - x\|_H < \delta/2.$$

Thanks to the invariance of ν_ϵ, we have

$$\begin{aligned}\nu_\epsilon(B_H(x,\delta)) &= \int_H \mathbb{P}(\|u_\epsilon^y(T)-x\|_H<\delta)\ d\nu_\epsilon(y)\\ &\geq \int_H \mathbb{P}(\|u_\epsilon^y - v^y\|_{C([0,T];H)}<\delta/2)\ d\nu_\epsilon(y)\\ &\geq \int_{B_H(0,R)} \mathbb{P}(\|u_\epsilon^y - v^y\|_{C([0,T];H)}<\delta/2)\ d\nu_\epsilon(y)\\ &\geq \nu_\epsilon(B_H(0,R)) \inf_{y\in B_H(0,R)} \mathbb{P}(\|u_\epsilon^y - v^y\|_{C([0,T];H)}<\delta/2).\end{aligned}$$

Since the invariant measures are becoming concentrated around 0, as $\epsilon\downarrow 0$, we have

$$\lim_{\epsilon\to 0}\nu_\epsilon(B_H(0,R))=1,$$

for any $R>0$. Thus, we can pick $\epsilon_1(R)>0$ small enough that

$$\nu_\epsilon(B_H(0,R))\geq \frac{1}{2},\qquad \epsilon\leq\epsilon_1(R).$$

Thanks to Theorem 5.3.3, a Freidlin-Wentzell uniform large deviations principle holds. Then, for every $s_0>0$ there exists $\epsilon_2(R)>0$ such that for any $v^y\in C([0,T];H)$ with $I_T^y(v^y)\leq s_0$,

$$\inf_{y\in B_H(0,R)}\mathbb{P}\left(\|u_\epsilon^y-v^y\|_{C([0,T];H)}<\delta/2\right)\geq \inf_{y\in B_H(0,R)}\exp\left(-\frac{1}{\epsilon}\left[I_T^y(\varphi^y)+\gamma/2\right]\right),$$

for every $\epsilon\leq\epsilon_2(R)$. Therefore, to complete the proof, it remains to find a T large enough that for each $y\in B_H(0,R)$, there exists a path $v^y\in C([0,T];H)$ with $v^y(0)=y$ that satisfies

(a) $I_T(v^y)\leq U(x)+\gamma/2$,

(b) $\|v^y(T) - x\|_H < \delta/2$.

The paths we choose are the solutions u^y_φ to the controlled Navier Stokes equations, equation (5.3.14), with initial condition $y \in H$ and control $\varphi \in L^2(0, T; H)$, defined by

$$\varphi(t) = \begin{cases} 0 & \text{if } 0 \le t \le T_1, \\ \\ \bar{\varphi}(t - T_1) & \text{if } T_1 \le t \le T_1 + T_2, \end{cases}$$

with T_1 and T_2 to be chosen. Here, $\bar{\varphi} \in C([0, T_2]; H)$ is a path such that $u^0_{\bar{\varphi}}(0) = 0$ and $u^0_{\bar{\varphi}}(T_2) = x$ with $I_{T_2}(u^0_{\bar{\varphi}}) \le U(x) + \gamma/2$. Such a T_2 and $\bar{\varphi}$ exist by the definition of the quasipotential U. Meanwhile, $T_1 = T_1(\lambda)$ is taken large enough that the solutions $\{u^y_0\}_{y \in B_H(0,R)}$ to the unforced Navier-Stokes equations satisfy

$$\sup_{y \in B_H(0,R)} \|u^y_0(T_1)\|_H < \lambda,$$

for some small λ. Clearly point (a) is satisfied since the path contributes nothing to the action integral on the interval $[0, T_1]$. Point (b) follows by noting that the controlled Navier Stokes equations are continuous with respect to initial conditions. Indeed, since $u^y_0(T_1) \in B_H(0, \lambda)$, we have by a standard estimate (for example see Proposition 2.1.25 of [39]) that

$$\begin{aligned} \left\|u^y_\varphi(T_1 + T_2) - x\right\|_H &\le \sup_{z \in B_H(0,\lambda)} \left\|u^z_{\bar{\varphi}}(T_2) - u^0_{\bar{\varphi}}(T_2)\right\|_H \\ &\le \sup_{z \in B_H(0,\lambda)} \|z\|_H \exp\Big(c\,\|z\|^2_H + c\,\|\bar{\varphi}\|^2_{L^2(0,T_2;H)}\Big). \end{aligned}$$

This implies point (b) if λ is taken small enough. We conclude the proof upon taking $\epsilon_0 := \min(\epsilon_1, \epsilon_2)$.

$\square$

5.4.2 Upper bound

Proposition 5.4.2. *Under the assumptions of Theorem 5.2.1, the family of invariant measures $\{\nu_\epsilon\}_{\epsilon>0}$ of equation (5.2.6) satisfies the large deviations principle upper bound in H with rate function $U(x)$. That is, for any $s \geq 0$, $\delta > 0$ and $\gamma > 0$, there exists $\epsilon_0 > 0$ such that*

$$\nu_\epsilon\left(\{h \in H : \mathrm{dist}_H(h, \Phi(s)) > \delta\}\right) \leq \exp\left(-\frac{s-\gamma}{\epsilon}\right), \qquad \epsilon \leq \epsilon_0.$$

where

$$\Phi(s) := \{y \in H : U(y) \leq s\}.$$

The proof requires the following three lemmas.

Lemma 5.4.1 (*Exponential Estimate*)**.** *Assume that Q has the form given in (5.2.4) for some $\beta > 2$. Moreover, suppose that*

$$\lim_{\epsilon \to 0} \delta(\epsilon) = 0, \qquad \lim_{\epsilon \to 0} \epsilon \delta(\epsilon)^{-2/\beta} = 0.$$

Then for any $s > 0$ there exist $\epsilon_s > 0$ and $R_s > 0$ such that

$$\nu_\epsilon(B_V(0, R_s)) \geq 1 - \exp\left(-\frac{s}{\epsilon}\right), \qquad \epsilon \leq \epsilon_s.$$

Proof. Fix $R > 0$, $\epsilon > 0$ and $\gamma > 0$ and let u_ϵ^0 be the solution of equation (5.2.6). Thanks to the ergodicity of ν_ϵ, we have

$$\nu_\epsilon(B_V^c(0,R)) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}(u_\epsilon^0(s) \in B_V^c(0,R))\, ds$$

$$\leq \exp\Big(-\frac{R^2}{2\epsilon}\Big)\frac{1}{T}\limsup_{T\to\infty}\int_0^T \mathbb{E}\exp\left(\frac{\|u_\epsilon^0(s)\|_V^2}{2\epsilon}\right) ds. \tag{5.4.3}$$

To estimate the expectation of the exponential, we apply the Ito formula to the functional $F : \mathbb{R}\times V \to \mathbb{R}$ defined by

$$F(t,v) = \exp\left(t + \frac{\|v\|_V^2}{2\epsilon}\right),$$

whose derivatives are given by

$$D_t F(t,u) = F(t,u),$$

and

$$D_u F(t,u) = \frac{1}{\epsilon}F(t,u)u, \qquad D_u^2 F(t,u) = \frac{1}{\epsilon^2}F(t,u)u\otimes u + \frac{1}{\epsilon}F(t,u)I.$$

Formal application of the Ito formula to the solution u_ϵ^x to equation (5.2.6) implies that

$$\mathbb{E}F(t,u_\epsilon^x(t)) = F(0,x) + \mathbb{E}\int_0^t \Big[D_tF(s,u_\epsilon^x(s)) + \langle D_uF(s,u_\epsilon^x(s)), -Au_\epsilon^x(s) - B(u_\epsilon^x(s))\rangle_V$$

$$+ \frac{\epsilon}{2}\sum_{k\in\mathbb{Z}_0^2}^{\infty} \langle D_u^2F(s,u_\epsilon^x(s))Q_\epsilon e_k, e_k\rangle_V\Big]ds$$

$$= F(0,x) + \mathbb{E}\int_0^t F(s,u_\epsilon^x(s))\Big[1 - \frac{1}{\epsilon}\|u_\epsilon^x(s)\|_2^2$$

$$+ \frac{\epsilon}{2}\sum_{k\in\mathbb{Z}_0^2}\frac{1}{\epsilon^2}\left(|\langle u_\epsilon^x(s), Q_\epsilon e_k\rangle_V|^2 + \frac{1}{\epsilon}\langle Q_\epsilon e_k, e_k\rangle_V\right)\Big]ds$$

$$= F(0,x) + \mathbb{E}\int_0^t F(s,u_\epsilon^x(s))\Big[1 - \frac{1}{\epsilon}\|u_\epsilon^x(s)\|_2^2$$

$$+ \frac{1}{2}\sum_{k\in\mathbb{Z}_0^2}\left(\frac{1}{\epsilon}\sigma_{\epsilon,k}^2|\langle u_\epsilon^x(s), Ae_k\rangle_H|^2 + |k|^2\sigma_{\epsilon,k}^2\right)\Big]ds$$

$$\leq F(0,x) + \mathbb{E}\int_0^T F(s,u_\epsilon^x(s))\Big[1 - \frac{1}{2\epsilon}\|u_\epsilon^x(s)\|_2^2 + \frac{1}{2}\sum_{k\in\mathbb{Z}_0^2}|k|^2\sigma_{\epsilon,k}^2\Big]ds,$$

where in the second line we used identity (5.2.3) to dispose of the nonlinearity and in the fourth line we used that $|\sigma_{\epsilon,k}| \leq 1$ for any $k \in \mathbb{Z}_0^2$ and $\epsilon > 0$.

Now, since $\beta > 2$, we have

$$P_\epsilon := \sum_{k\in\mathbb{Z}_0^2}|k|^2\sigma_{\epsilon,k}^2 = \sum_{k\in\mathbb{Z}_0^2}\frac{|k|^2}{1+\delta(\epsilon)|k|^{2\beta}} \leq c\int_1^\infty \frac{r}{1+\delta(\epsilon)r^\beta}\,dr \leq c\ \delta(\epsilon)^{-2/\beta}.$$

Therefore, thanks to the Poincaré inequality and the fact that $e^x(a-x) \leq \exp(a-1)$, for every $a > 1$ and $x \geq 0$, it follows that

$$\mathbb{E}F(t, u_\epsilon^x(t)) \leq \exp\Big(\frac{\|x\|_V^2}{2\epsilon}\Big) + \mathbb{E}\int_0^t \exp(s)\exp\Big(\frac{\|u_\epsilon^x(s)\|_V^2}{2\epsilon}\Big)\Big(1 + \frac{1}{2}P_\epsilon - \frac{\|u_\epsilon^x(s)\|_V^2}{2\epsilon}\Big)ds$$

$$\leq \exp\left(\frac{\|x\|_V^2}{2\epsilon}\right) + \int_0^t \exp(s)\exp\left(\frac{1}{2}P_\epsilon\right)ds.$$

Hence,

$$\mathbb{E}\exp\Big(\frac{\|u_\epsilon^x(t)\|_V^2}{2\epsilon}\Big) \leq \exp\Big(-t + \frac{\|x\|_V^2}{2\epsilon}\Big) + \exp\Big(\frac{1}{2}P_\epsilon\Big).$$

Finally, using equation (5.4.3), we see that

$$\nu_\epsilon(B_V^c(0,R)) \leq \exp\Big(-\frac{R^2}{\epsilon}\Big)\limsup_{T\to\infty}\frac{1}{T}\int_0^T \Big[e^{-t} + \exp\Big(\frac{1}{2}P_\epsilon\Big)\Big]dt$$

$$= \exp\Big(-\frac{R^2}{\epsilon} + \frac{P_\epsilon}{2}\Big) \leq \exp\Big(-\frac{R^2 - C\epsilon\,\delta_\epsilon^{-2/\beta}}{\epsilon}\Big),$$

which completes the proof of the lemma, since $\epsilon\,\delta(\epsilon)^{-2/\beta} \to 0$, as $\epsilon \downarrow 0$.

$\square$

Lemma 5.4.2. *For any $\delta > 0$ and $s > 0$, there exist $\lambda > 0$ and $T > 0$ such that for any $t \geq T$ and $z \in C([0,t];H)$,*

$$\|z(0)\|_H < \lambda, \quad I_T(z) \leq s \implies \operatorname{dist}_H(z(t), \Phi(s)) < \delta,$$

where $\Phi(s) := \{x \in H : U(x) \leq s\}$.

Lemma 5.4.3. *For any $s > 0$, $\delta > 0$ and $r > 0$, let λ be as in Lemma 5.4.2. Then there exists $N \in \mathbb{N}$ large enough that*

$$u \in H_{r,s,\delta}(N) \implies I_T(u) \geq s,$$

where the set $H_{r,s,\delta}(n)$ is defined for $N \in \mathbb{N}$ by

$$H_{r,s,\delta}(N) := \{u \in C([0,N];H),\ \|u(0)\|_H \leq r,\ \|u(j)\|_H \geq \lambda,\ j = 1, ..., N\}.$$

The proofs of Lemma 5.4.2 and 5.4.3 depend only on the properties of the deterministic Navier-Stokes equation and can be found in [8] (see Lemmas 7.2 and 7.3).

Proof of Proposition 5.4.2. Fix any $s > 0$, $\delta > 0$ and $\gamma > 0$ and let R_s be as in Lemma 5.4.1. Due to the invariance of ν_ϵ, for any $t \geq 0$ we have

$$\nu_\epsilon\left(\{h \in H : \mathrm{dist}_H(h, \Phi(s)) \geq \delta\}\right) = \int_H \mathbb{P}\left(\mathrm{dist}_H(u_\epsilon^y(t), \Phi(s)) \geq \delta\right) d\nu_\epsilon(y)$$

$$= \int_{B_V^c(0,R_s)} \mathbb{P}\left(\mathrm{dist}_H(u_\epsilon^y(t), \Phi(s)) \geq \delta\right) d\nu_\epsilon(y)$$

$$+ \int_{B_V(0,R_s)} \mathbb{P}\left(\mathrm{dist}_H(u_\epsilon^y(t), \Phi(s)) \geq \delta, u_\epsilon^y \in H_{R_s,s,\delta}(N)\right) d\nu_\epsilon(y)$$

$$+ \int_{B_V(0,R_s)} \mathbb{P}\left(\mathrm{dist}_H(u_\epsilon^y(t), \Phi(s)) \geq \delta, u_\epsilon^y \notin H_{R_s,s,\delta}(N)\right) d\nu_\epsilon(y)$$

$$=: K_1 + K_2 + K_3.$$

Now, thanks to Lemma 5.4.1 we know that

$$K_1 \leq \nu_\epsilon(B_V^c(0,R_s)) \leq \exp\Big(-\frac{s}{\epsilon}\Big).$$

Next, let N be as in Lemma 5.4.3. Since $H_{R_s,s,\delta}(N)$ is a closed set in $C([0,N];H)$ and $B_V(0,R_s)$ is a compact subset of H, the Dembo-Zeitouni uniform large deviation principle over compact sets, Corollary 5.3.1, implies that there exists $\epsilon_0 > 0$ such that

$$K_2 \leq \sup_{y \in B_V(0,R_s)} \mathbb{P}(u_\epsilon^y \in H_{R_s,s,\delta}(N))$$

$$\leq \exp\Big(-\frac{1}{\epsilon}\Big[\inf_{z\in B_V(0,R_s)} \inf_{h \in H_{R_s,s,\delta}(N)} I_T^z(h) - \gamma\Big]\Big),$$

for any $\epsilon \leq \epsilon_0$. Hence, by Lemma 5.4.3,

$$K_2 \leq \exp(-\frac{1}{\epsilon}[s-\gamma]).$$

To address K_3, we use the Markov property of u_ϵ to stop the process at integer times. We then have

$$\begin{aligned}
K_3 &= \int_{B_V(0,R_s)} \mathbb{P}\left(\bigcup_{j=1}^N \{\|u_\epsilon^y(j)\|_H < \lambda\} \bigcap \{\mathrm{dist}_H(u_\epsilon^y(t), \Phi(s)) \geq \delta\}\right) d\nu_\epsilon(y) \\
&\leq \sum_{j=1}^N \int_{B_V(0,R_s)} \mathbb{P}\left(\{\|u_\epsilon^y(j)\|_H < \lambda\} \bigcap \{\mathrm{dist}_H(u_\epsilon^y(t), \Phi(s)) \geq \delta\}\right) d\nu_\epsilon(y) \\
&\leq \sum_{j=1}^N \sup_{y \in B_H(0,\lambda)} \mathbb{P}(\mathrm{dist}_H(u_\epsilon^y(t-j), \Phi(s)) \geq \delta).
\end{aligned}$$

In order to use the uniform LDP of Theorem 5.3.3, we must convert this event at time $t-j$ to an event in $C([0,t-j];H)$. To do so, we pick t large enough that Lemma 5.4.2 applies for $\delta/2$. Then, if $y \in B_H(\lambda)$

$$\begin{aligned}
&\mathrm{dist}_H(u_\epsilon^y(t-j), \Phi(s)) \geq \delta \\
&\qquad \Longrightarrow \inf\Big\{\|u_\epsilon^y - v\|_{C([0,t-j];H)} : \|v(0)\|_H < \lambda, I_T(v) \leq s\Big\} \geq \frac{\delta}{2} \\
&\qquad \Longrightarrow \mathrm{dist}_{C([0,t-j];H)}\Big(u_\epsilon^y, \Psi^y(s)\Big) \geq \delta/2,
\end{aligned}$$

where

$$\Psi^y(s) := \{v \in C([0, t-j]; H) : v(0) = y, I_T(v) \leq s\}.$$

Then, by Theorem 5.3.3, there exists $\epsilon_{0,j}$ such that for any $\epsilon \leq \epsilon_{0,j}$,

$$\begin{aligned}
&\sup_{y \in B_H(0,\lambda)} \mathbb{P}(\mathrm{dist}_H(u_\epsilon^y(t-j), \Phi(s)) \geq \delta) \\
&\qquad \leq \sup_{y \in B_H(0,\lambda)} \mathbb{P}(\mathrm{dist}_{C([0,t-j];H)}\Big(u_\epsilon^y, \Psi^y(s)\Big) \geq \delta/2) \\
&\qquad \leq \exp\Big(-\frac{s-\gamma}{\epsilon}\Big).
\end{aligned}$$

Hence, for any $\epsilon < \min(\epsilon_0, \epsilon_{0,1}, ..., \epsilon_{0,N})$ it follows that

$$K_3 \leq N \exp\Big(-\frac{s-\gamma}{\epsilon}\Big),$$

which implies the result. □

Bibliography

[1] M. Bardi, I. Capuzzo-Dolcetta, Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations, Birkhauser, 1997.

[2] N. Berglund, *Kramers' law: Validity, derivations and generalisations*, Markov Processes and Related Fields 19 (2013), pp. 459-490.

[3] H. Bessaih, A. Millet, *Large deviation principle and inviscid shell models*, Electronic Journal of Probability 14 (2009), pp. 2551-2579.

[4] H. Bessaih, A. Millet, *Large deviations and the zero viscosity limit for 2D stochastic Navier–Stokes equations with free boundary*, SIAM Journal on Mathematical Analysis 44 (2012), pp. 1861–1893.

[5] D. Blömker, W. W. Mohammed, *Fast diffusion limit for reaction-diffusion systems with stochastic Neumann boundary conditions*, SIAM Journal on Mathematical Analysis 48 (2016), pp. 3547–3578.

[6] F. Bouchet, J. Reygner, *Generalisation of the Eyring-Kramers transition rate formula to irreversible diffusion processes*, Annales Henri Poincaré 17 (2016), pp. 3499-3532.

[7] M. Boué, P. Dupuis, *A variational representation for certain functionals of Brownian motion*, Annals of Probability 26 (1998), pp. 1641-1659.

[8] Z. Brzeźniak, S. Cerrai, *Large deviations principle for the invariant measures of the 2D stochastic Navier-Stokes equations on a torus,* Journal of Functional Analysis 273 (2017), pp. 1891-1930.

[9] Z. Brzeźniak, S. Cerrai, M. Freidlin, *Quasipotential and exit times for 2D Stochastic Navier-Stokes equations driven by space-time white noise*, Probability Theory and Related Fields 162 (2015), pp. 739–793.

[10] Z. Brzeźniak, Y. Li, *Asymptotic compactness and absorbing sets for 2D stochastic Navier-Stokes equations on some unbounded domains*, Transactions of the American Mathematical Society 358 (2006), pp. 5587-5629.

[11] J. Bucklew, Introduction to Rare Event Simulation, Springer Series in Statistics, 2004.

[12] A. Budhiraja, P. Dupuis *A Variational Representation for Positive Functionals of Infinite Dimensional Brownian Motion*, Probability and Mathematical Statistics 20 (2000), pp. 39-61.

[13] A. Budhiraja, P. Dupuis, V. Maroulas, *Large deviations for infinite dimensional stochastic dynamical systems*, The Annals of Probability 36 (2008), pp. 1390–1420.

[14] M. Cameron, *Finding the Quasipotential for Nongradient SDEs*, Physica D: Nonlinear Phenomena 241 (2012), pp. 1532–1550.

[15] S. Cerrai, A. Debussche, *Large deviations for the two-dimensional stochastic Navier-Stokes equation with vanishing noise correlation*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 55 (2019), pp. 211–236.

[16] S. Cerrai, M. Freidlin, *Fast transport asymptotics for stochastic RDEs with boundary noise*, The Annals of Probability, 39 (2011), pp. 369–405.

[17] S. Cerrai, M. Freidlin, *Approximation of quasi-potentials and exit problems for multidimensional RDE's with noise*, Transactions of the AMS 363 (2011), pp. 3853-3892.

[18] S. Cerrai, N. Paskal, *Large deviations for fast transport stochastic RDEs with applications to the exit problem*, Annals of Applied Probability 29 (2019), pp. 1993-2032.

[19] S. Cerrai, N. Paskal, *Large deviations principle for the invariant measures of the 2D stochastic Navier-Stokes equations with vanishing noise correlation*, arXiv:2012.14953, 2020.

[20] S. Cerrai, M. Salins, *Smoluchowski-Kramers approximation and large deviations for infinite dimensional non-gradient systems with applications to the exit problem*, The Annals of Probability 44 (2016), pp. 2591–2642.

[21] D. Dahiya, M. Cameron, *Ordered Line Integral Methods for Computing the Quasi-potential*, Journal of Scientific Computing 75 (2018), pp. 1351-1384.

[22] D. Dahiya, M. Cameron, *An Ordered Line Integral Method for computing the quasi-potential in the case of variable anisotropic diffusion*, Physica D: Nonlinear Phenomena 382-383 (2018), pp. 33-45.

[23] G. Da Prato, A. Debussche, *Two-dimensional Navier-Stokes equations driven by a space-time white noise*, Journal of Functional Analysis 196 (2002), pp. 180-210.

[24] G. Da Prato, J. Zabczyk, *Evolution equations with white-noise boundary conditions*, Stochastics and Stochastics Reports 42 (1993), pp. 167–182.

[25] G. Da Prato, J. Zabczyk, Ergodicity for Infinite Dimensional Systems, London Mathematical Society Lecture Note Series, Cambridge University Press, 1996.

[26] G. Da Prato, J. Zabczyk, Stochastic equations in infinite dimensions, Second Edition, Cambridge University Press, 2012.

[27] E. B. Davies, Heat Kernels and Spectral Theory, Cambridge Tracts in Mathematics, Cambridge University Press, 1990.

[28] A. Dembo, O. Zeitouni, Large Deviations Techniques and Applications, Springer Verlag, 2009.

[29] E. Dijkstra, *A note on two problems in connexion with graphs*, Numerische Mathematik 1 (1959), pp. 269-271.

[30] P. Dupuis, R. Ellis, Weak Convergence Approach to the Theory of Large Deviations, Wiley series in probability and statistics, 2011.

[31] W. E, W. Ren, E. Vanden-Eijnden, *Minimum action method for the study of rare events*, Communications on Pure and Applied Mathematics 57 (2004), pp. 637-656.

[32] B. Ferrario, *Stochastic Navier-Stokes equations: analysis of the noise to have a unique invariant measure*, Annali di Matematica Pura ed Applicata 177 (1999), pp. 331-347.

[33] W. Fleming, *Exit probabilities and optimal stochastic control*, Applied Mathematics and Optimization 4 (1977), pp. 329-346.

[34] M. Freidlin, *A mixed boundary-value problem for second-order elliptic differential equations with a small parameter*, Doklady Akademii Nauk 143 (1962), pp. 1300–1303.

[35] M. Freidlin, *A priori estimates of solutions of degenerating elliptic equations*, Soviet Mathematics Doklady 5 (1964), pp. 1231—1234.

[36] M. Freidlin, A. Wentzell, *Reaction-diffusion equations with randomly perturbed boundary conditions*, Annals of Probability 20 (1992), pp. 963–986.

[37] M. Freidlin, A. Wentzell, Random Perturbations of Dynamical Systems, third edition, Springer Verlag, 2012.

[38] M. Heymann, E. Vanden-Eijnden, E., *Pathways of Maximum Likelihood for Rare Events in Nonequilibrium Systems: Application to Nucleation in the Presence of Shear*, Physical Review Letters 100 (2008), p. 140601.

[39] S. Kuksin, A. Shirikyan, Mathematics of two-dimensional turbulence, Cambridge University Press, Cambridge, 2012.

[40] R. E. Langer. *A problem in diffusion or in the flow of heat for a solid in contact with a fluid*, Tohoku Mathematical Journal 35 (1932), pp.260-275.

[41] I. Lasiecka, *Unified theory for abstract parabolic boundary problems—a semigroup approach*, Applied Mathematics and Optimization 6 (1980), pp. 287-333.

[42] L. Lin, H. Yu, X. Zhou, *Quasi-potential Calculation and Minimum Action Method for Limit Cycle*, Journal of Nonlinear Science 29 (2019), pp. 961-991.

[43] J. L. Lions, E. Magenes, Non-homogeneous Boundary Value Problems and Applications, Springer-Verlag, 1972.

[44] R. Maier, D. Stein, *A Scaling Theory of Bifurcations in the Symmetric Weak-Noise Escape Problem*, Journal of Statistical Physics 83 (1996), pp. 291-357.

[45] A. Martin-Lof, *Harald cramer and insurance mathematics*, Applied Stochastic Models and Data Analysis 11 (1995), pp. 271-276.

[46] D. Martirosyan, *Large deviations for invariant measures of the white-forced 2D Navier-Stokes equation*, Journal of Evolution Equations 18 (2018), pp. 1245-1265.

[47] J. Mirebeau, *Efficient Fast Marching with Finsler metrics*, Numerische Mathematik 126 (2014), pp. 515-557.

[48] J. Nave, R. Rosales, B. Seibold, *A gradient-augmented level set method with an optimally local, coherent advection scheme*, Journal of Computational Physics 229 (2010), pp. 3802-3827.

[49] B. Nolting, K. Abbott, *Balls, cups, and quasi-potentials: quantifying stability in stochastic systems*, Ecology 97 (2016), no. 4, pp. 850-864.

[50] J. Peixoto and A. Oort. PHYSICS OF CLIMATE, Springer, 1992.

[51] S. Peszat, *Exponential tail estimates for infinite-dimensional stochastic convolutions*, Bulletin of the Polish Academy of Sciences, Mathematics 40 (1992), pp. 323–333.

[52] H. Pham, *Large deviations in mathematical finance*, Lecture Notes (2010).

[53] G. Poppe, T. Schaefer, *Computation of minimum action paths of the stochastic nonlinear Schrodinger equation with dissipation*, J. Phys. A: Math. Theor. 51 (2018), no. 33, 335102.

[54] S. Potter, M. Cameron, *Jet Marching Methods for Solving the Eikonal Equation*, arXiv:2009.05490, 2020.

[55] M. Salins, *Equivalences and counterexamples between several definitions of the uniform large deviations principle,* Probability Surveys 16 (2019), pp. 99-142.

[56] R. Schnaubelt, M. Veraar, *Stochastic Equations with Boundary Noise*, Progress in Nonlinear Differential Equations and Their Applications 80 (2011), pp. 609–629.

[57] J. A. Sethian, LEVEL SET METHODS AND FAST MARCHING METHODS EVOLVING INTERFACES IN COMPUTATIONAL GEOMETRY, FLUID MECHANICS, COMPUTER VISION, AND MATERIALS SCIENCE, Cambridge University Press, 1999.

[58] J. Sethian, A. Vladimirsky, *Ordered Upwind Methods for static Hamilton-Jacobi*

equations, Proceedings of the National Academy of Sciences 98 (2001), pp. 11069-11074.

[59] Sethian, J., Vladimirsky, A., *Ordered Upwind Methods for static Hamilton-Jacobi equations: theory and algorithms*, SIAM Journal of Numerical Analysis 41 (2003), pp. 325-363.

[60] R. Sowers, *Large deviations for the invariant measure of a reaction-diffusion equation with non-Gaussian perturbations*, Probability Theory and Related Fields 92 (1992), pp. 393-421.

[61] R. Sowers, *Multidimensional reaction-diffusion equations with white noise boundary perturbations*, Annals of Probability 22 (1994), pp. 2071–2121.

[62] P. Talkner, *Mean first passage time and the lifetime of a metastable state*, Zeitschrift für Physik B Condensed Matter 68 (1987), pp. 201-207.

[63] M. Tao, Hyperbolic periodic orbits in nongradient systems and small-noise-induced metastable transitions, Physica D: Nonlinear Phenomena 363 (2018), pp. 1-17.

[64] R. Temam, Navier-Stokes equations and nonlinear functional analysis, CBMS-NSF Regional Conference Series in Applied Mathematics, 41, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1983.

[65] S. Varadhan, *Asymptotic probabilities and differential equations*, Communications on Pure and Applied Mathematics 19 (1966), pp. 261-286.

[66] S. Yang, S. Potter, M. Cameron, *Computing the quasipotential for nongradient SDEs in 3D*, Journal of Computational Physics 379 (2019), pp. 325-350.

[67] X. Zhou, W. Ren, W. E, *Adaptive minimum action method for the study of rare events*, Journal of Chemical Physics 128 (2008), p. 104111.