PhD-FSTC-2018-32
The Faculty of Science, Technology and Communication

DISSERTATION

Defence held on 18/04/2018 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN MATHÉMATIQUES

by

Bob PEPIN

Born on 24 September 1983 in Luxembourg (Luxembourg)

TIME AVERAGES OF DIFFUSION PROCESSES AND APPLICATIONS TO TWO-TIMESCALE PROBLEMS

Dissertation defence committee

Dr. Anton THALMAIER, dissertation supervisor
Professor, University of Luxembourg

Dr. Ivan NOURDIN, chairman
Professor, University of Luxembourg

Dr. Giovanni PECCATI, vice chairman
Professor, University of Luxembourg

Dr. Tony LELIÈVRE
Professor, École des Ponts Paris Tech

Dr. Andreas EBERLE
Professor, Bonn Institute for Applied Mathematics

Acknowledgements

First of all, this thesis would not have been possible without the financial support from the Luxembourg National Research Fund. I would in particular like to thank the anonymous reviewers who agreed to fund me for doing mathematics research, despite the fact that I had an education in electrical engineering and biology. By extension I would also like to thank the people at the FNR for creating a culture that made such bold decisions possible.

The research proposal owed much to the valuable advice and support from my PhD supervisor Anton Thalmaier and the great people at the Luxembourg Centre for Systems Biomedicine: Rudi Balling, Alex Skupin and Jorge Gonçalves. I would also like to thank Frank Noé in Berlin for taking the trouble to reply to my email when I was fishing in the dark for research subjects.

It is only with hindsight that I can fully appreciate the degree to which my supervisor Anton Thalmaier trusted me during all the phases of this thesis, and the great intellectual freedom that I got to enjoy as a result. He always had the right advice at the right time, which he patiently repeated until I would eventually listen and understand.

During the four years I worked on this project, I saw many colleagues come and go, and I would like to thank all of them for entertaining and insightful office and lunch conversations. Greetings to Antoine, Robert, Erwan, Guillaume, Yannick, Sinan, François, Jacob, Mareille, Samuele, Johannes, Xiaolei, Angelica, Amirhossein, Frederik, Emil, Kang, Niels, Maj-Britt, Rune, Søren, Thomas, Jeffrey, Giovanni, Jostein, Anja, David, Trine, Jonas, Olivier, Pierre and all the other wonderful people I met during these years.

Finally, I would also like to thank Sofie and our two children Niels and Petra for always being there for me and encouraging me to keep going. And especially Niels and Petra for the good play times we had, which cheered me up whenever the work became too frustrating. This work is dedicated to our little family.

Contents

1 Introduction **4**

1.1 Background . . . 4

1.2 Applications of Averaging . . . 7

1.2.1 Simulation . . . 8

1.2.2 TAMD . . . 10

2 Concentration Inequalities and Averaging Principle **11**

2.1 Concentration Inequalities . . . 11

2.1.1 Ergodic Theory and Poisson Problems . . . 11

2.1.2 From Functional to Concentration Inequalities . . . 14

2.1.3 Regeneration methods . . . 17

2.2 Averaging Principle . . . 18

2.2.1 Introduction . . . 18

2.2.2 Freezing the slow process . . . 20

2.2.3 Asymptotic expansion of the generator . . . 23

2.2.4 Effective dynamics using conditional expectations . . . 27

3 Time Averages of Diffusion Processes **30**

3.1 Introduction . . . 30

3.2 Martingale Representation . . . 31

3.2.1 Discrete time . . . 35

3.3 Concentration inequalities from exponential gradient bounds . . . 36

3.4 Averaging: Two-timescale Ornstein-Uhlenbeck . . . 38

3.5 Averaging: Exact gradients in the linear case . . . 39

3.6 Averaging: Conditioning on the slow component . . . 42

3.7 Approximation by Averaged Measures . . . 46

4 Towards a Quantitative Averaging Principle for Stochastic Differential Equations **48**

4.1 Introduction and notation . . . 48

4.1.1 Motivation and main result . . . 48

4.1.2 Setting and notation . . . 50

4.2 Approximation by conditional expectations . . . 52

4.3 Distance between conditional and averaged measures 57
4.4 Decoupling 61
4.5 Proof of the main theorem 65
4.6 Applications 69
4.6.1 Averaging 69
4.6.2 Temperature-Accelerated Molecular Dynamics 72

Chapter 1

Introduction

1.1 Background

In the second half of the 18th century, scientists studying the many body problem in celestial mechanics realised that they could considerably simplify their equations by noting that the periodic perturbations of the other planets on the earth and sun approximately cancel out over time. This became later known as the averaging principle[SVM07].

One hundred years later, when Boltzmann was laying the foundations for statistical mechanics, he noted that the probability of a collection of particles being in a certain state was closely related to the time a typical particle with a sufficiently long trajectory spent in that state. This was the beginning of ergodic theory [Von91].

Both ergodic theory and the averaging principle rely on the idea that on a sufficiently long trajectory perturbations will eventually cancel out. When combined they lead to stochastic averaging principles, the main topic of this work.

The first main object we will study are integral functionals of the form

$$S_T = \int_0^T f(t, X_t)dt, \quad f \in C_c^\infty$$

where X is a solution to an SDE on $\mathbb{R}^n$ with possibly time-dependent coefficients

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = x$$

with $b(t, x)$, $\sigma(t, x)$ continuous in t and locally Lipschitz continuous in x, σ taking values in the space of nondegenerate $n \times n$ matrices and B being a standard Brownian motion on $\mathbb{R}^n$.

The second main object are the solutions (X, Y) to the following SDE on $\mathbb{R}^n \times \mathbb{R}^m$:

$$dX_t^\alpha = \alpha b_X(X_t^\alpha, Y_t^\alpha)dt + \sqrt{\alpha}\sigma_X dB_t^X, \quad X_0 = x$$
$$dY_t^\alpha = b_Y(X_t^\alpha, Y_t^\alpha)dt + \sigma_Y dB_t^Y, \quad Y_0 = y$$

where b_X, b_Y are locally Lipschitz continuous functions from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, σ_X, σ_Y are constant non-degenerate $n \times n$ and $m \times m$ matrices and B^X, B^Y standard Brownian motions on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. The parameter $\alpha > 0$ expresses the "acceleration" of X with respect to Y, which is why the process (X^α, Y^α) is interpreted as a two-timescale process.

Both objects share the common property that random fluctuations "average out". In the first case, we will develop a novel martingale argument to quantify the distance between S_T and its expectation via concentration inequalities. For the two-timescale process, we will extend and quantify existing results and introduce a new approach to show that as $\alpha \to \infty$ the slow component Y_t^α converges to a process $\bar{Y}$ adapted to B^Y. This can be interpreted as the random fluctuations of B^X being averaged out, since $\mathcal{F}^X = \sigma(B^X)$ provides no additional information on Y^α in the limit as $\alpha \to \infty$.

Both the ergodic theorem and the averaging principle have a long history in physics and engineering, where they are typically used to derive reduced models of complex physical systems. The values of T and α are imposed by the underlying physics and usually sufficiently large that the asymptotic results hold with very high precision.

More recently, these ideas have become an important ingredient in computer algorithms. For example, Stochastic Approximation and Stochastic Gradient Descent [LTE17], nowadays arguably the most important algorithms for optimisation and machine learning problems, are built on the principle that it is possible to optimize an objective by iteratively optimizing over randomly perturbed versions of it since the perturbations eventually cancel out. The averaging principle underlies multiscale simulation algorithms in physics [ERV09] and the implementation of important artificial intelligence algorithms such as actor-critic methods in reinforcement learning [Bha+09] and generative adversarial neural networks. What these problems have in common is that T and α become tuneable parameters, which should be chosen as small as possible since the running time of the algorithms increases with T and α. A quantitative, non-asymptotic understanding of both the ergodic theorem and the averaging principle now becomes essential for guiding the choice of T and α in practical situations.

The next subsection in this introductory chapter will elaborate on the application of averaging to the simulation of two-timescale systems by presenting in more detail multiscale methods and the Temperature-Accelerated Molecular Dynamics method.

The first part of Chapter 2 contains a presentation of the state of the art for concentration inequalities for functionals like S_t in the time-homogeneous case. The main purpose is to anchor the results from Chapter 3 and show that they can lead to new mathematical results, since such concentration inequalities are their most direct application. The second part of Chapter 2 gives a short taxonomy of different approaches to the averaging principle and serves to establish a context for Chapter 4 which presents a new approach to the Averaging Principle.

The first main contribution of this work is in Chapter 3, where we show that for T fixed the time average S_T defined above can be decomposed as

$$S_T = \mathbb{E}S_T + \int_0^T \int_t^T \nabla P_{t,s} f(X_t) ds \cdot \sigma(t, X_t) dB_t$$

where

$$P_{t,s} f(x) = \mathbb{E}[f(s, X_s) | X_t = x], \quad t \leq s$$

is the transition operator associated to X. In the time-homogeneous case $P_{t,s} = P_{s-t}$ is the familiar one-parameter semigroup associated to X.

To the author's knowledge, this is the first time that functionals such as S_T have been systematically studied beyond the time-homogeneous setting. It is also the first time that the martingale part in the martingale representation theorem for S_T is explicitly investigated via differentiation formulas for 2-parameter Markov semigroups.

In particular, when there exist positive constants C and λ such that $\nabla P_{t,s} f \leq C e^{-\lambda(s-t)}$ for all $s \in [t, T]$ then we get a Gaussian concentration inequality for S_T.

Chapter 3 concludes with an application of the martingale representation above to averaging.

The proof of the decomposition in Chapter 3 is elementary and we now give a quick preview. The decomposition follows in fact directly from the application of Itô's formula to

$$R_t^T f(x) = \int_t^T P_{t,s} f(x) ds$$

together with the observation that

$$(\partial_t + L_t) R_t^T f(x) = -f(t, x), \quad x \in \mathbb{R}^n, t \leq T$$

so that

$$f(t, X_t) dt = -d R_t^T f(X_t) + \nabla R_t^T f(X_t) \cdot \sigma(t, X_t) dB_t.$$

This gives the martingale representation shown above for S_T since $R_T^T = 0$. As a special case, this result also includes the well-known argument based on Poisson equations in the time-homogeneous ergodic setting. If $P_{s,t} = P_{t-s}$ then for a bounded continuous function $f(x)$ on $\mathbb{R}^n$, centered with respect to the invariant measure, we can let T go to ∞ and $-L R_t^\infty (f - \mu(f)) = f - \mu(f)$ for arbitrary $t > 0$, where L is the generator and μ the invariant measure.

The second main contribution is Chapter 4. It describes a new approach to the averaging principle, based on freezing the whole trajectory of the slow process. In a first step, we show how the mutual interaction of feedback between X^α and Y^α can be broken. Then we condition on a trajectory of the slow process, and due to the previous decoupling operation we can study X^α as a time-inhomogeneous diffusion process for each fixed value of $Y^\alpha \in C([0, T], \mathbb{R}^m)$. Finally we integrate over all values of Y^α to obtain the result. In the process, we need to estimate a functional of the form of S_T, which we do using a forward-backward martingale argument under the assumption of a Poincaré inequality holding for all time marginals of X^α.

1.2 Applications of Averaging

Averaging methods are well established in the analysis and control of physical systems in the engineering and natural sciences. In fact, many physical systems exhibit a natural hierarchy of timescales where some components change slowly while others evolve rapidly. The physical timescales often differ by several orders of magnitude, so that asymptotic results can be used to derive approximate models of the phenomena under investigation. Experiments can then validate and complete these reduced models. Typical examples in the natural sciences include physico-chemical systems such as large molecules, where physical processes often operate on timescales of femtoseconds (10^{-15} s) whereas chemical properties evolve on a timescale of milliseconds (10^{-3} s). In fact, the 2013 Nobel prize in chemistry was awarded for the development of multiscale models for complex chemical systems. Examples in the engineering sciences include the longitudinal movements in airplane dynamics, which exhibit fast short-period and slow phugoid oscillations, and the control of a DC-Motor [KKO99].

As computers became more powerful, there has been a growing interest in the simulation of complex systems with multiple timescales, the prototypical example being molecular dynamics simulations of large macromolecules such as proteins. The underlying challenge is that the timestep of the numerical integration is dictated by the fastest timescale in the system, so that simulations of extended time periods become prohibitively expensive. In situations where there is a clear separation into fast and slow degrees of freedom, many numerical methods have been proposed for approximating the evolution of the slow degrees of freedom without resorting to a complete simulation of the fast components. A popular approach proceeds by an alternance of slow and fast steps. In each fast step, the fast degrees of freedom are evolved until they reach stationarity while keeping the values of the slow coordinates frozen. The slow steps then evolve the slow degrees of freedom with the value of the fast components replaced by their stationary values. This corresponds in some sense to a numerical simulation of the approximation used in the proof of the averaging principle in Section 2.2.2.

Another approach keeps the original dynamics, except that the fast dynamics are now slowed down by a certain factor which is such that the dynamics of the slow variables remains intact while significantly increasing the step size of the simulation, and thus decreasing the number of steps necessary for a given accuracy. In order to determine how much we can slow down the fast dynamics, it is crucial to have a quantitative understanding and non-asymptotic understanding of the averaging principle. See [ERV09] for more complete descriptions and further references on the methods described in this and the previous paragraph.

A similar principle can be used when the coordinates can not be clearly separated according to their timescale, but we are only interested in the approximate dynamics of certain macroscopic quantities. This method, called Temperature-Accelerated Molecular dynamics (TAMD), was the original motivation for this work and is described in more detail in the next section. Here too, we have to choose the magnitude of the separation between the timescales.

In a different spirit, averaging also plays an important role in stochastic approximation algorithms such as stochastic gradient descent. In particular, when considering a variation of stochastic gradient descent for two-stage optimisation problems, we are again in a situation where we have to choose the separation of timescales so that it is important to have quantitative results on averaging.

1.2.1 Simulation

In this section we will take a brief look at the heterogeneous and seamless multiscale methods for simulating systems with disparate timescales as described in [ERV09].

For concreteness, suppose that the system we want to simulate is described by an SDE of the form

$$\begin{aligned} dX_t^\alpha &= \alpha b_X(X_t^\alpha, Y_t^\alpha)dt + \sqrt{\alpha} dB_t^X, \quad X_0 = x \\ dY_t^\alpha &= b_Y(X_t^\alpha, Y_t^\alpha)dt + dB_t^Y \quad Y_0 = y \end{aligned}$$

with B^X, B^Y standard Brownian motions on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, $\alpha > 0$ and b_X, b_Y bounded Lipschitz continuous functions.

A standard Euler-Maruyama numerical scheme with step size Δ is: For $k = 1 \dots N$ set

$$\begin{aligned} X_{k+1} &= X_k + \Delta\alpha b_X(X_k, Y_k) + \sqrt{\Delta\alpha} N_k^X \\ Y_{k+1} &= Y_k + \Delta b_Y(X_k, Y_k) + \sqrt{\Delta} N_k^Y \end{aligned}$$

with N_k^X, N_k^Y sequences of standard independent n- and m-dimensional normal variables. From the expression for X_k we can see that in order to keep a fixed approximation error as α increases, we need to choose Δ on the order of $1/\alpha$, meaning that the number of simulation steps N needed to simulate (X, Y) on a fixed time intervals is of order α.

For $y \in \mathbb{R}^m$ let X^y be the solution to

$$dX_t^y = b_Y(X_t, y)dt + dB_t^X$$

and suppose that X^y has a stationary measure μ^y. By the averaging principle as $\alpha \to \infty$ Y^α converges in probability to $\bar{Y}$ solution to

$$d\bar{Y}_t = \bar{b}(\bar{Y}_t)dt + dB_t^Y, \quad \bar{b}(y) = \int b_Y(x, y)\mu^y(dx).$$

This observation leads to the following numerical scheme for directly simulating $\bar{Y}$: For

$k = 1 \dots N$ and $i = 1 \dots M$ set

$$\begin{aligned} X_{i+1}^k &= X_i^k + \delta b_X(X_i^k, Y_k) + \sqrt{\delta} N_{k,i}^X, \quad X_1^k = X_M^{k-1} \\ \bar{b}_k &= \sum_{i=1}^{M} b_Y(X_i^k, \bar{Y}_k) \\ \bar{Y}_{k+1} &= \bar{Y}_k + \Delta \bar{b}_k + \sqrt{\Delta} N_k^Y. \end{aligned}$$

The total number of simulation steps is of order NM, where N can be much smaller than for the Euler-Maruyama scheme when α is large and M depends on how fast X^y converges to equilibrium.

The idea behind the seamless multiscale method is the following: Our goal is to numerically approximate a trajectory of Y^α where the approximation error is at most c with probability of at least $1 - \varepsilon$. Suppose that for c fixed we know a continuous decreasing function F such that for all α

$$\mathbb{P}(\sup|Y_t^\alpha - \bar{Y}_t| > c/2) < F(\alpha).$$

Then when α is sufficiently large we can choose $\beta < \alpha$ such that for given $\varepsilon > 2F(\alpha)$

$$\mathbb{P}(\sup|Y_t^\alpha - Y_t^\beta| > c) < \varepsilon.$$

because

$$\mathbb{P}(\sup|Y_t^\alpha - Y_t^\beta| > c) \leq \mathbb{P}(\sup|Y_t^\alpha - \bar{Y}_t| > c/2) + \mathbb{P}(\sup|Y_t^\beta - \bar{Y}_t| > c/2) \leq F(\alpha) + F(\beta)$$

so that we can choose $\beta < \alpha$ such that $F(\beta) = \varepsilon - F(\alpha)$. Now we can use an Euler-Maruyama scheme for X^β, Y^β which requires a number of simulation steps N on the order of β. We will see later that $F(\alpha)$ is usually on the order of $1/\sqrt{\alpha}$. If $F(\alpha) = K/\sqrt{\alpha}$, then $\beta = (\varepsilon/K - 1/\sqrt{\alpha})^{-2}$ and the acceleration relative to the Euler-Maruyama scheme for X^α is $\alpha/\beta = (\sqrt{\alpha}\varepsilon/K - 1)^2 = \mathcal{O}(\alpha)$.

1.2.2 TAMD

Following [MV06] (see also [SV17]) consider the TAMD (Temperature-Accelerated Molecular Dynamics) process (X_t, Y_t) and its averaged version $\bar{Y}_t$ defined by

$$\begin{aligned} dX_t^\alpha &= -\alpha \nabla_x U(X_t^\alpha, Y_t^\alpha)dt + \sqrt{\alpha}\sqrt{2\beta^{-1}}dB_t^X, \quad X_0 \sim e^{-\beta U(x,y_0)}dx \\ dY_t^\alpha &= -\tfrac{1}{\bar{\gamma}}\kappa(Y_t^\alpha - \theta(X_t^\alpha))dt + \sqrt{2(\bar{\beta}\bar{\gamma})^{-1}}dB_t^Y, \quad Y_0 = y_0 \\ d\bar{Y}_t &= \bar{b}(\bar{Y}_t)dt + \sqrt{2(\bar{\beta}\bar{\gamma})^{-1}}dB_t^Y, \quad \bar{Y}_0 = y_0 \\ U(x,y) &= V(x) + \tfrac{\kappa}{2}|y - \theta(x)|^2, \\ \bar{b}(y) &= Z(y)^{-1}\int -\bar{\gamma}^{-1}\kappa(y - \theta(x))e^{-\frac{\kappa}{2}|y-\theta(x)|^2}e^{-V(x)}dx, \\ Z(y) &= \int e^{-\frac{\kappa}{2}|y-\theta(x)|^2}e^{-V(x)}dx \end{aligned}$$

with $X_t \in \mathbb{R}^n$, $Y_t, \bar{Y}_t \in \mathbb{R}^m$, Lipschitz-continuous functions $V(x)$ and a map $\theta(x) = (\theta_1(x), \ldots, \theta_m(x))$, constants $\kappa, \alpha, \beta, \bar{\beta}, \bar{\gamma} > 0$ and independent standard Brownian motions B^X, B^Y on $\mathbb{R}^n$ and $\mathbb{R}^m$.

By the averaging principle in [PV03] Y^α converges weakly to $\bar{Y}$ on the space of trajectories.

Let $\mu(dx) = e^{-\beta V(x)}dx$, $\nu(dy) = \theta_\# \mu(dy)$ be the image measure of μ by θ and $\nu_\kappa = \nu * N(0, \kappa^{-1})$ be the convolution of ν with a centered Gaussian measure with variance κ^{-1}. Suppose that $\nu_\kappa = e^{-W(y)}dy$. It can be shown that

$$\bar{b}(y) = \nabla_y W(y).$$

This means that $\bar{Y}$ is a reversible diffusion process with invariant measure $e^{-\bar{\beta}W(y)}$, and so is Y^α in the limit $\alpha \to \infty$. In particular, if we choose $\bar{\beta}$ small then the "energy landscape" is flattened out and the mixing properties of $\bar{Y}$ are improved. The key for applications in molecular dynamics is now that μ is still the energy landscape corresponding to V at the original temperature β, and so the trajectories of $\bar{Y}$ can provide some insights on the topography of V at β.

Chapter 2

Concentration Inequalities and Averaging Principle

2.1 Concentration Inequalities

The next chapter presents a novel approach for studying functionals of the form

$$S_t := \int_0^t f(s, X_s) - \mathbb{E}f(s, X_s)ds.$$

The goal of this section is to build up some context for these results from a well-studied particular case. This case is that of an ergodic Markov process X with stationary measure μ. When f is independent of time and the initial measure of X is μ, then S_t/t takes the form

$$\frac{1}{t}\int_0^t f(X_s)ds - \int f(x)\mu(dx). \tag{2.1.1}$$

It is well-known from ergodic theory that this quantity goes to 0 as $t \to \infty$. In the first subsection, we will review a classical method based on Poisson equations that can be used to obtain some quantitative estimates for (2.1.1) in the limit as $t \to \infty$. In the next two subsections we summarize known results on upper bounds for quantities of the form

$$\mathbb{P}\left(\frac{1}{t}\int_0^t f(X_s)ds - \mu(f) \geq R\right).$$

2.1.1 Ergodic Theory and Poisson Problems

Consider a Markov diffusion process X_t on $\mathbb{R}^n$ with generator $(L, \mathcal{D}(L))$

$$Lf(x) = b(x) \cdot \nabla f(x) + \Delta f(x), \quad f \in \mathcal{D}(L)$$

where b is locally Lipschitz continuous. Suppose that X_t has a unique invariant probability measure μ such that

$$\int Lf d\mu = 0 \text{ for all } f \in \mathcal{D}(L).$$

Assume furthermore that X is ergodic: For every $f \in \mathcal{D}(L)$, $Lf = 0$ implies that f is constant. Then it is well known that for bounded measurable f

$$\frac{1}{T}\int_0^T f(X_t)dt \to \int f d\mu \tag{2.1.2}$$

a.s. for every initial distribution of X [Kal02, Theorem 20.21].

A common method of analysing functionals such as the left-hand side of (2.1.2) is via a solution to the Poisson problem. We say that a function g solves the Poisson problem on $\mathbb{R}^n$ for L and f if for all $x \in \mathbb{R}^n$

$$-Lg(x) = f(x) - \mu(f). \tag{2.1.3}$$

We will see below sufficient conditions on L and f that give existence and uniqueness of a bounded solution g with bounded first derivative.

Now fix a function f such that g solves the Poisson problem associated to L and f such that g is bounded with bounded first derivative. Then by Itô's formula

$$g(X_T) = g(X_0) + \int_0^T Lg(X_t)dt + M_T^g$$

where M^g is a continuous local martingale with quadratic variation $d\langle M^g\rangle_t = |\nabla g|^2 dt$. Rearranging and using that $-Lg = f - \mu(f)$, we get

$$\int_0^T f(X_t) - \mu(f)dt = g(X_0) - g(X_T) + M_T^g.$$

This leads to the following estimate on the L^2 distance between $\frac{1}{T}\int_0^T f(X_t)dt$ and $\int f d\mu$:

$$\mathbb{E}\left|\frac{1}{T}\int_0^T f(X_t)dt - \int f d\mu\right|^2 \le 2\frac{\mathbb{E}\,|g(X_0) - g(X_T)|^2}{T^2} + 2\frac{\mathbb{E}\langle M^g\rangle_T}{T^2} \le 8\frac{\|g\|_\infty^2}{T^2} + 2\frac{\|\nabla g\|_\infty^2}{T}.$$

We also have the well-known resolvent formula for $g = (-L^{-1})(f - \mu(f))$:

$$g(x) = \int_0^\infty \mathbb{E}f(X_t) - \mu(f)dt. \tag{2.1.4}$$

At the time of this writing, the most general results on existence and uniqueness to solutions to the Poisson equation can be found in [PV01]. They also give some bounds on the growth rate of the solutions. The next two propositions, which are simplified versions of the more general theorems found in the reference, give the flavour of the results available.

They are stated in the context of a second order partial differential operator $(L, \mathcal{D}(L))$

$$Lf(x) = b(x) \cdot \nabla f(x) + \Delta f(x), \quad f \in \mathcal{D}(L)$$

where b is a locally bounded Borel vector function.

Proposition 2.1.1. *Suppose that b satisfies the recurrence condition*

$$b(x) \cdot x \leq -r|x|^\alpha, \quad |x| > M_0$$

with $M_0 \geq 0$, $\alpha > 0$ and $r > 0$. Suppose that $f \in L^1(\mu)$ is such that $|f(x)| \leq C_1 + C_2|x|^\beta$ for some positive constants C_1, C_2 and $\beta \geq 0$. Then the resolvent formula (2.1.4) *defines a continuous function g which is a solution to the Poisson problem* (2.1.3) *for L and f.*

Proposition 2.1.2. *Assume that the conditions of Proposition 2.1.1 are in force.*

- *If there exists $C > 0$ and $\beta < 0$ such that*

  $$|f(x)| \leq C(1 + |x|)^{\beta+\alpha-2}$$

  then g is bounded.

- *If there exists $C > 0$ and $\beta > 0$ such that*

  $$|f(x)| \leq C(1 + |x|)^{\beta+\alpha-2}$$

  then there exist positive constants C', C'' such that

  $$|g(x)| \leq C'(1 + |x|)^{\beta+\alpha-2}$$

  and

  $$|\nabla g(x)| \leq C''(1 + |x|^{\beta+\alpha-2} + |x|^\beta).$$

Although the previous results give existence and growth rate for solutions to a large class of instances of the Poisson problem, they provide no information on the constants involved.

If we restrict to the class of Poisson problems where the right-hand side f is Lipschitz continuous and the operator L satisfies a certain dissipativity condition, we can get some explicit estimates on the solutions in terms of an auxiliary function $\kappa(r)$. With the same operator $(L, \mathcal{D}(L))$ as above, suppose that the invariant measure μ is such that $\int |x|^2 d\mu < \infty$. Define $\kappa(r)$ as

$$\kappa(r) = \inf_{|x-y|=r} \left\{ -\frac{(x-y) \cdot (b(x) - b(x))}{|x-y|} \right\}$$

so that

$$\frac{(x-y)}{|x-y|} \cdot (b(x) - b(x)) \leq -\kappa(|x-y|)$$

for all $x, y \in \mathbb{R}^n$. Let

$$\|f\|_{\mathrm{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|}.$$

Then we have the following result from [Wu09] Theorem 1.1 and Remark 3.5:

Proposition 2.1.3. *Suppose that*

$$c_L = \int_0^\infty r e^{-\int_0^r \kappa(s) ds} dr < \infty.$$

Then for a Lipschitz function f, the resolvent formula (2.1.4) *defines a Lipschitz continous function g that solves the Poisson problem for L and f on $\mathbb{R}^n$ and we have the bound*

$$\|g\|_{\mathrm{Lip}} \leq c_L \|f\|_{\mathrm{Lip}}.$$

The following result from [CCG12, Corollary 3.2] gives an idea on the convergence rates of X to stationarity that are necessary in order for solutions to the Poisson equation to exist:

Proposition 2.1.4. *Let*

$$L = -\nabla V \cdot \nabla + \Delta$$

for a smooth function V on $\mathbb{R}^n$ and denote P_t the associated semigroup. For $f \in L^2(\mu)$ such that $\int f d\mu = 0$ we have $f \in \mathcal{D}(L^{-1})$ iff

$$\int_0^\infty \|P_s f\|^2_{L^2(\mu)} ds < \infty$$

and in this case the Poisson equation has a unique solution g given by (2.1.4).

2.1.2 From Functional to Concentration Inequalities

Consider a stationary ergodic continuous-time Markov process X_t on a Polish space $\mathcal{X}$ with invariant measure μ and semigroup P_t. Assume that P_t is strongly continuous on $L^2(\mu)$ and denote L its generator with domain $\mathcal{D}_2(L)$ in $L^2(\mu)$.

We will start with some results that apply both to reversible and non-reversible processes. For that, we need to assume that the quadratic form $\mathcal{E}$ on $\mathcal{D}_2(L)$ defined by

$$\mathcal{E}(f) = \mathcal{E}(f, f) := -\int f L f d\mu$$

is closeable in $L^2(\mu)$. Denote $(\mathcal{E}, \mathcal{D}(E))$ its closure, which is the symmetrized Dirichlet form associated with X. When X is reversible $\mathcal{E}$ is always closeable.

For a measure ν on $\mathcal{X}$ define the Fisher-Donsker-Varadhan information of ν with respect to μ by

$$I(\nu|\mu) := \begin{cases} \mathcal{E}(\sqrt{d\nu/d\mu}), & \nu \ll \mu, \sqrt{d\nu/d\mu} \in \mathcal{D}(\mathcal{E}) \\ \infty \text{ else.} \end{cases}$$

For $f \in L^1(\mu)$ define $J'_f(r)$ and its lower semi-continuous regularisation $J_f(r)$ by

$$J'_f(r) = \inf\left\{I(\nu|\mu);\ \nu(|f|) < +\infty; \nu(f) = r\right\}, \quad J_f(r) = \lim_{\varepsilon \to 0^+} J'_f(r - \varepsilon).$$

The following theorem from [Wu00] forms the basis for all subsequent results in this section.

Proposition 2.1.5. *For any initial measure ν of X such that $\nu \ll \mu$ and $d\nu/d\mu \in L^2(\mu)$ we have for all $t > 0$, $R > 0$ and $f \in L^1(\mu)$*

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_s)ds - \mu(f) > R\right) \le \left\|\frac{d\nu}{d\mu}\right\|_{L^2(\mu)} \exp\left(-t\, J_f(R - \mu(f))\right).$$

Now, showing that certain functional inequalities for μ imply bounds on J_f leads to concentration inequalities for $\frac{1}{t}\int_0^t f(X_s)ds$.

We start with the following observation from [GGW14]: If there is a non-decreasing left-continuous convex function α_f with $\alpha_f(0) = 0$ such that for all measures ν on $\mathcal{X}$ with $f \in L^1(\nu)$

$$\alpha_f(\nu(f) - \mu(f)) \le I(\nu|\mu)$$

then $-J_f(r - \mu(f)) \le -\alpha_f(r)$ by the definition of J_f and the left-continuity of α_f.

This leads to the definition of transportation-information inequalities as in [Gui+09]. For instance, taking $\alpha_f(r) = r^2$ for 1-Lipschitz functions f leads to the definition of the L^1 transportation-information inequality $W_1 I$.

We say that μ satisfies a $W_1 I(c)$ inequality if for all measures ν on $\mathcal{X}$

$$W_1^2(\nu, \mu) \le 4c^2 I(\nu|\mu).$$

By the preceding observation together with the Kantorovich-Rubinstein duality we get the next result.

Proposition 2.1.6. *If μ satisfies a $W_1 I(c)$ inequality, then for all Lipschitz functions f on $\mathcal{X}$, $R > 0$ and initial measure ν such that $\nu \ll \mu$ and $d\nu/d\mu \in L^2(\mu)$*

$$\mathbb{P}_\nu\left(\frac{1}{t}\int_0^t f(X_s)ds - \mu(f) > R\right) \le \left\|\frac{d\nu}{d\mu}\right\|_{L^2(\mu)} \exp\left(-\frac{tR^2}{4c^2\|f\|_{\text{Lip}}^2}\right).$$

If X is reversible the last concentration bound is actually equivalent to the $W_1 I$ inequality [Gui+09].

A sufficient condition for $W_1 I(c)$ is a bound on solutions of the Poisson problem for Lipschitz continuous functions as in Proposition 2.1.3, see Corollary 2.2 in [Wu09].

We say that a Logarithmic Sobolev inequality with constant C holds for μ if for all $f \in \mathcal{D}_2(L)$

$$\int f^2 \log f^2 d\mu - \int f^2 d\mu \log \int f^2 d\mu \leq -2C \int fLf d\mu.$$

Still in the irreversible setting, the following result was shown in [Wu00].

Proposition 2.1.7. *Suppose that μ satisfies a Logarithmic Sovolev inequality with constant C. Then for all $f \in L^1(\mu)$*

$$\mathbb{P}_\nu \left(\frac{1}{t} \int_0^t f(X_s) ds - \mu(f) > R \right) \leq \left\| \frac{d\nu}{d\mu} \right\|_{L^2(\mu)} \exp\left(-\frac{t}{2C} H^*(R) \right)$$

where

$$H(\lambda) = \log \int e^{\lambda f} d\mu - \lambda \mu(f)$$

and

$$H^*(r) = \sup \left\{ \lambda r - H(\lambda); \lambda \in \mathbb{R} \right\}.$$

For the rest of this section, suppose that X is reversible, i.e.

$$\int fLg d\mu = \int gLf d\mu, \quad f, g \in \mathcal{D}_2(L).$$

What follows is a summary of the result from [GGW14] where the authors investigated the following Bernstein-type inequality for different classes of functions f and constants M under various ergodicity assumptions on X:

$$\begin{aligned} \mathbb{P}_\nu \left(\frac{1}{t} \int_0^t f(X_s) ds - \mu(f) > R \right) &\leq \left\| \frac{d\nu}{d\mu} \right\|_{L^2(\mu)} \exp\left(-\frac{tR^2}{\sigma^2 \left(\sqrt{1 + 2MR/\sigma^2} + 1 \right)^2} \right) \\ &\leq \left\| \frac{d\nu}{d\mu} \right\|_{L^2(\mu)} \exp\left(-\frac{tR^2}{2\left(\sigma^2 + MR\right)} \right) \end{aligned}$$

with

$$\sigma^2(f) = \lim_{t \to \infty} \mathrm{Var}_\mu \left(\int_0^t f(X_s) ds \right).$$

This inequality was first proved by Lezaud [Lez01] for bounded measurable f under a Poincaré inequality on μ. The standing assumption for the following results is that μ satisfies a Poincaré inequality with constant c_P in the sense that

$$\mathrm{Var}_\mu(f) \leq c_P \mathcal{E}(f), \quad f \in \mathcal{D}(\mathcal{E}).$$

1. (Proposition 2.1 [GGW14]) If only the Poincaré inequality holds, then the Bernstein inequality holds for bounded measurable f with

$$M = c_P \|f^+\|_\infty.$$

2. (Theorem 3.1 [GGW14]) If a Logarithmic Sobolev inequality holds with constant c_{LS}, then the Bernstein inequality holds for $f \in L^2(\mu)$ such that

$$\Lambda(\lambda) := \log \int e^{\lambda f} d\mu < \infty$$

with

$$M = \int_{\lambda > 0} \frac{1}{\lambda}[c_P \Lambda(\lambda) + 2c_{LS}].$$

3. (Corollary 3.1 [GGW14]) If a Logarithmic Sobolev inequality holds with constant c_{LS} and X has continuous sample paths, then the Bernstein inequality holds for $f \in \mathcal{D}(\mathcal{E})$ such that $\|\Gamma(f)\|_\infty < \infty$ with

$$M = 2c_{LS}\sqrt{c_P \|\Gamma(f)\|_\infty}.$$

4. (Theorem 4.1 [GGW14]) If there exist functions $U > 1$, $\phi > 0$ and a constant $b > 0$ such that

$$\frac{-LU}{U} \geq \phi - b$$

then the Bernstein inequality holds for all $f \in L^2(\mu)$ such that

$$K_\phi(f^+) := \inf\{C \geq 0 : |f^+| \leq C\phi\} < \infty$$

with

$$M = K_\phi(f^+)(bc_P + 1).$$

5. For all $f \in L^2(\mu)$ such that the Poisson problem $-LF = f$ has a solution F in $L^2(\mu)$ with $\|\Gamma(F)\|_\infty < \infty$ the Bernstein inequality holds with

$$M = 2\sqrt{c_P \|\Gamma(f)\|_\infty}.$$

2.1.3 Regeneration methods

Following [LL13], consider a positive Harris recurrent continuous-time strong Markov process X on a Polish space $\mathcal{X}$ with invariant measure μ. Denote L the associated infinitesimal generator and P_t the semigroup, which we assume to have a density with respect to some σ-finite positive measure on $\mathcal{X}$.

Suppose that the following Lyapunov-type condition holds: There exists a continuous function $V : E \to [1, \infty)$, constants $b, c > 0$ and a closed measurable set B such that for some $0 \le \alpha < 1$

$$LV \le -cV^{\alpha} + b\mathbb{1}_B.$$

In [LL13] the authors show that for bounded functions $f \in L^1(\mu)$ and $\varepsilon < \|f\|_\infty, t \ge 1$ it holds that

$$\mathbb{P}_x\left(\left|\frac{1}{T}\int_0^T f(X_t)dt - \mu(f)\right| \ge \varepsilon\right) \le K(\alpha)V(x)\left(\frac{\|f\|_\infty^2}{t\varepsilon^2}\right)^{\alpha/(1-\alpha)}. \tag{2.1.5}$$

The proof relies on the so-called Nummelin splitting technique in continuous time, which yields a sequence of stopping times R^n, called regeneration times, such that the random variables

$$\xi_n := \int_{R^n}^{R^{n+1}} f(X_t) - \mu(f)dt$$

are identically distributed and two-dependent, meaning ξ_n and ξ_{n+k} are independent for all $k \ge 2$. It can then be shown that the left-hand side of (2.1.5) can be estimated by

$$\mathbb{P}_x\left(\sum_{n=1}^{N}|\xi_n| > \varepsilon t/3\right)$$

plus some remainder terms. A deviation inequality for two-dependent identically distributed random variables then yields a bound in terms of $\mathbb{E}|\xi_1|^p$ for $p > 1$. Using

$$\mathbb{E}|\xi_1|^p \le 2\|f\|_\infty^p \mathbb{E}(R_2 - R_1)^p$$

the result then follows from an estimate on the expectation in the right-hand side, which is an analogue of excursion times.

2.2 Averaging Principle

2.2.1 Introduction

The theory of averaging is concerned with processes that admit a decomposition into fast and slow degrees of freedom. A typical example would be the solution to an SDE of the form

$$dX_t = \alpha b_X(X_t, Y_t)dt + \sqrt{\alpha}dB_t^X, \quad X_0 = x \tag{2.2.1a}$$

$$dY_t = b_Y(X_t, Y_t)dt + dB_t^X, \quad Y_0 = y \tag{2.2.1b}$$

where b_X, b_Y are bounded Lipschitz continuous real-valued coefficients and B^X, B^Y standard independent Brownian motions on $\mathbb{R}$. The parameter $\alpha \ge 0$ represents the acceleration of X relative to Y. We say that an averaging principle holds if there exists an averaged process $\bar{Y}$ that is independent of B^X such that Y converges to $\bar{Y}$ as $\alpha \to \infty$. Intuitively we expect the fluctuations of X to average out due to the separation of timescales. We will below review a number of averaging results from the literature, together with the conditions they impose on the process and the sense in which the convergence is to be understood.

But first, let us have a closer look at $\bar{Y}$. The intuition is that due to the separation of time scales Y sees X as quasi-static whereas X sees Y as almost equilibrated. Consider the following process X^y, corresponding to X with the Y component frozen at a fixed value y:

$$dX_t^y = b_X(X_t^y, y)dt + dB_t^X, \quad X_0^y = x$$

Suppose that X^y has a stationary distribution μ^y for each y, and let

$$\bar{b}(y) = \int b_Y(x, y)\mu^y(dx).$$

Then it is part of the statement of an averaging principle that the averaged process $\bar{Y}$ solves the SDE

$$d\bar{Y}_t = \bar{b}(\bar{Y}_t)dt + dB_t^Y.$$

We will concentrate on averaging principles for stochastic differential equations. The most general SDEs for which an averaging principle is known to hold are of the form

$$\begin{aligned} dX_t &= \alpha b_X(X_t, Y_t)dt + \sqrt{\alpha}\sigma_X(X_t, Y_t)dB_t^X, \quad X_0 = x \\ dY_t &= b_Y(X_t, Y_t)dt + \sqrt{\alpha}c_Y(X_t, Y_t) + \sigma_Y(X_t, Y_t)dB_t^X, \quad Y_0 = y \end{aligned}$$

and, depending on the regularity of the coefficients, averaging principles have been shown for strong convergence on the space of trajectories and both strong and weak convergence of the time marginals.

Compared to the simple example (2.2.1), the coefficient $\sigma_X(x, y)$ does not cause any extra difficulties as long as the process remains elliptic. Neither does $\sigma_Y = \sigma_Y(y)$ as long as it only depends on the y variable. The situation with an extra coefficient c_Y requires some additional technical tools but is generally well-understood. However, $\sigma_Y = \sigma_Y(x, y)$ introduces a qualitative difference and a simple counterexample to the strong convergence of Y to $\bar{Y}$ can be found in [Liu10]. In this case only weak convergence of the time marginals on compact state spaces has been shown.

Large and moderate deviation principles for averaging are in general well understood, and there has been some progress on quantitative non-asymptotic estimates under assumptions of strong contractivity or the existence of a reversible stationary distribution for the process (X, Y), see [Liu10] and Section 2.2.4 below.

We will now proceed to present the main approaches to averaging for SDEs found in the mathematical literature and give an outline of their proofs in the setting of the example (2.2.1).

2.2.2 Freezing the slow process

This method essentially consists in formalising the intuition of the frozen slow process. It consists of dividing a fixed time interval $[0, T]$ into subintervals of size $\Delta(\alpha)$ on which Y is kept constant. On each interval of length Δ an ergodic theorem is then applied to a process still running on the time scale α. This imposes the requirement that $\alpha\Delta(\alpha) \to \infty$ as $\alpha \to \infty$. Since the approximating process still needs to converge to the original process, we also need $\Delta(\alpha) \to 0$ as $\alpha \to \infty$.

We now outline a proof of an averaging principle using a frozen slow process for (2.2.1), following [FW12, Section 7.9]. Since the estimates hold only asymptotically we make the dependence of X and Y on α explicit by using the notation X^α, Y^α whenever it is necessary for clarity of presentation. The statement is:

Proposition 2.2.1. *Suppose that there exists a constant C independent of y such that for any $y \in \mathbb{R}$, $T > 0$*

$$\mathbb{E}\left|\frac{1}{T}\int_0^T b_Y(X_t^y, y)dt - \int b_Y(x, y)\mu^y(dx)\right| \leq \frac{C}{\sqrt{T}}.$$

Then for any $T > 0, \delta > 0$ and any initial values x, y of X_t^α, Y_t^α it holds

$$\lim_{\alpha\to\infty} \mathbb{P}\left(\sup_{0\leq t\leq T} \left|X_t^\alpha - \bar{X}_t\right| > \delta\right) = 0$$

Outline of Proof. Consider a partition of $[0, T]$ into intervals of length Δ. Construct a decoupled process $\hat{X}, \hat{Y}$ piecewise such that for $t \in [k\Delta, (k+1)\Delta]$

$$\hat{X}_t = X_{k\Delta} + \alpha \int_{k\Delta}^t b_X(\hat{X}_s, Y_{k\Delta})ds + \sqrt{\alpha}\int_{k\Delta}^t dB_s^X$$
$$\hat{Y}_t = \hat{Y}_{k\Delta} + \int_{k\Delta}^t b_Y(\hat{X}_s, Y_{k\Delta})ds + \int_{k\Delta}^t dB_s^Y$$

Note that for $t \in [k\Delta, (k+1)\Delta]$, given initial conditions $X_{k\Delta}, Y_{k\Delta}$ the SDE for $\hat{X}_t$ is closed and can be solved in isolation from $\hat{Y}_t$. We have passed from a system of mutually dependent SDEs to one in which the dependence only goes in one direction.

Now the proof consists in showing that (X, Y) converges in probability to $(\hat{X}, \hat{Y})$ and that $\hat{Y}$ converges in probability to $\bar{Y}$.

We have for $t \in [k\Delta, (k+1)\Delta]$

$$\begin{aligned}\mathbb{E}\left|X_t - \hat{X}_t\right|^2 &= \alpha^2\mathbb{E}\left|\int_{k\Delta}^t b_X(X_s, Y_s) - b_X(\hat{X}_s, Y_{k\Delta})ds\right|^2 \\ &\leq \alpha^2(t - k\Delta)\|b_X\|_{\text{Lip}}\int_{k\Delta}^t \mathbb{E}\left|X_s - \hat{X}_s\right|^2 + \mathbb{E}\left|Y_s - Y_{k\Delta}\right|^2 ds.\end{aligned}$$

Now the key to decoupling is that by using the boundedness of b_Y we can get the estimate

$$\mathbb{E}\left|Y_t - Y_{k\Delta}\right|^2 = \mathbb{E}\left|\int_{k\Delta}^t b_Y(X_s, Y_s)ds\right|^2 \leq \|b_Y\|_\infty \Delta^2$$

without studying the interdependence of X and Y.

We get

$$\mathbb{E}\left|X_t - \hat{X}_t\right|^2 \leq \alpha^2 \Delta \|b_X\|_{\text{Lip}} \int_{k\Delta}^t \mathbb{E}\left|X_s - \hat{X}_s\right|^2 ds + \alpha^2 \Delta^4 \|b_X\|_{\text{Lip}} \|b_Y\|_\infty$$

so that by Gronwall's inequality

$$\mathbb{E}\left|X_t - \hat{X}_t\right|^2 \leq \alpha^2 \Delta^4 \|b_X\|_{\text{Lip}} \|b_Y\|_\infty \exp\left(\alpha^2 \Delta^2 \|b_X\|_{\text{Lip}}\right).$$

Now

$$\begin{aligned}
&\mathbb{E}\left|\int_0^T \left|b_Y(X_s, Y_s) - b_Y(\hat{X}_s, Y_{\lfloor s/\Delta \rfloor \Delta}) ds\right|^2\right| \\
&\leq T\|b_Y\|_{\text{Lip}} \int_0^T \mathbb{E}\left|X_s - \hat{X}_s\right|^2 + \mathbb{E}\left|Y_s - Y_{\lfloor s/\Delta \rfloor \Delta}\right|^2 ds \\
&\leq T^2 \|b_Y\|_{\text{Lip}} \left(\Delta^4 \alpha^2 \|b_X\|_{\text{Lip}} \|b_Y\|_\infty \exp\left(\alpha^2 \Delta^2 \|b_X\|_{\text{Lip}}\right) + \Delta^2 \|b_Y\|_\infty\right).
\end{aligned}$$

If we put $\Delta(\alpha) = \frac{\sqrt{\log \alpha}}{\alpha}$ then this expression goes to 0 as $\alpha \to \infty$.

From this we get that as $\alpha \to \infty$

$$\mathbb{P}\left(\sup_{0 \leq t \leq T} \left|Y_t - \hat{Y}_t\right| > \delta\right) \to 0$$

and

$$\sup_{0 \leq t \leq T} \mathbb{E}\left|Y_t - \hat{Y}_t\right|^2 \to 0. \tag{2.2.2}$$

Recall the definition of $\bar{Y}$

$$\bar{Y}_t = y_0 + \int_0^t \bar{b}(\bar{Y}_s) ds + B_t^Y$$

so that

$$\begin{aligned}
&\mathbb{E}\left|\hat{Y}_t - \bar{Y}_t\right|^2 \\
&= \mathbb{E}\left|\int_0^t b_Y(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(\bar{Y}_s)ds\right|^2 \\
&\leq 3\mathbb{E}\left|\int_0^t b_Y(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_s)ds\right|^2 + 3\mathbb{E}\left|\int_0^t \bar{b}(Y_s) - \bar{b}(\hat{Y}_s)ds\right|^2 \\
&\quad + 3\mathbb{E}\left|\int_0^t \bar{b}(\hat{Y}_s) - \bar{b}(\bar{Y}_s)ds\right|^2 \\
&\leq 3\mathbb{E}\left|\int_0^t b_Y(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_s)ds\right|^2 + 3T\|\bar{b}\|_{\text{Lip}}\int_0^t \mathbb{E}\left|Y_s - \hat{Y}_s\right|^2 ds \\
&\quad + 3T\|\bar{b}\|_{\text{Lip}}\int_0^t \mathbb{E}\left|\hat{Y}_s - \bar{Y}_s\right|^2 ds
\end{aligned}$$

and by Gronwall's inequality

$$\begin{aligned}
\mathbb{E}\left|\hat{Y}_t - \bar{Y}_t\right|^2 \leq \Bigg(&3\mathbb{E}\left|\int_0^t b_Y(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_s)ds\right|^2 \\
&+ 3T\|\bar{b}\|_{\text{Lip}}\int_0^t \mathbb{E}\left|Y_s - \hat{Y}_s\right|^2 ds\Bigg)e^{3T^2\|\bar{b}\|_{\text{Lip}}}.
\end{aligned} \tag{2.2.3}$$

We now use our ergodic assumption to obtain

$$\begin{aligned}
&\mathbb{E}\left|\int_{k\Delta}^{(k+1)\Delta} b_Y(\hat{X}_t, Y_{k\Delta}) - \bar{b}(Y_{k\Delta})ds\right| \\
&= \Delta\mathbb{E}\left|\frac{1}{\alpha\Delta}\int_0^{\alpha\Delta} b_Y(X_t^{Y_{k\Delta}}, Y_{k\Delta}) - \bar{b}(Y_{k\Delta})dt\right| \\
&\leq C\sqrt{\Delta}/\sqrt{\alpha}
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{E} \sup_{0\le t\le T}\left|\int_0^t b_Y(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_s)\right| \\
&\quad\le \mathbb{E} \sup_{0\le t\le T}\left|\int_0^t b_Y(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_{\lfloor s/\Delta\rfloor\Delta})\right| + \mathbb{E} \sup_{0\le t\le T}\left|\int_0^t \bar{b}(Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_s)\right| \\
&\quad\le \mathbb{E} \max_{0\le n\le T/\Delta}\left|\sum_{k=0}^{n}\int_{k\Delta}^{(k+1)\Delta} b(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_{\lfloor s/\Delta\rfloor\Delta})ds\right| \\
&\qquad + \|\bar{b}\|_{\text{Lip}}\mathbb{E}\int_0^T \left|Y_s - Y_{\lfloor s/\Delta\rfloor\Delta}\right| ds \\
&\quad\le \sum_{k=0}^{T/\Delta-1}\mathbb{E}\left|\int_{k\Delta}^{(k+1)\Delta} b(\hat{X}_s, Y_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_{\lfloor s/\Delta\rfloor\Delta})ds\right| + \|\bar{b}\|_{\text{Lip}}T\Delta \\
&\quad\le \frac{TC}{\sqrt{\alpha\Delta}} + \|\bar{b}\|_{\text{Lip}}T\Delta.
\end{aligned}
$$

Together with (2.2.3) we finally obtain

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{0\le t\le T}\left|\hat{Y}_t^\alpha - \bar{Y}\right| > \delta\right) \\
&\le \mathbb{P}\left(\sup_{0\le t\le T}\left|\int_0^t b_Y(\hat{X}_s, Y^\alpha_{\lfloor s/\Delta\rfloor\Delta}) - \bar{b}(Y_s^\alpha)\right| > \delta/3\right) \\
&\quad + \mathbb{P}\left(\int_0^T \left|\bar{b}(\bar{Y}_s) - \bar{b}(\hat{Y}_s^\alpha)\right| > \delta/3\right) \\
&\quad + \mathbb{P}\left(\int_0^T \left|\bar{b}(\hat{Y}_s^\alpha) - \bar{b}(Y_s^\alpha)\right| > \delta/3\right)
\end{aligned}
$$

which converges to 0 as $\alpha \to \infty$ by the preceding results.

□

2.2.3 Asymptotic expansion of the generator

In this section we present the asymptotic expansion approach from [KY04]. The goal is to establish a weak averaging principle by identifying a process $\bar{Y}$ independent of α such that for appropriate test functions f and $t > 0$ fixed

$$
\mathbb{E}[f(Y_t^\alpha)] \to \mathbb{E}[f(\bar{Y}_t)] \text{ as } \alpha \to \infty.
$$

The technique consists of a formal series expansion in α^{-1} of the one-parameter semigroup associated to (X^α, Y^α), which is justified by making use of the compactness of the state space, and then letting $\alpha \to \infty$ in the resulting expansion.

We assume that (X^α, Y^α) takes values in a compact C^∞-manifold $\mathcal{K} = \mathcal{K}_X \times \mathcal{K}_Y$ where $\mathcal{K}_X$ and $\mathcal{K}_Y$ are connected and compact C^∞-manifolds. Now, the idea behind the asymptotic expansion approach is to note that the generator L^X for (X, Y) can be split into two operators L^X and L^Y such that $L = L^Y + \alpha L^X$. Suppose that $u(t, x, y) = P_t f(x, y)$ can be written as a power series in α^{-1} such that $u(t, x, y) = u_0(t, y) + \sum_{i=1}^\infty \alpha^{-i} u_i(t, x, y)$. Then we can formally substitute this series in the backward Kolmogorov equation $\partial_t u = L^Y u + \alpha L^X u$ and equate coefficients of matching powers on both sides.

Define the operators L^X and L^Y by

$$L^X f(x,y) = b_X(x,y)\partial_x f(x,y) + \partial_x^2 f(x,y)$$
$$L^Y f(x,y) = b_Y(x,y)\partial_y f(x,y) + \partial_y^2 f(x,y).$$

Let $u_\alpha(t, x, y)$ be the solution to the following Cauchy problem on $\mathcal{K}$:

$$\partial_t u_\alpha = L^Y u_\alpha + \alpha L^X u_\alpha \tag{2.2.4a}$$
$$u_\alpha(0, x, y) = f(x, y) \tag{2.2.4b}$$

or equivalently $u_\alpha(t, x, y) = \mathbb{E}^{x,y} f(X_t^\alpha, Y_t^\alpha)$.

Due to the compactness of $\mathcal{K}_X$ and the regularity of the coefficients there exists for each y fixed an invariant measure μ^y associated to L^X such that for all smooth test functions f on $\mathcal{K}_X$

$$\int_{\mathcal{K}_X} L^X f(x,y)\mu^y(dx) = 0, \quad y \in \mathcal{K}_Y.$$

For functions $\varphi(x, y)$ on $\mathcal{K}$ we will use the notation

$$\bar{\varphi}(y) := \int_{\mathcal{K}_X} \varphi(x,y)\mu^y(dx).$$

Define the truncated series

$$u_\alpha^n(t,x,y) = u_0(t,y) + \sum_{i=1}^n \alpha^{-i} u_i(t,x,y) + \sum_{i=0}^n \alpha^{-i} v_i(\alpha t, x, y)$$

and the associated error term

$$e_{\alpha,n} = u_\alpha^n - u_\alpha.$$

To justify the series expansion up to order n, we need to show that the growth of $e_{\alpha,n}$ in α is of order $O(\alpha^{-(n+1)})$. We will outline in the following the procedure for $n = 0$, see [KY04] for a complete proof and the extension to arbitrary powers n.

We begin by substituting $u_0(t, y) + \sum_{i=1}^{n+1} \alpha^{-i} u_i(t, x, y)$ into (2.2.4) and equating powers of α so that

$$\partial_t u_k = L^X u_{k+1} + L^Y u_k \tag{2.2.5}$$

and similarly for the boundary layer term $v_k(\tau, x, y)$ with $\tau = \alpha t$

$$\partial_\tau v_0 = L^X v_0 \tag{2.2.6a}$$

$$\partial_\tau v_k = L^X v_k + L^Y v_{k-1}. \tag{2.2.6b}$$

Since the leading term consists of the sum $u_0 + v_0$ we can choose the following initial condition for u_0, which is independent of x,

$$u_0(0, y) = \mu^y(f)$$

and let

$$v_0(0, x, y) = f(x, y) - \mu^y(f).$$

Using the fact that u_0 is a function of t and y only and integrating (2.2.5) against μ^y for $k = 0$ yields

$$\begin{aligned} \partial_t u_0(t, y) &= \int L^X u_1(t, x, y)\mu^y(dx) + \int L^Y u_0(t, x, y)\mu^y(dx) \\ &= \bar{L}^Y u_0(t, y), \quad \bar{L}^Y = \bar{b}(y)\partial_y + \partial_y^2 \end{aligned} \tag{2.2.7}$$

where we used the fact that μ^y is an invariant measure for L^X. We have the probabilistic representation

$$u_0(t, y) = \mathbb{E}^y \mu^{\bar{Y}_t}(f).$$

We also chose above

$$v_0(0, x, y) = f(x, y) - \mu^y(f)$$

so that from (2.2.6) we get

$$v_0(\tau, x, y) = \mathbb{E}^x f(X_\tau^y, y) - \mu^y(f)$$

where X^y is the process with "frozen" y solution to the SDE

$$dX_t^y = b_X(X_t^y, y)dt + dB_t^X, \quad X_0^y = x.$$

It can be shown using compactness of the state space that (using multi-index notation)

$$\left| \frac{\partial^{|\nu|} v_0}{\partial_y^{\nu_1} \dots \partial_{y_d}^{\nu_d}} \right| \le c_1 e^{-c_2 \tau}, \quad |\nu| = 0 \dots 4$$

for some finite constants c_1 and c_2. This implies an exponential bound of the same form on $L^Y v_0$.

In order to get an error estimate, we also need to determine u_1 and v_1. Substracting (2.2.7) from (2.2.5) we get

$$L^X u_1 = (\bar{L}^Y - L^Y)u_0. \tag{2.2.8}$$

The solution writes as

$$u_1(t,x,y) = U_1(t,y) + \tilde{u}_1(t,x,y)$$

where $\tilde{u}_1(t,x,y)$ is a particular centered solution of the Poisson equation (2.2.8).

In order to determine U_1, we integrate (2.2.5) against μ^y to get

$$\begin{aligned}\partial_t U_1(t,y) &= \bar{L}^Y U_1 + \mu^y(L^Y \tilde{u}_1)\\ U_1(0,y) &= \int u_1(0,x,y)\mu^y(dx).\end{aligned}$$

In order to determine the initial condition, first observe that since there were no terms in α in the initial conditions for our original Cauchy problem, we have

$$u_k(0,x,y) + v_k(0,x,y) = 0 \text{ for all } k \geq 1.$$

We then impose $v_1(\tau,x,y) \to 0$ as $\tau \to \infty$ and integrate (2.2.6) in time from 0 to ∞ and in space against μ^y to get

$$\int u_1(0,x,y)\mu^y(dx) = -\int v_1(0,x,y)\mu^y(dx) = \int_0^\infty \int_{\mathcal{K}_X} L^Y v_0(s,x,y)\mu^y(dx)ds$$

where the right hand side is finite due to the exponential decay of $L^Y v_0$. This initial condition together with (2.2.6) uniquely determines v_1.

It can again be shown that there exists constants c_1 and c_2 (possibly different from the ones in the estimate for v_0) such that

$$\left| \frac{\partial^{|\nu|} v_1}{\partial_y^{\nu_1} \ldots \partial_{y_d}^{\nu_d}} \right| \leq c_1 e^{-c_2 \tau}, \quad |\nu| = 0 \ldots 4.$$

This implies in particular that $L^Y v_1$ is bounded and exponentially decaying. It can be noted that the proof in [KY04] relies heavily on the compactness of the state space. Unlike in the case of v_0, there is no obvious probabilistic interpretation of v_1 and it is unclear how the result could be extended to a non-compact setting.

We now proceed to bound the growth of $e_{\alpha,0}$ so that the series expansion to order 0 is justified. Let

$$L_\alpha f = \partial_t f - \alpha L^X f - L^Y f.$$

Now

$$\begin{aligned}L_\alpha e_{\alpha,0} &= L_\alpha u_0 + L_\alpha v_0\\ &= \partial_t u_0 - \alpha L^X u_0 - L^Y u_0 + \alpha(\partial_\tau v_0 - L^X v_0) - L^Y v_0\\ &= \partial_t u_0 - L^Y u_0 - L^Y v_0\end{aligned}$$

is of order $O(1)$, since we saw above that $L^Y v_0$ is bounded. Similarly, using that $L^Y v_1$ is bounded, it can be shown that $L_\alpha e_{\alpha,1}$ is of order $O(\alpha^{-1})$. This implies that $e_{\alpha,1}$ is also of order $O(\alpha^{-1})$.

We conclude by noting that

$$e_{\alpha,0} = e_{\alpha,1} - \alpha^{-1} u_1 - \alpha^{-1} v_1$$

where all the terms on the right hand side are of order $O(\alpha^{-1})$.

2.2.4 Effective dynamics using conditional expectations

We give a short description of some of the ideas employed in [LL10] and [LLO16]. Let us introduce a process $\tilde{Y}$ solution to

$$d\tilde{Y}_t = \tilde{b}(\tilde{Y}_t)dt + dB^Y_t$$

with

$$\tilde{b}(y) = \mathbb{E}[b_Y(X_t, Y_t)|Y_t = y].$$

Now suppose furthermore that (X, Y) has a joint stationary probability measure $\mu = e^{-V(x,y)}dxdy$ for a smooth function V and write $\mathbb{E}_\mu$ for the expectation with initial measure μ. Then we have for all $f \in C_c^\infty(\mathbb{R} \times \mathbb{R}), g \in C_c^\infty(\mathbb{R})$

$$\begin{aligned}
\mathbb{E}_\mu g(Y_t) f(X_t, Y_t) &= \int g(y) \int f(x,y) e^{-V(x,y)} dxdy \\
&= \int g(y) \left(\frac{\int f(x,y) e^{-V(x,y)} dx}{\int e^{-V(x,y)} dx} \right) e^{-V(x,y)} dxdy \\
&= \mathbb{E}_\mu g(Y_t) \tilde{\mu}^{Y_t}(f)
\end{aligned}$$

with

$$\tilde{\mu}^y(dx) = \frac{e^{-V(x,y)}}{\int e^{-V(x,y)} dx}$$

so that by Kolmogorov's characterisation of conditional expectation

$$\mathbb{E}_\mu[f(X_t, Y_t)|Y_t = y] = \tilde{\mu}^y(f)$$

and therefore

$$\tilde{b}(y) = \tilde{\mu}^y(b_Y).$$

Now suppose furthermore that $b_X = -\partial_x V(x,y)$. Then for each y, X^y is a reversible diffusion with invariant probability measure $\tilde{\mu}^y$ so that $\tilde{\mu}^y = \mu^y$ and $\tilde{b}(y) = \bar{b}(y)$.

Finally, suppose that $b_Y(x,y) = -\partial_y V(x,y)$ so that (X,Y) is reversible with respect to μ. Then we can use a forward-backward martingale argument to show that for $T > t$ arbitrary but fixed and f centered, meaning $\mu(f) = 0$,

$$2\int_0^t f(X_s, Y_s)ds = M_t + (\tilde{M}_T - \tilde{M}_{T-t})$$

where M and $\tilde{M}$ are martingale with quadratic variation

$$\langle M\rangle_t = \langle \tilde{M}\rangle_t = \int_0^t \Gamma((-L)^{-1}f)(X_s, Y_s)ds.$$

Now we make the observation that, since μ is an invariant measure for both L and L^X and f is centered with respect to μ, $f \in \mathcal{D}((-L^X)^{-1})$ and we can write

$$f = (-L^X)(-L^X)^{-1}f$$

so that

$$\begin{aligned}
\int \Gamma((-L)^{-1}f)d\mu &= \int fL^{-1}fd\mu = -\int L^{-1}fL^X(-L^X)^{-1}fd\mu \\
&= \int \Gamma^X(L^{-1}f, (-L^X)^{-1}f)d\mu \\
&\leq \left(\int \Gamma^X(-L^{-1}f)d\mu\right)^{1/2} \left(\int \Gamma^X((-L^X)^{-1}f)d\mu\right)^{1/2} \\
&\leq \left(\int \Gamma(-L^{-1}f)d\mu\right)^{1/2} \left(\int \Gamma^X((-L^X)^{-1}f)d\mu\right)^{1/2}
\end{aligned}$$

so that after dividing both sides by the square root of the left hand side

$$\int \Gamma((-L)^{-1}f)d\mu \leq \int \Gamma^X((-L^X)^{-1}f)d\mu.$$

Remark 2.2.2 (Probabilistic interpretation).

$$\int_0^\infty \mathrm{Cov}_\mu(X_0, X_t)dt \leq \int_0^\infty \mathrm{Cov}_\mu(X_0^{Y_0}, X_t^{Y_0})dt$$

Freezing the slow component degrades the mixing. To be investigated.

We have the following dual formulation of the Poincaré inequality (see [BGL14] Proposition 4.8.3): A measure ν satisfies a Poincaré inequality with constant c_P with respect to a carré du champs Γ and associated reversible generator L iff

$$\int \Gamma(f)d\nu \leq c_P \int (Lf)^2 d\nu.$$

Suppose now that a Poincaré inequality with constant c_P holds for μ^y uniformly in y with respect to the standard carré du champs in the sense that

$$\int f^2 d\mu^y \leq c_P \int |\nabla f|^2 d\mu^y.$$

Since $\Gamma^X = \alpha|\nabla f|^2$ this implies a Poincaré inequality with constant c_P/α with respect to Γ^X. We can now combine the two previous results to obtain

$$\begin{aligned}\mathbb{E}_\mu \langle M \rangle_t = \int_0^t \int \Gamma((-L)^{-1} f)(x,y) \mu(dx,dy) \\ \leq \int_0^t \int \Gamma^X((-L^X)^{-1} f)(x,y) \mu^y(dx)\mu(dy) \leq \frac{c_P}{\alpha} \|f\|_{L^2(\mu)} t.\end{aligned}$$

Together with the forward-backward martingale decomposition, using Doob's maximal inequality this gives

$$\mathbb{E} \left| \sup_{0 \leq t \leq T} \int_0^t b_Y(X_s, Y_s) - \mu^{Y_s} ds \right|^2 \leq \frac{27}{4} \frac{c_P}{\alpha} \|f\|_{L^2(\mu)} T.$$

We have

$$Y_t - \bar{Y}_t = \int_0^t b_Y(X_s, Y_s) - \mu^{Y_s}(b_Y) ds + \int_0^t \bar{b}(Y_s) - \bar{b}(\bar{Y}_s) ds.$$

We saw above how to estimate the first term. For the second term, we can assume that $\bar{b}$ is Lipschitz so that by Gronwall's inequality

$$|Y_t - \bar{Y}_t| \leq \left(\int_0^t b_Y(X_s, Y_s) - \mu^{Y_s}(b_Y) ds \right) e^{\|\bar{b}\|_{\mathrm{Lip}} t}$$

and finally

$$\mathbb{E} \sup_{0 \leq t \leq T} |Y_t - \bar{Y}_t|^2 \leq \frac{27}{4} \frac{c_P}{\alpha} \|b_Y\|_{L^2(\mu)} T e^{2\|\bar{b}\|_{\mathrm{Lip}} T}.$$

In [LLO16] it is shown how the Lipschitz condition on $\bar{b}$ can be relaxed to a one-sided Lipschitz condition in the one-dimensional case.

Chapter 3

Time Averages of Diffusion Processes

3.1 Introduction

For a Markov process $(X_t)_t$ with $t \in [0, T]$ or $t = 0, 1, \ldots, T$ let

$$S_T f = \int_0^T f(t, X_t) dt$$

in the continuous-time case or

$$S_T f = \sum_{t=0}^{T-1} f(t, X_t)$$

in discrete time.

In the first part of this work, we will show a decomposition of the form

$$S_T f = \mathbb{E} S_T f + M_T^{T,f}$$

where $M^{T,f}$ is a martingale depending on T and f for which we will give an explicit representation in terms of the transition operator or semigroup associated to X.

We then proceed to illustrate how the previous results can be used to obtain Gaussian concentration inequalities for S_T when X is the solution to an Itô SDE.

The last part of the work showcases a number of results on two-timescale processes that follow from our martingale representation.

3.2 Martingale Representation

Consider the following SDE with time-dependent coefficients on $\mathbb{R}^n$:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = x$$

where B is a standard Brownian motion on $\mathbb{R}^n$ with filtration $(\mathcal{F}_t)_{t\geq 0}$ and $b(t,x), \sigma(t,x)$ are continuous in t and locally Lipschitz continuous in x. We assume that X_t does not explode in finite time.

Denote C_c^∞ the set of smooth compactly supported space-time functions on $\mathbb{R}_+ \times \mathbb{R}^n$.

Let $P_{s,t}$ be the evolution operator associated to X,

$$P_{s,t}f(x) = \mathbb{E}\left[f(t, X_t)|X_s = x\right], \quad f \in C_c^\infty.$$

For $T > 0$ fixed consider the martingale

$$M_t = \mathbb{E}^{\mathcal{F}_t} \int_0^T f(s, X_s)ds.$$

and observe that since X is adapted and by the Markov property

$$M_t = \int_0^t f(s, X_s)ds + \mathbb{E}^{\mathcal{F}_t} \int_t^T f(s, X_s)ds = \int_0^t f(s, X_s)ds + R_t^T f(X_t)$$

with

$$R_t^T f(x) = \int_t^T P_{t,s} f(x)ds.$$

By applying the Itô formula to $R_t^T f$ we can identify the martingale M. This is the content of the following short theorem.

Theorem 3.2.1. *For $T > 0$ fixed, $t \in [0, T]$ and $f \in C_c^\infty$*

$$\int_0^t f(s, X_s)ds + R_t^T f(X_t) = \mathbb{E} \int_0^T f(s, X_s)ds + M_t^{T,f}$$

with

$$M_t^{T,f} = \int_0^t \nabla R_s^T f(X_s) \cdot \sigma(s, X_s)dB_s.$$

Proof. From the Kolmogorov backward equation $\partial_t P_{t,s} f = -L_t P_{t,s} f$ and since $P_{t,t} f = f$ we have

$$\partial_t R_t^T f(x) = -f(t, x) - \int_t^T L_t P_{t,s} f(x)ds = -f(t, x) - L_t R_t^T f(x).$$

By Itô's formula

$$\begin{aligned}R_t^T f(X_t) &= R_0^T f(X_0) + \int_0^t \partial_s R_s^T f(X_s) ds + \int_0^t L_s R_s^T f(X_s) ds \\ &\quad + \int_0^t \nabla R_s^T f(X_s) \cdot \sigma(s, X_s) dB_s \\ &= \mathbb{E} \int_0^T f(t, X_t) dt - \int_0^t f(s, X_s) ds + \int_0^t \nabla R_s^T f(X_s) \cdot \sigma(s, X_s) dB_s\end{aligned}$$

and we are done.

□

Remark 3.2.2 (Poisson Equation). In the time-homogeneous case $P_{t,s} = P_{s-t}$ and when the limit below is finite then it is independent of t and we have

$$R^\infty f := \lim_{T\to\infty} R_t^T f = \lim_{T\to\infty} \int_t^T P_{s-t} f ds = \lim_{T\to\infty} \int_0^{T-t} P_s f ds = \int_0^\infty P_s f ds.$$

This is the resolvent formula for the solution to the Poisson equation $-Lg = f$ with $g = R^\infty f$.

By taking $t = T$ in Theorem 3.2.1 we can identify the martingale part in the martingale representation theorem for $\int_0^T f(t, X_t) dt$.

Corollary 3.2.3. *For $T > 0$ fixed, $f \in C_c^\infty$*

$$\int_0^T f(t, X_t) dt - \mathbb{E} \int_0^T f(t, X_t) dt = \int_0^T \nabla \int_t^T P_{t,s} f(X_t) ds \cdot \sigma(t, X_t) dB_t.$$

By applying the Itô formula to $P_{t,T} f(X_t)$ we obtain for $T > 0$ fixed

$$dP_{t,T} f(X_t) = \nabla P_{t,T} f(X_t) \cdot \sigma(t, X_t) dB_t \tag{3.2.1}$$

and by integrating from 0 to T

$$f(T, X_T) = \mathbb{E}[f(T, X_T)] + \int_0^T \nabla P_{t,T} f(X_t) \cdot \sigma(t, X_t) dB_t.$$

This was observed at least as far back as [EK89] and is commonly used in the derivation of probabilistic formulas for $\nabla P_{s,t}$.

Combining the formula (3.2.1) with Theorem 3.2.1 we obtain the following expression for $S_t - \mathbb{E} S_t$ in terms of $\nabla P_{s,t} f$.

Corollary 3.2.4. *For $f \in C_c^\infty$, $T > 0$ fixed and any $t < T$*

$$\int_0^t f(s, X_s) - \mathbb{E} f(s, X_s) ds = M_t^{T,f} - Z_t^{T,f}$$

with

$$Z_t^{T,f} = \int_t^T \int_0^t \nabla P_{r,s} f(r, X_r) \cdot \sigma(r, X_r) \, dB_r \, ds$$

$$M_t^{T,f} = \int_0^t \int_r^T \nabla P_{r,s} f(r, X_r) \, ds \cdot \sigma(r, X_r) \, dB_r.$$

Proof. Let $f_0(t,x) = f(t,x) - \mathbb{E} f(t, X_t) = f(t,x) - P_{0,t}(x_0)$. We have

$$\begin{aligned} R_t^T f_0(X_t) &= \int_t^T P_{t,s} f_0(X_t) ds \\ &= \int_t^T P_{t,s} f(X_t) - P_{0,s} f(X_0) ds \\ &= \int_t^T \int_0^t \nabla P_{r,s} f(r, X_r) \cdot \sigma(r, X_r) dB_r \, ds \end{aligned}$$

where the last equality follows by integrating (3.2.1) from 0 to t (with $T = s$). Since $R_0^T f_0 = 0$ and $\nabla P_{t,s} f_0 = \nabla P_{t,s} f$ we get from Theorem 3.2.1 that

$$\int_0^t f_0(s, X_s) ds = M_t^{T,f} - R_t^T f_0(X_t)$$

and the result follows with $Z_t^{T,f} = R_t^T f_0(X_t)$. □

Remark 3.2.5 (Carré du Champs and Mixing). For differentiable functions f, g let

$$\Gamma_t(f,g)(x) = \tfrac{1}{2} \nabla f(t,x)(\sigma\sigma^\top)(t,x) \nabla g(t,x).$$

Then we have the following expression for the quadratic variation of $M^{T,f}$:

$$\begin{aligned} d\langle M^{T,f} \rangle_t &= \left| \int_t^T \sigma(t, X_t)^\top \nabla P_{t,s} f(X_t) \, ds \right|^2 dt \\ &= \left(4 \int_{t \le s \le r \le T} \Gamma_t(P_{t,s} f, P_{t,r} f)(X_t) \, dr \, ds \right) dt. \end{aligned}$$

Furthermore, since

$$\partial_s P_{r,s}(P_{s,t} f P_{s,t} g) = 2 P_{r,s}(\Gamma_s(P_{s,t} f, P_{s,t} g))$$

and setting $g(t,x) = \int_t^T P_{t,s}f(x)ds$ we have

$$\begin{aligned}
E\langle M^{T,f}\rangle_t &= 2\int_0^T \int_t^T 2P_{0,t}\Gamma_t(P_{t,s}f, P_{t,s}g)ds\,dt \\
&= 2\int_0^T \int_t^T \partial_t P_{0,t}(P_{t,s}fP_{t,s}g)ds\,dt \\
&= 2\int_0^T \partial_t \int_t^T P_{0,t}(P_{t,s}fP_{t,s}g)ds\,dt + 2\int_0^T P_{0,t}(fg)dt \\
&= 2\int_0^T P_{0,t}(fg) - P_{0,t}fP_{0,t}g\,dt \\
&= 2\int_{0\le t\le s\le T} \mathrm{Cov}(f(t,X_t), f(s,X_s))ds\,dt.
\end{aligned}$$

This shows how the expressions we obtain in terms of the gradient of the semigroup relate to mixing properties of X.

Remark 3.2.6 (Pathwise estimates). We would like to have a similar estimate for

$$\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t f(X_s) - \mathbb{E}f(X_s)ds\right|.$$

Setting

$$f_0(t,x) = f(x) - \mathbb{E}f(X_t) = f(x) - P_{0,t}f(x_0)$$

we have

$$\begin{aligned}
\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t f(X_s) - \mathbb{E}f(X_s)ds\right| &\le \mathbb{E}\sup_{0\le t\le T}|M_t^{T,f_0}| + \mathbb{E}\sup_{0\le t\le T}|R_t^T f_0(X_t)| \\
&\le 2\left(\mathbb{E}\langle M^{T,f_0}\rangle_T\right)^{1/2} + \mathbb{E}\sup_{0\le t\le T}|R_t^T f_0(X_t)|
\end{aligned}$$

and

$$\begin{aligned}
R_t^T f_0(X_t) &= \int_t^T P_{t,s}f(X_t) - P_{0,s}f(x_0)ds \\
&= \int_t^T \int_0^t \nabla P_{r,s}f(X_r)\cdot\sigma(r,X_r)dB_r ds
\end{aligned}$$

where the last equality follows from (for s fixed)

$$dP_{t,s}f(X_t) = \nabla P_{t,s}f(X_t)\cdot\sigma(t,X_t)dB_t.$$

3.2.1 Discrete time

Consider a discrete-time Markov process $(X_n)_{n=1\dots N}$ with transition operator

$$P_{m,n}f(x) = \mathbb{E}[f_n(X_n)|X_m = x]$$

and generator

$$L_n f(x) = P_{n,n+1}f(x) - f_n(x).$$

As in the continuous-time setting

$$M_n := f_n(X_n) - f_0(X_0) - \sum_{m=0}^{n-1} L_m f(X_m)$$

is a martingale (by the definition of L) and by direct calculation

$$M_n - M_{n-1} = f_n(X_n) - P_{n-1,n}f(X_{n-1}).$$

Let

$$R_n^N f(x) = \sum_{m=n}^{N-1} P_{n,m}f(x)$$

and observe that

$$L_n R^N f(x) = \sum_{m=n+1}^{N} P_{n,n+1}P_{n+1,m}f(x) - \sum_{m=n}^{N-1} P_{n,m}f(x) = -f_n(x).$$

Note that

$$R_N^N f(x) = 0 \text{ and } R_0^N f(x) = \mathbb{E}\left[\sum_{m=n}^{N-1} f(X_m)\middle| X_0 = x\right].$$

It follows that

$$\sum_{m=0}^{n-1} f_m(X_m) + R_n^N f(X_n) = -\sum_{m=0}^{n-1} L_m R^N f(X_m) + R_n^N f(X_n) = R_0^N f(X_0) + M_n^{N,f}$$

with

$$M_n^{N,f} - M_{n-1}^{N,f} = \sum_{m=n}^{N-1} P_{n,m}f(X_n) - P_{n-1,m}f(X_{n-1}).$$

Analogous to the continuous-time case, we define the carré du champs

$$\begin{aligned}\Gamma_n(f,g) &:= L_n(fg) - g_n L_n f - f_n L_n g \\ &= P_{n,n+1}(fg) - f_n P_{n,n+1}g - g_n P_{n,n+1}f + f_n g_n \\ &= \mathbb{E}\left[(f_{n+1}(X_{n+1}) - f_n(X_n))(g_{n+1}(X_{n+1}) - g_n(X_n))|\mathcal{F}_n\right]\end{aligned}$$

and using the summation by parts formula

$$
\begin{aligned}
&\langle M^{N,f}\rangle_n - \langle M^{N,f}\rangle_{n-1} = \mathbb{E}[(M_n^{N,f} - M_{n-1}^{N,f})^2|\mathcal{F}_{n-1}] \\
&= 2 \sum_{n\le k\le m<N-1} \mathbb{E}\left[(P_{n,m}f(X_n) - P_{n-1,m}f(X_{n-1}))(P_{n,k}f(X_n) - P_{n-1,k}f(X_{n-1}))|\mathcal{F}_{n-1}\right] \\
&\quad + \sum_{m=n}^{N-1} \mathbb{E}\left[(P_{n,m}f(X_n) - P_{n-1,m}f(X_{n-1}))^2|\mathcal{F}_{n-1}\right] \\
&= 2\sum_{m=n}^{N-1}\sum_{k=m}^{N-1} \Gamma_{n-1}(P_{n-1,m}f, P_{n-1,k}f)(X_{n-1}) + \sum_{m=n}^{N-1} \Gamma_{n-1}(P_{n-1,m}f)(X_{n-1}).
\end{aligned}
$$

3.3 Concentration inequalities from exponential gradient bounds

In this section we focus on the case where we have uniform exponential decay of $\nabla P_{s,t}$ so that

$$|\sigma(s,x)^\top \nabla P_{s,t} f(x)| \le C_s e^{-\lambda_s (t-s)} \quad (0 \le s \le t \le T) \tag{3.3.1}$$

for all $x \in \mathbb{R}^n$ and some class of functions f.

We first show that exponential gradient decay implies a concentration inequality.

Proposition 3.3.1. *For $T > 0$ fixed and all functions f such that (3.3.1) holds we have*

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(t,X_t) - \mathbb{E}f(t,X_t)dt > R\right) \le e^{-\frac{R^2 T}{V_T}}, \quad V_T = \frac{1}{T}\int_0^T \left(\frac{C_t}{\lambda_t}\left(1 - e^{-\lambda_t(T-t)}\right)\right)^2 dt$$

Proof. By (3.3.1)

$$
\begin{aligned}
d\langle M^{T,f}\rangle_t &= \left|\int_t^T \sigma(t,X_t)^\top \nabla P_{t,s} f(X_t) ds\right|^2 dt \\
&\le \left(\int_t^T C_t e^{-\lambda_t (s-t)} ds\right)^2 dt = \left(\frac{C_t}{\lambda_t}\left(1 - e^{-\lambda_t(T-t)}\right)\right)^2 dt
\end{aligned}
$$

so that $\langle M^{T,f}\rangle_T \le V_T T$.

By Corollary 3.2.3 and since Novikov's condition holds trivially due to $\langle M^{T,f}\rangle$ being bounded by a deterministic function we get

$$
\begin{aligned}
&\mathbb{E}\exp\left(a\int_0^T f(t,X_t) - \mathbb{E}f(t,X_t)dt\right) = \mathbb{E}\exp\left(aM_T^{T,f}\right) \\
&\qquad \le \mathbb{E}\left[\exp\left(aM_T^{T,f} - \frac{a^2}{2}\langle M^{T,f}\rangle_T\right)\right]\exp\left(\frac{a^2}{2}\langle M^{T,f}\rangle_T\right) \le \exp\left(\frac{a^2}{2}V_T T\right).
\end{aligned}
$$

By Chebyshev's inequality

$$\mathbb{P}\left(\frac{1}{T}\int_0^T f(t,X_t)-\mathbb{E}f(t,X_t)dt>R\right)\leq \exp(-aRT)\exp\left(\frac{a^2}{2}V_T T\right)$$

and the result follows by optimising over a. □

The corresponding lower bound is obtained by replacing f by $-f$.

For the rest of this section, suppose that $\sigma = \mathrm{Id}$ and that we are in the time-homogeneous case so that $P_{s,t} = P_{t-s}$. An important case where bounds of the form (3.3.1) hold is when there is exponential contractivity in the L^1 Kantorovich (Wasserstein) distance W_1. If for any two probability measures μ, ν on $\mathbb{R}^n$

$$W_1(\mu P_t, \nu P_t) \leq C e^{-\lambda t} W_1(\mu,\nu). \tag{3.3.2}$$

then (3.3.1) holds for all Lipschitz functions f with $C_s = C$, $\lambda_s = \lambda$.

Here the distance W_1 between two probability measures μ and ν on $\mathbb{R}^n$ is defined by

$$W_1(\mu,\nu) = \inf_\pi \int |x-y|\pi(dx\,dy)$$

where the infimum runs over all couplings π of μ. We also have the Kantorovich-Rubinstein duality

$$W_1(\mu,\nu) = \sup_{\|f\|_{\mathrm{Lip}}\leq 1} \int f d\mu - \int f d\nu \tag{3.3.3}$$

and we use the notation

$$\|f\|_{\mathrm{Lip}} = \sup_{x\neq y} \frac{f(x)-f(y)}{|x-y|}.$$

We can see that (3.3.2) implies (3.3.1) from

$$\begin{aligned}|\nabla P_t f|(x) = \lim_{y\to x}\frac{|P_t f(y) - P_t f(x)|}{|y-x|} &\leq \lim_{y\to x}\frac{W_1(\delta_y P_t, \delta_x P_t)}{|y-x|}\\ &\leq \|f\|_{\mathrm{Lip}} C e^{-\lambda t}\lim_{y\to x}\frac{W_1(\delta_y,\delta_x)}{|y-x|} = \|f\|_{\mathrm{Lip}} C e^{-\lambda t}\end{aligned}$$

where the first inequality is due to the Kantorovich-Rubinstein duality (3.3.3) and the second is (3.3.1).

Bounds of the form (3.3.2) have been obtained using coupling methods in [Ebe16; EGZ16; Wan16] under the condition that there exist positive constants κ, R_0 such that

$$(x-y)\cdot(b(x)-b(y)) \leq -\kappa|x-y|^2 \text{ when } |x-y| > R_0.$$

Similar techniques lead to the corresponding results for kinetic Langevin diffusions [EGZ17].

Using a different approach, in [CO16] the authors directly show uniform exponential contractivity of the semigroup gradient for bounded continuous functions, focusing on situations beyond hypoellipticity.

Besides gradient bounds, exponential contractivity in W_1 also implies the existence of a stationary measure μ_∞ [Ebe16]. Proposition 3.3.1 now leads to a simple proof of a deviation inequality that was obtained in a similar setting in [Jou09] via a tensorization argument.

Proposition 3.3.2. *If* (3.3.2) *holds then for all Lipschitz functions f and all initial measures μ_0*

$$\mathbb{P}_{\mu_0}\left(\frac{1}{T}\int_0^T f(X_t)dt - \int f d\mu_\infty > R\right) \leq \exp\left(-\left(\frac{\lambda\sqrt{T}\,R}{C\|f\|_{\mathrm{Lip}}(1-e^{-\lambda T})} - \frac{W_1(\mu_0,\mu_\infty)}{\sqrt{T}}\right)^2\right)$$

Proof. We start by applying Proposition 3.3.1 so that

$$\begin{aligned}
&\mathbb{P}_{\mu_0}\left(\frac{1}{T}\int_0^T f(X_t)dt - \int f d\mu_\infty > R\right) \\
&\quad = \mathbb{P}_{\mu_0}\left(\frac{1}{T}\int_0^T f(X_t) - \mathbb{E}f(X_t)dt > R + \frac{1}{T}\int_0^T \mu_\infty(f) - \mu_0 P_t(f)dt\right) \\
&\quad \leq \exp\left(-\left(R - \left|\frac{1}{T}\int_0^T \mu_\infty(f) - \mu_0 P_t(f)dt\right|\right)^2 \frac{T}{V_T}\right), \quad V_T = \left(\frac{\|f\|_{\mathrm{Lip}} C(1-e^{-\lambda T})}{\lambda}\right)^2.
\end{aligned}$$

By the Kantorovich-Rubinstein duality

$$\begin{aligned}
\left|\frac{1}{T}\int_0^T \mu_\infty(f) - \mu_0 P_t(f)dt\right| &\leq \left|\frac{1}{T}\int_0^T \|\nabla f\|_\infty W_1(\mu_\infty P_t, \mu_0 P_t)dt\right| \\
&\leq \frac{\|\nabla f\|_\infty C}{\lambda}\frac{(1-e^{-\lambda T})}{T} W_1(\mu,\mu_0) = \frac{\sqrt{V_T}}{T} W_1(\mu,\mu_0).
\end{aligned}$$

from which the result follows immediately. □

3.4 Averaging: Two-timescale Ornstein-Uhlenbeck

Consider the following linear multiscale SDE on $\mathbb{R}\times\mathbb{R}$ where the first component is accelerated by a factor $\alpha \geq 0$:

$$\begin{aligned}
dX_t &= -\alpha(X_t - Y_t)dt + \sqrt{\alpha}dB_t^X, \quad X_0 = x_0 \\
dY_t &= -(Y_t - X_t)dt + dB_t^Y, \quad Y_0 = y_0
\end{aligned}$$

with B^X, B^Y independent Brownian motions on $\mathbb{R}$. Denote P_t and L the associated semigroup and infinitesimal generator respectively.

Let $f(x,y) = x - y$ and note that $Lf = -(\alpha+1)f$. We have by the regularity of P_t and the Kolmogorov forward equation

$$\partial_t \partial_x P_t f = \partial_x P_t L f = -(\alpha+1)\partial_x P_t f$$

so that

$$\partial_x P_t f = \partial_x f e^{-(\alpha+1)t} = e^{-(\alpha+1)t}.$$

Repeating the same reasoning for $\partial_y P_t$ and P_t gives

$$\partial_y P_t f = -e^{-(\alpha+1)t} \text{ and } P_t f(x,y) = (x-y)e^{-(\alpha+1)t}.$$

From Corollary 3.2.3

$$\int_0^T X_t - Y_t \, dt = R_0^T f(x_0, y_0) + M_T^{T,f}$$

with

$$\begin{aligned}
R_t^T f(x,y) &= \int_t^T P_{s-t} f(x,y) ds = (x-y)\frac{1-e^{-(\alpha+1)(T-t)}}{\alpha+1}, \\
M_T^{T,f} &= \int_0^T \int_t^T \partial_x P_{s-t} f(X_t, Y_t) ds \sqrt{\alpha} dB_t^X + \int_0^T \int_t^T \partial_y P_{s-t} f(X_t, Y_t) ds \, dB_t^Y \\
&= \int_0^T \frac{1-e^{-(\alpha+1)(T-t)}}{\alpha+1} (\sqrt{\alpha} dB_t^X - dB_t^Y).
\end{aligned}$$

This shows that for each T fixed

$$Y_T - (B_T^Y + y_0) = \int_0^T X_t - Y_t dt$$

is a Gaussian random variable with mean

$$R_0^T = (x_0 - y_0)\frac{1-e^{-(\alpha+1)T}}{\alpha+1}$$

and variance

$$\langle M^{T,f} \rangle_T = \frac{1}{(\alpha+1)} \int_0^T \left(1 - e^{-(\alpha+1)(T-t)}\right)^2 dt.$$

3.5 Averaging: Exact gradients in the linear case

Consider

$$\begin{aligned}
dX_t &= -\alpha(X_t - Y_t)dt + \sqrt{\alpha} dB_t^X, \quad X_0 = x_0 \\
dY_t &= -(Y_t - X_t)dt - \beta Y_t + dB_t^Y, \quad Y_0 = y_0
\end{aligned}$$

Denote $Z_t((x,y)) = (X_t(x), Y_t(x))$ the solution for $X_0 = x, Y_0 = y$ and let $V_t(z,v) = Z_t(z+v) - Z_t(z)$. Then

$$dV_t = -AV_t\,dt \text{ with } A = \begin{pmatrix} \alpha & -\alpha \\ -1 & (1+\beta) \end{pmatrix}.$$

The solution to the linear ODE for V_t is

$$V_t(z,v) = e^{-At}v$$

Since V_t does not depend on z we drop it from the notation. Now for any continuously differentiable function f on $\mathbb{R}^2$ and $v \in \mathbb{R}^2, z \in \mathbb{R}^2$ we obtain the following expression for the gradient of $P_t f(z)$ in the direction v:

$$\begin{aligned} \nabla_v P_t f(z) &= \lim_{\varepsilon\to 0} \frac{P_t f(z+\varepsilon v) - P_t f(z)}{\varepsilon} = \lim_{\varepsilon\to 0} \frac{\mathbb{E} f(Z_t(z+\varepsilon v)) - f(Z_t(z))}{\varepsilon} \\ &= \lim_{\varepsilon\to 0} \frac{\mathbb{E}\nabla f(Z_t(z)) \cdot V_t(\varepsilon v) + o(|V_t(\varepsilon v)|)}{\varepsilon} \\ &= \mathbb{E}\nabla f(Z_t(z)) \cdot e^{-At} v. \end{aligned}$$

Since $\nabla_v P_t f = \nabla P_t f \cdot v$ we can identify $\nabla P_t f(z) = \mathbb{E}^z (e^{-At})^\top \nabla f(Z_t)$.

The eigenvalues of A are $(\lambda_0, \alpha\lambda_1)$ with

$$\begin{aligned} \lambda_0 &= \frac{1}{2}\left(\alpha+\beta+1 - \sqrt{(\alpha+\beta+1)^2 - 4\alpha\beta}\right), \\ \lambda_1 &= \frac{1}{2\alpha}\left(\alpha+\beta+1 + \sqrt{(\alpha+\beta+1)^2 - 4\alpha\beta}\right). \end{aligned}$$

By observing that

$$(\alpha+\beta+1)^2 - 4\alpha\beta = (\alpha - (1+\beta))^2 + 4\alpha = (\beta - (\alpha+1))^2 + 4\beta$$

we see that asymptotically as $\alpha \to \infty$

$$\begin{aligned} \lambda_0 &= \beta + O\left(\frac{1}{\alpha}\right) \\ \lambda_1 &= 1 + \frac{1}{\alpha} + O\left(\frac{1}{\alpha^2}\right). \end{aligned}$$

We can compute the following explicit expression for e^{-At}

$$\begin{aligned} e^{-At} &= c_0(t)\,\mathrm{Id} - \frac{c_1(t)}{\alpha} A \\ &= \begin{pmatrix} \frac{c_2(t)}{\alpha} & c_1(t) \\ \frac{c_1(t)}{\alpha} & c_0(t) - \frac{1+\beta}{\alpha} c_1(t) \end{pmatrix} \end{aligned}$$

with

$$c_0(t) = \frac{\alpha\lambda_1 e^{-\lambda_0 t} - \lambda_0 e^{-\alpha\lambda_1 t}}{\alpha\lambda_1 - \lambda_0} = \frac{(1+\alpha)e^{-\lambda_0 t} - \beta e^{-\alpha\lambda_1 t}}{\alpha\lambda_1 - \lambda_0} + O\left(\frac{1}{\alpha^2}\right),$$
$$c_1(t) = \frac{\alpha}{\alpha\lambda_1 - \lambda_0}\left(e^{-\lambda_0 t} - e^{-\alpha\lambda_1 t}\right),$$
$$c_2(t) = \alpha(c_0(t) - c_1(t)) = \frac{\alpha}{\alpha\lambda_1 - \lambda_0}\left(e^{-\lambda_0 t} - (\beta-\alpha)e^{-\alpha\lambda_1 t}\right) + O\left(\frac{1}{\alpha}\right).$$

Note that $\lambda_0, \lambda_1, c_0, c_1$ and c_2 are all of order $O(1)$ as $\alpha \to \infty$.

We obtain

$$\begin{aligned}\sigma^\top \nabla P_t f(z) &= \mathbb{E}\left[\begin{pmatrix} \frac{c_2(t)}{\sqrt{\alpha}} & \frac{c_1(t)}{\sqrt{\alpha}} \\ c_1(t) & c_0(t) - \frac{1+\beta}{\alpha}c_1(t)\end{pmatrix}\nabla f(Z_t)\right] \\ &= \frac{\alpha}{1+\alpha}\left(G_0 e^{-\lambda_0 t} + G_1 \alpha e^{-\alpha\lambda_1 t}\right) P_t \nabla f(z)\end{aligned}$$

with

$$G_0 = \begin{pmatrix} \frac{1}{\sqrt{\alpha}} & \frac{1}{\sqrt{\alpha}} \\ 1 & 1 \end{pmatrix} + O\left(\frac{1}{\alpha}\right)$$
$$G_1 = \begin{pmatrix} \frac{1}{\sqrt{\alpha}} - \frac{\lambda_0}{\alpha\sqrt{\alpha}} & -\frac{1}{\alpha\sqrt{\alpha}} \\ -\frac{1}{\alpha} & -\frac{1+\lambda_0+\beta}{\alpha^2}\end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\alpha}} & 0 \\ 0 & 0 \end{pmatrix} + O\left(\frac{1}{\alpha}\right)$$

The expression for G_0 shows that $|\sigma^\top \nabla P_t f(z)|$ can be of order $1/\sqrt{\alpha}$ only for functions $f_\alpha(z)$ such that $\mathbb{E}^z[\partial_x f_\alpha(Z_t) + \partial_y f_\alpha(Z_t)] = O(1/\sqrt{\alpha})$.

Furthermore, for any function $f \in C_c^\infty$ we have

$$\mathrm{Cov}\left(f(Z_t), B_t^X\right) = O\left(\frac{1}{\sqrt{\alpha}}\right)$$

and

$$\mathrm{Cov}\left(\int_0^t f(s, Z_s)ds, B_t^X\right) = O\left(\frac{1}{\sqrt{\alpha}}\right).$$

Since by Itô's formula $dP_{s,t}f(Z_s) = \nabla P_{s,t}f(Z_s)\cdot \sigma dB_s$ we have

$$\begin{aligned} f(Z_t) - \mathbb{E}f(Z_t) &= \int_0^t \nabla P_{s,t} f(Z_s)\cdot \sigma dB_s \\ &= \int_0^t \nabla_x P_{s,t} f(Z_s)\sqrt{\alpha}dB_s^X + \int_0^t \nabla_y P_{s,t} f(Z_s) dB_s^Y\end{aligned}$$

we have

$$
\begin{aligned}
\operatorname{Cov}\left(f(Z_t), B_t^X\right) &= \mathbb{E}\left[(f(Z_t) - \mathbb{E}f(Z_t))B_t^X\right] \\
&= \mathbb{E}\left[\int_0^t \nabla_x P_{s,t} f(Z_s)\sqrt{\alpha} ds\right] \\
&= \left(\frac{1}{\sqrt{\alpha}} + O\left(\frac{1}{\alpha}\right)\right) \frac{\alpha}{1+\alpha} \int_0^t e^{-\lambda_0 s} P_{s,t}(\nabla_x f + \nabla_y f)(Z_s) ds.
\end{aligned}
$$

The result for $\int_0^t f(s, Z_s)ds$ follows by the same arguments from the martingale representation for $\int_0^t f(s, Z_s)ds - \mathbb{E}\int_0^t f(s, Z_s)ds$.

3.6 Averaging: Conditioning on the slow component

Consider the following linear multiscale SDE on $\mathbb{R} \times \mathbb{R}$ accelerated by a factor α:

$$
\begin{aligned}
dX_t &= -\alpha\kappa_X(X_t - Y_t)dt + \sqrt{\alpha}\sigma_X dB_t^X, \quad X_0 = 0 \\
dY_t &= -\kappa_Y(Y_t - X_t)dt + \sigma_Y dB_t^Y, \quad Y_0 = 0
\end{aligned}
$$

where B^X, B^Y are independent Brownian motions and $\alpha, \kappa_X, \kappa_Y, \sigma_X, \sigma_Y$ are strictly positive constants and we are interested in the solution on a fixed inverval $[0, T]$.

We define the corresponding averaged process to be the solution to

$$
d\bar{X}_t = -\alpha\kappa_X(\bar{X}_t - \bar{Y}_t)dt + \sqrt{\alpha}\sigma_X dB_t^X, \quad \bar{X}_0 = 0 \tag{3.6.1a}
$$

$$
d\bar{Y}_t = \mathbb{E}\left[-\kappa_Y(\bar{Y}_t - \bar{X}_t)\middle|\mathcal{F}_t^{\bar{Y}}\right] dt + \sigma_Y dB_t^Y, \quad \bar{Y}_0 = 0 \tag{3.6.1b}
$$

where $\mathcal{F}_t^{\bar{Y}}$ is the σ-algebra generated by $(\bar{Y}_s)_{s \le t}$.

The conditional measure $\mathbb{P}(\cdot|\mathcal{F}_T^{\bar{Y}})$ has a regular conditional probability density $u \mapsto \mathbb{P}(\cdot|\bar{Y} = u)$, $u \in C([0, T], \mathbb{R})$. Now observe that B^X remains unchanged under $\mathbb{P}(\cdot|\bar{Y} = u)$ since $\bar{Y}$ and B^X are independent. This means that for all $u \in C([0, T], \mathbb{R})$ and $f \in C_c^\infty(\mathbb{R})$, $\mathbb{P}(\cdot|\bar{Y} = u)$ solves the same martingale problem as the measure associated to

$$
dX_t^u = -\alpha\kappa_X(X_t^u - u(t))dt + \sqrt{\alpha}\sigma_X dB_t^X, \quad X_0^u = 0. \tag{3.6.2}
$$

It follows that the conditional expectation given $\mathcal{F}_T^{\bar{Y}}$ of any functional involving $\bar{X}$ equals the usual expectation of the same functional with $\bar{X}$ replaced by X^u evaluated at $u = Y$.

For example, since

$$
\mathbb{E}X_t^u = \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X(t-s)}\, u(s)\, ds
$$

the drift coefficient of $\bar{Y}$ is

$$\begin{aligned}\mathbb{E}\left[-\kappa_Y(\bar{Y}_t - \bar{X}_t)\Big|\mathcal{F}_t^{\bar{Y}}\right] &= -\kappa_Y(\bar{Y}_t - \mathbb{E}[\bar{X}_t|\mathcal{F}_T^{\bar{Y}}]) = -\kappa_Y(\bar{Y}_t - \mathbb{E}X_t^u|_{u=\bar{Y}}) \\ &= -\kappa_Y\left(\bar{Y}_t - \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X(t-s)}\,\bar{Y}_s\,ds\right)\end{aligned}$$

so that $\bar{Y}$ solves the SDE

$$dZ_t = -\alpha\kappa_X(Z_t - \bar{Y}_t)dt \tag{3.6.3a}$$

$$d\bar{Y}_t = -\kappa_Y(\bar{Y}_t - Z_t)dt + \sigma_Y dB_t^Y. \tag{3.6.3b}$$

The key step in our estimate for $Y_t - \bar{Y}_t$ is the application of the results from the first section to

$$\int_0^T h(t)(X_t^u - \mathbb{E}X_t^u)dt$$

for a certain function $h(t)$.

We begin with a gradient estimate for the evolution operator $P_{s,t}^u$ associated to X^u.

Lemma 3.6.1. *Let* $\mathrm{id}(x) = x$ *be the identity function and* $h(t) \in C([0,T],\mathbb{R})$*. We have for all* $x \in \mathbb{R}$

$$\partial_x P_{s,t}^u(h\,\mathrm{id})(x) = h(t)e^{-\alpha\kappa_X(t-s)}.$$

Proof. Denote $X_t^{s,x}$ the solution to (3.6.2) with $X_s^u = x$. Then

$$d(X_t^{s,x+\varepsilon} - X_t^{s,x}) = -\alpha\kappa_X(X_t^{s,x+\varepsilon} - X_t^{s,x})dt$$

so that

$$X_t^{s,x+\varepsilon} - X_t^{s,x} = \varepsilon e^{-\kappa_X\alpha(t-s)}$$

and

$$\partial_x P_{s,t}(h\,\mathrm{id})(x) = \lim_{\varepsilon\to 0}\varepsilon^{-1}\mathbb{E}\left[h(t)X_t^{s,x+\varepsilon} - h(t)X_t^{s,x}\right] = h(t)e^{-\kappa_X\alpha(t-s)}.$$

□

Theorem 3.6.2.

$$\begin{aligned}\mathbb{E}|Y_T - \bar{Y}_T|^2 &= \frac{\alpha\kappa_Y^2\sigma_X^2}{(\alpha\kappa_X + \kappa_Y)^2}\int_0^T \left(1 - e^{-\alpha\kappa_X(T-t)}\left(2 - e^{-\kappa_Y(T-t)}\right)\right)^2 dt \\ &\le \frac{T}{\alpha}\frac{\kappa_Y^2\sigma_X^2}{\kappa_X^2}\end{aligned} \tag{3.6.4}$$

and

$$\begin{aligned}\mathbb{E}|\bar{Y}_T - \sigma_Y B_T^Y|^2 &= \frac{\kappa_Y^2\sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)^2}\int_0^T \left(1 - e^{-(\alpha\kappa_X+\kappa_Y)t}\right)^2 dt \\ &\le \frac{T}{\alpha^2}\frac{\kappa_Y^2\sigma_Y^2}{\kappa_X^2}.\end{aligned} \tag{3.6.5}$$

Proof of Theorem 3.6.2. We now proceed to show the equality (3.6.4). We decompose

$$Y_T - \bar{Y}_T = \int_0^T \kappa_Y (X_t - Y_t)dt - \int_0^T \kappa_Y (\mathbb{E}[\bar{X}_t|\bar{Y}] - \bar{Y}_t)dt$$
$$= -\kappa_Y \int_0^T (\mathbb{E}[\bar{X}_t|\bar{Y}] - \bar{X}_t)dt - \kappa_Y \int_0^T (Y_t - \bar{Y}_t) - (X_t - \bar{X}_t)dt. \qquad (3.6.6)$$

Using linearity, we can rewrite this as

$$Y_T - \bar{Y}_T = -\kappa_Y \int_0^T h(T-t)(\mathbb{E}[\bar{X}_t|\bar{Y}] - \bar{X}_t)dt$$

for some function h.

Since

$$d(X_t - \bar{X}_t) = -\alpha\kappa_X (X_t - \bar{X}_t)dt + \alpha\kappa_X (Y_t - \bar{Y}_t)dt$$

we have

$$X_t - \bar{X}_t = \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X (t-s)} (Y_t - \bar{Y}_t)ds.$$

With the notation

$$f(t) = Y_t - \bar{Y}_t, \quad g(t) = \bar{X}_t - \mathbb{E}[\bar{X}_t|\bar{Y}]$$

equation (3.6.6) reads as

$$\frac{1}{\kappa_Y} f'(t) + f(t) - \int_0^t \alpha\kappa_X e^{-\alpha\kappa_X (t-s)} f(s)ds = g(t).$$

Using capital letters for the Laplace transform, this writes as

$$\frac{s}{\kappa_Y} F(s) + F(s) - \frac{\alpha\kappa_X}{s + \alpha\kappa_X} F(s) = G(s)$$

or, after rearranging,

$$F(s) = \kappa_Y \frac{s + \alpha\kappa_X}{s(s + \alpha\kappa_X + \kappa_Y)} G(s) = \kappa_Y H(s) G(s).$$

Inverting the Laplace transform, we find that

$$h(t) = \frac{\alpha\kappa_X}{\alpha\kappa_X + \kappa_Y} + \frac{\kappa_Y}{\alpha\kappa_X + \kappa_Y} e^{-(\alpha\kappa_X + \kappa_Y)t}$$

so that

$$Y_T - \bar{Y}_T = \kappa_Y \int_0^T h(T-s) \left(\bar{X}_s - \mathbb{E}[\bar{X}_s|\bar{Y}]\right) ds.$$

By the properties of conditional expectation and Corollary 3.2.3 we have for any integrable function Φ that

$$\mathbb{E}\Phi(Y_T - \bar{Y}_T) = \mathbb{E}[\mathbb{E}\Phi(Y_T - \bar{Y}_T)|\mathcal{F}_T^{\bar{Y}}] = \mathbb{E}[\mathbb{E}(\Phi(Y_T - \bar{Y}_T)|u = \bar{Y})] = \mathbb{E}[(\mathbb{E}\Phi(M_T^u))|_{u=\bar{Y}}]$$

with

$$\begin{aligned}
M_T^u &= \kappa_Y \int_0^T \int_t^T \partial_x P_{t,s}^u(h(T-\cdot)\,\mathrm{id})(X_t)\,ds\,\sqrt{\alpha}\sigma_X dB_t \\
&= \kappa_Y \sqrt{\alpha}\sigma_X \int_0^T \int_t^T h(T-s) e^{-\alpha\kappa_X(s-t)}\,ds\,dB_t \\
&= \frac{\kappa_Y\sqrt{\alpha}\sigma_X}{\alpha\kappa_X + \kappa_Y} \int_0^T \int_t^T \alpha\kappa_X e^{-\alpha\kappa_X(s-t)} ds \\
&\quad + \int_t^T \kappa_Y e^{-\kappa_Y(T-s)} e^{-\alpha\kappa_X(T-s)} e^{-\alpha\kappa_X(s-t)}\,ds\,dB_t \\
&= \frac{\sqrt{\alpha}\kappa_Y\sigma_X}{\alpha\kappa_X + \kappa_Y} \int_0^T 1 - e^{-(\alpha\kappa_X + \kappa_Y)(T-t)}\,dB_t.
\end{aligned}$$

Since M_t^u is independent of u we can let $M_t = M_t^u$ for an arbitrary u so that

$$\mathbb{E}\Phi(Y_T - \bar{Y}_T) = \mathbb{E}\Phi(M_T).$$

Now we can compute

$$\mathbb{E}\left|Y_T - \bar{Y}_T\right|^2 = \mathbb{E}\langle M\rangle_T = \frac{\alpha\kappa_Y^2\sigma_X^2}{(\alpha\kappa_X + \kappa_Y)^2} \int_0^T \left(1 - e^{-(\alpha\kappa + \kappa_Y)(T-t)}\right)^2 dt.$$

We now turn to the computation of $\mathbb{E}|\bar{Y}_t - \sigma_Y B_t^Y|^2$.

From equation (3.6.3) we have

$$d(\bar{Y}_t - Z_t) = -(\alpha\kappa_X + \kappa_Y)(\bar{Y}_t - Z_t)dt + \sigma_Y B_t^Y$$

so that

$$\bar{Y}_t - Z_t = \sigma_Y \int_0^t e^{-(\alpha\kappa_X + \kappa_Y)(t-s)} dB_s^Y. \tag{3.6.7}$$

is an Ornstein-Uhlenbeck process. This means that

$$\mathbb{E}(\bar{Y}_t - Z_t)(\bar{Y}_s - Z_s) = \frac{\sigma_Y^2 e^{-(\alpha\kappa_X + \kappa_Y)t}}{\alpha\kappa_X + \kappa_Y}\sinh((\alpha\kappa_X + \kappa_Y)s), \quad s \le t.$$

so that

$$
\begin{aligned}
\mathbb{E}|\bar{Y}_t - \sigma_Y B_t^Y|^2 &= \kappa_Y^2 \left| \int_0^t \bar{Y}_s - Z_s ds \right|^2 \\
&= 2\kappa_Y^2 \int_0^t \int_0^s \mathbb{E}(\bar{Y}_s - Z_s)(\bar{Y}_r - Z_r) dr ds \\
&= \frac{2\kappa_Y^2 \sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)} \int_0^t e^{-(\alpha\kappa_X+\kappa_Y)s} \int_0^s \sinh((\alpha\kappa_X + \kappa_Y)r) dr ds \\
&= \frac{2\kappa_Y^2 \sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)^2} \int_0^t e^{-(\alpha\kappa_X+\kappa_Y)s} \left(\cosh((\alpha\kappa_X + \kappa_Y)s) - 1\right) ds \\
&= \frac{\kappa_Y^2 \sigma_Y^2}{(\alpha\kappa_X + \kappa_Y)^2} \left(\int_0^t 1 + e^{-2(\alpha\kappa_X+\kappa_Y)s} - 2e^{-(\alpha\kappa_X+\kappa_Y)s} ds \right)
\end{aligned}
$$

□

3.7 Approximation by Averaged Measures

In the previous section, the computation for $\mathbb{E}|\bar{Y}_t - \sigma_Y B_t^Y|^2$ relied on the fact that we had an explicit expression for $\mathbb{E}[\bar{X}_t - \bar{Y}_t|Y]$. Here we will see a method that can be used to obtain similar estimates in more general situations.

Consider a diffusion process (X_t, Y_t) on $\mathbb{R}^n \times \mathbb{R}^m$

$$
\begin{aligned}
dX_t &= b_X(X_t, Y_t)dt + \sigma_X(X_t, Y_t)dB_t^X \\
dY_t &= b_Y(Y_t)dt + \sigma_Y(Y_t)dB_t^Y
\end{aligned}
$$

where B^X and B^Y are standard independent Brownian motions. Denote L the generator of (X, Y) and $\mathcal{F}^Y$ the filtration of B^Y.

Let

$$
Q_t f = \mathbb{E}^{\mathcal{F}_t^Y} f(X_t, Y_t)
$$

so that, by the Itô formula and since Y is adapted to $\mathcal{F}^Y$ and B^X and B^Y are independent, we have

$$
\begin{aligned}
Q_t f = \mathbb{E}^{\mathcal{F}_t^Y} \Bigg[& f(X_0, Y_0) + \int_0^t Lf(X_s, Y_s) ds + \int_0^t \nabla_x f(X_s, Y_s) \cdot \sigma_X(X_s, Y_s) dB_s^X \\
& + \int_0^t \nabla_y f(X_s, Y_s) \cdot \sigma_Y(Y_s) dB_s^Y \Bigg] \\
= \mathbb{E}^{\mathcal{F}_0^Y}[f(X_0, Y_0)] & + \int_0^t \mathbb{E}^{\mathcal{F}_s^Y} Lf(X_s, Y_s) ds + \int_0^t (\mathbb{E}^{\mathcal{F}_s^Y} \nabla_y f(X_s, Y_s)) \cdot \sigma_Y(Y_s) dB_s^Y.
\end{aligned}
$$

In other words,

$$
dQ_t f = Q_t Lf dt + (Q_t \nabla_y f) \cdot \sigma_Y(Y_t) dB_t^Y.
$$

Example 3.7.1 (Averaged Ornstein-Uhlenbeck). Consider again the process $(\bar{X}, \bar{Y})$ from the previous section. In this case, $f(x, y) = x - y$ is an eigenfunction of $-L$ with eigenvalue $\alpha\kappa_X + \kappa_Y$ and we have $\partial_y f = -1$. Therefore

$$dQ_t f = -(\alpha\kappa_X + \kappa_Y)Q_t f dt - \sigma_Y dB_t^Y$$

so that we retrieve the result from (3.6.7)

$$\mathbb{E}[\bar{X}_t - \bar{Y}_t | \bar{Y}] = Q_t f = -\sigma_Y \int_0^t e^{-(\alpha\kappa_X + \kappa_Y)(t-s)} dB_s^Y.$$

Chapter 4

Towards a Quantitative Averaging Principle for Stochastic Differential Equations

This chapter is a reproduction of the preprint [Pep17], which is why the notation differs slightly from the rest of the thesis.

4.1 Introduction and notation

4.1.1 Motivation and main result

We are interested in stochastic differential equations of the form

$$\begin{aligned} dX_t &= \varepsilon^{-1} b_X(X_t, Y_t)dt + \varepsilon^{-1/2}\sigma_X(X_t, Y_t)dB_t^X, \quad X_0 = x_0, \\ dY_t &= b_Y(X_t, Y_t)dt + \sigma_Y(Y_t)dB_t^Y, \quad Y_0 = y_0 \end{aligned}$$

for some $\varepsilon > 0$ and $x_0 \in \mathbb{R}^n$, $y_0 \in \mathbb{R}^m$. The precise assumptions on the coefficients are stated in Assumption 4.1.1 and they essentially amount to b_X being one-sided Lipschitz outside a compact set, b_Y being differentiable with bouded derivative, σ_X being bounded and the process being elliptic.

It is well known (see for example [FW12]) that when all the coefficients and their first derivatives are bounded, Y (which depends on ε) can be approximated by a process $\bar{Y}$ on $\mathbb{R}^m$ in the sense that for all $T > 0$ fixed

$$\mathbb{P}\left(\sup_{0\le t\le T} |Y_t - \bar{Y}_t| > \varepsilon\right) \to 0 \text{ as } \varepsilon \to 0.$$

The process $\bar{Y}$ solves the SDE

$$d\bar{Y}_t = \bar{b}_Y(\bar{Y}_t)dt + \sigma_Y(\bar{Y}_t)dB_t^Y, \quad \bar{Y}_0 = y_0$$

with

$$\bar{b}(y) = \int_{\mathbb{R}^n} b_Y(x, y)\mu^y(dx).$$

Here $(\mu^y)_{y\in\mathbb{R}^m}$ is a family of measures on $\mathbb{R}^n$ such that for each y, μ^y is the unique stationary measure of X^y with

$$dX_t^y = b_X(X_t^y, y)dt + \sigma_X(X_t^y, y)dB_t^X.$$

The work [Liu10] replaces the boundedness assumption on b_X and σ_X by a dissipativity condition and shows the following rate of convergence of the time marginals:

$$\sup_{0\le t\le T} \mathbb{E}|Y_t - \bar{Y}_t| \le C\varepsilon^{1/2}$$

for some constant C independent of ε.

In [LLO16] the authors relax the growth conditions on the coefficients of the SDE and show that when (X_t, Y_t) is a reversible diffusion process with stationary measure $\mu = e^{-V(x,y)}dxdy$ such that for each y, a Poincaré inequality holds for $e^{-V(x,y)}dx$, then there exists a constant C independent of ε such that

$$\mathbb{E} \sup_{0\le t\le T} |Y_t - \bar{Y}_t| \le C\varepsilon^{1/2}.$$

The present work extends the approach from [LLO16] to the non-stationary case and drops the boundedness assumption on b_Y, σ_Y commonly found in the averaging literature. The general setting and notation will be outlined in Section 4.1.2. Section 4.2 presents a forward-backward martingale argument under the assumption of a Poincaré inequality for the regular conditional probability density ρ_t^y of X_t given $Y_t = y$. By dropping the stationarity assumption, we have to deal with the fact that ρ_t^y is no longer equal to μ^y defined above. This is done in Section 4.3 by developing the relative entropy between ρ_t^y and μ^y along the trajectories of Y. Dropping the boundedness assumption on b_Y forces us to consider the mutual interaction between X_t and Y_t. In Section 4.4 we address this problem when the timescales of X and Y are sufficiently separated. The main theorem is proven in Section 4.5. Section 4.6 applies the theorem to a particular class of SDEs to obtain sufficient conditions such that for any $T > 0$ and ε sufficiently small

$$\mathbb{E} \sup_{0\le t\le T} \left|Y_t - \bar{Y}_t\right| \le C\varepsilon^{1/2}$$

where C will be explicitly given in terms of the coefficients of the SDE and the Poincaré constant for ρ_t^y.

4.1.2 Setting and notation

The results in Sections 2 to 5 will be stated in the setting of an SDE on $\mathcal{X}\times\mathcal{Y}=\mathbb{R}^n\times\mathbb{R}^m$ of the form

$$\begin{aligned} dX_t &= b_X(X_t,Y_t)dt+\sigma_X(X_t,Y_t)dB_t^X,\ X_0=x \\ dY_t &= b_Y(X_t,Y_t)dt+\sigma_Y(X_t,Y_t)dB_t^Y,\ Y_0=y \end{aligned}$$

where $x\in\mathcal{X}=\mathbb{R}^n$, $y\in\mathcal{Y}=\mathbb{R}^m$, B^X,B^Y are independent standard Brownian motions on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively and $b_X=(b_X^i)_{1\le i\le n}$, $b_Y=(b_Y^i)_{1\le i\le m}$, σ_X and σ_Y are continuous mappings from $\mathcal{X}\times\mathcal{Y}$ to $\mathcal{X}$, $\mathcal{Y}$, $\mathbb{R}^{n\times n}$ and $\mathbb{R}^{m\times m}$ respectively.

The matrices $A_X=\left(a_X^{ij}(x,y)\right)_{i,j\le n}$ and $A_Y=\left(a_Y^{ij}(x,y)\right)_{i,j\le m}$ are defined by

$$A_X(x,y)=\tfrac{1}{2}\sigma_X(x,y)\sigma_X(x,y)^T,\qquad A_Y(x,y)=\tfrac{1}{2}\sigma_Y(x,y)\sigma_Y(x,y)^T$$

and the infinitesimal generator L of (X,Y) has a decomposition $L=L^X+L^Y$ such that

$$\begin{aligned} L^Xf &= \sum_{i=1}^n b_X^i\partial_{x_i}f+\sum_{i,j=1}^n a_X^{ij}\partial^2_{x_ix_j}f, \\ L^Yf &= \sum_{i=1}^m b_Y^i\partial_{x_i}f+\sum_{i,j=1}^m a_Y^{ij}\partial^2_{x_ix_j}f, \\ Lf &= (L^X+L^Y)f. \end{aligned}$$

We will also make use of the square field operators Γ and Γ^X, defined by

$$\begin{aligned} \Gamma(f,g) &= \tfrac{1}{2}(L(fg)-gLf-fLg)=\sum_{i,j=1}^n a_X^{ij}\partial_{x_i}f\partial_{x_j}g+\sum_{i,j=1}^m a_Y^{ij}\partial_{y_i}f\partial_{y_j}g, \\ \Gamma^X(f,g) &= \tfrac{1}{2}(L^X(fg)-gL^Xf-fL^Xg)=\sum_{i,j=1}^n a_X^{ij}\partial_{x_i}f\partial_{x_j}g. \end{aligned}$$

We denote $\rho_t(dx,dy)$ the marginal distribution of (X,Y) at time t, i.e. for $\varphi\in C_c^\infty$

$$\mathbb{E}[\varphi(X_t,Y_t)]=\int_{\mathcal{X}\times\mathcal{Y}}\varphi(x,y)\rho_t(dx,dy)$$

and we let $\rho_t^y(dx)$ be the regular conditional probability density of $P(X_t\in dx|Y_t=y)$.

If a measure $\mu(dx,dy)$ is absolutely continuous with respect to Lebesgue measure we will make a slight abuse of notation and denote $\mu(x,y)$ its density.

We will also make use of a family of auxiliary processes $(X^y)_{y\in\mathcal{Y}}$ defined by

$$dX_t^y=b_X(X_t,y)dt+\sigma_X(X_t,y)dB_t^X,\ X_0^y=x$$

which we assume to be uniformly ergodic and we denote μ^y the unique stationary invariant measure of X^y.

We will furthermore use another auxiliary process $\tilde{X}$ solution to

$$d\tilde{X}_t = b_X(\tilde{X}_t, Y_t)dt + \sigma_X(\tilde{X}_t, Y_t)d\tilde{B}_t^X,\ \tilde{X}_0 = x$$

where $\tilde{B}^X$ is an n-dimensional Brownian motion independent of B^X and B^Y and we denote $\tilde{\rho}_t^y$ the regular conditional probability density of $P(\tilde{X}_t \in dx | Y_t = y)$.

For the section on decoupling and the main theorem we need in addition to a separation of timescales the following regularity conditions on the coefficients of (X, Y):

Assumption 4.1.1. *Regularity of the coefficients:*

- b_X *verifies a one-sided Lipschitz condition with constant* κ_X *and perturbation* α*:*

  $$(x_1 - x_2)^T(b_X(x_1, y) - b_X(x_2, y)) \le -\kappa_X|x_1 - x_2|^2 + \alpha \text{ for all } x_1, x_2 \in \mathcal{X}, y \in \mathcal{Y}$$

- b_Y *has a bounded first derivative in* x*:*

  $$\kappa_Y{}^2 := \frac{1}{m}\sum_{i=1}^{m}\sup_{x,y}|\nabla_x b_Y^i(x, y)|^2 < \infty$$

- A_X *is nondegenerate uniformly with respect to* (x, y)*, i.e. there exist two constants* $0 < \lambda_X \le \Lambda_X < \infty$ *such that the following matrix inequalities hold (in the sense of nonnegative definiteness):*

  $$\lambda_X \operatorname{Id} \le A_X(x, y) \le \Lambda_X \operatorname{Id}$$

- σ_Y *is invertible and* A_Y *is uniformly elliptic with respect to* (x, y)*, i.e. there exists a constant* $\lambda_Y > 0$ *such that the following matrix inequality holds (in the sense of nonnegative definiteness):*

  $$\lambda_Y \operatorname{Id} \le A_Y(x, y)$$

Assumption 4.1.2. *Regularity of the time marginals:*

- *There exists* M_0 *such that for* $|x|^2 + |y|^2 > M_0$*,* $r > 0$*,* $\alpha > 0$

  $$\nabla_x \log \rho_t(x, y)^T x + \nabla_y \log \rho_t(x, y)^T y \le -r(|x|^2 + |y|^2)^{\alpha/2}.$$

- *The regular conditional probability densities* $\tilde{\rho}_t^y$ *of* $P(\tilde{X}_t \in dx | Y_t = y)$ *satisfy Poincaré inequalities with constants* $c_P(y)$ *independent of* ε*:*

  $$\int (f - \tilde{\rho}_t^y(f))^2 d\tilde{\rho}_t^y \le c_P(y) \int |\sigma_X \nabla_x f|^2 d\tilde{\rho}_t^y.$$

In order to characterise the separation of timescales, we introduce a parameter γ defined by

$$\gamma = \frac{\kappa_X{}^2 \lambda_Y}{\Lambda_X \kappa_Y{}^2}.$$

4.2 Approximation by conditional expectations

We will start with a Lemma for a form of the Lyons-Meyer-Zheng forward-backward martingale decomposition.

Lemma 4.2.1 (Forward-backward martingale decomposition). *For a diffusion process ξ_t with generator L_t and square field operator Γ_t we have for $f(s,\cdot) \in \mathcal{D}(L_s + \tilde{L}_{T-s})$ and $1 \le p \le 2$*

$$\mathbb{E} \sup_{0 \le t \le T} \left| \int_0^t -(L_s + \tilde{L}_{T-s}) f(s, \xi_s) ds \right|^p \le 3^{p-1}(2C_p + 1) \left(\mathbb{E} \int_0^T 2\Gamma_t(f)(\xi_t) dt \right)^{p/2}$$

where $\tilde{L}_s$ is the generator of the time-reversed process $\tilde{\xi}_t = \xi_{T-t}$ and C_p is the constant in the upper bound of the Burkholder-Davis-Gundy inequality for L^p.

Proof. First, suppose that $f(t,x)$ is once differentiable in t and twice differentiable in x so that we can apply the Itô formula.

We express $f(t,\xi_t) - f(0,\xi_0)$ in two different ways, using the fact that $\xi_t = \tilde{\xi}_{T-t}$:

$$f(t,\xi_t) - f(0,\xi_0) = \int_0^t (\partial_s + L_s) f(s,\xi_s) ds + M_t \tag{1}$$

$$\begin{aligned} f(0,\xi_0) - f(t,\xi_t) &= (f(0,\tilde{\xi}_T) - f(T,\tilde{\xi}_0)) - (f(t,\tilde{\xi}_{T-t}) - f(T,\tilde{\xi}_0)) \\ &= \int_{T-t}^T (-\partial_s + \tilde{L}_s) f(T-s, \tilde{\xi}_s) ds + \tilde{M}_T - \tilde{M}_{T-t} \\ &= \int_0^t (-\partial_s + \tilde{L}_{T-s}) f(s, \tilde{\xi}_{T-s}) ds + \tilde{M}_T - \tilde{M}_{T-t} \\ &= \int_0^t (-\partial_s + \tilde{L}_{T-s}) f(s, \xi_s) ds + \tilde{M}_T - \tilde{M}_{T-t} \end{aligned} \tag{2}$$

where M and $\tilde{M}$ are martingales with

$$\langle M \rangle_T = \int_0^T 2\Gamma_s(f)(s,\xi_s) ds,$$

$$\langle \tilde{M} \rangle_T = \int_0^T 2\Gamma_{T-s}(f)(T-s,\tilde{\xi}_s) ds = \int_0^T 2\Gamma_s(f)(s,\xi_s) ds = \langle M \rangle_T.$$

Summing (1) and (2), we get

$$\int_0^t -(L_s + \tilde{L}_{T-s}) f(s,\xi_s) = M_t + \tilde{M}_T - \tilde{M}_{T-t}.$$

We have by the Burkholder-Davis-Gundy L^p-inequality that

$$\mathbb{E} \sup_{0 \le t \le T} |M_t|^p \le C_p \mathbb{E}[\langle M \rangle_T^{p/2}]$$

$$\mathbb{E} \sup_{0 \le t \le T} |\tilde{M}_{T-t}|^p = \mathbb{E} \sup_{0 \le t \le T} |\tilde{M}_t|^p \le C_p \mathbb{E}[\langle \tilde{M} \rangle_T^{p/2}] = C_p \mathbb{E}[\langle M \rangle_T^{p/2}]$$

so that

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t -(L_s+\tilde{L}_{T-s}f)(s,\xi_s)\right|^p ds\right] = \mathbb{E}\sup_{0\le t\le T}\left|M_t+\tilde{M}_T-\tilde{M}_{T-t}\right|^p$$
$$\le 3^{p-1}\left(\mathbb{E}\sup_{0\le t\le T}|M_t|^p+\mathbb{E}|\tilde{M}_T|^p+\mathbb{E}\sup_{0\le t\le T}|\tilde{M}_{T-t}|^p\right)$$
$$\le 3^{p-1}(2C_p+1)(\mathbb{E}\langle M\rangle_T)^{p/2}$$
$$\le 3^{p-1}(2C_p+1)\left(\mathbb{E}\int_0^T 2\Gamma_t(f)(t,\xi_t)dt\right)^{p/2}$$

For a general $f(t,x)$, C^2 in x and locally integrable in t, we approximate first in space by stopping ξ_t and then in time by mollifying $f(\cdot,x)$.

For $R>0,\varepsilon>0$ and a function $f(t,x)$ we will use the notation

$$(f)^R(t,x) = f(t,x\frac{|x|\wedge R}{|x|}),$$
$$(f)_\varepsilon(t,x) = \int_{-\infty}^{+\infty} f(s,x)\phi_\varepsilon(t-s)ds$$

where ϕ_ε is a mollifier. In particular, $(f)^R(t,\cdot)$ is bounded and $(f)_\varepsilon(\cdot,x)$ is differentiable.

Let $K_t = L_t+\tilde{L}_{T-t}$. K_t is a second order partial differential operator and so can be written as

$$K_t f(t,x) = \sum b^i(t,x)\partial_{x_i}f(t,x)+\sum a^{ij}(t,x)\partial^2_{x_ix_j}f(t,x)$$

for some functions b^i and a^{ij}.

Define the stopping times $\tau_R = \inf\{t>0:|\xi_t|\ge R\}$. Then

$$\mathbb{E}\sup_{0\le t\le T}\left|\int_0^{t\wedge\tau_R} K_s(f)_\varepsilon(s,\xi_s)ds\right|^p = \mathbb{E}\sup_{0\le t\le\tau_R\wedge T}\left|\int_0^t (K_s(f)_\varepsilon)^R(s,\xi_s)ds\right|^p$$
$$\le \mathbb{E}\sup_{0\le t\le T}\left|\int_0^t (K_s(f)_\varepsilon)^R(s,\xi_s)ds\right|^p$$
$$\le 3^{p-1}(2C_p+1)\left(\mathbb{E}\int_0^T 2(\Gamma_t((f)_\varepsilon)^R(t,\xi_t)dt\right)^{p/2}. \tag{4.2.1}$$

By differentiating inside the integral for $(f)_\varepsilon$ we get

$$\int_0^{t\wedge\tau_R} K_s(f-(f)_\varepsilon)(s,\xi_s)ds \le \sup_{0\le t\le T,|x|\le R}|b^i(t,x)|\int_0^T\sup_{|x|\le R}|(\partial_{x_i}f-(\partial_{x_i}f)_\varepsilon)(s,x)|ds$$
$$+\sup_{0\le t\le T,|x|\le R}|a^{ij}(t,x)|\int_0^T\sup_{|x|\le R}|(\partial^2_{x_ix_j}f-(\partial^2_{x_ix_j}f)_\varepsilon)(s,x)|ds.$$

As $\varepsilon \to 0$, $(g)_\varepsilon \to g$ in $L^1([0,T], L^\infty(B_R))$ and the integrals on the right hand side go to 0. We now let first $\varepsilon \to 0$ with dominated convergence and then $R \to \infty$ with monotone convergence to get

$$\mathbb{E} \sup_{0 \le t \le T} \left| \int_0^t K_s f(s, \xi_s) ds \right|^p = \lim_{R \to \infty} \lim_{\varepsilon \to 0} \mathbb{E} \sup_{0 \le t \le T} \left| \int_0^{t \wedge \tau_R} K_s (f)_\varepsilon (s, \xi_s) ds \right|^p$$

For the right hand side of (4.2.1), note that

$$(\Gamma_t(f) - \Gamma_t((f)_\varepsilon))^R = \Gamma_t(f - (f)_\varepsilon, f + (f)_\varepsilon)^R = (a^{ij})^R (\partial_{x_i} f - (\partial_{x_i} f)_\varepsilon)^R (\partial_{x_j} f + (\partial_{x_j} f)_\varepsilon)^R$$

so that

$$\int_0^T |(\Gamma_t(f) - \Gamma_t((f)_\varepsilon))^R| \le \sup_{0 \le t \le T, |x| \le R} a^{ij}(t,x)(\partial_{x_j} f + (\partial_{x_j} f)_\varepsilon) \int_0^T \sup_{|x| \le R} |\partial_{x_i} f - (\partial_{x_i} f)_\varepsilon| dt.$$

Now the convergence follows again by first letting $\varepsilon \to 0$ with dominated convergence and then $R \to \infty$ with monotone convergence. □

Lemma 4.2.2. *Let L and $\hat{L}$ be generators of diffusion processes with common invariant measure μ and square field operators Γ and $\hat{\Gamma}$ respectively. Let f, g be a pair of functions such that*

$$Lf = \hat{L} g \text{ and } \int \hat{\Gamma}(f) d\mu \le \int \Gamma(f) d\mu.$$

Then

$$\int \Gamma(f) d\mu \le \int \hat{\Gamma}(g) d\mu.$$

Proof.

$$\begin{aligned}
\int \Gamma(f) d\mu = \int f L f d\mu = \int f \hat{L} g d\mu &= \int \hat{\Gamma}(f,g) d\mu \\
&\le \left(\int \hat{\Gamma}(f) d\mu \right)^{1/2} \left(\int \hat{\Gamma}(g) d\mu \right)^{1/2} \\
&\le \left(\int \Gamma(f) d\mu \right)^{1/2} \left(\int \hat{\Gamma}(g) d\mu \right)^{1/2}.
\end{aligned}$$

The result follows by dividing both sides by $\left(\int \Gamma(f) d\mu \right)^{1/2}$. □

Lemma 4.2.3. *Consider a generator L with invariant measure μ and associated square field operator Γ. Assume that the following Poincaré inequality holds:*

$$\int (\varphi - \mu(\varphi))^2 d\mu \le c_P \int \Gamma(\varphi) d\mu.$$

Then for any sufficiently nice f

$$\int \Gamma(f) d\mu \le c_P \int (-Lf)^2 d\mu \le {c_P}^2 \int \Gamma(-Lf) d\mu$$

Proof. Since both Γ and L are differential operators, we can assume that $\mu(f) = 0$. Now,

$$\Big(\int \Gamma(f)d\mu\Big)^2 = \Big(-\int fLfd\mu\Big)^2 \leq \int f^2 d\mu \int (-Lf)^2 d\mu \leq c_P \int \Gamma(f)d\mu \int (-Lf)^2 d\mu$$

and the first inequality follows after dividing both sides by $\int \Gamma(f)d\mu$. For the second inequality, we apply the Poincaré inequality again with $\varphi = (-Lf)$. $\square$

Proposition 4.2.4. *In the general setting of section 4.1.2 with Assumption 4.1.2 let $\nu_t^\eta(dx)$ be the regular conditional probability density of $\mathbb{P}(X_t \in dx|\phi(Y_t) = \eta)$ for a measurable function $\phi\colon \mathcal{Y} \to \mathbb{R}^l$. If ν_t^η satisfies a Poincaré inequality with constant $c_P(\eta)$ independent of t with respect to Γ^X then for any function $f_t(x, y)$ with at most polynomial growth in x and y such that $f_t(\cdot) \in C^2(\mathcal{X} \times \mathcal{Y})$, $\int_{\mathcal{X}} f_t(x, y)\nu_t^{\phi(y)}(dx) = 0$ and $1 \leq p \leq 2$*

$$\mathbb{E} \sup_{0\leq t\leq T} \left|\int_0^t f_s(X_s, Y_s)ds\right|^p \leq 3^{p-1}2^{-p/2}(2C_p + 1)\Big(\mathbb{E}\int_0^T c_P(\phi(Y_t))f_t^2(X_t, Y_t)dt\Big)^{p/2}$$

where C_p is the constant in the upper bound of the Burkholder-Davis-Gundy inequality for L^p.

Proof of Proposition 4.2.4. The generator of the time-reversed process $(X, Y)_{T-t}$ is [HP86]

$$\begin{aligned}\tilde{L}_t\varphi = &-\sum_{i=1}^n b_X^i \partial_{x_i}\varphi - \sum_{i=1}^m b_Y^i \partial_{y_i}\varphi + \sum_{i,j=1}^n a_X^{ij}\partial^2_{x_i x_j}\varphi + \sum_{i,j=1}^m a_Y^{ij}\partial^2_{y_i y_j}\varphi \\ &+ \frac{1}{p_{T-t}}\sum_{i,j=1}^n \partial_{x_j}(2a_X^{ij}p_{T-t})\partial_{x_i}\varphi + \frac{1}{p_{T-t}}\sum_{i,j=1}^m \partial_{y_j}(2a_Y^{ij}p_{T-t})\partial_{y_i}\varphi\end{aligned}$$

so that the symmetrized generator is

$$\begin{aligned}K_t\varphi &:= \frac{(L + \tilde{L}_{T-t})\varphi}{2} \\ &= \frac{1}{p_t}\sum_{i,j=1}^n \partial_{x_j}(a_X^{ij}p_t)\partial_{x_i}\varphi + \sum_{i,j=1}^n a_X^{ij}\partial^2_{x_i x_j}\varphi + \frac{1}{p_t}\sum_{i,j=1}^m \partial_{y_j}(a_Y^{ij}p_t)\partial_{y_i}\varphi + \sum_{i,j=1}^m a_Y^{ij}\partial^2_{y_i y_j}\varphi \\ &= \sum_{i,j=1}^n \frac{1}{p_t}\partial_{x_i}(p_t a_X^{ij}\partial_{x_j}\varphi) + \sum_{i,j=1}^m \frac{1}{p_t}\partial_{y_i}(p_t a_Y^{ij}\partial_{y_j}\varphi).\end{aligned}$$

For fixed $\tau \geq 0$, we see from the expression for K that $p_\tau(dx, dy)$ is an invariant measure for K_τ (use integration by parts).

By the properties of conditional expectation $\int f_\tau dp_\tau = 0$. From Assumption 4.1.2 and Theorem 1 in [PV01] it follows that for each τ there exists a unique solution $F_\tau \in C^2(\mathcal{X} \times \mathcal{Y})$ to the Poisson Problem $K_\tau F_\tau = f_\tau$.

We can now apply the forward-backward martingale decomposition via Lemma 4.2.1 to obtain

$$\begin{aligned}\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t f_s(X_s,Y_s)ds\right|^p &= \mathbb{E}\sup_{0\le t\le T}\left|\int_0^t K_sF_s(X_s,Y_s)ds\right|^p\\ &= 2^{-p}\mathbb{E}\sup_{0\le t\le T}\left|\int_0^t (L+\tilde{L}_{T-s})F_s(X_s,Y_s)ds\right|^p\\ &\le 2^{-p}3^{p-1}(2C_p+1)\left(\mathbb{E}\int_0^T 2\Gamma(F_s)(X_s,Y_s)ds\right)^{p/2}.\end{aligned}$$

Now, we want to pass from Γ to Γ^X in order to use our Poincaré inequality for ν_t^η.

For $\varphi\in C^2(\mathcal{X})$ and $y\in\mathcal{Y},\tau\ge 0$ fixed let $\hat{K}^{\tau,y}\varphi$ be the the reversible generator associated to $\Gamma^X(\varphi)(\cdot,y)$ and $\nu_\tau^{\phi(y)}$.

Since $\nu_\tau^{\phi(y)}$ satisfies a Poincaré inequality and $\int f_\tau(x,y)\nu_\tau^{\phi(y)}(dx)=0$ by assumption,

$$\hat{K}_\tau\hat{F}^{\tau,y}(x)=f_\tau(x,y)$$

has a unique solution $\hat{F}^{\tau,y}(x)$.

If we set $\hat{K}_\tau\varphi(x,y)=(\hat{K}^{\tau,y}\varphi(\cdot,y))(x)$ and $\hat{F}_\tau(x,y)=\hat{F}^{\tau,y}(x)$ then

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{Y}}\hat{K}_\tau\varphi(x,y)p_t(dx,dy)=\int_{\mathcal{Y}}\int_{\mathcal{X}}(\hat{K}^{\tau,y}\varphi(\cdot,y))(x)\nu_t^{\phi(y)}p_t(\mathcal{X},dy)=0 \text{ and}\\ \hat{K}_\tau\hat{F}_\tau(x,y)=f_\tau(x,y)=K_\tau F_\tau(x,y).\end{aligned}$$

By Lemma 4.2.2 we get that

$$\int_{\mathcal{X}\times\mathcal{Y}}\Gamma(F_t)dp_t\le\int_{\mathcal{X}\times\mathcal{Y}}\Gamma^X(\hat{F}_t)dp_t.$$

Since $\hat{K}\hat{F}_t=f_t$ and $\hat{K}_t$ is the generator associated with Γ^X and $\nu_t^{\phi(y)}$, we can use the Poincaré inequality on $\nu_t^{\phi(y)}$ in Lemma 4.2.3 to estimate the right hand side by

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{Y}}\Gamma^X(\hat{F}_t)(x,y)p_t(dx,dy)&=\int_{\mathcal{Y}}\int_{\mathcal{X}}\Gamma^X(\hat{F}_t)(x,y)\nu_t^{\phi(y)}(dx)p_t(\mathcal{X},dy)\\ &\le\int_{\mathcal{Y}}c_P(\phi(y))\int_{\mathcal{X}}{f_t}^2(x,y)\nu_t^{\phi(y)}(dx)p_t(\mathcal{X},dy)\\ &=\int_{\mathcal{X}\times\mathcal{Y}}c_P(\phi(y)){f_t}^2(x,y)p_t(dx,dy)\end{aligned}$$

which completes the proof.

□

4.3 Distance between conditional and averaged measures

We will first show a general result on the relative entropy between $\rho_t^{Y_t}$ and μ^{Y_t} by studying the relative entropy along the trajectories of Y_t. We are still in the setting of section 4.1.2.

Proposition 4.3.1. *Let $f_t(x,y) = \frac{d\rho_t^y}{d\mu^y}(x)$. If μ^y satisfies a Logarithmic Sobolev inequality with constant c_L uniformly in y with respect to Γ^X then for $r \in \mathbb{R}$*

$$\mathbb{E}\,\mathrm{H}(\rho_t^{Y_t}|\mu^{Y_t})e^{rt} \le \mathbb{E}\,\mathrm{H}(\rho_0^{Y_0}|\mu^{Y_0}) - \left(\frac{2}{c_L} - r\right)\int_0^t \mathbb{E}\,\mathrm{H}(\rho_s^{Y_s}|\mu^{Y_s})e^{rs}ds$$
$$+ \int_0^t \mathbb{E}[L^Y \log f_s(X_s, Y_s)]e^{rs}ds.$$

Proof. We have

$$H(\rho_t^y|\mu^y) = \int_{\mathcal{X}} f_t \log f_t \mu^y(dx) = \mathbb{E}[\log f_t(X_t, Y_t)|Y_t = y]$$

so that the quantity we want to estimate is

$$\mathbb{E}H(\rho_t^{Y_t}|\mu^{Y_t}) = \mathbb{E}[\log f_t(X_t, Y_t)].$$

Now by Itô's formula

$$\begin{aligned} de^{rt}\log f_t(X_t,Y_t) &= \left((\partial_t + L)\log f_t(X_t,Y_t) + r\log f_t(X_t,Y_t)\right)e^{rt}dt + dM_t \\ &= \big((\partial_t \log\rho_t^y(x))(X_t,Y_t) + L^X\log f_t(X_t,Y_t) + L^Y\log f_t(X_t,Y_t) \\ &\quad + r\log f_t(X_t,Y_t)\big)e^{rt}dt + dM_t \end{aligned}$$

where M_t is a local martingale.

Since $\rho_t^y dx$ is a probability measure, we have

$$\begin{aligned} \mathbb{E}[\partial_t \log\rho_t^y(x)(X_t,Y_t)|Y_t = y] &= \int_{\mathcal{X}} (\partial_t \log\rho_t^y(x))\rho_t^y(x)dx \\ &= \int_{\mathcal{X}} \partial_t\rho_t^y(x)dx \\ &= \partial_t \int_{\mathcal{X}} \rho_t^y(x)dx = 0. \end{aligned}$$

By the definition of μ^y as an invariant measure for X^y we have for all φ in the domain of L^X

$$\int_{\mathcal{X}} L^X\varphi(x,y)d\mu^y = 0. \tag{4.3.1}$$

From the Logarithmic Sobolev inequality for μ^y we get

$$\mathrm{H}(\rho_t^y|\mu^y) \leq \tfrac{1}{2}c_L\,\mathrm{I}(\rho_t^y|\mu^y) = \tfrac{1}{2}c_L \int_{\mathcal{X}} \frac{\Gamma^X(f_t)(x,y)}{f_t(x)}\mu^y(x)dx.$$

Together with the formula $L^X(g \circ f) = g'(f)L^X f + g''(f)\Gamma^X(f)$ this implies

$$\begin{aligned}
\mathbb{E}[L^X \log f_t(X_t,Y_t)|Y_t = y] &= \int_{\mathcal{X}} L^X(\log f_t)(x,y)\rho_t^y(x)dx \\
&= \int_{\mathcal{X}} L^X f_t(x,y)\mu^y(x)dx - \int_{\mathcal{X}} \frac{\Gamma^X(f_t)(x,y)}{f_t(x)}\mu^y(x)dx \\
&= -\,\mathrm{I}(\rho_t^y|\mu^y) \\
&\leq -\frac{2}{c_L}\,\mathrm{H}(\rho_t^y|\mu^y).
\end{aligned}$$

By the tower property for conditional expectation and the preceding results, $\mathbb{E}[(\partial_t \log \rho_t^y(x))(X_t,Y_t)] = 0$ and $\mathbb{E}[L^X \log f_t(X_t,Y_t)] \leq -\frac{2}{c_L}\mathbb{E}\,\mathrm{H}(\rho_t^{Y_t}|\mu^{Y_t})$ so that

$$\begin{aligned}
&\mathbb{E}\,\mathrm{H}(\rho_t^{Y_t}|\mu^{Y_t})e^{rt} = E[\log f_t(X_t,Y_t)e^{rt}] \\
&\quad\leq \mathbb{E}\,\mathrm{H}(\rho_0^{Y_0}|\mu^{Y_0}) - \left(\frac{2}{c_L} - r\right)\int_0^t \mathbb{E}\,\mathrm{H}(\rho_s^{Y_s}|\mu^{Y_s})e^{rs}ds + \int_0^t \mathbb{E}[L^Y \log f_s(X_s,Y_s)]e^{rs}ds.
\end{aligned}$$

□

We now proceed to estimate the term $\mathbb{E}[L^Y \log f_s(X_t,Y_t)]$ in a restricted setting where the coefficients of L^Y are independent of x and μ^y has a density $\mu^y(x) = Z(y)^{-1}e^{-V(x,y)}$ where V has bounded first and second derivatives in y.

Lemma 4.3.2. *If the coefficients b_Y^i and a_Y^{ij} of L^Y only depend on y then for $f_t(x,y) = \frac{d\rho_t^y}{d\mu^y}(x)$*

$$\int_{\mathcal{X}} L^Y \log f_t d\rho_t^y \leq -\int_{\mathcal{X}} L^Y \log \mu^y d\rho_t^y$$

Proof. Let $g_t(x,y) = \rho_t^y(x)$. Provided that all the integrals exist, we have

$$\begin{aligned}
\int_{\mathcal{X}} L^Y(\log g_t)(x,y)\rho_t^y(dx) &= \int_{\mathcal{X}} L^Y(\log g_t(x,\cdot))(y)\rho_t^y(dx) \\
&= \int_{\mathcal{X}} L^Y(g_t(x,\cdot))(y)dx - \int_{\mathcal{X}} \frac{\Gamma^Y(g_t(x,\cdot))(y)}{g_t(x,y)}\rho_t^y(dx) \\
&\leq \int_{\mathcal{X}} L^Y(g_t(x,\cdot))(y)dx \\
&= L^Y\left(\int_{\mathcal{X}} g_t(x,\cdot)dx\right)(y) = 0
\end{aligned}$$

since $g_t(x,y)dx$ is a probability measure. Now the result follows since

$$L^Y \log f_t = L^Y \log g_t - L^Y \log \mu^y.$$

$\square$

Lemma 4.3.3. *Consider a probability measure $\mu(dx,dy)$ with density $\mu(x,y)$ on $\mathcal{X}\times\mathcal{Y}$ and let $Z(y)=\int_{\mathcal{X}}\mu(x,y)dx$, $\mu^y(dx)=\mu(dx,y)/Z(y)$. We have the identities*

$$\begin{aligned}
\partial_{y_i}\log Z(y) &= \int_{\mathcal{X}} \partial_{y_i}\log\mu(x,y)\mu^y(dx),\\
\partial^2_{y_iy_j}\log Z(y) &= \int_{\mathcal{X}} \partial^2_{y_iy_j}\log\mu(x,y)\mu^y(dx) + \mathrm{Cov}_{\mu^y}(\partial_{y_i}\log\mu, \partial_{y_j}\log\mu).
\end{aligned}$$

Proof. By differentiating under the integral

$$\partial_{y_i}\log Z(y) = \frac{\partial_{y_i}Z(y)}{Z(y)} = \int_{\mathcal{X}}\partial_{y_i}\mu(x,y)\frac{dx}{Z(y)} = \int_{\mathcal{X}}\frac{\partial_{y_i}\mu(x,y)}{\mu(x,y)}\frac{\mu(x,y)dx}{Z(y)} = \int_{\mathcal{X}}\partial_{y_i}\log\mu(x,y)\mu^y(dx)$$

and

$$\begin{aligned}
&\partial^2_{y_iy_j}\log Z(y)\\
&= \partial_{y_i}\int_{\mathcal{X}}\partial_{y_j}\log\mu(x,y)\mu^y(dx)\\
&= \int_{\mathcal{X}}\partial_{y_i}\partial_{y_j}\log\mu(x,y)\mu^y(dx) + \int_{\mathcal{X}}\partial_{y_j}\log\mu(x,y)\frac{\partial_{y_i}\mu(x,y)}{Z(y)}dx - \int_{\mathcal{X}}\partial_{y_j}\log\mu(x,y)\mu(x,y)\frac{\partial_{y_i}Z(y)}{Z(y)^2}dx\\
&= \int_{\mathcal{X}}\partial^2_{y_iy_j}\log\mu(x,y)\mu^y(dx)\\
&\quad + \int_{\mathcal{X}}\partial_{y_j}\log\mu(x,y)\partial_{y_i}\log\mu(x,y)\mu^y(dx) - \partial_{y_i}\log Z(y)\int_{\mathcal{X}}\partial_{y_j}\log\mu(x,y)\mu^y(dx)\\
&= \int_{\mathcal{X}}\partial^2_{y_iy_j}\log\mu(x,y)\mu^y(dx) + \mathrm{Cov}_{\mu^y}(\partial_{y_i}\log\mu, \partial_{y_j}\log\mu).
\end{aligned}$$

$\square$

Lemma 4.3.4. *For any Lipschitz function f*

$$\left|\int f d\mu^y - \int f d\rho_t^y\right|^2 \le \|f\|^2_{\mathrm{Lip}}\Lambda_X c_L H(\rho_t^y|\mu^y)$$

uniformly in $y\in\mathcal{Y}$.

Proof. By the Logarithmic Sobolev inequality of μ^y with respect to Γ^X and the uniform boundedness of A we have

$$\mathrm{Ent}_{\mu^y}(f^2) \le 2c_L\int\Gamma^X(f)d\mu^y = 2c_L\int(\nabla_x f)^T A(\cdot,y)(\nabla_x f)d\mu^y \le 2c_L\Lambda_X\int|\nabla_x f|^2 d\mu^y$$

which says that μ^y satisfies a Logarithmic Sobolev inequality with respect to the usual square field operator $|\nabla_x|^2$ with constant $c_L \Lambda_X$. By the Otto-Villani theorem, this implies a T_2 inequality with the same constant:

$$W_2(\rho_t^y, \mu^y)^2 \leq c_L \Lambda_X H(\rho_t^y | \mu^y).$$

By the Kantorovich duality formulation of W_1 and monotonicity of Kantorovich norms it follows from the preceding T_2 inequality that

$$\left| \sup_{\|f\|_{\text{Lip}} \leq 1} \int f d(\rho_t^y - \mu^y) \right|^2 = W_1(\rho_t^y, \mu^y)^2 \leq W_2(\rho_t^y, \mu^y)^2 \leq c_L \Lambda_X H(\rho_t^y | \mu^y)$$

from which the result follows. □

Proposition 4.3.5. *If b_Y, σ_Y depend only on y and $\mu^y(dx) = Z(y)^{-1} e^{-V(x,y)} dx$ such that $\|\partial_{y_i} V(\cdot, y)\|_{\text{Lip}} < \infty$, $\|\partial^2_{y_i y_j} V(\cdot, y)\|_{\text{Lip}} < \infty$ for all y then*

$$\begin{aligned} \mathbb{E} L^Y f_t(X_t, Y_t) \leq \frac{\Lambda_X c_L}{2} \mathbb{E} \left(\sum_{i=1}^m \|\partial_{y_i} V(\cdot, Y_t)\|^2_{\text{Lip}} + \sum_{i,j=1}^m \|\partial^2_{y_i y_j} V(\cdot, Y_t)\|^2_{\text{Lip}} \right) H(\rho_t^{Y_t} | \mu^{Y_t}) \\ + \mathbb{E} \Phi(Y_s) \end{aligned}$$

where

$$\Phi(y) = \tfrac{1}{2} \sum_{i=1}^m b_Y^i(y)^2 + \tfrac{1}{2} \sum_{i,j=1}^m a_Y^{ij}(y)^2 + \sum_{i,j=1}^m a_Y^{ij}(y) \operatorname{Cov}_{\mu^y}(\partial_{y_i} V, \partial_{y_j} V).$$

Proof. Using Lemmas Lemma 4.3.2, 4.3.3 and 4.3.4 together with the inequality $2ab \leq a^2 + b^2$ we get

$$\begin{aligned} &\int_{\mathcal{X}} L^Y \log f_t d\rho_t^y \\ &\quad = -\int_{\mathcal{X}} L^Y \log \mu^y d\rho_t^y \\ &\quad = L^Y \log Z(y) - \int_{\mathcal{X}} L^Y \log \mu d\rho_t^y \\ &\quad = b_Y^i(y) \int_{\mathcal{X}} \partial_{y_i} \log \mu \, d(\mu^y - \rho_t^y) + a_Y^{ij}(y) \int_{\mathcal{X}} \partial^2_{y_i y_j} \log \mu \, d(\mu^y - \rho_t^y) \\ &\qquad + a_Y^{ij}(y) \operatorname{Cov}_{\mu^y}(\partial_{y_i} \log \mu, \partial_{y_j} \log \mu) \\ &\quad \leq \tfrac{1}{2} b_Y^i(y)^2 + \tfrac{1}{2} \|\partial_{y_i} \log \mu\|^2_{\text{Lip}} \Lambda_X c_L H(\rho_t^y | \mu^y) + \tfrac{1}{2} a_Y^{ij}(y)^2 \\ &\qquad + \tfrac{1}{2} \|\partial^2_{y_i y_j} \log \mu\|^2_{\text{Lip}} \Lambda_X c_L H(\rho_t^y | \mu^y) + a_Y^{ij}(y) \operatorname{Cov}_{\mu^y}(\partial_{y_i} \log \mu, \partial_{y_j} \log \mu). \end{aligned}$$

The result now follows from the tower property of conditional expectation. □

4.4 Decoupling

We are still in the general setting of Section 4.1.2. We also require that $\sigma_Y(x,y) = \sigma_Y(y)$ only depends on y and that Assumption 4.1.1 is in force. The key requirement for the results in this section is a sufficient separation of timescales expressed by assumptions on γ.

The goal in this subsection is to estimate expressions of the type $\mathbb{E}F(X,Y)$ by $\mathbb{E}F(\tilde{X},Y)$ for any functional F on $\mathcal{W}_{\mathcal{X}} \times \mathcal{W}_{\mathcal{Y}}$.

Denoting $\mathbb{P}$ the Wiener measure on $C([0,T], \mathcal{X} \times \mathcal{Y})$, define a new probability measure $\mathbb{Q} = \mathcal{E}(M)\mathbb{P}$ with

$$dM_t = \Big(\sigma_Y(Y_t)^{-1}(b_Y(\tilde{X}_t, Y_t) - b_Y(X_t, Y_t))\Big)^T dB_t^Y.$$

Corollary 4.4.5 will show in particular that under our assumption on γ $\mathcal{E}(M)$ is a true martingale so that $\mathbb{Q}$ is indeed a probability measure.

Under this conditions, there is a $\mathbb{Q}$-Brownian motion $\tilde{B}^Y$ such that

$$dY_t = b_Y(\tilde{X}_t, Y_t)dt + \sigma_Y(Y_t)d\tilde{B}_t^Y$$

with

$$d\tilde{B}_t^Y = dB_t^Y - \sigma_Y(Y_t)^{-1}(b_Y(\tilde{X}_t, Y_t) - b_Y(X_t, Y_t))dt.$$

The following Proposition 4.4.2 states the key property of $\mathbb{Q}$ which we are going to use.

Lemma 4.4.1. *Under $\mathbb{Q}$, B^X, $\tilde{B}^X$ and $\tilde{B}^Y$ are independent Brownian motions.*

Proof. Girsanov's theorem states that if L is a continuous $\mathbb{P}$-local martingale, then $L - \langle L, M\rangle$ is a continuous $\mathbb{Q}$-local martingale. Thus $\tilde{B}^Y = B^Y - \langle B^Y, M\rangle$ is a continuous $\mathbb{Q}$-local martingale by definition, and $B^X, \tilde{B}^X$ are continuous $\mathbb{Q}$-local martingales since $\langle B^X, M\rangle = 0$ and $\langle \tilde{B}^X, M\rangle = 0$. Since the quadratic variation process is invariant under a change of measure we can conclude using Lévy's characterisation theorem. □

Proposition 4.4.2. *The laws of $(X, Y, \tilde{X})$ under $\mathbb{P}$ and of $(\tilde{X}, Y, X)$ under $\mathbb{Q}$ are equal.*

Proof. (X, Y) solves the martingale problem for L under $\mathbb{P}$, and $(\tilde{X}, Y)$ solves the martingale problem for L under $\mathbb{Q}$. Since b_X and b_Y are locally Lipschitz, the martingale problem has a unique solution. □

Note in particular that under $\mathbb{Q}$ B_t^X and Y are independent.

The rest of this section is dedicated to show that we can estimate expectations under $\mathbb{P}$ by expectations under $\mathbb{Q}$ when we have a sufficient separation of timescales.

Lemma 4.4.3. *For any $p > 1, q > 1$ and $\mathcal{F}_t$-measurable variable X*

$$\Big(\mathbb{E}X\Big)^p \le \Big(\mathbb{E}_{\mathbb{Q}}X^p\Big)\Big(\mathbb{E}e^{\lambda(p,q)\langle M\rangle_t}\Big)^{\frac{p-1}{q}} \quad \text{with } \lambda(p,q) = \frac{q}{2(p-1)^2}\left(p + \frac{1}{q-1}\right)$$

Proof. We have

$$\begin{aligned}\mathbb{E}X &= \mathbb{E}[X\mathcal{E}(M)^{1/p}\mathcal{E}(M)^{-1/p}] \le (\mathbb{E}X^p\mathcal{E}(M))^{1/p}(\mathbb{E}\mathcal{E}(M)^{-p'/p})^{1/p'} \\ &= (\mathbb{E}_{\mathbb{Q}}X^p)^{1/p}(\mathbb{E}\mathcal{E}(M)^{-p'/p})^{1/p'} \text{ with } \tfrac{1}{p} + \tfrac{1}{p'} = 1\end{aligned}$$

Furthermore, using that for any $\alpha \in \mathbb{R}$ we have $\mathcal{E}(M)^{-\alpha} = \mathcal{E}^{\alpha}(-M)e^{\alpha(1+\alpha)/2}$, we get

$$\begin{aligned}\mathbb{E}[\mathcal{E}(M)^{-p'/p}] &= \mathbb{E}\left[\Big(\mathcal{E}(M)^{-q'p'/p}\Big)^{1/q'}\right] = \mathbb{E}\left[\Big(\mathcal{E}^{q'p'/p}(-M)\Big)^{1/q'}\Big(e^{\frac{q'p'}{2p}(\frac{q'p'}{p}+1)\langle M\rangle}\Big)^{1/q'}\right] \\ &\le \Big(\mathbb{E}\mathcal{E}^{q'p'/p}(-M)\Big)^{1/q'}\Big(\mathbb{E}e^{\frac{qp'}{2p}(\frac{q'p'}{p}+1)\langle M\rangle}\Big)^{1/q} \text{ with } \tfrac{1}{q} + \tfrac{1}{q'} = 1 \\ &\le \Big(\mathbb{E}e^{\frac{q}{2(p-1)^2}\left(p+\frac{1}{q-1}\right)\langle M\rangle_t}\Big)^{1/q}\end{aligned}$$

The first expectation in the second line is ≤ 1 since $\mathcal{E}^{q'p'/p}(-M)$ is a positive local martingale and therefore a supermartingale. Expressing q' and p' in terms of p and q in the second expectation, we pass to the last line and conclude. ☐

Lemma 4.4.4. *Under Assumption 4.1.1 for*

$$\beta \le \frac{\gamma}{4}$$

we have

$$\mathbb{E}\exp\left(\beta\langle M\rangle_t\right) \le \exp\left(\frac{2\beta\kappa_X(\alpha + n\bar{\lambda}_X)t}{\Lambda_X\gamma}\right)$$

Proof. From the definition of M_t we have

$$\begin{aligned}d\langle M\rangle_t &= \left|\sigma_Y{}^{-1}\Big(b(X_t,Y_t) - b(\tilde{X}_t,Y_t)\Big)\right|^2 dt \le \frac{1}{\lambda_Y}\left|b(X_t,Y_t) - b(\tilde{X}_t,Y_t)\right|^2 dt \\ &\le \frac{\kappa_Y}{\lambda_Y}\left|X_t - \tilde{X}_t\right|^2 dt.\end{aligned}$$

We also have

$$\begin{aligned}d|X_t - \tilde{X}_t|^2 &= 2\Big(X_t - \tilde{X}_t\Big)^T\Big(b_X(X_t,Y_t) - b_X(\tilde{X}_t,Y_t)\Big)dt \\ &\quad + 2\Big(X_t - \tilde{X}_t\Big)^T\Big(\sigma_X(X_t,Y_t)dB_t^X - \sigma_X(\tilde{X}_t,Y_t)d\tilde{B}_t^X\Big) \\ &\quad + 2\operatorname{Tr}(A_X(X_t,Y_t))dt + 2\operatorname{Tr}(A_X(\tilde{X}_t,Y_t))dt \\ &\overset{(m)}{\le} -2\kappa_X|X_t - \tilde{X}_t|^2dt + 2(\alpha + n\bar{\lambda}_X)dt\end{aligned}$$

where $\overset{(m)}{\leq}$ means inequality modulo local martingales, and

$$\begin{aligned} d\langle |X_t - \tilde{X}_t|^2 \rangle &= 4(X_t - \tilde{X}_t)^T \left(A_X(X_t, Y_t) + A_X(\tilde{X}_t, Y_t) \right) (X_t - \tilde{X}_t) \\ &\leq 8\Lambda_X |X_t - \tilde{X}_t|^2 \end{aligned}$$

so that

$$\begin{aligned} de^{\frac{r}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} &= \left(\frac{r}{2} d|X_t - \tilde{X}_t|^2 + \beta d\langle M \rangle_t + \frac{r^2}{8} d\langle |X_t - \tilde{X}_t|^2 \rangle \right) e^{\frac{r}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} \\ &\overset{(m)}{\leq} \left(\left(r^2 \Lambda_X - r\kappa_X + \frac{\beta \kappa_Y^2}{\lambda_Y^2} \right) |X_t - \tilde{X}_t|^2 + r(\alpha + n\bar{\lambda}_X) \right) e^{\frac{r}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} dt \\ &= \left(\Lambda_X (r - r_-)(r - r_+) |X_t - \tilde{X}_t|^2 + r(\alpha + n\bar{\lambda}_X) \right) e^{\frac{r}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} dt \end{aligned}$$

with

$$r_\pm = \frac{\kappa_X}{2\Lambda_X} \left(1 \pm \sqrt{1 - 4\beta/\gamma} \right).$$

According to our assumptions, $1 - 4\beta/\gamma > 0$ and we have, choosing $r = r_-$

$$de^{\frac{r_-}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} \overset{(m)}{\leq} r_-(\alpha + n\bar{\lambda}_X) e^{\frac{r_-}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} dt$$

so that

$$e^{\frac{r_-}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} \overset{(m)}{\leq} e^{r_-(\alpha + n\bar{\lambda}_X)t}$$

and

$$\mathbb{E} e^{\beta \langle M \rangle_t} \leq \mathbb{E} e^{\frac{r_-}{2}|X_t - \tilde{X}_t|^2} e^{\beta \langle M \rangle_t} \leq e^{r_-(\alpha + n\bar{\lambda}_X)t}.$$

Since $1 - \sqrt{1 - x} \leq x$ for $0 \leq x \leq 1$ we have furthermore

$$r_- \leq \frac{\kappa_X}{2\Lambda_X} \frac{4\beta}{\gamma}$$

so that

$$\mathbb{E} e^{\beta \langle M \rangle_t} \leq \exp \left(\frac{2\beta \kappa_X (\alpha + n\bar{\lambda}_X) t}{\Lambda_X \gamma} \right).$$

□

Corollary 4.4.5. *If* $\gamma > 2$ *then*

$$\mathcal{E}(M)_t \text{ is a true martingale.}$$

Proof. Since $\frac{1}{2} < \frac{\gamma}{4}$ by our assumption we get from the previous Proposition that

$$\mathbb{E} \left[e^{\frac{1}{2} \langle M \rangle_t} \right] < \infty$$

and Novikov's criterion leads directly to the conclusion. □

Proposition 4.4.6. *Under assumption 4.1.1 for any $\mathcal{F}_t$-measurable random variable Z and* $1+\frac{2}{\gamma}+2\sqrt{\frac{2}{\gamma}} \le p \le 2$

$$(\mathbb{E}Z)^p \le \mathbb{E}_{\mathbb{Q}}\left[Z^p\right] \exp\left(\frac{p\,\kappa_X(\alpha+n\bar{\lambda}_X)\,t}{\left(p-1-\sqrt{2/\gamma}\right)\Lambda_X\,\gamma}\right)$$

Proof. We would like to apply Lemmas 4.4.3 and 4.4.4, so we need to find conditions that ensure the existence of a q such that $\lambda(p,q) \le \frac{\gamma}{4}$.

After some straightforward computations we get the identities

$$\begin{aligned}
\lambda(p,q) - \frac{\gamma}{4} &= \frac{p(q-q_-)(q-q_+)}{2(p-1)^2(q-1)}, \\
q_\pm &= \frac{\gamma(p-1)}{4p}\left(p-1+\frac{2}{\gamma} \pm \sqrt{(p-p_-)(p-p_+)}\right), \\
p_\pm &= 1+\tfrac{2}{\gamma} \pm 2\sqrt{\tfrac{2}{\gamma}}.
\end{aligned}$$

Our assumption on p implies that $1+\frac{2}{\gamma}+2\sqrt{\frac{2}{\gamma}} \le 2 \iff \gamma \ge \frac{1}{(\sqrt{3}-\sqrt{2})^2} > 2$ so that $p-p_- > p-1+\frac{2}{\gamma} > 0$ and by our assumption on p, $p-p_+ > 0$ as well so that $q_\pm$ is real and $\lambda(p,q_+) = \frac{\gamma}{4}$.

For our particular values of p_- and p_+ we have furthermore $(p-p_-)(p-p_+) \ge (p-p_+)^2$ so that

$$q_+ \ge \frac{\gamma(p-1)(p-1-\sqrt{\frac{2}{\gamma}})}{2p}$$

Now, apply Lemma 4.4.3 with $q = q_+$ to obtain

$$\mathbb{E}[Z]^p \le \mathbb{E}_{\mathbb{Q}}\big[Z^p\big]\mathbb{E}\big[e^{\frac{\gamma}{4}\langle M\rangle_t}\big]^{\frac{p-1}{q_+}}.$$

We estimate the second expectation on the right hand side using Proposition 4.4.4

$$\begin{aligned}
\mathbb{E}\big[e^{\frac{\gamma}{4}\langle M\rangle_t}\big]^{\frac{p-1}{q_+}} &\le \exp\left(\frac{(p-1)}{q_+}\frac{\kappa_X(\alpha+n\bar{\lambda}_X)t}{2\Lambda_X}\right) \\
&\le \exp\left(\frac{p\,\kappa_X(\alpha+n\bar{\lambda}_X)\,t}{\left(p-1-\sqrt{2/\gamma}\right)\Lambda_X\,\gamma}\right)
\end{aligned}$$

which leads to our result. ∎

4.5 Proof of the main theorem

Lemma 4.5.1. *If $\bar{b}_Y$ is Lipschitz then*

$$\sup_{0\le t\le T}|Y_t-\bar{Y}_t| \le \sup_{0\le t\le T}\left|\int_0^t b_Y(X_s,Y_s)-\bar{b}(Y_s)ds\right| e^{\|\bar{b}\|_{\mathrm{Lip}}T}$$

Proof.

$$\begin{aligned}\sup_{0\le t\le T}|Y_t-\bar{Y}_t| &= \sup_{0\le t\le T}\left|\int_0^t b_Y(X_s,Y_s)-\bar{b}_Y(\bar{Y}_s)ds\right| \\ &\le \sup_{0\le t\le T}\left|\int_0^t b_Y(X_s,Y_s)-\bar{b}_Y(Y_s)ds\right| + \|\bar{b}\|_{\mathrm{Lip}}\int_0^T \sup_{0\le s\le t}|Y_s-\bar{Y}_s|ds\end{aligned}$$

and the conclusion follows from Gronwall's inequality. □

Theorem 4.5.2. *Under Assumption 4.1.1 if $\sigma_Y(x,y)=\sigma_Y(y)$, a Poincaré inequality with constant c_P holds for $\tilde{\rho}_t^y$, a Logarithmic Sobolev inequality with constant c_L holds for $\mu^y(dx)=Z(y)^{-1}e^{-V(x,y)}dx$ both with respect to Γ^X, $X_0\sim\mu^{Y_0}$ and $\bar{b}$ is Lipschitz then for $1\le p\le \frac{2}{1+\frac{2}{\gamma}+2\sqrt{\frac{2}{\gamma}}}$ we have the estimate*

$$\mathbb{E}\left[\sup_{0\le t\le T}|Y_t-\bar{Y}_t|^p\right]^{2/p} \le m{\kappa_Y}^2\Lambda_X\left(27{c_P}^2T+\frac{2{c_L}^2}{4-{c_L}^2\Lambda_X(m{\kappa_Y}^2+3{c_V}^2)}\mathbb{E}\int_0^T\Psi(Y_t)dt\right)$$
$$\exp\left(\frac{2p'\kappa_X(\alpha+n\bar{\lambda}_X)T}{p\gamma\Lambda_X}+2\|\bar{b}\|_{\mathrm{Lip}}T\right)$$

with

$$\Psi(y)=\frac{3m{\kappa_Y}^2(\alpha+n\bar{\lambda}_X)}{2\kappa_X}+\tfrac{3}{2}|\bar{b}(y)|^2+\tfrac{1}{2}\sum_{i,j=1}^m a_Y^{ij}(y)^2+\sum_{i,j=1}^m a_Y^{ij}(y)\,\mathrm{Cov}_{\mu^y}(\partial_{y_i}V,\partial_{y_j}V),$$

$$p'=\frac{1}{1-\frac{p}{2}\left(1+\sqrt{\frac{2}{\gamma}}\right)}>\frac{2}{2-p}$$

and

$${c_V}^2=\sup_y\left(\sum_{i=1}^m\|\partial_{y_i}V(\cdot,y)\|_{\mathrm{Lip}}^2+\sum_{i,j=1}^m\|\partial^2_{y_iy_j}V(\cdot,y)\|_{\mathrm{Lip}}^2\right).$$

Proof. By Lemma 4.5.1 we have

$$\mathbb{E}\left[\sup_{0\le t\le T}|Y_t-\bar{Y}_t|^p\right]\le\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t b_Y(X_s,Y_s)-\bar{b}(Y_s)ds\right|^p\right]e^{p\|\bar{b}\|_{\mathrm{Lip}}T}.$$

Using Proposition 4.4.6 we get for $1 \leq p \leq \frac{2}{1+\frac{2}{\gamma}+2\sqrt{\frac{2}{\gamma}}}$ that

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y(X_s,Y_s)-\bar{b}(Y_s)ds\right|^p\right]$$
$$\leq \mathbb{E}_{\mathbb{Q}}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y(X_s,Y_s)-\bar{b}(Y_s)ds\right|^2\right]^{p/2}\exp\left(\frac{p'\kappa_X(\alpha+n\bar{\lambda_X})T}{\gamma\Lambda_X}\right)$$

with

$$0<p'=\frac{1}{1-\frac{p}{2}\left(1+\sqrt{\frac{2}{\gamma}}\right)}<\infty.$$

By Proposition 4.4.2

$$\mathbb{E}_{\mathbb{Q}}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y(X_s,Y_s)-\bar{b}(Y_s)ds\right|^2\right]=\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y(\tilde{X}_s,Y_s)-\bar{b}(Y_s)ds\right|^2\right].$$

Now we decompose

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y(\tilde{X}_s,Y_s)-\bar{b}(Y_s)ds\right|^2\right]$$
$$\leq 2\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y(\tilde{X}_s,Y_s)-\mathbb{E}[b_Y(\tilde{X}_s,Y_s)|(X_s,Y_s)]ds\right|^2\right]$$
$$+2\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t \mathbb{E}[b_Y(\tilde{X}_s,Y_s)|(X_s,Y_s)]-\bar{b}(Y_s)ds\right|^2\right]. \quad (4.5.1)$$

For the rest of the proof we put ourselves in the setting of section 4.1.2 where we substitute $\tilde{X}$ for X and (X,Y) for Y.

For $1\leq i\leq m$ we now apply Proposition 4.2.4 with $\phi:(x,y)\mapsto y$, $\nu_t^y=\tilde{\rho}_t^y$ and $f_t(\tilde{x},x,y)=b_Y^i(\tilde{x},y)-\mathbb{E}[b_Y^i(\tilde{X}_s,Y_s)|(X_s,Y_s)=(x,y)]$. Since $\tilde{\rho}_t^y$ satisfies a Poincaré inequality by assumption and $\int f_t(\cdot,y)d\tilde{\rho}_t^y=0$ by the properties of conditional expectation, we get

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y^i(\tilde{X}_s,Y_s)-\mathbb{E}[b_Y^i(\tilde{X}_s,Y_s)|(X_s,Y_s)]ds\right|^2\right]$$
$$\leq\frac{27}{2}\int_0^T\mathbb{E}\left[c_P(Y_t)f_t^2(\tilde{X}_t,X_t,Y_t)\right]dt$$
$$\leq\frac{27}{2}\int_0^T\mathbb{E}\left[c_P(Y_t)^2\Gamma^X(b_Y^i)(\tilde{X}_t,Y_t)\right]dt$$
$$\leq\frac{27c_P{}^2\Lambda_X\|\nabla_x b_Y^i\|_\infty^2T}{2}$$

where the second inequality follows from the tower property of conditional expectation and applying the Poincaré inequality a second time to $\tilde{\rho}_t^y$ and the last line from $\Gamma^X(b_Y^i) = \nabla_x {b_Y^i}^T A_X \nabla_x b_Y^i \leq \Lambda_X |\nabla_x b_Y^i|^2 \leq \Lambda_X \|\nabla_x b_Y^i\|_\infty^2$. Summing over the components b_Y^i we get

$$\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t b_Y(\tilde{X}_s, Y_s) - \mathbb{E}[b_Y(\tilde{X}_s, Y_s)|(X_s, Y_s)]ds\right|^2\right] \leq \frac{27{c_P}^2\Lambda_X T}{2}\sum_{i=1}^m \|\nabla_x b_Y^i\|_\infty^2$$
$$= \frac{27{c_P}^2\Lambda_X m{\kappa_Y}^2 T}{2}$$

We now turn to the second term on the right hand side in the decomposition (4.5.1). First, note that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t \mathbb{E}[b_Y(\tilde{X}_s, Y_s)|(X_s, Y_s)] - \bar{b}(Y_s)ds\right|^2\right] \\
&\leq \mathbb{E}\int_0^T \left|\mathbb{E}[b_Y(\tilde{X}_s, Y_s)|(X_s, Y_s)] - \bar{b}(Y_s)ds\right|^2 \\
&= \sum_{i=1}^n \int_0^T \mathbb{E}\left|\int_{\mathcal{X}} b_Y^i(\tilde{x}, Y_t)\tilde{\rho}^{X_t,Y_t}(d\tilde{x}) - \int_{\mathcal{X}} b_Y^i(\tilde{x}, Y_t)\mu^{Y_t}(d\tilde{x})\right|^2
\end{aligned}$$

By Lemma 4.3.4 we have

$$\sum_{i=1}^n \mathbb{E}\left|\int_{\mathcal{X}} b_Y^i(\tilde{x}, Y_t)\tilde{\rho}^{X_t,Y_t}(d\tilde{x}) - \int_{\mathcal{X}} b_Y^i(\tilde{x}, Y_t)\mu^{Y_t}(d\tilde{x})\right|^2$$
$$\leq \sum_{i=1}^n \|b_Y^i\|_{\mathrm{Lip}}^2 \Lambda_X c_L \mathbb{E}H(\tilde{\rho}_t^{X_t,Y_t}|\mu^{Y_t}) \leq m{\kappa_Y}^2\Lambda_X c_L \mathbb{E}H(\tilde{\rho}_t^{X_t,Y_t}|\mu^{Y_t}).$$

Suppose that uniformly in y

$$\left(\sum_{i=1}^m \|\partial_{y_i} V(\cdot, y)\|_{\mathrm{Lip}}^2 + \sum_{i,j=1}^m \|\partial_{y_i y_j}^2 V(\cdot, y)\|_{\mathrm{Lip}}^2\right) < {c_V}^2.$$

Now, for some $r \in \mathbb{R}$ to be fixed later, use Propositions 4.3.1 and 4.3.5 to get

$$\mathbb{E}H(\tilde{\rho}_t^{X_t,Y_t}|\mu^{Y_t})e^{rt} \leq -\left(\frac{2}{c_L} - \frac{{c_V}^2\Lambda_X c_L}{2} - r\right)\int_0^t \mathbb{E}H(\rho_s^{X_s,Y_s}|\mu^{Y_s})e^{rs}ds + \int_0^t \mathbb{E}\Phi(X_s, Y_s)e^{rs}ds. \tag{4.5.2}$$

We have

$$\mathbb{E}\Phi(X_s, Y_s) = \mathbb{E}\left[\tfrac{1}{2}\sum_{i=1}^m b_Y^i(X_s, Y_s)^2 + \tfrac{1}{2}\sum_{i,j=1}^m a_Y^{ij}(Y_s)^2 + \sum_{i,j=1}^m a_Y^{ij}(Y_s)\,\mathrm{Cov}_{\mu^{Y_s}}(\partial_{y_i}V, \partial_{y_j}V)\right]$$

and we estimate the first term on the right hand side as follows:

$$\mathbb{E} b_Y^i(X_s, Y_s)^2 \leq 3\mathbb{E}[b_Y^i(X_s, Y_s) - b_Y^i(\tilde{X}_s, Y_s)]^2 + 3[\mathbb{E} b_Y^i(\tilde{X}_s, Y_s) - \int_{\mathcal{X}} b_Y^i(x, Y_s)\mu^{Y_s}(dx)]^2 + 3\left|\int_{\mathcal{X}} b_Y^i(x, Y_s)\mu^{Y_s}(dx)\right|^2.$$

Since b_Y is Lipschitz in the first variable we get for the first term

$$\sum_{i=1}^m \mathbb{E}[b_Y^i(X_s, Y_s) - b_Y^i(\tilde{X}_s, Y_s)]^2 = \mathbb{E}\left|b_Y(X_s, Y_s) - b_Y(\tilde{X}_s, Y_s)\right|^2 \leq m{\kappa_Y}^2 \mathbb{E}|X_s - \tilde{X}_s|^2 \leq \frac{m{\kappa_Y}^2(\alpha + n\bar{\lambda}_X)}{\kappa_X}.$$

Still using the Lipschitzness of b_Y, we use Lemma 4.3.4 together with the tower property for conditional expectation on the second term to get

$$\sum_{i=1}^m [\mathbb{E} b_Y^i(\tilde{X}_s, Y_s) - \int_{\mathcal{X}} b_Y^i(x, Y_s)\mu^{Y_s}(dx)]^2 \leq m{\kappa_Y}^2 c_L \Lambda_X \mathbb{E}\,\mathrm{H}(\tilde{\rho}_s^{X_s,Y_s}|\mu^{Y_s}).$$

This leads us to

$$\mathbb{E}\Phi(X_s, Y_s) \leq \tfrac{3}{2}\left(m{\kappa_Y}^2 c_L \Lambda_X \mathbb{E}\,\mathrm{H}(\tilde{\rho}_s^{X_s,Y_s}|\mu^{Y_s}) + \frac{m{\kappa_Y}^2(\alpha + n\bar{\lambda}_X)}{\kappa_X} + \mathbb{E}|\bar{b}(Y_s)|^2\right) + \tfrac{1}{2}(a_Y^{ij}(Y_s) + a_Y^{ij}(Y_s)\,\mathrm{Cov}_{\mu^{Y_s}}(\partial_{y_i}V, \partial_{y_j}V).$$

Substituting Φ in (4.5.2) we get

$$\mathbb{E}H(\tilde{\rho}_t^{X_t,Y_t}|\mu^{Y_t})e^{rt} \leq -\left(\frac{2}{c_L} - \frac{\Lambda_X c_L(m{\kappa_Y}^2 + 3{c_V}^2)}{2} - r\right)\int_0^t \mathbb{E}H(\rho_s^{X_s,Y_s}|\mu^{Y_s})e^{rs}ds + \mathbb{E}\int_0^t e^{rs}\frac{3m{\kappa_Y}^2(\alpha + n\bar{\lambda}_X)}{2\kappa_X} + \tfrac{3}{2}|\bar{b}(Y_s)|^2 + \tfrac{1}{2}\sum_{i,j=1}^m a_Y^{ij}(Y_s)^2 + \sum_{i,j=1}^m a_Y^{ij}(Y_s)\,\mathrm{Cov}_{\mu^{Y_s}}(\partial_{y_i}V, \partial_{y_j}V)ds.$$

Now we choose

$$r = \frac{2}{c_L} - \frac{\Lambda_X c_L(m{\kappa_Y}^2 + 3{c_V}^2)}{2}$$

so that

$$\mathbb{E}H(\tilde{\rho}_t^{X_t,Y_t}|\mu^{Y_t}) \leq \mathbb{E}\int_0^t e^{-r(t-s)}\Psi(Y_s)ds.$$

with

$$\Psi(y) = \frac{3m{\kappa_Y}^2(\alpha + n\bar{\lambda}_X)}{2\kappa_X} + \tfrac{3}{2}|\bar{b}(y)|^2 + \tfrac{1}{2}\sum_{i,j=1}^m a_Y^{ij}(y)^2 + \sum_{i,j=1}^m a_Y^{ij}(y)\,\mathrm{Cov}_{\mu^y}(\partial_{y_i}V, \partial_{y_j}V).$$

By the preceding inequality and the Young inequality for convolutions on $L^1([0,T])$

$$\begin{aligned}\int_0^T \mathbb{E}H(\tilde{\rho}_t^{X_t,Y_t}|\mu^{Y_t})dt &\leq \mathbb{E}\int_0^T\int_0^t e^{-r(t-s)}\Psi(Y_s)dsdt \\ &\leq \mathbb{E}\int_0^T e^{-rt}dt\int_0^T \Psi(Y_t)dt \\ &= \frac{1}{r}(1-e^{-rT})\mathbb{E}\int_0^T \Psi(Y_t)dt\end{aligned}$$

so that finally

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq t\leq T}\left|\int_0^t \mathbb{E}[b_Y(\tilde{X}_s,Y_s)|(X_s,Y_s)] - \bar{b}(Y_s)ds\right|^2\right] &\leq \frac{m{\kappa_Y}^2\Lambda_X c_L}{r}(1-e^{-rT})\mathbb{E}\int_0^T \Psi(Y_t)dt \\ &= \frac{2{c_L}^2\Lambda_X m{\kappa_Y}^2}{4-{c_L}^2\Lambda_X(m{\kappa_Y}^2+3{c_V}^2)}(1-e^{-rT})\mathbb{E}\int_0^T \Psi(Y_t)dt \\ &\leq \frac{2{c_L}^2\Lambda_X m{\kappa_Y}^2}{4-{c_L}^2\Lambda_X(m{\kappa_Y}^2+3{c_V}^2)}\mathbb{E}\int_0^T \Psi(Y_t)dt.\end{aligned}$$

Assembling the previous results, we obtain

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq t\leq T}|Y_t-\bar{Y}_t|^p\right]^{2/p} \leq m{\kappa_Y}^2\Lambda_X &\left(27{c_P}^2T + \frac{2{c_L}^2}{4-{c_L}^2\Lambda_X(m{\kappa_Y}^2+3{c_V}^2)}\mathbb{E}\int_0^T \Psi(Y_t)dt\right) \\ &\exp\left(\frac{2p'\kappa_X(\alpha+n\bar{\lambda}_X)T}{p\gamma\Lambda_X} + 2\|\bar{b}\|_{\mathrm{Lip}}T\right).\end{aligned}$$

□

4.6 Applications

4.6.1 Averaging

For $\varepsilon > 0$ fixed consider an SDE of the form

$$dX_t = -\varepsilon^{-1}\nabla_x V(X_t,Y_t)dt + \varepsilon^{-1/2}\sqrt{2\beta_X^{-1}}dB_t^X \tag{4.6.1}$$

$$dY_t = b_Y(X_t,Y_t)dt + \sqrt{2\beta_Y^{-1}}dB_t^Y \tag{4.6.2}$$

with $Y_0 = y_0 \in \mathbb{R}^m$ and $X_0 \sim \mu^{y_0} = e^{-\beta V(x,y_0)}dx$ and $V(x,y)$ is of the form

$$V(x,y) = \frac{1}{2}(x-g(y))Q(x-g(y)) + h(x,y)$$

where h is uniformly bounded in both arguments and both $\partial_y h$ and $\partial_y^2 h$ are Lipschitz in x uniformly in y. Under these conditions

$$\mu^y(dx) = Z(y)^{-1} e^{-\beta_X V(x,y)} dx \text{ with } Z(y) = \int_{\mathcal{X}} e^{-\beta_X V(x,y)} dx$$

is a Gaussian measure with covariance matrix $\beta_X Q$ and mean $g(y)$ perturbed by a bounded factor $e^{-\beta_X h(x,y)}$. As such it satisfies a Logarithmic Sobolev inequality with respect to the usual square field operator $|\nabla|^2$ with constant

$$c_L^0 = (\beta_X \lambda_Q)^{-1} e^{\beta_X \operatorname{osc}(h)} \quad \text{with } \operatorname{osc}(h) = \sup h - \inf h$$

and λ_Q is the smallest eigenvalue of Q. In particular, μ^y satisfies a Logarithmic Sobolev inequality with constant

$$c_L = \varepsilon \lambda_Q^{-1} e^{\beta_X \operatorname{osc}(h)}$$

with respect to $\Gamma^X = \varepsilon^{-1} \beta_X^{-1} |\nabla|^2$.

We have

$$\begin{aligned}
-(x_1 - x_2)^T &(\nabla_x V(x_1, y) - \nabla_x V(x_2, y)) \\
&= -(x_1 - x_2)^T Q (x_1 - x_2) - (x_1 - x_2)^T (\nabla_x h(x_1, y) - \nabla_x h(x_2, y)) \\
&\leq -\lambda_Q |x_1 - x_2|^2 + |x_1 - x_2| \|\nabla_x h\|_\infty \\
&\leq -\lambda_Q |x_1 - x_2|^2 + \frac{\|\nabla_x h\|_\infty}{4\lambda_Q}
\end{aligned}$$

so that we can choose

$$\kappa_X = \varepsilon^{-1} \lambda_Q, \qquad \alpha = \varepsilon^{-1} \frac{\|\nabla_x h\|_\infty}{4\lambda_Q}.$$

We also have trivially

$$\lambda_X = \Lambda_X = \bar{\lambda}_X = \varepsilon^{-1} \beta_X^{-1}, \qquad \Lambda_Y = \beta_Y^{-1}, \qquad \kappa_Y = \|\nabla_x b_Y\|_\infty$$

and the separation of timescales is

$$\gamma = \frac{{\kappa_X}^2 \lambda_Y}{\Lambda_X m {\kappa_Y}^2} = \varepsilon^{-1} \frac{{\lambda_Q}^2 \beta_Y^{-1}}{\|\nabla_x b_Y\|_\infty^2 \beta_X^{-1}}.$$

If $\gamma > \frac{1}{(\sqrt{3}-\sqrt{2})^2} \approx 9.899$ we can apply Theorem 4.5.2 with $p = 1$ to get

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} |Y_t - \bar{Y}_t|\right]^2 \leq \varepsilon C_1 \left(27(c_P(\varepsilon)/c_L)^2 T + C_2 \mathbb{E} \int_0^T \Psi(Y_t) dt\right) \exp\left(2p' C_3 T + 2\|\bar{b}\|_{\text{Lip}} T\right)$$

with

$$C_1 = \varepsilon^{-1} m {\kappa_Y}^2 \Lambda_X {c_L}^2 = m {\kappa_Y}^2 \beta_X^{-1} \lambda_Q^{-2} e^{2\beta_X \operatorname{osc}(h)},$$

$$C_2 = \frac{2}{4 - {c_L}^2 \Lambda_X (m {\kappa_Y}^2 + 3 {c_V}^2)} \leq 1 \text{ for } \varepsilon \leq \frac{2\lambda_Q e^{-\beta_X \operatorname{osc}(h)} \beta_X}{\|\nabla_X b_Y\|_\infty^2 + 3 {c_V}^2},$$

$$C_3 = \frac{\kappa_X(\alpha + n\bar{\lambda}_X)}{\gamma \Lambda_X} = \frac{\|\nabla_x b_Y\|_\infty^2 (\frac{\|\nabla_x h\|_\infty}{4\lambda_Q} + n\beta_X^{-1})}{\beta_Y^{-1} \lambda_Q},$$

$$\begin{aligned}
\Psi(y) &= \frac{3m{\kappa_Y}^2(\alpha + n\bar{\lambda}_X)}{2\kappa_X} + \tfrac{3}{2}|\bar{b}(y)|^2 + \tfrac{1}{2}\sum_{i,j=1}^m a_Y^{ij}(y)^2 + \sum_{i,j=1}^m a_Y^{ij}(y) \operatorname{Cov}_{\mu^y}(\partial_{y_i}\beta_X V, \partial_{y_j}\beta_X V) \\
&= \frac{3\|\nabla_x b_Y\|_\infty^2 (\frac{\|\nabla_x h\|_\infty}{4\lambda_Q} + n\beta_X^{-1})}{2\lambda_Q} + \tfrac{1}{2} m\beta_Y^{-2} + \beta_Y^{-1}\beta_X^2 \sum_i \operatorname{Var}_{\mu^y}(\partial_{y_i} V) + \tfrac{3}{2}|\bar{b}(y)|^2 \\
&\leq \frac{3\|\nabla_x b_Y\|_\infty^2 (\frac{\|\nabla_x h\|_\infty}{4\lambda_Q} + n\beta_X^{-1})}{2\lambda_Q} + \tfrac{1}{2} m\beta_Y^{-2} + \beta_Y^{-1}\beta_X^2 c_L^0 \sum_i \|\partial_{y_i} V\|_{\mathrm{Lip}}^2 + \tfrac{3}{2}|\bar{b}(y)|^2 \\
&= \frac{3\|\nabla_x b_Y\|_\infty^2 (\frac{\|\nabla_x h\|_\infty}{4\lambda_Q} + n\beta_X^{-1})}{2\lambda_Q} + \tfrac{1}{2} m\beta_Y^{-2} + \beta_X(\beta_Y\lambda_Q)^{-1} e^{\beta_X \operatorname{osc}(h)} \sum_i \|\partial_{y_i} V\|_{\mathrm{Lip}}^2 + \tfrac{3}{2}|\bar{b}(y)|^2,
\end{aligned}$$

$$2 < p' = \frac{1}{1 - \frac{1}{2}\left(1 + \sqrt{\frac{2}{\gamma}}\right)} < \frac{2}{3 - \sqrt{2}\sqrt{3}} \approx 3.633$$

and

$${c_V}^2 = \sup_y \left(\sum_{i=1}^m \|\partial_{y_i} V(\cdot, y)\|_{\mathrm{Lip}}^2 + \sum_{i,j=1}^m \|\partial_{y_i y_j}^2 V(\cdot, y)\|_{\mathrm{Lip}}^2 \right).$$

If we suppose that $c_P(\varepsilon)/c_L$ converges to a finite limit as $\varepsilon \to 0$ and that

$$\mathbb{E} \int_0^T \bar{b}(Y_t)^2 dt < \infty$$

then there exists a constant C depending on T, V, β_X, b_Y and β_Y such that for ε sufficiently small

$$\mathbb{E} \sup_{0 \leq t \leq T} |Y_t - \bar{Y}_t| \leq \sqrt{\varepsilon} C.$$

In other words, we obtain a strong averaging principle of order $1/2$ in ε.

4.6.2 Temperature-Accelerated Molecular Dynamics

In [MV06] the authors introduced the TAMD process (X_t, Y_t) and its averaged version $\bar{Y}_t$ defined by

$$
\begin{aligned}
&dX_t = -\tfrac{1}{\varepsilon}\nabla_x U(X_t, Y_t)dt + \sqrt{2(\beta\varepsilon)^{-1}}dB_t^X, \quad X_0 \sim e^{-\beta U(x,y_0)}dx \\
&dY_t = -\tfrac{1}{\bar{\gamma}}\kappa(Y_t - \theta(X_t))dt + \sqrt{2(\bar{\beta}\bar{\gamma})^{-1}}dB_t^Y, \quad Y_0 = y_0 \\
&d\bar{Y}_t = \bar{b}(\bar{Y}_t)dt + \sqrt{2(\bar{\beta}\bar{\gamma})^{-1}}dB_t^Y, \quad \bar{Y}_0 = y_0 \\
U(x,y) &= V(x) + \tfrac{\kappa}{2}|y - \theta(x)|^2, \\
&\bar{b}(y) = Z(y)^{-1}\int -\bar{\gamma}^{-1}\kappa(y - \theta(x))e^{-\frac{\kappa}{2}|y-\theta(x)|^2}e^{-V(x)}dx, \quad Z(y) = \int e^{-\frac{\kappa}{2}|y-\theta(x)|^2}e^{-V(x)}dx
\end{aligned}
$$

with $X_t \in \mathbb{R}^n$, $Y_t, \bar{Y}_t \in \mathbb{R}^m$, a Lipschitz-continuous function $V(x)$, constants $\kappa, \varepsilon, \beta, \bar{\beta}, \bar{\gamma} > 0$ and independent standard Brownian motions B^X, B^Y on $\mathbb{R}^n$ and $\mathbb{R}^m$.

Let $D \subset \mathbb{R}^m$ be a compact set and define the stopping time $\tau = \inf\{t \geq 0 : Y_t \notin D\}$.

We will show that under some additional assumptions, a strong averaging principle with rate 1/2 holds in the sense that for any fixed T and ε sufficiently small but fixed, there exists a constant C not depending on ε such that

$$\sup_{0\leq t\leq T} |Y_{t\wedge\tau} - \bar{Y}_{t\wedge\tau}| \leq C\varepsilon^{1/2}.$$

We need the following extra assumptions on the TAMD process:

$$
\begin{gathered}
0 < \lambda_\theta \operatorname{Id}_m < D\theta(x)D\theta(x)^T < \Lambda_\theta \operatorname{Id}_m < \infty, \\
-(x_1 - x_2)^T(\nabla_x(\theta(x_1) - y)^2 - \nabla_x(\theta(x_2) - y)^2) \leq -\kappa_\theta|x_1 - x_2|^2 + \alpha_\theta \\
\lim_{|x|\to\infty} |\theta(x)| = \infty \\
\lambda_\theta\kappa > \Lambda_\theta\beta^{-1}.
\end{gathered}
$$

In order to apply Theorem 4.5.2 we also need to suppose that Assumption 4.1.2 holds for the TAMD process.

We will now briefly comment on the form of $\bar{Y}_t$. Let

$$\mu(dx) = Z_0^{-1}e^{-V(x)}dx, \quad Z_0 = \int e^{-V(x)}dx$$

so that

$$\begin{aligned}\bar{b}(y) &= \frac{Z_0}{Z(y)} \int -\bar{\gamma}^{-1}\kappa(\theta(x) - y)e^{-\frac{\kappa}{2}|\theta(x)-y|^2}\mu(dx) \\ &= \frac{Z_0}{Z(y)}\bar{\gamma}^{-1} \int -\kappa(z - y)e^{-\frac{\kappa}{2}|z-y|^2}\theta_{\#}\mu(dz) \\ &= \frac{Z_0}{Z(y)}\bar{\gamma}^{-1}\nabla_y \int e^{-\frac{\kappa}{2}|z-y|^2}\theta_{\#}\mu(dz)\end{aligned}$$

where $\theta_{\#}\mu$ denotes the image measure of μ by θ. Now note that

$$\frac{Z(y)}{Z_0} = \int e^{-\frac{\kappa}{2}|\theta(x)-y|^2}\mu(dx) = \int e^{-\frac{\kappa}{2}|z-y|^2}\theta_{\#}\mu(dz)$$

so that

$$\bar{b}(y) = \bar{\gamma}^{-1}\nabla_y \log \int e^{-\frac{\kappa}{2}|z-y|^2}\theta_{\#}\mu(dz) = \nabla_y \log(\theta_{\#}\mu * \mathcal{N}(0, \kappa^{-1}))(y).$$

In the last expression, $*$ denotes convolution, $\mathcal{N}(0, \kappa^{-1})$ denotes the Gaussian measure with mean 0 and variance κ^{-1} and we identify through an abuse of notation measures and their densities which we suppose to exist.

Thus,

$$d\bar{Y}_t = \bar{\gamma}^{-1}\nabla_y \log(\theta_{\#}\mu * \mathcal{N}(0, \kappa^{-1}))(\bar{Y}_t)dt + \sqrt{2(\bar{\beta}\bar{\gamma})^{-1}}dB_t^Y.$$

In physical terms, $\bar{Y}_t$ evolves at an inverse temperature of $\bar{\beta}$ on the energy landscape corresponding to the image measure of μ by θ convolved with a Gaussian measure of variance κ^{-1}.

We proceed to establish a Logarithmic Sobolev inequality for μ^y via the Lyapunov function method. From [CG17] Theorem 1.2 it follows that a sufficient condition for a Logarithmic Sobolev inequality to hold for an elliptic, reversible diffusion process with generator L and reversible measure μ is: there exist constants $\lambda > 0$, $b > 0$, a function $W \geq w > 0$, a function $V(x)$ such that V goes to infinity at infinity, $|\nabla V(x)| \geq v > 0$ for $|x|$ large enough and such that $\mu(e^{aV}) < \infty$ verifying

$$LW(x) \leq -\lambda V(x)W(x) + b.$$

Fix y and let $F(x, y) = \frac{1}{2}|\theta(x) - y|^2$. In order to establish a Logarithmic Sobolev inequality for μ^y we are going to show that the preceding condition holds for $V(x) = F(x, y)$ and $W(x) = e^{F(x,y)}$. We have

$$\begin{aligned}\nabla_x F(x) &= D\theta(x)^T(\theta(x) - y), \\ \lambda_\theta|\theta(x) - y|^2 \leq |\nabla_x F|^2 &\leq \Lambda_\theta|\theta(x) - y|^2, \\ \Delta F &= n\bar{\lambda}_\theta + (\Delta\theta)^T(\theta - y).\end{aligned}$$

Furthermore

$$\begin{aligned}
\varepsilon L^X F &= -\nabla_x V_0^T \nabla_x F - \kappa|\nabla_x F|^2 + \beta^{-1}\Delta F \\
&= -\nabla_x V_0^T D\theta^T(\theta - y) - \kappa|D\theta^T(\theta-y)|^2 + \beta^{-1} n\bar{\lambda}_\theta + \beta^{-1}\Delta\theta^T(\theta - y) \\
&\leq |\nabla_x V_0|\sqrt{\Lambda_\theta}|\theta - y| - \kappa\lambda_\theta|\theta-y|^2 + \beta^{-1}n\bar{\lambda}_\theta + \beta^{-1}|\Delta\theta||\theta - y| \\
&\leq -\kappa\lambda_\theta F + \frac{(|\nabla_x V_0|\sqrt{\Lambda_\theta} + \beta^{-1}|\Delta\theta|)^2}{2\kappa\lambda_\theta} + \beta^{-1}n\bar{\lambda}_\theta \\
&= -\kappa\lambda_\theta F + G(x)
\end{aligned}$$

where we used the fact that $-ax^2 + bx + c \leq -\frac{1}{2}ax^2 + \frac{b^2}{2a} + c$ for the second inequality.

Let $W(x,y) = e^{F(x,y)}$. Now,

$$\begin{aligned}
\varepsilon L^X W(x,y) &= \varepsilon L^X F(x,y)W(x,y) + \beta^{-1}|\nabla_x F(x,y)|^2 W(x,y) \\
&\leq -(\lambda_\theta\kappa - \Lambda_\theta\beta^{-1})F(x,y)W(x,y) + \|G\|_\infty W(x,y) \\
&= -((\lambda_\theta\kappa - \Lambda_\theta\beta^{-1})F(x,y) - \|G\|_\infty)W(x,y).
\end{aligned}$$

Since F goes to infinity at infinity, for x outside a compact set

$$-(\lambda_\theta\kappa - \Lambda_\theta\beta^{-1})F(x,y) + \|G\|_\infty \leq -\tfrac{1}{2}(\lambda_\theta\kappa - \Lambda_\theta\beta^{-1})F(x,y)$$

so that

$$\varepsilon L^X W(x,y) \leq -\tfrac{1}{2}(\lambda_\theta\kappa - \Lambda_\theta\beta^{-1})F(x,y)W(x,y) + K$$

for some constant K. This establishes a Log-Sobolev inequality for the measure μ^y with respect to $\varepsilon\Gamma^X$ in the sense that

$$\int f^2 \log f^2 d\mu^y \leq 2c_L^y \int \varepsilon\Gamma^X d\mu^y$$

for some constant c_L^y depending on y. Let $c_L = \sup_{y\in D} c_L^y$ so that

$$\int f^2 \log f^2 d\mu^y \leq 2\varepsilon c_L \int \Gamma^X d\mu^y.$$

This shows that a Log-Sobolev inequality with a constant εc_L holds for each measure $\mu^y, y \in D$.

It remains to estimate $\kappa_X, \kappa_Y, \|\partial_{y_i} U\|_{\text{Lip}}^2, \|\partial_{y_i}^2 U\|_{\text{Lip}}^2$ and $\bar{b}(y)^2$.

We have $b_X = -\varepsilon^{-1}\nabla_x V(x) - \varepsilon^{-1}\frac{\kappa}{2}\nabla_x|\theta(x) - y|^2$ and we want to find κ_X such that

$$(x_1 - x_2)^T(b_X(x_1,y) - b_X(x_2,y)) \leq -\kappa_X|x_1 - x_2|^2 + \alpha \text{ for all } x_1, x_2 \in \mathbb{R}^n, y \in \mathbb{R}^m.$$

Since $|\nabla_x V|$ is bounded and using the assumption on θ, we get

$$\begin{aligned}
&(x_1 - x_2)^T (b_X(x_1, y) - b_X(x_2, y)) \\
&\quad = -\varepsilon^{-1}(x_1 - x_2)^T (\nabla_x V(x_1) - \nabla_x V(x_2)) - \varepsilon^{-1}\tfrac{\kappa}{2}(x_1 - x_2)^T (\nabla_x |\theta(x_1) - y|^2 - \nabla_x |\theta(x_2) - y|^2) \\
&\quad \leq -\varepsilon^{-1}\tfrac{\kappa}{2}\kappa_\theta |x_1 - x_2|^2 + 2\varepsilon^{-1}|x_1 - x_2| \|\nabla_x V(x)\|_\infty + \varepsilon^{-1}\alpha_\theta \\
&\quad \leq -\varepsilon^{-1}\frac{\kappa\kappa_\theta}{4}|x_1 - x_2|^2 + 4\varepsilon^{-1}\frac{\|\nabla_x V\|_\infty}{\kappa\kappa_\theta} + \varepsilon^{-1}\alpha_\theta
\end{aligned}$$

so that we can identify

$$\kappa_X = \varepsilon^{-1}\frac{\kappa\kappa_\theta}{4} \qquad\qquad \alpha = 4\varepsilon^{-1}\frac{\|\nabla_x V\|_\infty}{\kappa\kappa_\theta} + \varepsilon^{-1}\alpha_\theta.$$

We have

$$b_Y^i(x, y) = -\nabla_{y_i} U(x, y) = -\kappa(y_i - \theta_i(x))$$

so that

$$\nabla_x b_Y^i(x, y) = \kappa \nabla_x \theta_i(x)$$

and

$$\kappa_Y{}^2 = \frac{1}{m}\sum_{i=1}^m \kappa^2 \|\nabla_x \theta_i(x)\|_\infty^2 \leq \kappa^2 \Lambda_\theta.$$

We also have

$$\|\partial_{y_i} U\|_{\mathrm{Lip}}^2 = \|b_Y^i\|_{\mathrm{Lip}}^2 \leq \kappa^2 \|\nabla_x \theta_i\|_\infty^2 \leq \kappa^2 \Lambda_\theta$$

and

$$\|\partial_{y_i}^2 U\|_{\mathrm{Lip}}^2 = \|\partial_{y_i} \kappa\theta(x)\|_{\mathrm{Lip}}^2 = 0$$

so that

$$c_V{}^2 = \sup_y \left(\sum_{i=1}^m \|\partial_{y_i} U(\cdot, y)\|_{\mathrm{Lip}}^2 + \sum_{i,j=1}^m \|\partial_{y_i y_j}^2 U(\cdot, y)\|_{\mathrm{Lip}}^2 \right) \leq m\kappa^2 \Lambda_\theta.$$

From the expression for $\varepsilon L^X F$ we get that

$$F \leq -\frac{\varepsilon}{\kappa\lambda_\theta} L^X F + \frac{G(x)}{\kappa\lambda_\theta}.$$

Now

$$\begin{aligned}
\bar{b}^2(y) &= \left(\int -\kappa(y-\theta(x))\mu^y(dx)\right)^2 \\
&\leq \kappa^2 \int F(x,y)\mu^y(dx) \\
&\leq -\frac{\kappa\varepsilon}{\lambda_\theta}\int L^X F(x,y)\mu^y(dx) + \frac{\kappa}{\lambda_\theta}\int G(x)\mu^y(dx) \\
&= \frac{\kappa}{\lambda_\theta}\int G(x)\mu^y(dx)
\end{aligned}$$

since μ^y is invariant for $L^X(\cdot, y)$.

The separation of timescales is

$$\gamma = \frac{{\kappa_X}^2\lambda_Y}{\Lambda_x {\kappa_Y}^2} \geq \varepsilon^{-1}\frac{{\kappa_\theta}^2(\bar{\beta}\bar{\gamma})^{-1}}{16\Lambda_\theta(\beta^{-1})}.$$

If $\gamma > \frac{1}{(\sqrt{3}-\sqrt{2})^2}$ we can now apply Theorem 4.5.2 as in the previous section to show that an averaging principle holds for the stopped TAMD process with rate $\varepsilon^{1/2}$, i.e. there exists a constant C depending on T, V, β_X, b_Y and β_Y such that for

$$\varepsilon \leq \frac{16(\sqrt{3}-\sqrt{2})^2\Lambda_\theta\bar{\gamma}\beta^{-1}}{\kappa_\theta^2\bar{b}^1}$$

we have

$$\mathbb{E}\sup_{0\leq t\leq T}|Y_{t\wedge\tau} - \bar{Y}_{t\wedge\tau}| \leq \sqrt{\varepsilon}C.$$

Bibliography

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators.* Vol. 348. Grundlehren der mathematischen Wissenschaften. Cham: Springer International Publishing, 2014.

[Bha+09] Shalabh Bhatnagar, Richard S. Sutton, Mohammad Ghavamzadeh, and Mark Lee. “Natural actor–critic algorithms”. In: *Automatica* 45.11 (Nov. 2009), pp. 2471–2482.

[CCG12] Patrick Cattiaux, Djalil Chafaı, and Arnaud Guillin. “Central limit theorems for additive functionals of ergodic Markov diffusions processes”. In: *ALEA Lat. Am. J. Prob. Math. Stat* 9.2 (2012), pp. 337–382.

[CG17] Patrick Cattiaux and Arnaud Guillin. “Hitting times, functional inequalities, lyapunov conditions and uniform ergodicity”. In: *Journal of Functional Analysis* 272.6 (2017), pp. 2361–2391.

[CO16] Dan Crisan and Michela Ottobre. “Pointwise gradient bounds for degenerate semigroups (of UFG type)”. In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* 472.2195 (Nov. 2016).

[Ebe16] Andreas Eberle. “Reflection couplings and contraction rates for diffusions”. In: *Probability Theory and Related Fields* 166.3 (2016), pp. 851–886.

[EGZ16] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. “Quantitative Harris type theorems for diffusions and McKean-Vlasov processes”. In: *arXiv preprint arXiv:1606.06012* (2016).

[EGZ17] Andreas Eberle, Arnaud Guillin, and Raphael Zimmer. “Couplings and quantitative contraction rates for Langevin dynamics”. In: *ArXiv e-prints* (Mar. 2017).

[EK89] Robert J. Elliott and Michael Kohlmann. “Integration by Parts, Homogeneous Chaos Expansions and Smooth Densities”. In: *The Annals of Probability* 17.1 (1989), pp. 194–207.

[ERV09] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. “A general strategy for designing seamless multiscale methods”. In: *Journal of Computational Physics* 228.15 (Aug. 2009), pp. 5437–5453.

[FW12] Mark I. Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems.* 3rd edition. Springer Berlin Heidelberg, 2012.

[GGW14] Fuqing Gao, Arnaud Guillin, and Liming Wu. "Bernstein-type Concentration Inequalities for Symmetric Markov Processes". In: *Theory of Probability & Its Applications* 58.3 (Jan. 2014), pp. 358–382.

[Gui+09] Arnaud Guillin, Christian Léonard, Liming Wu, and Nian Yao. "Transportation-information inequalities for Markov processes". In: *Probability Theory and Related Fields* 144.3 (July 2009), pp. 669–695.

[HP86] U. G. Haussmann and Étienne Pardoux. "Time Reversal of Diffusions". In: *The Annals of Probability* 14.4 (1986), pp. 1188–1205.

[Jou09] Aldéric Joulin. "A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature". In: *Bernoulli* 15.2 (May 2009), pp. 532–549.

[Kal02] Olav Kallenberg. *Foundations of modern probability.* Springer Science & Business Media, 2002.

[KKO99] Petar V. Kokotović, Hassan K. Khalil, and J. O'Reilly. *Singular perturbation methods in control: analysis and design.* Classics in applied mathematics 25. Philadelphia: Society for Industrial and Applied Mathematics, 1999. 371 pp.

[KY04] Rafail Z Khasminskii and George Yin. "On averaging principles: An asymptotic expansion approach". In: *SIAM journal on mathematical analysis* 35.6 (2004), pp. 1534–1560.

[Lez01] Pascal Lezaud. "Chernoff and Berry–Esséen inequalities for Markov processes". In: *ESAIM: Probability and Statistics* 5 (2001), pp. 183–201.

[Liu10] Di Liu. "Strong convergence of principle of averaging for multiscale stochastic dynamical systems". In: *Commun. Math. Sci* 8.4 (2010), pp. 999–1020.

[LL10] Frédéric Legoll and Tony Lelièvre. "Effective dynamics using conditional expectations". In: *Nonlinearity* 23.9 (Sept. 1, 2010), pp. 2131–2163.

[LL13] Eva Löcherbach and Dasha Loukianova. "Polynomial deviation bounds for recurrent Harris processes having general state space". In: *ESAIM: Probability and Statistics* 17 (2013), pp. 195–218.

[LLO16] Frédéric Legoll, Tony Lelièvre, and Stefano Olla. "Pathwise estimates for an effective dynamics". In: *arXiv preprint arXiv:1605.02644* (2016).

[LTE17] Qianxiao Li, Cheng Tai, and Weinan E. "Stochastic Modified Equations and Adaptive Stochastic Gradient Algorithms". In: *International Conference on Machine Learning.* 2017, pp. 2101–2110.

[MV06] Luca Maragliano and Eric Vanden-Eijnden. "A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations". In: *Chemical Physics Letters* 426.1 (July 2006), pp. 168–175.

[Pep17] Bob Pepin. “Towards a Quantitative Averaging Principle for Stochastic Differential Equations”. In: *ArXiv e-prints* (Sept. 2017).

[PV01] Étienne Pardoux and Alexander Yu. Veretennikov. “On the Poisson equation and diffusion approximation I”. In: *The Annals of Probability* (2001), pp. 1061–1085.

[PV03] Étienne Pardoux and Alexander Yu. Veretennikov. “On the Poisson equation and diffusion approximation II”. In: *The Annals of Probability* 31.3 (2003), pp. 1166–1192.

[SV17] Gabriel Stoltz and Eric Vanden-Eijnden. “Longtime convergence of the Temperature-Accelerated Molecular Dynamics Method”. In: *arXiv preprint arXiv:1708.08800* (2017).

[SVM07] J. A. Sanders, F. Verhulst, and James A. Murdock. *Averaging methods in nonlinear dynamical systems*. 2nd ed. Applied mathematical sciences (Springer-Verlag New York Inc.) v. 59. New York: Springer, 2007. 431 pp.

[Von91] Jan Von Plato. “Boltzmann’s ergodic hypothesis”. In: *Archive for History of Exact Sciences* 42.1 (1991), pp. 71–89.

[Wan16] Feng-Yu Wang. “Exponential Contraction in Wasserstein Distances for Diffusion Semigroups with Negative Curvature”. In: *arXiv preprint arXiv:1603.05749* (2016).

[Wu00] Liming Wu. “A deviation inequality for non-reversible Markov processes”. In: *Annales de l’IHP Probabilités et statistiques*. Vol. 36. 2000, pp. 435–445.

[Wu09] Liming Wu. “Gradient estimates of Poisson equations on Riemannian manifolds and applications”. In: *Journal of Functional Analysis* 257.12 (Dec. 2009), pp. 4015–4033.