Perturbation of Conservation Laws and Averaging on Manifolds

Xue-Mei Li

Abstract We prove a stochastic averaging theorem for stochastic differential equations in which the slow and the fast variables interact. The approximate Markov fast motion is a family of Markov process with generator $\mathscr{L}_x$ for which we obtain a quantitative locally uniform law of large numbers and obtain the continuous dependence of their invariant measures on the parameter x. These results are obtained under the assumption that $\mathscr{L}_x$ satisfies Hörmander's bracket conditions, or more generally $\mathscr{L}_x$ is a family of Fredholm operators with sub-elliptic estimates. For stochastic systems in which the slow and the fast variable are not separate, conservation laws are essential ingredients for separating the scales in singular perturbation problems, we demonstrate this by a number of motivating examples, from mathematical physics and from geometry, where conservation laws taking values in non-linear spaces are used to deduce slow-fast systems of stochastic differential equations.

1 Introduction

A deterministic or random system with a conservation law is often used to approximate the motion of an object that is also subjected to many other smaller deterministic or random influences. The latter is a perturbation of the former. To describe the evolution of the dynamical system, we begin with these conservation laws. A conservation law is a quantity which does not change with time, for us it is an equivariant map on a manifold, i.e. a map which is invariant under an action of a group. They describe the orbit of the action. Quantities describing the perturbed systems have their natural scales, the conservations laws can be used to determine the different components of the system which evolve at different speeds. Some components may move at a much faster speed than some others, in which case we either ignore

Xue-Mei Li
Department of Mathematics, Imperial College London, London SW7 2AZ, U.K. e-mail: xue-mei.li@imperial.ac.uk

the slow components, in other words we approximate the perturbed system by the unperturbed one, or ignore the fast components and describe the slow components for which the key ingredient is ergodic averaging. It is a standard assumption that the fast variable moves so fast that its influence averaged over any time interval, of the size comparable to the natural scale of our observables, is effectively that of an averaged vector field. The averaging is with respect to a probability measure on the state space of the fast variable. Depending on the object of the study, we will need to neglect either the small perturbations or quantities too large (infinities) to fit into the natural scale of things. To study singularly perturbation operators, we must discard the infinities and at the same time retain the relevant information on the natural scale. In Hamiltonian formulation, for example, the time evolution of an object, e.g. the movements of celestial bodies, is governed by a Hamiltonian function. If the magnitude of the Hamiltonian is set to be of order '1', the magnitude of the perturbation (the collective negligible influences) is of order ε, then the perturbation is negligible on an interval of any fixed length. This ratio in magnitudes translates into time scales. If the original system is on scale 1, we work on a time interval of length $\frac{1}{\varepsilon}$ to see the deviation of the perturbed trajectories. Viewed on the time interval $[0,1]$ the perturbation is not observable. On $[0,\frac{1}{\varepsilon}]$ the perturbation is observable, the natural object to study is the evolution of the energies while the dynamics of the Hamiltonian dynamics becomes too large. See [34, 4, 18].

If the state space of our dynamical system has an action by a group, the orbit manifold is a fundamental object. We use the projection to the orbit manifold as a conservation law and use it to separate the slow and the fast variables in the system. The slow variables lie naturally on a quotient manifold. In many examples we can further reduce this system of slow-fast stochastic differential equations (SDEs) to a product manifold $N\times G$, which we describe later by examples. From here we proceed to prove an averaging principle for the family of SDEs with a parameter ε. In these SDEs the slow and the fast variables are already separate, but they interact with each other.

This can then be applied to a local product space such as a principal bundle. In [65, 67, 66], the slow variables in the reduced system are random ODEs, where we study the system on the scale of $[1,\frac{1}{\varepsilon^2}]$ to obtain results of the nature of diffusion creation. In these studies we bypassed stochastic averaging and went straight for the diffusion creation. In [63, 49, 39] stochastic averaging are studied, but they are computed in local coordinates. Here the slow variables solve a genuine SDE with a stochastic integral and the computations are global. We first prove an averaging theorem for these SDEs and then study some examples where we deduce a slow-fast system of SDEs from non-linear conservation laws, to which our main theorems apply.

Throughout the article $(\Omega,\mathscr{F},\mathscr{F}_t,P)$ is a probability space satisfying the usual assumptions. Let (B_t,W_t) be a Brownian motion on $\mathbb{R}^{m_1}\times\mathbb{R}^{m_2}$ where $m_1,m_2\in\mathscr{N}$. We write $B_t=(B_t^1,\dots,B_t^{m_1})$ and $W_t=(W_t^1,\dots,W_t^{m_1})$. Let N and G be two complete connected smooth Riemannian manifolds, let $x_0\in N$ and $y_0\in G$. Let ε

denote a small positive number and let m_1 and m_2 be two natural numbers. Let $X : N\times G\times \mathbb{R}^{m_1} \to TN$ and $Y : N\times G\times \mathbb{R}^{m_2} \to TG$ be C^3 smooth maps linear in the last variable. Let X_0 and Y_0 be C^2 smooth vector fields on N and on G respectively, with a parameter taking its values in the other manifold. Let us consider the SDEs,

$$\begin{cases} dx_t^\varepsilon = X(x_t^\varepsilon, y_t^\varepsilon)\circ dB_t + X_0(x_t^\varepsilon, y_t^\varepsilon)\,dt, & x_0^\varepsilon = x_0,\\ dy_t^\varepsilon = \dfrac{1}{\sqrt{\varepsilon}} Y(x_t^\varepsilon, y_t^\varepsilon)\circ dW_t + \dfrac{1}{\varepsilon} Y_0(x_t^\varepsilon, y_t^\varepsilon)\,dt, & y_0^\varepsilon = y_0. \end{cases} \tag{1}$$

The symbol $\circ$ is used to denote Stratonovich integrals. By choosing an orthonormal basis $\{e_i\}$ of $\mathbb{R}^{m_1}\times\mathbb{R}^{m_2}$, we obtain a family of vector fields $\{X_1,\dots,X_{m_1},Y_1,\dots,Y_{m_2}\}$ as following: $X_i(x) = X(x)(e_i)$ for $1\le i\le m_1$ and $Y_i(x) = Y(x)(e_i)$ for $i = m_1+1,\dots,m_1+m_2$. Then the system of SDEs (1) is equivalent to the following

$$\begin{cases} dx_t^\varepsilon = \sum_{k=1}^{m_1} X_k(x_t^\varepsilon, y_t^\varepsilon)\circ dB_t^k + X_0(x_t^\varepsilon, y_t^\varepsilon)\,dt, & x_0^\varepsilon = x_0,\\ dy_t^\varepsilon = \dfrac{1}{\sqrt{\varepsilon}} \sum_{k=1}^{m_2} Y_k(x_t^\varepsilon, y_t^\varepsilon)\circ dW_t^k + \dfrac{1}{\varepsilon} Y_0(x_t^\varepsilon, y_t^\varepsilon)\,dt, & y_0^\varepsilon = y_0. \end{cases}$$

If V is a vector field, by Vf we mean $df(V)$ or $L_V f$, the Lie differential of f in the direction of V. Then $(x_t^\varepsilon, y_t^\varepsilon)$ is a sample continuous Markov process with generator $\mathscr{L}^\varepsilon := \frac{1}{\varepsilon}\mathscr{L} + \mathscr{L}^{(1)}$ where

$$\mathscr{L} = \frac{1}{2}\sum_{k=1}^{m_2} Y_i^2 + Y_0, \quad \mathscr{L}^{(1)} = \frac{1}{2}\sum_{k=1}^{m_1} X_i^2 + X_0.$$

In other words if $f : N\times G \to \mathbb{R}$ is a smooth function then

$$\mathscr{L}^\varepsilon f(x,y) := \frac{1}{\varepsilon}\mathscr{L}_x(f(\cdot,y))(x) + \mathscr{L}_y^{(1)}(f(x,\cdot))(y),$$

where

$$\mathscr{L}_x f(x,\cdot) = \left(\frac{1}{2}\sum_{k=1}^{m_2} Y_i^2(x,\cdot) + Y_0(x,\cdot)\right) f(x,\cdot),$$
$$\mathscr{L}_y^{(1)} f(\cdot,y) = \left(\frac{1}{2}\sum_{k=1}^{m_1} X_i^2(\cdot,y) + X_0(\cdot,y)\right) f(\cdot,y).$$

The result we seek is the weak convergence of the slow variables x_t^ε to a Markov process $\bar{x}_t$ whose Markov generator $\bar{\mathscr{L}}$ is to be described.

Let T be a positive number and let $C([0,T];N)$ denote the family continuous functions from $[0,T]$ to N, the topology on $C([0,T];N)$ is given by the uniform distance. A family of continuous stochastic processes x_t^ε on N is said to converge to a continuous process $\bar{x}_t$ if for every bounded continuous function $F : C([0,T];N) \to \mathbb{R}$, as ε converges to zero,

$$\mathbf{E}[F(x^\varepsilon_\cdot)] \to \mathbf{E}[F(\bar{x}_\cdot)].$$

In particular, if $u^\varepsilon(t,x,y)$ is a bounded regular solution to the Cauchy problem for the PDE (for example C^3 in space and C^1 in time) $\frac{\partial u^\varepsilon}{\partial t} = \mathscr{L}^\varepsilon u$ with the initial value f in L_∞, then $u^\varepsilon(t,x_0,y_0) = \mathbf{E}[f(x_t^\varepsilon, y_t^\varepsilon)]$. Suppose that the initial value function f is independent of the second variable so $f: N \to \mathbb{R}$. Then the weak convergence will imply that

$$\lim_{\varepsilon\to 0} u^\varepsilon(t,x_0,y_0) = u(t,x_0)$$

where $u(t,x)$ is the bounded regular solution to the Cauchy problem

$$\frac{\partial u}{\partial t} = \bar{\mathscr{L}} u, \quad u(0,x) = f(x).$$

Stochastic averaging is a procedure of equating time averages with space averages using a form of Birkhoff's ergodic theorem or a law of large numbers. Birkhoff's pointwise ergodic theorem states that if $T: E \to E$ is a measurable transformation preserving a probability measure μ on the metric space E then for any $F \in L^1(\mu)$,

$$\frac{1}{n}\sum_{k=1}^n F(T^k x) \to \mathbf{E}(F|\mathscr{I})$$

for almost surely all x, as $n \to \infty$, and where $\mathscr{I}$ is the invariant σ-algebra of T. Suppose that (z_t) is a sample continuous ergodic stochastic process with values in E, stationary on the space of paths $C([0,1];E)$. Denote by μ its one time probability distribution. Then for any real valued function $f \in L^1(\mu)$,

$$\frac{1}{t}\int_0^t f(z_s)ds \to \int f(z)\mu(dz).$$

This is simply Birkhoff's theorem applied to the shift operator and to the function $F(\omega) = \int_0^1 f(z_s(\omega))ds$. If z_t is not stationary, but a Markov process with initial value a point, conditions are needed to ensure the convergence of the Markov process to equilibrium with sufficient speed.

We explain below stochastic averaging for a random field whose randomness is introduced by a fast diffusion. Let $(x_t^\varepsilon, y_t^\varepsilon)$ be solution to the SDE on $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$:

$$dx_t^\varepsilon = \sum_{k=1}^{m_1} \sigma_k(x_t^\varepsilon, y_t^\varepsilon)\, dB_t^k + b(x_t^\varepsilon, y_t^\varepsilon)\, dt, \quad dy_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2} \theta_k(x_t^\varepsilon, y_t^\varepsilon) dW_t^k + \frac{1}{\varepsilon} b(x_t^\varepsilon, y_t^\varepsilon)\, dt.$$

with initial values $x_0^\varepsilon = x_0$, and $y_0^\varepsilon = y_0$. Here the stochastic integrations are Itô integrals. A sample averaging theorem is as following. Let z_t^x denote the solution to the SDE

$$dz_t^x = \sum_{k=1}^{m_2} \theta_k(x, z_t^x) dW_t^k + b(x, z_t^x)\, dt$$

with initial value z_0^x. Suppose that the coefficients are globally Lipschitz continuous and bounded. Suppose that $\sup_{t\in[0,T]}\sup_{\varepsilon\in(0,1]}\mathbf{E}|y_t^\varepsilon|^2$ and $\sup_x \sup_{t\in[0,T]}\mathbf{E}|z_t^x|^2$ are finite. Also suppose that there exist functions $\bar a_{i,j}$ and $\bar b$ such that

$$\begin{aligned} \left|\frac{1}{t}\mathbf{E}\int_0^t b(x,z_s^x)ds-\bar b(x)\right| &\le C(t)(|x|^2+|z|^2+1),\\ \left|\frac{1}{t}\mathbf{E}\int_0^t \sum_k \sigma_k^i\sigma_k^j(x,z_s^x)ds-\bar a_{i,j}(x)\right| &\le C(t)(|x|^2+|z|^2+1). \end{aligned} \tag{2}$$

Then the stochastic processes x_t^ε converge weakly to a Markov process with generator $\frac{1}{2}\bar a_{i,j}\frac{\partial^2}{\partial x_i\partial x_j}+\bar b_k\frac{\partial^2}{\partial x_k}$, see [81, 82, 47, 17, 90]. See also [74, 60, 54, 48, 76]. See [46, 8, 38, 37, 24, 68, 78, 27, 18, 36, 43, 20, 52, 41] for a range of more recent related work. We also refer to the following books [56, 77, 80, 59, 12]

Averaging of stochastic differential equations on manifolds has been studied in the following articles [53], [63], [64], and [39]. In these studies either one restricts to local coordinates, or has a set of convenient coordinates, or one works directly with local coordinates. We will be using a global approach.

We will first deduce a quantitative locally uniform Birkhoff's ergodic theorem for $\mathscr{L}_x$, then prove an averaging theorem for (1). Finally we study a number of examples of singular perturbation problems.

The main assumptions on $\mathscr{L}_x$ is a Hörmander's (bracket) condition.

Definition 1. Let $X_0, X_1, \dots, X_k$ be smooth vector fields.

1. The differential operator $\sum_{k=1}^m (X_i)^2+X_0$ is said to satisfy *Hörmander's condition* if $\{X_k, k=0,1,\dots,m\}$ and their iterated Lie brackets generate the tangent space at each point.
2. The differential operator $\sum_{k=1}^m (X_i)^2+X_0$, is said to satisfy *strong Hörmander's condition* if $\{X_k, k=1,\dots,m\}$ and their iterated Lie brackets generate the tangent space at each point.

Outline of the paper. In §3 we study the regularity of invariant probability measures μ^x of $\mathscr{L}_x$ with respect to the parameter x and prove the local uniform law of large numbers with rate. We may assume that each $\mathscr{L}_x$ satisfies Hörmander's condition. What we really need is that $\mathscr{L}_X$ is a family of Fredholm operators satisfying the sub-elliptic estimates and with zero Fredholm index. In §4 we give estimates for SDEs on manifolds. It is worth noticing that we do not assume that the transition probabilities have densities. We use an approximating family of distance functions to overcome the problem that the distance function is not smooth. These estimates lead easily to the tightness of the slow variables. In §5 we prove the convergence of the slow variables, for which we first prove a theorem on time averaging of path integrals of the slow variables. This is proved under a law of large numbers with any uniform rate. In §2 we study some examples of singular perturbation problems. Finally, we pose a number of open questions, one of which is presented in the next section, the others are presented in §2.

1.1 Description of results.

The following law of large numbers with a locally uniform rate is proved is section 3.

Theorem 1 (Quantitative Locally Uniform Law of Large Numbers). *Let G be a compact manifold. Suppose that Y_i are bounded, C^∞ with bounded derivatives. Suppose that each*

$$\mathscr{L}_x = \frac{1}{2}\sum_{i=1}^m Y_i^2(x,\cdot) + Y_0(x,\cdot)$$

satisfies Hörmander's condition (Def. 1), and has a unique invariant probability measure μ_x. Then the following statements hold for μ_x.

(a) $x \mapsto \mu_x$ is locally Lipschitz continuous in the total variation norm.
(b) For every $s > 1 + \frac{\dim(G)}{2}$ there exists a positive constant $C(x)$, depending continuously in x, such that for every smooth function $f : G \to \mathbb{R}$,

$$\left| \frac{1}{T}\int_t^{t+T} f(z_r^x)\, dr - \int_G f(y)\mu_x(dy) \right|_{L_2(\Omega)} \le C(x)\|f\|_s \frac{1}{\sqrt{T}}, \tag{3}$$

where z_r denotes an $\mathscr{L}_x$-diffusion.

Remark 1. Let $P^x(t,y,\cdot)$ denote the transition probability of $\mathscr{L}_x$. Suppose $\mathscr{L}_x$ satisfies Doeblin's condition. Then $\mathscr{L}_x$ has a unique invariant probability measure. This holds in particular if $\mathscr{L}_x$ satisfies the strong Hörmander's condition and G is compact. The uniqueness follows from the fact that it has a smooth strictly positive density. (Hörmander's condition ensures that any invariant measure has a smooth kernel and the kernel of its L_2 adjoint $\mathscr{L}^*$ contains a non-negative function. The density is however not necessarily positive.) Suppose that each $\mathscr{L}_x$ satisfies the strong Hörmander's condition (c.f. Def. 1) and G is compact. It is well know that the transition probability measures $P^x(t,y_0,\cdot)$, with any initial value y_0, converges to the unique invariant probability measure μ^x with an exponential rate which we denote by $C(x)e^{\gamma(x)t}$. If x takes values also in a compact space N, the exponential rate and the constant in front of the exponential rate can be taken to be independent of x. When N is non-compact, we obviously need to make further assumptions on $\mathscr{L}_x$ for a uniform estimate. There have been work on ergodicity of this type. We refer to : [23, 10, 91, 70],

Set

$$\tilde{X}_0(\cdot,y) = \frac{1}{2}\sum_{i=1}^{m_1} \nabla X_i(X_i)(y,\cdot) + X_0(y,\cdot),$$
$$\tilde{Y}_0(x,\cdot) = \frac{1}{2}\sum_{i=1}^{m_2} \nabla Y_i(Y_i)(x,\cdot) + Y_0(x,\cdot).$$

Let O be a reference point in N and ρ is the Riemannian distance from O.

Assumption 1 (Assumptions on X_i and N) *Suppose that $\tilde{X}_0$ and X_i are C^1, where $i = 1,\dots,m$. Suppose that* **one** *of the following two statements holds.*

(i) The sectional curvature of N is bounded. There exists a constant K such that

$$\sum_{i=1}^m |X_i(x,y)|^2 \le K(1+\rho(x)), \quad |X_0(x,y)| \le K(1+\rho(x)), \quad \forall x \in N, \forall y \in G.$$

(ii) Suppose that the square of the distance function on N is smooth. Suppose that

$$\frac{1}{2}\sum_{i=1}^m \nabla d\rho^2(X_i(\cdot,y),X_i(\cdot,y)) + d\rho^2(\tilde{X}_0(\cdot,y)) \le K + K\rho^2(\cdot), \quad \forall y \in G.$$

Assumption 2 (Assumptions on Y_i and G) *We suppose that G has bounded sectional curvature. Suppose that $\tilde{Y}_0$ and Y_j are C^2 and bounded with bounded first order derivatives.*

The following is extracted from Theorem 5.6.

Theorem 2 (Averaging Theorem). *Suppose that there exists a family of invariant probability measure μ_x on G that satisfies the conclusions of Theorem 1. Suppose the assumptions on X_i, N, Y_i and G hold (Assumptions 1 and 2). Then as $\varepsilon \to 0$, the stochastic processes x_t^ε converges weakly on $C([0,T],N)$ to a Markov process with generator $\bar{\mathscr{L}}$.*

Remark 2. (i) If f is a smooth function on N with compact support then

$$\bar{\mathscr{L}} f(x) = \int_G \left(\frac{1}{2}\sum_{i=1}^{m_1} X_i^2(\cdot,y) f + X_0(\cdot,y) f \right)(x)\, \mu_x(dy). \tag{4}$$

See the Appendix in §5 for a sum of squares of vector fields decomposition of $\bar{\mathscr{L}}$.

(ii) Under Assumptions 1 there exists a unique global solution x_t^ε for each initial value (x,y). We also have uniform estimates on the distance $\rho(x_s^\varepsilon, x_t^\varepsilon)$ which leads to the conclusion that the family $\{x_\cdot^\varepsilon, \varepsilon > 0\}$ is tight. Also we may conclude that the moments of the solutions are bounded uniformly on any compact time interval and in ε for $\varepsilon \in (0,1]$. Such estimates are given in §4.

(iii) Under Assumptions 1- 2 we may approximate the fast motion, on sub-intervals $[t_i, t_{i+1}]$, by freezing the slow variables and obtain a family of Markov processes with generator $\mathscr{L}_x$. The size of the sub-intervals must be of size $o(\varepsilon)$ for the error of the approximation to converge to zero as $\varepsilon \to 0$, and large on the scale of $\frac{1}{\varepsilon}$ for the ergodic average to take effect.

Problem 1. Suppose that $\mathscr{L}_x$ satisfies Hörmander's condition. Then the kernel of $\mathscr{L}_x^*$ is finite dimensional. Without assuming the uniqueness of the invariant probability measures, it is possible to define a projection to the kernel of $\mathscr{L}_x$, by pairing up a basis $\{u_i(x)\}$ of $\ker(\mathscr{L}_x)$ with a dual basis $\pi^i(x)$ of $\ker(\mathscr{L}_x^*)$ and this leads to a family of projection operators $\Pi(x)$. To obtain a locally uniform version of this, we should consider the continuity of Π with respect to x. Let us consider the simple case of a family of Fredhom operators $T(x)$ from a Hilbert space E to a Hilbert space F. It is well known that the dimension of their kernels may not be a continuous function of x, but the Fredholm index of $T(x)$ is a continuous function if x in the space of bounded linear operators [5]. See also [88, 89] for non-elliptic operators. Given that the projection $\pi(x)$ involves both the kernel and the co-kernel, it is reasonable to expect that $\Pi(x)$ is continuous in x. The question is whether this is true and more importantly whether in this situation there is a local uniform Law of large numbers.

2 Examples

We describe some motivating examples, the first being dynamical descriptions for Brownian motions, the second being the convergence of metric spaces. The overarching question concerning the second is: given a family of metric spaces converging to another in measured Gromov Hausdorff topology, can we give a good dynamical description for their convergence? What can one say about the associated singular operators? These will considered in terms of stochastic dynamics. See [51, 73] and [64] concerning collapsing of Riemannian manifolds. The third example is a model on the principal bundle. These singular perturbation models were introduced in [63, 65, 67, 64], where the perturbations were chosen carefully for diffusion creation. The reduced systems are random ODEs for which a set of limit theorems are available, and the perturbations are chosen so that one could bypass the stochastic averaging procedure and work directly on the faster scale for diffusion creation $[0, \frac{1}{\varepsilon}]$. Theorem 5.6 allows us to revisit these models to include more general perturbations, in which the effective limits on $[0,1]$ are not trivial. It also highlights from a different angle the choice of the perturbation vector in the models which we explain below.

2.1 *A dynamical description for Brownian motions*

In 1905, Einstein, while working on his atomic theory, proposed the diffusion model for describing the density of the probability for finding a particle at time t in a position x. A similar model was proposed by Smoluchowski with a force field. Some years later Langevin (1908) [62] and Ornstein-Uhlenbeck (1930) [86] proposed a dynamical model for Brownian motion for time larger than the relaxation time $\frac{1}{\beta}$:

$$\begin{cases} \dot{x}(t) = v(t) \\ \dot{v}(t) = -\beta v(t)dt + \sqrt{D}\beta dB_t + \beta b(x(t))dt \end{cases}$$

where B_t a one dimensional Brownian motion and b a vector field. This equation is stated for $\mathbb{R}$ with β, D constants and was studied by Kramers [58] and Nelson [72].

The model is on the real line, there exists only one direction for the velocity field. The magnitude of $v(t)$ together with the sign changes rapidly.

The second order differential equations for unit speed geodesics, on a manifold M, are equivalent to first order ODEs on the space of orthonormal frames of M, this space will be denoted by OM. Suppose that we rotate the direction of the geodesic uniformly, according to the probability distribution of a Brownian motion on $SO(n)$, while keeping its magnitude fixed to be 1, and suppose that the rotation is at the scale of $\frac{1}{\varepsilon}$ then the projections to M of the solutions of the equations on OM converge to a fixed point as $\varepsilon \to 0$. But if we further tune up the speed of the rotation, these motions converge to a scaled Brownian motion, whose scale is given by an eigenvalue of the fast motion on $SO(n)$. See [65]. An extension to manifolds was first studied [25] followed by [16]. That in [25] is different from that in Li-geodesic, which is also followed up in [2] where the authors iremoved the geometric curvature restrictions in [65]. See also [14] for a local coordinate approach and more recently [26]. Assume the dimension of M is greater than 1. The equations, [65], describing this are as following:

$$\begin{cases} du_t^\varepsilon = H_{u_t^\varepsilon}(e_0)\,dt + \dfrac{1}{\sqrt{\varepsilon}} \displaystyle\sum_{k=1}^N A_k^*(u_t^\varepsilon) \circ dW_t^k + A_0^*(u_t^\varepsilon)\,dt, \\ u_0^\varepsilon = u_0. \end{cases} \tag{1}$$

where $\{A_1, \dots, A_N\}$ is an o.n.b. of $\mathfrak{so}(n)$, and $A_0 \in \mathfrak{so}(n)$. The star sign denotes the corresponding vertical fundamental vector fields and $H(u)(e_0)$ is the horizontal vector field corresponding to a unit vector e_0 in $\mathbb{R}^n$. This following theorem is taken from [65].

Theorem 2A. The position part of $u^\varepsilon_{\frac{t}{\varepsilon}}$, which we denote by $(x^\varepsilon_{\frac{t}{\varepsilon}})$, converges to a Brownian motion on M with generator $\frac{4}{n(n-1)}\Delta$. Furthermore the parallel translations along these smooth paths of $(x^\varepsilon_{\frac{t}{\varepsilon}})$ converge to stochastic parallel translations along the Hörder continuous sample paths of the effective scaled Brownian motion.

The conservation law in this case is the projection π, taking a frame to its base point, using which we obtain the following reduced system of slow-fast SDEs:

$$\begin{cases} \dfrac{d}{dt}\tilde{x}_t^\varepsilon = H_{\tilde{x}_t^\varepsilon}(g_{\frac{t}{\varepsilon}} e_0), \quad \tilde{x}_0^\varepsilon = u_0, \\ dg_t = \displaystyle\sum_{k=1}^m g_t A_k \circ dw_t^k + g_t A_0\,dt, \quad g_0 = Id. \end{cases}$$

The slow variable does not have a stochastic part, the averaging equation is given by the average vector field $\int_{SO(n)} H(ug)(e)dg$, where dg is the Haar measure, and vanishes. Hence we may observe the slow variable on a faster scale and consider $x^\varepsilon_{\frac{t}{\varepsilon}}$.

In section 6.1 we use the general results obtained later to study two generalised models.

2.2 Collapsing of manifolds

Our overarching question is how the stochastic dynamics describe the convergence of metric spaces. Let us consider a simple example: $SU(2)$ which can be identified with the sphere S^3. The Lie algebra of $SU(2)$ is given by the Pauli matrices

$$X_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

By declaring $\{\frac{1}{\sqrt{\varepsilon}}X_1, X_2, X_3\}$ an orthonormal frame we define Berger's metrics g^ε. Thus (S^3, g^ε) converges to S^2. They are the first known family of manifolds which collapse to a lower dimensional one, while keeping the sectional curvatures uniformly bounded (J. Cheeger). Then all the operators in the sum

$$\Delta^\varepsilon_{S^3} = \frac{1}{\varepsilon}(X_1)^2 + (X_2)^2 + (X_3)^2 = \frac{1}{\varepsilon}\Delta_{S^1} + \Delta_H$$

commute, the eigenvalues satisfy the relation $\lambda_3(\Delta^\varepsilon_{S^3}) = \frac{1}{\varepsilon}\lambda_1(\Delta_{S^1}) + \lambda_2(\Delta_H)$. The non-zero eigenvalues of Δ_{S^1} flies away and the eigenfunctions of $\lambda_1 = 0$ are function on the sphere $S^2(\frac{1}{2})$ of radius $\frac{1}{2}$, the convergence of the spectrum of $\Delta^\varepsilon_{S^3}$ follows. See [85], [11] [87] for discussions on the spectrum of Laplacians on spheres, on homogeneous Riemannian manifolds and on Riemannian submersions with totally geodesic fibres.

We study

$$\mathscr{L}^\varepsilon := \frac{1}{\varepsilon}\Delta_{S^1} + Y_0$$

in which Δ_{S^1} and Y_0 do not commute. Take for example, $Y_0 = aX_2 + bX_3$ where $|Y_0| = 1$. Let $\pi(z,w) = \frac{1}{2}(|w|^2 - |z|^2, z\bar{w})$ be the Hopf map. Let u^ε_t be an $\mathscr{L}^\varepsilon$-diffusion with the initial value u_0. Then $\pi(u^\varepsilon_{\frac{t}{\varepsilon}})$ converges to a BM on $S^2(\frac{1}{2})$, scaled by $\frac{1}{2}$. See [67]. See also [73] for related studies. It is perhaps interesting to observe that $\mathscr{L}^\varepsilon$ satisfies Hörmander's condition for any $Y_0 \neq 0$. Later we see that this fact is not an essential feature of the problem. The model on S^3 in [64] is a variation of this one.

2.3 Inhomogeneous scaling of Riemannian metrics

If a manifold is given a family of Riemannian metrics depending on a small parameter $\varepsilon > 0$, the Laplacian operators Δ^ε is a family of singularly perturbed operators. We might ask the question whether their spectra converge. More generally let us consider a family of second order differential operators $\mathscr{L}^\varepsilon = \frac{1}{\varepsilon}\mathscr{L}_0 + \mathscr{L}_1$, each in the form of a finite sum of squares of smooth vector fields with possibly a first order term. As $\varepsilon \to 0$, the corresponding Markov process does not converge in general. In the spirit of Noether's theorem, to see a convergent slow component we expect to see some symmetries for the system $\mathscr{L}_0$. On the other hand, by a theorem of S. B.

Myers and N. E. Steenrod [71], the set of all isometries of a Riemannian manifold M is a Lie group under composition of maps, and furthermore the isotropy subgroup $\mathrm{Iso}_o(M)$ is compact. See also S. Kobayashi and K. Nomizu [55]. We are led to study homogeneous manifolds G/H, where G is a smooth Lie group and H is a compact sub-group of G. We denote by $\mathfrak{g}$ and $\mathfrak{h}$ their respective Lie algebras.

Let $\mathfrak{g}$ be endowed an $\mathrm{Ad}(H)$-invariant inner product and take $\mathfrak{m} = \mathfrak{h}^\perp$. Then G/H is a reductive homogeneous manifold, in the sense of Nomizu, by which we mean $\mathrm{Ad}(H)(\mathfrak{m}) \subset \mathfrak{m}$. This is a different from the concept of a reductive Lie group, where the adjoint representation of the Lie group G is completely reducible. (Bismut studied a natural deformation of the standard Laplacian on a compact Lie group G into a hypoelliptic operator on TG see [15].) We assume that the real Lie group G is smooth, connected, not necessarily compact, of dimension n and H a closed connected proper subgroup of dimension at least one. We identify elements of the Lie algebra with left invariant vector fields.

We generate a family of Riemannian metrics on G by scaling the $\mathfrak{h}$ directions by ε. Let $\{A_1,\dots,A_p,Y_{p+1},\dots,X_N\}$ be an orthonormal basis of $\mathfrak{g}$ for an inner product extending an orthonormal basis $\{A_1,\dots,A_p\}$ of $\mathfrak{h}$ with the remaining vectors from $\mathfrak{m}$. By declaring

$$\left\{\frac{1}{\sqrt{\varepsilon}}A_1,\dots,\frac{1}{\sqrt{\varepsilon}}A_p,Y_{p+1},\dots Y_N\right\}$$

an orthonormal frame, we obtain a family of left invariant Riemannian metrics. Let us consider the following second order differential operator, related to the re-scaled metric:

$$\mathscr{L}^\varepsilon = \frac{1}{2\varepsilon}\sum_{k=1}^{m_2}(A_k)^2 + \frac{1}{\varepsilon}A_0 + Y_0,$$

where $A_k \subset \mathfrak{h}$ and $Y_0 \in \mathfrak{m}$ is a unit vector. This leads to the following family of equations, where $\varepsilon \in (0,1]$,

$$dg_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2}A_k(g_t^\varepsilon)\circ dB_t^k + \frac{1}{\varepsilon}A_0(g_0^\varepsilon)dt + Y_0(g_t^\varepsilon)dt, \quad g_0^\varepsilon = g_0.$$

These SDEs belong to the following family of equations

$$dg_t = \sum_{k=1}^{m_2}\gamma A_k(g_t)\circ dB_t^k + \gamma A_0(g_0)\,dt + \delta Y_0(g_t)dt.$$

The solutions of the latter family of equations, with parameters γ and δ real numbers, interpolate between translates of a one parameter subgroups of G and diffusions on H. Our study of $\mathscr{L}^\varepsilon$ is related to the concept o 'taking the adiabatic limit' [13, 69].

Let (g_t^ε) be a Markov processes with Markov generator $\mathscr{L}^\varepsilon$, and set $x_t^\varepsilon = \pi(g_t^\varepsilon)$ where π is the map taking an element of G to the coset gH. Then $\mathscr{L}^\varepsilon = \frac{1}{\varepsilon}\mathscr{L}_0 + Y_0$ where $\mathscr{L}_0 = \frac{1}{2}\sum_{k=1}^{m_2}(A_k)^2 + A_0$. We will assume that $\{A_k\} \subset \mathfrak{h}$ are bracket generating. Scaled by $1/\varepsilon$, the Markov generator of $(g_{\frac{t}{\varepsilon}}^\varepsilon)$ is precisely $\frac{1}{\varepsilon}\mathscr{L}^\varepsilon$.

The operators $\mathscr{L}^\varepsilon$ are not necessarily hypo-elliptic in G, and they will not expected to converge in the standard sense. Our first task is to understand the nature of the perturbation and to extract from them a family of first order random differential operators, $\tilde{\mathscr{L}}^\varepsilon$, which converge and which have the same orbits as $\mathscr{L}^\varepsilon$, the 'slow motions'. The reduced operators, $\frac{1}{\varepsilon}\tilde{\mathscr{L}}^\varepsilon$, describe the motion of the orbits under 'perturbation'.

Their effective limit is either a one parameter sub-groups of G in which case our study terminate, or a fixed point in which case we study the fluctuation dynamics on the time scale $[0, \frac{1}{\varepsilon}]$. On the Riemmanian homogeneous manifold, if G is compact, the effect limit on G is a geodesic at level one and a fixed point at level two. On the scale of $[0, \frac{1}{\varepsilon}]$ we would consider $\frac{1}{\varepsilon}\mathscr{L}_0$ as perturbation. It is counter intuitive to consider the dominate part as the perturbation. But the perturbation, although very large in magnitude, is fast oscillating. The large oscillating motion get averaged out, leaving an effective motion corresponding to a second order differential operator on G.

This problems breaks into three parts: separate the slow and the fast variable, which depends on the principal bundle structure of the homogeneous space, and determine the natural scales; the convergence of the solutions of the reduced equations which is a family of random ODEs; finally the buck of the interesting study is to determine the effective limit, answering the question whether it solves an autonomous equation.

It is fairly easy to see that x_t^ε moves relatively slowly. The speed at which x_t^ε crosses M is expected to depend on the specific vector Y_0, however in the case of $\{A_1, \dots, A_p\}$ is an o.n.b. of $\mathfrak{h}$ and $A_0 = 0$, they depend only on the $\mathrm{Ad}(H)$-invariant component of Y_0.

The separation of slow and fast variables are achieved by first projecting the motion down to G/H and then horizontally lift the paths back (a non-Markovian procedure), exposing the action in the fibre directions. The horizontal process thus obtained is the 'slow part' of g_t^ε and will be denoted by u_t^ε. It is easy to see that the reduced dynamic is given by

$$\dot{u}_t^\varepsilon = \mathrm{Ad}(h_{\frac{t}{\varepsilon}})(Y_0)(u_t^\varepsilon).$$

where h_t has generator $\frac{1}{2}\sum(A_i)^2 + A_0$.

If $\{A_0, A_1, \dots, A_m\}$ generates the vector space $\mathfrak{h}$, the differential operator $\frac{1}{2}\sum(A_i)^2 + A_0$ satisfies Hörmander's condition in which case the invariant probability measure is the normalised Haar measure. Then u_t^ε converges to the solution of the ODE:

$$\frac{d}{dt}\bar{u}_t = \int_H \mathrm{Ad}(h)(Y_0)dh.$$

Let us take an $\mathrm{Ad}(H)$ invariant decomposition of $\mathfrak{m}$, $\mathfrak{m} = \mathfrak{m}_0 + \mathfrak{m}_1$ where $\mathfrak{m}_0$ is the vector space of invariant vectors and $\mathfrak{m}'$ is its orthogonal complement. Then

$$\int_H \mathrm{Ad}(h)(Y_0)dh = Y_0{}^{\mathfrak{m}_0}$$

where the superscript $\mathfrak{m}_0$ denote the $\mathfrak{m}_0$ component of Y. This means that the dynamics is a fixed point if and only if $Y_0^{\mathfrak{m}_0} = 0$.

In [67] we take $Y_0^{\mathfrak{m}_0} = 0$ and answered this question by a multi-scale analysis and studied directly the question concerning $Y_0 \in \tilde{\mathfrak{m}}$, without having to go through stochastic averaging. Theorem 1 makes this procedure easier to understand. Then we consider the dynamics on $[0, \frac{1}{\varepsilon}]$. The reduced first order random differential operators give rise to second order differential operators by the action of the Lie bracket.

2.4 Perturbed dynamical systems on Principal bundles

In the examples described earlier, we have a perturbed dynamical system on a manifold P. On P there is an action by a Lie group G, and the projection to $M = P/G$ is a conservation law. We then study the convergence of the slow motion, the projection to M, and their horizontal lifts. More precisely we have a principal bundle with fibre the Lie group G. To describe these motions we consider the kernels of the differential of the projection π: they are called the vertical tangent spaces and will be denoted by VT_uP. Any vector field taking values in the vertical tangent space is called a vertical vector field, the Lie-bracket of any two vertical vector fields is vertical. A smooth choice of the complements of the vertical spaces, that are right invariant, determines a connection. These complements are called the horizontal spaces. The ensemble is denoted by HT_uP and called the horizontal bundle. From now on we assume that we have chosen such a horizontal space. A vector field taking values in the horizon tangent spaces is said to be a horizontal vector field. Right invariant horizontal vector fields are specially interesting, they are precisely the horizontal lifts of vector fields on M.

Let $\pi : P \to M$ denote the canonical projection taking an element of the total space P to the corresponding element of the base manifold. Also let $R_g : P \to P$ denote the right action by g, for simplicity we also write ug, where $u \in P$, for $R_g u$. A connection on a principal bundle P is a splitting of the tangent bundle $T_uP = HT_uP + VT_uP$ where VT_uP is the kernel of the differential of tπ. Let $\mathfrak{g}$ denote the Lie algebra of G. For any $A \in \mathfrak{g}$ we define

$$A^*(u) = \lim_{t\to 0} R_{\exp(tA)} u.$$

The splitting mentioned earlier is in one to one correspondence with a connection 1-form, by which we mean a map $\varpi : T_uP \to \mathfrak{g}$ with the following properties:

$$(R_g)^*\varpi = \mathrm{ad}(g^{-1})\varpi, \quad \varpi(A^*) \equiv A.$$

This splitting also determines a horizontal lifting map $\mathfrak{h}_u$ at $u \in P$ and a family of horizontal vector fields H_i. If $\{e_1, \dots e_n\}$ is an orthonormal basis of $\mathbb{R}^n$, where $n = \dim(M)$, we set $H_i(u) = \mathfrak{h}_u(ue_i)$. If $\{A_1, \dots, A_N\}$ is an orthonormal basis of the

Lie algebra $\mathfrak{g}$, then at every point u, $\{H_1(u),\dots,H_n(u),A_1^*(u),\dots,A_N^*(u)\}$ is a basis of T_uP. We give P the Riemannian metric so that the basis is orthonormal.

Any stochastic differential equation (SDE) on P are of the following form, where β and γ are two real positive numbers and σ_j^k and θ_j^k are BC^3 functions on P.

$$\begin{aligned} du_t =&\beta \sum_{k=1}^{m_1}\left(\sum_{i=1}^n \sigma_k^i(u_t)H_i(u_t)\right)\circ dB_t^k + \beta^2\sum_{i=1}^n \sigma_0^i(u_t)H_i(u_t)dt\\ &+\gamma^2\sum_{k=1}^{m_2}\left(\sum_{j=1}^N \theta_k^j(u_t)A_j(u_t)\right)\circ dW_t^k+\gamma\sum_{j=1}^N \theta_0^j(u_t)A_j(u_t)dt. \end{aligned}$$

Set $X_k=\sum_{i=1}^n \sigma_k^i H_i$, and $Y_k=\sum_{j=1}^N \theta_k^j A_j$. Then the equation is of the form

$$du_t=\beta\sum_{k=1}^{m_1}X_k(u_t)\circ dB_t^k+\beta^2\, X_0(u_t)\, dt+\gamma\sum_{k=1}^{m_2}Y_k\circ dW_t^k+\gamma^2\, Y_0(u_t)dt.$$

The solutions are Markov processes with Markov generator

$$\beta^2\left(\sum_{k=1}^{m_1}(X_k)^2+X_0\right)+\gamma^2\left(\sum_{k=1}^{m_2}(Y_k)^2+Y_0\right).$$

We observe that the projection of the second factor vanishes, so if $\beta=0$, then $\pi(u_t)=\pi(u_0)$ and π is a conservation law. The equation with small β is a stochastic dynamic whose orbits deviate slightly from that of the initial value u_0. If on the other hand, X_i are vector fields invariant under the action of the group, and $\gamma=0$ then the projection $\pi(u_t)$ is an autonomous SDE on the manifold M.

Let us take $\beta=1$ and $\gamma=\frac{1}{\sqrt{\varepsilon}}$.

$$\begin{cases} du_t^\varepsilon=\displaystyle\sum_{k=1}^{m_1}\left(\sum_{i=1}^n \sigma_k^i(u_t^\varepsilon)H_i(u_t^\varepsilon)\right)\circ dB_t^k+\sum_{i=1}^n\sigma_0^i(u_t^\varepsilon)H_i(u_t^\varepsilon)\, dt\\ \qquad +\displaystyle\frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2}\left(\sum_{j=1}^N \theta_k^j(u_t^\varepsilon)A_j^*(u_t^\varepsilon)\right)\circ dW_t^k+\frac{1}{\varepsilon}\sum_{j=1}^N\theta_0^j(u_t^\varepsilon)A_j^*(u_t^\varepsilon)dt,\\ u_0^\varepsilon=u_0. \end{cases}$$

We proceed to compute the equations for the slow and for the fast variables. Let $x_t^\varepsilon=\pi(u_t^\varepsilon)$. Then x_t^ε has a horizontal lift, see e.g. [32, 3, 29]. See also [30] and [31]. Let TR_g denote the differential of R_g. For $k=0,1,\dots,m_1$, set

$$X_k(ug)=\sum_{i=1}^p \sigma_k^i(ug)\,TR_{g^{-1}}H_i(ug).$$

Below we deduce an equation for x_t^ε which is typically not autonomous.

Lemma 2.1 *The horizontal lift processes satisfy the following system of slow-fast SDE's:*

$$d\tilde{x}_t^\varepsilon = \sum_{k=1}^{m_1} X_k(\tilde{x}_t^\varepsilon g_t^\varepsilon)\circ dB_t^k + X_0(\tilde{x}_t^\varepsilon g_t^\varepsilon)\,dt, \quad \tilde{x}_0^\varepsilon = g_0$$
$$dg_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2}\left(\sum_{j=1}^N \theta_k^j(\tilde{x}_t^\varepsilon g_t^\varepsilon)A_j^*(g_t^\varepsilon)\right)\circ dW_t^k + \frac{1}{\varepsilon}\sum_{j=1}^N \theta_0^j(\tilde{x}_t^\varepsilon g_t^\varepsilon)A_j^*(g_t^\varepsilon)\,dt,\ g_0^\varepsilon = id. \tag{2}$$

Proof. Since $\tilde{x}_t^\varepsilon$ and u_t^ε belong to the same fibre we may define $g_t^\varepsilon \in G$ by $u_t^\varepsilon = \tilde{x}_t^\varepsilon g_t^\varepsilon$. If a_t is a C^1 curve in the lie group G

$$\frac{d}{dt}|_t u a_t = \frac{d}{dr}_{|_{r=0}} u a_t a_t^{-1} a_{r+t} = (a_t^{-1}\dot{a}_t)^*(u a_t).$$

It follows that

$$du_t^\varepsilon = TR_{g_t^\varepsilon} d\tilde{x}_t^\varepsilon + (TL_{(g_t^\varepsilon)^{-1}} dg_t^\varepsilon)^*(u_t^\varepsilon).$$

Since right translations of horizontal vectors are horizontal,

$$TL_{(g_t^\varepsilon)^{-1}} dg_t^\varepsilon = \varpi(du_t^\varepsilon) = \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_1}\left(\sum_{j=1}^N \theta_k^j(u_t^\varepsilon)A_j\right)\circ dW_t^k + \frac{1}{\varepsilon}\sum_{j=1}^N \theta_0^j(u_t^\varepsilon)A_j\,dt$$

Hence, denoting by A^* also the left invariant vector fields on G, we have an equation for g_t^ε:

$$dg_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2}\left(\sum_{j=1}^N \theta_k^j(u_t^\varepsilon)A_j^*(g_t^\varepsilon)\right)\circ dW_t^k + \frac{1}{\varepsilon}\sum_{j=1}^N \theta_0^j(u_t^\varepsilon)A_j^*(g_t^\varepsilon)\,dt.$$

Since $\pi_*(A_j) = 0$ and by the definition of H_i we also have,

$$dx_t^\varepsilon = \sum_{k=1}^{m_1}\left(\sum_{i=1}^p \sigma_k^i(u_t^\varepsilon)(u_t^\varepsilon e_i)\right)\circ dB_t^k + \sum_{i=1}^n \sigma_0^i(u_t^\varepsilon)(u_t^\varepsilon e_i)\,dt.$$

Its horizontal lift is given by $d\tilde{x}_t = \mathfrak{h}_{\tilde{x}_t}(\circ dx_t^\varepsilon)$ and so we have the following SDE

$$d\tilde{x}_t^\varepsilon = \sum_{k=1}^{m_1}\left(\sum_{i=1}^p \sigma_k^i(u_t^\varepsilon)\,\mathfrak{h}_{\tilde{x}_t^\varepsilon}(u_t^\varepsilon e_i)\right)\circ dB_t^k + \sum_{i=1}^n \sigma_0^i(u_t^\varepsilon)\mathfrak{h}_{\tilde{x}_t^\varepsilon}(u_t^\varepsilon e_i)\,dt.$$

Since $\mathfrak{h}_u(uge_i) = TR_{g^{-1}}\mathfrak{h}_{ug}(uge_i) = TR_{g^{-1}}H_i(ug)$, we may rewrite the above equation in the following more convenient form:

$$d\tilde{x}_t^\varepsilon = \sum_{k=1}^{m_1}\left(\sum_{i=1}^p \sigma_k^i(\tilde{x}_t^\varepsilon g_t^\varepsilon)\,TR_{g_t^\varepsilon}^{-1} H_i(\tilde{x}_t^\varepsilon g_t^\varepsilon)\right)\circ dB_t^k + \sum_{i=1}^n \sigma_0^i(\tilde{x}_t^\varepsilon g_t^\varepsilon)TR_{g_t^\varepsilon}^{-1} H_i(\tilde{x}_t^\varepsilon g_t^\varepsilon)\,dt. \tag{3}$$

Finally we also rewrite the equation for the fast variable in terms of the fast and slow splitting:

$$dg_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}} \sum_{k=1}^{m_2} \left(\sum_{j=1}^N \theta_k^j(\tilde{x}_t^\varepsilon g_t^\varepsilon) A_j^*(g_t^\varepsilon) \right) \circ dW_t^k + \frac{1}{\varepsilon} \sum_{j=1}^N \theta_0^j(\tilde{x}_t^\varepsilon g_t^\varepsilon) A_j^*(g_t^\varepsilon)\, dt. \tag{4}$$

This completes the proof.

If θ_k^j are lifts of functions from M, i.e. equi-variant functions, then the system of SDEs for g_t^ε do not depend on the slow variables. Define

$$\mathscr{L}_u f(g) = \frac{1}{2} \sum_{k=1}^{m_1} \left(\theta_k^j(ug) A_j^*(g) \right)^2 f(g) + \sum_{j=1}^N \theta_0^j(ug) A_j^*(g) f(g).$$

The matrix with entries $\Theta_{i,j} = \sum_{k=1}^{m_1} \theta_k^j \theta_k^i$ measures the ellipticity of the system.

In section 6.3 we state an averaging principle for this system of slow-fast equations.

2.5 *Completely integrable stochastic Hamiltonian systems*

In [63] a completely integrable Hamiltonian system (CISHS) in an $2n$ dimensional symplectic manifolds is introduced, which has n Poisson commuting Hamiltonian functions. After some preparation this reduces to a slow-fast system in the action angle components.

We begin comparing this model with the very well studied random perturbation problem $dx_t = (\nabla H)^\perp(x_t)dt + \varepsilon dB_t$ where B is a real valued Brownian motion, $H : \mathbb{R}^2 \to \mathbb{R}$, and $(\nabla H)^\perp$ is the skew gradient of H. In the more recent CISHS model, the energy function is assumed to be random and of the form $\dot{B}_t$ so we have the equation $dx_t = (\nabla H)^\perp(x_t) \circ dB_t$. In both cases $H(x_t) = H(x_0)$ for all time, so H is a conserved quantity for the stochastic system. Suppose that the CISHS system is perturbed by a small vector field, we have the family of equations

$$dx_t^\varepsilon = (\nabla H)^\perp(x_t^\varepsilon) \circ dB_t + \varepsilon V(x_t) dt.$$

Given a perturbation transversal to the energy surface of the Hamiltonians, one can show that the energies converge on $[1, \frac{1}{\varepsilon}]$ to the solution of a system of ODEs. If moreover the perturbation is Hamiltonian, the limit is a constant and one may rescale time and find an effective Markov process on the scale $1/\varepsilon^2$. The averaging theorem was obtained from studying a reduced system of slow and fast variables. The CISHS reduces to a system of equations in (H, θ), the action angle coordinates, where $H \in \mathbb{R}^n$ is the slow variable and $\theta \in S^n$ is the fast variables.

$$\frac{d}{dt}H_t^i = \varepsilon f(H_t^\varepsilon, \theta_t^\varepsilon),$$
$$d\theta_t^i = \sum_{i=1}^n X_i(H_t^\varepsilon, \theta_t^\varepsilon) \circ dW_t^i + \varepsilon X_0(H_t^\varepsilon, \theta_t^\varepsilon) dt.$$

This slow-fast system falls, essentially, into the scope of the article.

3 Ergodic theorem for Fredholm operators depending on a parameter

Birkhoff's theorem for a sample continuous Markov process is directly associated to the solvability of the elliptic differential equation $\mathscr{L}u = v$ where $\mathscr{L}$ is the diffusion operator (i.e. the Markov generator) of the Markov process and v is a given function. A function v for which $\mathscr{L}u = v$ is solvable should satisfy a number of independent constraints. The index of the operator $\mathscr{L}$ is the dimension of the solutions for the homogeneous problem minus the dimension of the independent constraints.

Definition 3.1 A linear operator $T : E \to F$, where E and F are Hilbert spaces, is said to be a Fredholm operator if both the dimensions of the kernel of T and the dimension of its cokernel $F/\text{Range}(T)$ are finite dimensional. The Fredholm index of a Fredholm operator T is defined to be

$$\text{index}(T) = \dim(\ker(T)) - \dim(\text{cokernel}(T)).$$

A Fredholm operator T has also closed range and $E_2/\text{Range}(T) = \ker(T^*)$.

A smooth elliptic diffusion operator on a compact space is Fredholm. It also has a unique invariant probability measure. The Poisson equation $\mathscr{L}u = v$ is solvable for a function $v \in L^2$ if and only if v has null average with respect to the invariant measure, the latter is the centre condition used in diffusion creations.

If we have a family of operators $\{\mathscr{L}_x : x \in N\}$ satisfying Hörmander's condition where x is a parameter taking values in a manifold N, the parameter space is typically the state space for the slow variable, we will need a continuity theorem on the projection operator $f \mapsto \bar{f}$. We give a theorem on this in case each $\mathscr{L}_x$ has a unique invariant probability measure. It is clear that for each bounded measurable function f, $\int f(z) d\mu_x(z)$ is a function of x. We study its smooth dependence on x.

For the remaining of the section, for $i = 0, 1, \ldots, m$, let $Y_i : N \times G \to TG$ be smooth vector fields and let $\mathscr{L}_x = \frac{1}{2}\sum_{i=1}^m Y_i^2(x, \cdot) + Y_0(x, \cdot)$.

Definition 3.2 If $\mathscr{L}_x$ satisfies Hörmander's condition, let $r(x,y)$ denote the minimum number for the vector fields and their iterated Lie brackets up to order $r(x,y)$ to span T_yG. Let $r(x) = \inf_{y \in G} r(x,y)$. If G is compact, $r(x)$ is a finite number and will be called the rank of $\mathscr{L}_x$.

Let $s \geq 0$, let dx denote the volume measure of a Riemannian manifold G and let Δ denote the Laplacian. If f is a C^∞ function we define its Sobolev norm to be

$$\|f\|_s = \left(\int_M f(x)(I+\Delta)^{s/2} f(x)\, dx \right)^{\frac{1}{2}}$$

and we let H_s denote the closure of C^∞ functions in this norm. This can also be defined without using a Riemannian structure. If $\{\lambda_i\}$ is a partition of unity subordinated to a system of coordinates $\{\phi_i, u_i\}$, then the above Sobolev norm is equivalent to the norm $\sum_i \|(\lambda_i f)\circ \phi_i\|_s$. For a compact manifold, the Sobolev spaces are independent of the choice of the Riemannian metric. Let us denote by $|T|$ the operator norm of a linear map T.

Suppose that $\mathscr{L}_x$ satisfies Hörmander's condition. Let us re-name the vector fields Y_i and their iterated Lie brackets up to order $r(x)$ as $\{Z_k\}$. Let us define the quadratic form

$$Q^x(y)(df, df) = \sum_i |df(Z_i(x,y))|^2.$$

Then $Q^x(y)$ measures the sub-ellipticity of the operator. Let

$$\gamma(x) = \inf_{|\xi|=1} Q^x(y)(\xi, \xi).$$

Then $\gamma(x)$ is locally bounded from below by a positive number.

We summarise the properties of Hörmander type operators in the proposition below. Let $\mathscr{L}_x^*$ denote the L_2 adjoint of $\mathscr{L}_x$. An invariant probability measure for $\mathscr{L}_x$ is a probability measure such that $\int_G \mathscr{L}_x f(y)\mu_x(dy) = 0$ for any f in the domain of the generator.

Proposition 3.3 *Suppose that each $\mathscr{L}_x$ satisfies Hörmander's condition and that G is compact. Then the following statements hold.*

(1) There exists a positive number $\delta(x)$ such that for every $s \in \mathbb{R}$ there exists a constant $C(x)$ such that for all $u \in C^\infty(G;\mathbb{R})$ the following sub-elliptic estimates hold,

$$\|u\|_{s+\delta} \leq C(\|\mathscr{L}_x u\|_s + |u|_{L_2}), \quad \|u\|_{s+\delta} \leq C(\|\mathscr{L}_x^* u\|_s + |u|_{L_2}).$$

We may and will choose $C(x)$ to be continuous and $\delta(x)$ to be locally bounded from below. If r is bounded there exists $\delta_0 > 0$ such that $\delta(x) \geq \delta_0$.

(2) $\mathscr{L}_x$ and $\mathscr{L}_x^*$ are hypo-elliptic.

(3) $\mathscr{L}_x$ and $\mathscr{L}_x^*$ are Fredholm and index=0.

(4) If the dimension of $\ker(\mathscr{L}_x)$ *is* 1*, then* $\ker(\mathscr{L}_x)$ *consists of constants.*

Proof. It is clear that Hörmander's condition still holds if we change the sign of the drift Y_0, or add a zero order term, or add a first order term which can be written as a linear combination of $\{Y_0, Y_1, \dots, Y_m\}$. Since

$$\mathcal{L}_x^* = \frac{1}{2}\sum_{i=1}^m (Y_i)^2 - Y_0 - \sum_i \mathrm{div}(Y_i)Y_i + \mathrm{div}(Y_0) - \frac{1}{2}\sum_i L_{Y_i}\mathrm{div}(Y_i) + \frac{1}{2}\sum_i [\mathrm{div}(Y_i)]^2,$$

$\mathcal{L}_x$ satisfies also Hörmander's condition.

By a theorem of Hörmander in [50], there exists a positive number $\delta(x)$, such that for every $s \in \mathbb{R}$ and all $u \in C^\infty(G;\mathbb{R})$,

$$\|u\|_{s+\delta(x)} \le C(x)(\|\mathcal{L}_x u\|_s + |u|_{L_2}).$$

The constant $C(x)$ may depend on s, the L_∞ bounds on the vector fields and their derivatives, and on the rank $r(x)$, and the sub-ellipticity constant $\gamma(x)$. The constant $\delta(x)$ in the sub-elliptic estimates depend only on how many number of brackets are needed for obtaining a basis of the tangent spaces, we can for example take $\delta(x)$ to be $\frac{1}{r(x)}$. The number of brackets needed to obtain a basis at T_yG is upper semi-continuous in y and is bounded for a compact manifold. Since $\mathcal{L}_x$ varies smoothly in x, then for $x \in D$ there is a uniform upper bound on the number of brackets needed. Also as indicated in Hörmander's proof [50], the constant $C(x)$ depends smoothly on the vector fields. If there exists a number k_0 such that $r(x) \le k_0$ for all x, then we can choose a positive δ that is independent of x. This proves the estimates in part (1) for both $\mathcal{L}_x$ and $\mathcal{L}_x^*$. The hypo-ellipticity of $\mathcal{L}_x$ and $\mathcal{L}_x^*$ is the celebrated theorem of Hörmander and follows from his sub-elliptic estimates, this is part (2).

For part (3) we only need to work with $\mathcal{L}_x$. We sketch a proof for $\mathcal{L}_x$ to be Fredholm as a bounded operator from its domain with the graph norm to L_2. From the sub-elliptic estimates it is easy to see that $\mathcal{L}_x$ has compact resolvents and that $\ker(\mathcal{L}_x)$ and $\ker(\mathcal{L}_x^*)$ are finite dimensional. Then a standard argument shows that $\mathcal{L}_x$ has closed range: If $\mathcal{L}_x f_n$ converges in L_2, then either the sequence $\{f_n\}$ is bounded in which case they are also bounded in H_δ the latter is compactly embedded in L_2, and therefore has a convergent sub-sequence. Let us denote g a limit point. Then since $\mathcal{L}_x$ if closed, g satisfies that $\mathcal{L} g = \lim_{n\to\infty} \mathcal{L}_x f_n$. If $\{f_n\}$ is not L_2 bounded, we can find another sequence $\{g_n\}$ in the kernel of $\mathcal{L}$ such that $f_n - g_n$ is bounded to which the previous argument produces a convergent sub-sequence. The dimension of the cokernel is the dimension of the kernel of $\mathcal{L}_x^*$, proving the Fredholm property. That it has zero index is another consequence of the sub-elliptic estimates and can be proved from the definition and is an elementary (using properties of the eigenvalues of the resolvents and their duals), see [92]. Part (4) is clear as constants are always in the kernel of $\mathcal{L}_x$. □

If μ_1 and μ_2 are two probability measures on a metric space M we denote by $|\mu - \nu|_{TV} = \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)|$ their total variation norm and W_1 their Wasserstein distance:

$$W_1(\mu_1,\mu_2) = \inf_\nu \int_{M\times M} \rho(x,y)\nu(x,y)$$

where ρ is the distance function and the infimum is taken over all couplings of μ_1 and μ_2. Suppose that $\mathcal{L}_x$ has an invariant probability measure $\mu_x(dy) = q(x,y)dy$. If for a constant K, $|q(x_1,y) - q(x_2,y)| \le K\rho(x_1,x_2)$ for all $x_1 \in M, x_2 \in M, y \in G$, then $|\mu^{x_1} - \mu^{x_2}|_{TV} \le K\rho(x_1,x_2)$.

Let μ_x be an invariant probability measure for $\mathscr{L}_x$. We study the regularity of the densities of the invariant probability measures with respect to the parameter, especially the continuity of the invariant probability measures in the total variation norm. This can be more easily obtained if $\mathscr{L}_x$ are Fredholm operators on the same Hilbert space and if there is a uniform estimate on the resolvent. For a family of uniformly strict elliptic operators, these are possible.

Remark 3.4 For the existence of an invariant probability measure, we may use Krylov-Bogoliubov theorem which is valid for a Feller semi-group: Let $P_t(x,\cdot)$ be the transition probabilities. If for some probability measure μ_0 and for a sequence of numbers T_n with $T_n \to \infty$, $\{Q_n(\cdot) = \frac{1}{T_n}\int_0^{T_n}\int_M P_t(x,\cdot)d\mu_0(x)dt, n \geq 1\}$ is tight, then any limit point is an invariant probability measure. The existence of an invariant probability measure is trivial for a Feller Markov process on a compact space. Otherwise, a Laypunov function is another useful tool. See [28, 45, 44, 42] for relevant existence and uniqueness theorems.

Remark 3.5 Our operators $\mathscr{L}_x$ are Fredholm from their domains to L_2. On a compact manifold $\mathscr{L}_x$ is a bounded operator from $W^{2,2}$ to L_2 but this is only an extension of $\mathscr{L}_x$, where $W^{2,2}$ denotes the standard Sobolev space of functions, twice weakly differentiable with derivatives in L_2. We have $W^{2,2} \subset \mathrm{Dom}(\mathscr{L}_x) \subset W^{\delta(x),2}$. Due to the directions of degeneracies the domain of $\mathscr{L}_x$, given by its graph norm, can be larger than $W^{2,2}$. Since the points of the degeneracies of $\mathscr{L}_x$ move, in general, with x, their domain also change with x. Suppose that $\mathscr{L}_x$ has zero Fredholm index, then $\mathscr{L}_x$ is an isometry from $[\ker(\mathscr{L}_x)]^\perp$ to its image and $\mathscr{L}_x^*$ is invertible on $N^\perp$, the annihilator of the kernel of $\mathscr{L}_x$. Set

$$A(x) = \left|(\mathscr{L}_x^*)^{-1}_{N_x^\perp}\right|_{op}.$$

In the following proposition we consider the continuity of μ^x.

Proposition 3.6 *Let G be compact. Suppose that $Y_i \in BC^\infty$ and the conclusions of Proposition 3.3. Suppose also that each $\mathscr{L}_x$ has a unique invariant probability measure $\mu^x(dy)$.*

(i) Let $q(x,y)$ denote the kernel of $\mu^x(dy)$. Then q and its derivatives in y are locally bounded in x.
If the rank r is bounded from above, γ is bounded from below, then q and its derivatives in y are bounded, i.e. $\sup_x |\nabla^{(k)}\rho(x,\cdot)|_\infty$ *is finite for any $k \in \mathscr{N}$.*
(ii) The kernel q is smooth in both variables.
(iii) Let D be a compact subset of N. There exists a number c such that for any $x_1, x_2 \in D$, $|\mu^{x_1} - \mu^{x_2}|_{TV} \leq c\rho(x_1,x_2)$.
(iv) Suppose furthermore that r is bounded from above, γ is bounded from below, and A is bounded, then μ^x is globally Lipschitz continuous in x and $q \in BC^\infty(N \times G)$.

Proof. Each function q solves the equation $\mathscr{L}_x^* q = 0$ where $\mathscr{L}_x^*$ is the L^2 adjoint of $\mathscr{L}_x$. Since $\mathscr{L}_x^*$ is hypo-elliptic, then for each x, $q(x,\cdot)$ is C^∞. In other words, $q(x,\cdot)$ is a function from M to $C^\infty(G,\mathbb{R})$. We observe that $q(x,\cdot)$ are probability densities, so bounded in L^1. If we take s to be a number smaller than $-n/2$, n being the dimension of the manifold, then $|q(x,\cdot)|_s \le C|q(x,\cdot)|_{L^1(G)}$. We apply the sub-elliptic estimates in part (1) of Proposition 3.3 to q:

$$\|u\|_{s+\delta(x)} \le c_0(x)(\|\mathscr{L}_x^* u\|_s + \|u\|_s),$$

where $\delta(x)$ and $c(x)$ are constants, and obtain that $|q(x,\cdot)|_{s+\delta(x)} \le C(x)$. Iterating this we see that for all s,

$$|q(x,\cdot)|_s \le C(\delta(x), r(x), \gamma(x), Y).$$

The function $C(x)$ depends on the L^∞ norms of the vector fields Y_i and their covariant derivatives, and also on $\gamma(x)$. Also, δ can be taken to be $\frac{1}{r(x)+1}$ and $r(x)$ is locally bounded. By the Sobolev embedding theorems, q and the norms of its derivatives in y are locally bounded in x. (If furthermore r and γ are bounded, Y_i and their derivatives in x are bounded, then both δ and C can be taken as a constant, in which case q and their derivatives in y are bounded.)

Since q is in L^1, its distributional derivative in the x-variable exists and will be denoted by $\partial_x q$. For each x, $\mathscr{L}_x^* q = 0$, and so the distributional derivative in x of $\mathscr{L}_x^* q$ vanishes and

$$\partial_x(\mathscr{L}_x^*)q(x,y) + \mathscr{L}_x^* \partial_x q(x,y) = 0.$$

Set

$$g(x,y) = -\left(\partial_x(\mathscr{L}_x^*)\right) q(x,y).$$

Then g is smooth in y, whose Sobolev norms in y are locally bounded in x. Since the distributional derivative of q in x satisfies $\int_G \mathscr{L}_x^*(\partial_x q)(x,y)\,dy = 0$ for every x, $\int_G g(x,y)\,dy$ vanishes also. Since the index of $\mathscr{L}_x$ is zero, the invariant measure is unique, the dimension of the kernel of $\mathscr{L}_x$ is 1. The kernel consists of only constants and so $g(x,\cdot)$ is an annihilator of the kernel of $\mathscr{L}$. By Fredholm's alternative, this time applied to $\mathscr{L}_x^*$, we see that for each x we can solve the Poisson equation

$$\mathscr{L}_x^* G(x,y) = g(x,y).$$

Furthermore, by the sub-elliptic estimates, $|G(x,\cdot)|_{L_2(G)} \le A(x)|g(x,\cdot)|_{L_2(G)}$ for some number $A(x)$. Since A is locally bounded, then $G(x,y)$ has distributional derivative in x. But $\partial_x q(x,y)$ also solves $\mathscr{L}_x^* \partial_x q(x,y) = g(x,y)$, by the uniqueness of solutions we see that $\partial_x q(x,y) = G(x,y)$. Thus the distributional derivative of q in x is a locally integrable function. Iterating this procedure and use sub-elliptic estimates to pass to the supremum norm we see that $q(x,y)$ is C^∞ in x with its derivatives in x locally bounded, in particular for a locally bounded function c_1,

$$\sup_{y\in G} |\partial_x q(x,y)| \le A(x)c_1(x).$$

Finally, let f be a measurable function with $|f| \le 1$. Then

$$\left|\int_G f(y)q(x_1,y)dy - \int_G f(y)q(x_2,y)dy\right| \le \sup_{x\in D} A(x) \sup_{x\in D} c_1(x)\, \rho(x_1,x_2),$$

where D is a relatively compact open set containing a geodesic passing through x_1 and x_2. We use the fact that the total variation norm between two probability measures μ and ν is $\frac{1}{2}\sup_{|g|\le 1}|\int g\,d\mu - \int g\,d\nu|$ where the supremum is taken over the family of measurable functions with values in $[-1,1]$ to conclude that $|\mu^{x_1} - \mu^{x_2}|_{TV} \le \sup_{x\in D} A(x)\, \rho(x_1,x_2)$ and conclude the proof. □

Example 3.7 *An example of a fast diffusion satisfying all the conditions of the proposition is the following on S^1 and take $x \in \mathbb{R}$:*

$$dy_t = \sin(y_t + x)dB_t + \cos(y_t + x)dt.$$

Then $\mathscr{L}_x = \cos(x+y)\frac{\partial}{\partial y} + \frac{1}{2}\sin^2(x+y)\frac{\partial^2}{\partial y^2}$ satisfies Hörmander's condition, has a unique invariant probability measure and $r(x) = 1$. Furthermore the resolvent of $\mathscr{L}_x$ is bounded in x.

Definition 3.8 The operator $\mathscr{L}_x$ is said to satisfy the parabolic Hörmander's condition if $\{Y_1(x,\cdot),\dots,Y_{m_2}(x,\cdot)\}$ together with the brackets and iterated the brackets of $\{Y_0(x,\cdot),Y_1(x,\cdot),\dots,Y_{m_2}(x,\cdot)\}$ spans the tangent space of N at every point. Let $P^x(t,y_0,y)$ denote the semigroup generated by $\mathscr{L}$.

Remark 3. Suppose that each $\mathscr{L}_x$ is symmetric, satisfies the parabolic Hörmander's condition and the following uniform Doeblin's condition: there exists a constant $c \in (0,1]$, $t_0 > 0$, and a probability measure ν such that

$$P_t^x(y_0,U) \ge c\nu(U),$$

for all $x \in N$, $y_0 \in G$ and for every Borel set U of G, Suppose that $Y_j \in BC^\infty$. Then $A(x)$ is bounded. In fact for any f with $\int f(y)\mu^x(dy) = 0$, the function $P_t^x f(y_0) = \int_G f(y)P^x(t,y_0,dy)$ converges to 0 as $t \to \infty$ with a uniform exponential rate. Since $\mathscr{L}_x$ satisfies the parabolic Hörmander's condition, $\mathscr{L} - \frac{\partial}{\partial x}$ satisfies Hörmander's condition on $M \times \mathbb{R}$. Then by the sub-elliptic estimates for $\mathscr{L} - \frac{\partial}{\partial t}$, $P_t^x f$ converges also in L_2. Let R^x denote the resolvent of $\mathscr{L}_x$. Since

$$\langle R^x f, f\rangle_{L_2} = \int_G \int_0^\infty P_t^x f(y) f(y)\,dt\,dy,$$

then R^x is uniformly bounded. Since $\mathscr{L}_x$ is symmetric, this gives a bound on $A(x)$. We refer to the book [6] for studies on Poincaré inequalities for Markov semigroups.

Corollary 3.9 Let G be compact. Then q is smooth in both variables and in $BC^\infty(N \times G)$. In particular $\mu^x = q(x,y)dy$ is globally Lipschitz continuous.

Just note that the semigroups P_t^x converges to equilibrium with uniform rate. The spectral gap of $\mathscr{L}_x$ is bounded from below by a positive number.

The following is a version of the law of large numbers.

Theorem 3.10 *Let G be compact. Suppose that $\sum_{j=1}^{m_2} |Y_j|_\infty$ is finite, and the conclusions of Proposition 3.3. Suppose that each $\mathscr{L}_x$ has a unique invariant probability measure μ_x.*

Let $s > 1 + \frac{\dim(G)}{2}$. Then there exists a constant $C(x)$ such that for every $x \in N$ and for every smooth real valued function $f : N \times G \to \mathbb{R}$ with compact support in the first variable (or independent of the first variable),

$$\sqrt{\mathbf{E}\left(\frac{1}{T}\int_t^{t+T} f(x,z_r^x)\,dr - \int_G f(x,y)\mu_x(dy)\right)^2} \le C(x)\|f(x,\cdot)\|_s \frac{1}{\sqrt{T}} \qquad (1)$$

where z_r^x is an $\mathscr{L}_x$ diffusion and $C(x)$ is locally bounded.

Proof. In the proof we take $t = 0$ for simplicity. We only need to work with a fixed $x \in N$. We may assume that $\int_G f(x,y)\mu_x(dy) = 0$. Since $\mathscr{L}_x$ is hypo-elliptic and since μ_x is the unique invariant probability measure then, for any smooth function f with $\int_G f(x,y)\,\mu_x(dy) = 0$, $\mathscr{L}_x g(x,\cdot) = f(x,\cdot)$ has a smooth solution. If f is compactly supported in the first variable, so is g. We may then apply Itô's formula to the smooth function $g(x,\cdot)$, allowing us to estimate $\frac{1}{T}\int_0^T f(y_r^x)dr$ whose $L^2(\Omega)$ norm is controlled by the norm of g in C^1 and the norms $|Y_j(x,\cdot)|_\infty$. The $\mathscr{L}_x$ diffusion satisfies the equation:

$$\frac{1}{T}\int_0^T f(x,z_r^x)dr = \frac{1}{T}\left(g(x,z_T^x) - g(x,y_0)\right) - \frac{1}{T}\left(\sum_{k=1}^{m_2}\int_0^T dg(x,\cdot)(Y_k(x,z_r^x))dW_r^k\right).$$

Since $|Y_j(x,\cdot)|_\infty$ is bounded, it is sufficient to estimate the stochastic integral term by Burkholder-Davis-Gundy inequality:

$$\mathbf{E}\left(\sum_{k=1}^{m_2}\int_0^T dg(x,\cdot)(Y_k(x,z_r^x))dW_r^k\right)^2 \le \sum_{k=1}^{m_1}|Y_k|_\infty^2\int_0^T \mathbf{E}|dg(x,z_r^x)|^2\,ds.$$

It remains to control the supremum norm of $dg(x,\cdot)$. By the Sobolev embedding theorem this is controlled by the L_2 Sobolev norms $\|f(x,\cdot)\|_s$ where $s > 1 + \frac{\dim(G)}{2}$. Let D be a compact set containing the supports of the functions $f(\cdot,y)$. We can choose a number $\delta > 0$, chosen according to $\sup_{x\in D} r(x)$, such that the sub-elliptic estimates holds for every $x \in D$. There exist constants c_1, c_2, c_3 such that for every $x \in D$,

$$|dg(x,\cdot)|_\infty \le c_1\,\|g(x,\cdot)\|_{s+\delta} \le c_2(x)\,(\|f(x,\cdot)\|_s + |g(x,\cdot)|_{L_2}) \le c_3(x)\,\|f(x,\cdot)\|_s.$$

The constant c_2 may depend on s. The constant $c_3(x)$ is locally bounded. We have used the following fact. The spectrum of $\mathscr{L}_x$ is discrete, the dimension of the kernel space of $\mathscr{L}_x$ is 1 and hence the only solutions to $\mathscr{L}_x h = 0$ are constants. We know that the spectral distance is continuous, which is not the right reason for $c_3(x)$ to be locally bounded. To see that we may assume that f is not a constant and observe that $|\mathscr{L}_x^{-1}|_{op} \le k(x)$ where $k(x)$ is a finite number. This number is locally bounded following the fact that the semi-group $P_t^x f$ converges to zero exponentially and the kernels for the probability distributions of $\mathscr{L}_x$ are smooth in the parameter x. □

For the study of the limiting process in stochastic averaging we would need to know the regularity of the average of a Lipschitz continuous function with respect to one of its variables. The following illustrates what we might need.

Proposition 3.11 *Let $\{\mu^x, x \in M\}$ be a family of probability measures on G. Let $f : N \times G \to \mathbb{R}$ be a measurable function.*

(1) Let f be a bounded function, Lipschitz continuous in the first variable, i.e. $|f(x_1,y) - f(x_2,y)| \le K_1(y)\rho(x_1,x_2)$ with $\sup_{x\in M} |K_1|_{L_1(\mu_x)} < \infty$. Then

$$\left|\int_G f(x_1,y)\,\mu^{x_1}(dy) - \int_G f(x_2,y)\,\mu^{x_2}(dy)\right| \le K_2\,\rho(x_1,x_2) + |f|_\infty|\mu^{x_1} - \mu^{x_2}|_{TV}.$$

(2) Suppose furthermore that μ_x depends continuously on x in the total variation metric. Let f be bounded continuous such that

$$|f(x_1,z) - f(x_2,z)| \le K_3\,\rho(x_1,x_2), \quad \forall z \in G, x_1, x_2 \in M,$$

for a positive number K_2. Then $\int_0^T \int_G f(x_s,z)\mu^{x_s}(dz)ds$ exists, and if D is the support of f then

$$\begin{aligned}&\left|\sum_{i=0}^{N-1} \Delta t_i \int_G f(x_{t_i},z)\,\mu^{x_{t_i}}(dz) - \int_0^T \int_G f(x_s,z)\mu^{x_s}(dz)ds\right| \\ \le & T K_3 \sup_{0\le i<N-1}\sup_{s\in[t_i,t_{i+1})} [\rho(x_s,x_{t_i})] + |f|_\infty \cdot \sup_{0\le i<N-1}\sup_{s\in[t_i,t_{i+1})} \left(|\mu^{x_s} - \mu^{x_{t_i}}|_{TV}\,\chi_{x_s\in D}\right).\end{aligned}$$

(3) Suppose that μ_x depends continuously on x in the Wasserstein 1-distance. Then for any bi-Lipschitz continuous f, $\int_0^T \int_G f(x_s,z)\mu^{x_s}(dz)ds$ exists and the estimate in part (1) holds with the total variation distance replaced by W_1, the Wasserstein 1-distance.

Proof. Just observe that:

$$\begin{aligned}&\left|\int_G f(x_1,y)\,\mu^{x_1}(dy) - \int_G f(x_2,y)\,\mu^{x_2}(dy)\right| \\ &\le \int K_1(y)\mu^{x_1}(dy)\,\rho(x_1,x_2) + |f|_\infty|\mu^{x_1} - \mu^{x_2}|_{TV},\end{aligned}$$

obtaining the required inequality in part (1) . For any non-negative numbers s, t,

$$\begin{aligned}&\left|\int_G f(x_t,z)\,\mu^{x_t}(dz)-\int_G f(x_s,z)\mu^{x_s}(dz)\right|\\&\le K_3\,\rho(x_t,x_s)+\left|\int_G f(x_s,z)\mu^{x_t}(dz)-\int_G f(x_s,z)\mu^{x_s}(dz)\right|.\end{aligned}$$

This holds pathwise. Since each function $f(x_s(\omega,\cdot))$ is bounded by $|f|_\infty$,

$$\left|\int_G f(x_t,z)\,\mu^{x_t}(dz)-\int_G f(x_s,z)\mu^{x_s}(dz)\right|\le K_3\,\rho(x_t,x_s)+|f|_\infty\,|\mu^{x_s}-\mu^{x_t}|_{TV}\,\chi_{x_s\in D}.$$

Since x_s is sample continuous, $x\mapsto\mu^x$ is continuous and f is a bounded and continuous, $\int_G f(x_s,z)\mu^{x_s}(dz)$ is continuous in s and so integrable in s. Consequently,

$$\begin{aligned}&\left|\sum_{i=0}^{N-1}\Delta t_i\int_G f(x_{t_i},z)\,\mu^{x_{t_i}}(dz)-\int_0^T\int_G f(x_s,z)\mu^{x_s}(dz)ds\right|\\&\le\sum_{i=0}^{N-1}\Delta t_i K_3[\rho(x_s,x_{t_i})]+\sum_{i=0}^{N-1}\Delta t_i|f|_\infty[\chi_{x_s\in D}|\mu^{x_s}-\mu^{x_{t_i}}|_{TV}]\\&\le TK_3\sup_{s\in[t_i,t_{i+1})}\mathbf{E}[\rho(x_s,x_{t_i})]+|f|_\infty\sup_{s\in[t_i,t_{i+1})}(\chi_{x_s\in D}|\mu^{x_s}-\mu^{x_{t_i}}|_{TV}).\end{aligned}$$

Finally we use the fact that f is Lipschitz in the second variable and the following dual formulation for the Wasserstein 1-distance $W_1(\mu,\nu)$ of two probability measures μ and ν,

$$W_1(\mu,\nu)=\sup_{|g|_{\mathrm{Lip}=1}}\left|\int g\,d\mu-\int g\,d\nu\right|,$$

where $|g|_{\mathrm{Lip}}$ denotes the Lipschitz constant of g. We obtain

$$\left|\int_G f(x_t,z)\,\mu^{x_t}(dz)-\int_G f(x_s,z)\mu^{x_s}(dz)\right|\le K_3\,\rho(x_t,x_s)+K_4\,W_1(\mu^{x_s},\mu^{x_t}).$$

The required assertion and estimate now follows by the argument in part (2). □

Put Proposition 3.6 and Proposition 3.10 together we obtain Theorem 1.

Finally we would like to refer to [7] for the convergence in total variation in the Law of large numbers for independent random variable, see also [57]. See the books [30, 9] for stochastic flows in sub-Riemmian geometry. It would be interesting to study problems in this section under the ‘uniformly finitely generated’ conditions, see e.g. [22, 61]. See also [1, 19].

4 Basic Estimates for SDEs on manifolds

To obtain an averaging theorem associated to a family of stochastic processes $\{x_t^\varepsilon,\varepsilon>0\}$ on a manifold N, we first prove that the family of stochastic processes is pre-compact and we then proceed to identify the limiting processes. To this end we first obtain uniform estimates on the family of slow variables, on the space of continuous functions on the manifold, and also obtain estimates on the limiting Markov processes. In this section we obtain essential estimates for a general SDE and these estimates will be in terms of bounds on the driving vector fields.

Throughout this section we assume that M is a connected smooth and complete Riemannian manifold, $B_t = (B_t^1, \dots, B_t^m)$ is an $\mathbb{R}^m$-valued Brownian motion. Let X_0 be a vector field and $X : M \times \mathbb{R}^m \to TM$ be a map linear in the second variable. For $x \in M$, let $\phi_t(x)$ denote the solution to the SDE

$$dx_t = \sum_{k=1}^{m} X_k(x_t) \circ dB_t^k + X_0(x_t)\,dt, \tag{1}$$

with initial value x. We also set $x_t = \phi_t(x_0)$.

The type of estimates we need are variation of the following $\mathbf{E}[\rho(x_s, x_t)]^2 \le C|t-s|$ where the constant C depends on the SDE only on specific bounds for the driving vector fields. Since no ellipticity is assumed, it is essential to deal with the problem that $\rho(x,y)$ is only C^1, when x and y are on the cut locus of each other, and we cannot apply Itô's formula to ρ directly. If we are only interested in obtaining tightness results, this problem can be overcome by choosing an auxiliary distance function. Otherwise, e.g. for the convergence of the stochastic processes, we work with the Riemannian distance function $\rho : M \times M \to \mathbb{R}$. Let $M \times M$ be given the product Riemannian metric. Let $|f|_\infty$ denote the L_∞ norm of a function f.

Lemma 4.1 *(1) Suppose that M is a complete Riemannian manifold with bounded sectional curvature. Then for each $\delta > 0$ there exists a smooth distance like function $f_\delta : M \times M \to \mathbb{R}$ and a constant K_1 independent of δ such that*

$$|f_\delta - \rho|_\infty \le \delta, \quad |\nabla f_\delta| \le K_1, \quad |\nabla^2 f_\delta| \le K_1.$$

If furthermore the curvature has a bounded covariant derivative, then we may also assume that $|\nabla^3 f_\delta| \le K_1$.

(2) If M is compact Riemannian manifold, there exists a smooth function $f : M \times M \to \mathbb{R}$ such that f agrees with ρ on a tubular neighbourhood of the diagonal set of the product manifold $M \times M$.

Proof. (1) For the distance function $\rho(\cdot, O)$, where O is a fixed point in M, this is standard, see [79, 84, 21]. To obtain the stated theorem it is sufficient to repeat the proof there for the distance function on the product manifold. The basic idea is as following. By a theorem of Greene and Wu [40], every Lipschitz continuous function with gradient less or equal to K can be approximated by C^∞ functions whose gradients are bounded by K. We apply this to the distance function ρ and obtain for each δ a smooth function $f_\delta : M \times M \to \mathbb{R}$ such that

$$|\rho - f_\delta|_\infty \le \delta, \qquad |\nabla f_\delta|_\infty \le 2.$$

We then convolve f_δ with the heat flow to obtain $f_\delta(x,y,t)$, apply Li-Yau heat kernel estimate for manifolds whose Ricci curvature is bounded from below and using harmonic coordinates on a a small geodesic ball of radius a/K where K is the upper bound of the sectional curvature and a is a universal constant. For part (ii), M is compact. We take a smooth cut off function $h:\mathbb{R}_+\to\mathbb{R}_+$ such that $h(t)=1$ for $t<a$ and vanishes for $t>2a$ where $2a$ is the injectivity radius of M and such that $|\nabla h|$ is bounded. The function $f:=h\circ\rho$ is as required. □

Set $\tilde{X}_0=\frac{1}{2}\sum_{i=1}^m \nabla X_i(X_i)+X_0$. We denote by ρ the Riemannian distance on M. Let T be a positive number and let $O\in M$. Let K', K, a_i and b_i denote constants.

Lemma 4.2 *Suppose that $\tilde{X}_0$ and X_i are C^1, where $i=1,\dots,m$. Suppose* **one** *of the following two conditions hold.*

(i) The sectional curvature of M is bounded by K', and for every $x\in M$,

$$|X_i(x)|^2\le K+K\rho(x,O),\quad |\tilde{X}_0(x)|\le K+K\rho(x,O).$$

(ii) Suppose that $\rho^2:M\times M\to\mathbb{R}$ is smooth and

$$\frac{1}{2}\sum_{i=1}^m \nabla d\rho^{2p}(X_i,X_i)+d\rho^{2p}(\tilde{X}_0)\le K+K\rho^{2p}.$$

Then, the following statements hold.

(a) There exists a constant c which depends only on K', T, p, and $\dim(M)$ such that for every pair of numbers s,t with $0\le s\le t\le T$,

$$\mathbf{E}\rho^{2p}(x_t,O)\le c(Kt+1+\rho^{2p}(x_0,O))e^{cKt},$$
$$\mathbf{E}\left\{\rho^{2p}(x_s,x_t)\Big|\mathscr{F}_s\right\}\le c|t-s|(1+K)e^{cK|t-s|}.$$

(b) Suppose that in addition $|X_i|$ is bounded for every $i=1,\dots,m$. Then, for every $p\ge 1$, there exists a constant C, which depends only on p, K', m, and $\dim(M)$ and a constant $c(T)$, such that for every $s<t\le T$,

$$\mathbf{E}\left(\sup_{s\le u\le t}\rho^{2p}(x_s,x_u)\right)\le c+KC(T)\,e^{C(T)K}.$$

Also, $\mathbf{E}\left(\sup_{s\le u\le t}\rho^{2p}(O,x_u)\right)\le c(\rho^{2p}(O,x_0)+Kc(T))\,e^{Kc(T)}$.

Proof. Let $\delta\in(0,1]$ and let $f_\delta:M\times M\to\mathbb{R}$ be a smooth function satisfying the estimates

$$|f_\delta-\rho|_\infty\le\delta,\quad |\nabla f_\delta|\le K_1,\quad |\nabla^2 f_\delta|\le K_1$$

where K_1 is a constant depending on K' and $\dim(M)$. If ρ^2 is smooth we take $f_\delta=\rho$.

(a) Either hypothesis (i) or (ii) implies that the SDE (1) is conservative. For any $x\in M$ fixed we apply Itô's formula to the second variable of the function $f_\delta^2(x,y)$ on the time interval $[s,t]$:

$$f_\delta^{2p}(x,x_t) = f_\delta^{2p}(x,x_s) + \int_s^t \mathscr{L} f_\delta^{2p}(x,x_r)\,dr + \int_s^t 2f_\delta^{2p-1}(x,x_r)(df_\delta)(X_i(x_r))\,dB_r^i, \tag{2}$$

where d and $\mathscr{L}$ are applied to the second variable. Let τ_n denote the first time after s that $f_\delta(x,x_t) \geq n$ and we take the expectation of the earlier identity to obtain

$$\mathbf{E}[f_\delta^{2p}(x,x_{t\wedge\tau_n})] = \mathbf{E}[f_\delta^{2p}(x,x_s)] + \int_s^t \mathbf{E}\left[\chi_{r<\tau_n}\mathscr{L} f_\delta^{2p}(x,x_r)\right] dr.$$

Under hypothesis (ii), we use ρ in place of f_δ and conclude by Gronwall's inequality that $\mathbf{E}\rho^{2p}(x_t,O) \leq (\rho^{2p}(x,O)+Kt)e^{Kt}$. The second estimate follows from Markov property and taking $O = x_s$.

Let us now assume hypothesis (i) and let $C_1, C_2, \dots$ denote a constant depending on p. In the formula below, ∇ denotes differentiation w.r.t. the second variable,

$$\begin{aligned}\mathscr{L}[f_\delta^{2p}](x,y) =& p(2p-1)\sum_{i=1}^m f_\delta^{2p-2}(x,y)|\nabla f_\delta(X_i(y))|^2 \\ &+ p\sum_{i=1}^m f_\delta^{2p-1}(x,y)|\nabla^2 f_\delta(X_i(y),X_i(y))| + 2f_\delta^{2p-1}(x,y)df_\delta(\tilde{X}_0(y)).\end{aligned}$$

We first take $x = O$ and $s = 0$, to see that $\mathscr{L} f_\delta^{2p}(O,y) \leq C_1 K f_\delta^{2p}(O,y) + C_1 K$. We may then apply Grownall's inequality followed by Fatou's lemma to obtain:

$$\mathbf{E} f_\delta^{2p}(x_t,O) \leq (f_\delta^{2p}(x_0,O) + C_1 Kt)e^{C_1 Kt}.$$

Take $\delta = 1$, we conclude the first estimate from the following inequality:

$$\mathbf{E}[\rho^{2p}(x_t,O)] \leq C_2 + C_2\mathbf{E} f_1^{2p}(x_t,O) \leq C_2 + C_3(\rho^{2p}(x_0,O)+1+Kt)e^{C_1 Kt}.$$

Let $s<t$. Using the flow property, we see that

$$\mathbf{E}\{\rho^{2p}(x_s,x_t)|\mathscr{F}_s\} \leq C_4\delta^{2p} + C_4\mathbf{E}\{f_\delta^{2p}(x_s,x_t)|\mathscr{F}_s\} \leq C_4\delta^{2p} + C_5 K(t-s)e^{C_5 K(t-s)}.$$

For any $s,t>0$ we may choose δ_0 such that $\delta_0^{2p} < |t-s|$ and conclude that

$$\mathbf{E}\{\rho^{2p}(x_s,x_t)|\mathscr{F}_s\} \leq C_6(1+K)|t-s|e^{C_6 K|t-s|}.$$

For part (b) we take $\delta = 1$ and take $p = 2$ in (2). Then

$$\begin{aligned}\mathbf{E}\sup_{u\leq t} f_1^{2p}(O,x_u) =& C_1 f_\delta^{2p}(O,x_0) + C_1\left(\int_0^t (K + K f_\delta^2(O,x_r))\,dr\right)^p \\ &+ C_1\sum_i \mathbf{E}\left(\int_0^t 2f_1^{2p-1}(O,x_r)(df_1)(X_i(x_r))dr\right)^p\end{aligned}$$

Since $|X_i|$ is bounded for $i = 1,\dots,m$, $|2f_1(x,y)(df_1)(X_i(y))| \leq 2|f_1(x,y)|\cdot|X_i(y)|$. We conclude that

$$\mathbf{E}\sup_{0\le u\le t} f_1^{2p}(O,x_u) \le C_2(f_\delta^{2p}(O,x_0)+KC(T))e^{KC(T)}.$$

This leads to the required estimates for $\mathbf{E}\left[\sup_{0\le u\le t}\rho^{2p}(O,x_u)\right]$. Similarly, for some constants c_1 and c, depending on m and the bound of the sectional curvature, for some constants c and $C(T)$,

$$\mathbf{E}\left[\sup_{s\le u\le t}\rho^{2p}(x_s,x_u)\right] \le c_1 + c_1 K\mathbf{E}\left[\sup_{s\le u\le t}(f_1)^{2p}(x_s,x_u)\right] \le c + cKC(T)e^{Kc(T)}.$$

We have completed the proof for part (b). □

These estimates will be applied in the next section to both of our slow and fast variables. For the slow variables, we have the uniform bounds on the driving vector fields and hence we obtain a uniform moment estimate (in ε) of the distance traveled by the solutions. For the fast variables, the vector fields are bounded by $\frac{1}{\varepsilon}$ and we expect that the evolution of the y-variable in an interval of size Δt_i to be controlled by the following quantity $\frac{\Delta t_i}{\varepsilon}e^{\frac{\Delta t_i}{\varepsilon}}$.

5 Proof of Theorem 2

We proceed to prove the main averaging theorem, this is Theorem 2 in section 1.1.

In this section N and G are smooth complete Riemannian manifolds and $N\times G$ is the product manifold with the product Riemannian metric. We use ρ to denote the Riemannian distance on N, or on G, or on $N\times G$. This will be clear in the context and without ambiguity. For each $y\in G$ let $X_i(\cdot,y)$ be smooth vector fields on N and for each $x\in N$ let $Y_i(x,\cdot)$ be smooth vector fields on G, as given in the introduction. Let $x_0\in N$ and $y_0\in G$. We denote by $(x_t^\varepsilon, y_t^\varepsilon)$ the solution to the equations:

$$\begin{cases} dx_t^\varepsilon = \sum_{k=1}^{m_1} X_k(x_t^\varepsilon,y_t^\varepsilon)\circ dB_t^k + X_0(x_t^\varepsilon,y_t^\varepsilon)\,dt, \quad x_0^\varepsilon = x_0;\\ dy_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2} Y_k(x_t^\varepsilon,y_t^\varepsilon)\circ dW_t^k + \frac{1}{\varepsilon}Y_0(x_t^\varepsilon,y_t^\varepsilon)\,dt, \quad y_0^\varepsilon = y_0. \end{cases} \tag{1}$$

Let us first study the slow variables $\{x_t^\varepsilon, \varepsilon\in(0,1]\}$. We use O to denote a reference point in N.

Lemma 5.1 *Under Assumption 1, the family of stochastic processes $\{x_t^\varepsilon, \varepsilon\in(0,1]\}$ is tight on any interval $[0,T]$ where T is a positive number. Furthermore there exists a number C such that for any $p>0$,*

$$\sup_{\varepsilon\in(0,1]}\sup_{s,t\in[0,T]}\mathbf{E}\rho^2(x_s^\varepsilon,x_t^\varepsilon)\le C|t-s|, \qquad \sup_{\varepsilon\in(0,1]}\sup_{s,t\in[0,T]}\mathbf{E}\rho^{2p}(x_s^\varepsilon,x_t^\varepsilon)<\infty.$$

Any limiting process of x_t^ε, which we denote by $\bar{x}_t$, has infinite life time and satisfies the same estimates: $\mathbf{E}\rho^2(\bar{x}_s,\bar{x}_t)\le C(t-s)$ *and* $\sup_{s,t\in[0,T]}\mathbf{E}\rho^{2p}(\bar{x}_s,\bar{x}_t)$ *is finite.*

Proof. Assumption 1 states that: the sectional curvature of N is bounded, $|X_i(x,y)|^2\le K+K\rho(x,O)$ and $|\tilde{X}_0(x,y)|\le K+K\rho(x,O)$. Or $\rho^2: N\times N\to\mathbb{R}$ is smooth, and

$$\frac{1}{2}\sum_{i=1}^m \nabla d\rho^2(X_i(\cdot,y),X_i(\cdot,y))+d\rho^2(\tilde{X}_0(\cdot,y))\le K+K\rho^2(\cdot,O).$$

In either case, the bounds are independent of the y-variable. We apply Lemma 4.2 to each x_t^ε to obtain estimates that are uniform in ε: there exists a constant C such that for all $0\le s\le t\le T$ and for every $\varepsilon>0$, $\mathbf{E}\rho^2(x_s^\varepsilon,x_t^\varepsilon)\le C|t-s|$. Then use a chaining argument we obtain the following estimate for some positive constant α: $\mathbf{E}\left[\sup_{|s-t|\neq 0}\frac{\rho(x_t^\varepsilon,x_s^\varepsilon)}{|t-s|^\alpha}\right]<\infty$, this proves the tightness. Since $\mathbf{E}[\rho(x_t^\varepsilon,O)^2]$ is uniformly bounded, we see x_t has infinite lifetime and $\mathbf{E}[\rho(x_t,O)^2]$ is finite. From the uniform estimates $\mathbf{E}\rho^2(x_s^\varepsilon,x_t^\varepsilon)\le C(t-s)$ and $\mathbf{E}\rho^4(x_s^\varepsilon,x_t^\varepsilon)^2\le C(t-s)^2$, we easily obtain $\mathbf{E}\rho^2(\bar{x}_s,\bar{x}_t)\le C(t-s)$ and the other required estimates for $\bar{x}_s$. □.

Let us fix $x\in N$. For $t\ge s$, let $\phi_{s,t}^x(y)$ denote the solution to the equation

$$dz_t=\sum_{k=1}^{m_2} Y_k(x,z_t)\circ dW_t^k+Y_0(x,z_t)\,dt,\quad z_s=y. \tag{2}$$

Write $z_t^x=\phi_{0,t}^x(z_0)$, its Markov generator is $\mathscr{L}_0^x=\frac{1}{2}\sum_{k=1}^{m_1}(Y_i(x,\cdot))^2+Y_0(x,\cdot)$. Let $\phi_{s,t}^{\varepsilon,x}$ denote the solution flow to the SDE:

$$dy_t=\frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2} Y_k(x,y_t)\circ dW_t^k+\frac{1}{\varepsilon}Y_0(x,y_t)\,dt,\quad y_s=y_0 \tag{3}$$

Observe that the time changed solution flow $\phi^x_{\frac{s}{\varepsilon},\frac{t}{\varepsilon}}(\cdot)$ agrees with $\phi_{s,t}^{\varepsilon,x}(\cdot)$. On each sub-interval $[t_i,t_{i+1})$ we set

$$z_t^{x_{t_i}^\varepsilon}=\phi^{x_{t_i}^\varepsilon}_{\frac{t_i}{\varepsilon},t}(y_{t_i}^\varepsilon),\quad y_t^{x_{t_i}^\varepsilon}=\phi^{x_{t_i}^\varepsilon}_{t_i/\varepsilon,t/\varepsilon}(y_{t_i}^\varepsilon). \tag{4}$$

In the following locally uniform law of large numbers (LLN), any rate of convergence $\lambda(t)$ is allowed.

Assumption 3 (Locally Uniform LLN) *Suppose that there exists a family of probability measures μ_x on G which is continuous in the total variation norm. Suppose that for any smooth function $g: G\to\mathbb{R}$ and for any initial point $z_0\in G$ and $t_0\ge 0$,*

$$\left|\frac{1}{t}\mathbf{E}\int_{t_0}^{t+t_0} g\left(\phi_{t_0,s}^x(z_0)\right)ds-\int_G g(z)\mu_x(dz)\right|_{L_2(\Omega)}\le\alpha(x)\,\|g\|_s\,\lambda(t).$$

Here $\lambda(t)$ is a constant such that $\lim_{t\to\infty}\lambda(t)=0$, s is a non-negative number, and $\alpha(x)$ is a real number locally bounded in x.

Remark 4. In Proposition 3.10 we proved that if each $\mathscr{L}_x$ satisfies Hörmander's condition and if μ_x is the invariant probability measure for $\mathscr{L}_x$ (assume uniqueness), the locally uniform LLN holds with $\lambda(t)=\frac{1}{\sqrt{t}}$.

Suppose that $f: N\times G\to\mathbb{R}$ is bounded measurable, we define $\bar f(x)=\int_G f(x,z)\,\mu^x(dz)$.

Lemma 5.2 *Suppose the locally uniform LLN assumption. Let $f: N\times G\to\mathbb{R}$ be a smooth function with compact support (it is allowed to be independent of the first variable). Let $t_0=0<t_1<\cdots<t_N=T$ be a partition of equal size Δt_i. Then, for some number c,*

$$\mathbf{E}\sum_{i=0}^{N-1}\left|\int_{t_i}^{t_{i+1}} f\left(x_{t_i}^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right) ds-\Delta t_i\, f\left(x_{t_i}^\varepsilon\right)\right|\le c\,T\,\lambda\Big(\frac{\Delta t_i}{\varepsilon}\Big)\sup_{x\in D}\left\|f(x,\cdot)-\bar f(x)\right\|_s.$$

Proof. Set $\bar\alpha=\sup_{x\in D}\alpha(x)$ ant $C=\sup_{x\in D}\left\|f(x,\cdot)-\bar f(x)\right\|_s$, both are finite numbers by the assumptions on f and on $\alpha(x)$. Firstly we observe that

$$\left|\mathbf{E}\left\{\frac{\varepsilon}{\Delta t_i}\int_{\frac{t_i}{\varepsilon}}^{\frac{t_{i+1}}{\varepsilon}} f(x_{t_i}^\varepsilon, y_r^{x_{t_i}^\varepsilon})\,dr-\bar f(x_{t_i}^\varepsilon)\middle|\mathscr{F}_{t_i}\right\}\right|\le\alpha(x_{t_i}^\varepsilon)\,\lambda\Big(\frac{\Delta t_i}{\varepsilon}\Big)\,\chi_{x_{t_i}^\varepsilon\in D}\left\|f(x_{t_i}^\varepsilon,\cdot)-\bar f(x_{t_i}^\varepsilon)\right\|_s.$$

Summing up over i and making a time change we obtain that

$$\begin{aligned}\left|\mathbf{E}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} f\left(x_{t_i}^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right) dr-\Delta t_i\bar f(x_{t_i}^\varepsilon)\right| &= \sum_{i=0}^{N-1}\mathbf{E}\left|\varepsilon\int_{t_i/\varepsilon}^{t_{i+1}/\varepsilon} f\left(x_{t_i}^\varepsilon, z_r^{x_{t_i}^\varepsilon}\right) dr-\Delta t_i\bar f(x_{t_i}^\varepsilon)\right|\\ &\le\bar\alpha C\lambda\Big(\frac{\Delta t_i}{\varepsilon}\Big)\sum_{i=0}^{N-1}\Delta t_i,\end{aligned}$$

and thus conclude the proof. □

For the application of the LLN, we must ensure the size of the sub-interval to be sufficiently large and we should consider $\Delta t_i/\varepsilon$ to be of order ∞ as $\varepsilon\to 0$. Then we must ensure that $z_{\frac{t}{\varepsilon}}^{x_{t_i}^\varepsilon}=y_r^{x_{t_i}^\varepsilon}$ is an approximation for the fast variable y_t^ε on the sub-interval $[t_i,t_{i+1}]$. A crude counting shows that the distance of the two, beginning with the same initial value, is bounded above by $\frac{\Delta t_i}{\varepsilon}$. To obtain better estimates, we must choose the size of the interval carefully and use the slower evolutions of the slow variables on the sub-intervals and the Lipschitz continuity of the driving vector fields Y_i. We describe the intuitive idea for $\mathbb{R}^n\times\mathbb{R}^d$, assuming all vector fields are in BC^∞. We use the Lipschitz continuity of the vector fields $\frac{1}{\varepsilon}Y_i$. On $[0,r]$, we have a pre-factor of $\frac{1}{\varepsilon}$ from the stochastic integrals and $\frac{r}{\varepsilon}$ from the deterministic interval (by Holder's inequality). Then there exists a constant C such that

$$\mathbf{E}\left|y_r^\varepsilon - y_r^{x_{t_i}^\varepsilon}\right|^2 \le C(\frac{1}{\varepsilon}+\frac{\Delta t_i}{\varepsilon^2})\int_{t_i}^r \mathbf{E}\left|x_s^\varepsilon - x_{t_i}^\varepsilon\right|^2 ds + C(\frac{1}{\varepsilon}+\frac{\Delta t_i}{\varepsilon^2})\int_{t_i}^r \mathbf{E}\left|y_s^\varepsilon - y_s^{x_{t_i}^\varepsilon}\right|^2 ds.$$

By Lemma 4.2, $\mathbf{E}\left|x_s^\varepsilon - x_{t_i}^\varepsilon\right|^2 \le \tilde{C}\Delta t_i$ on $[t_i, t_{i+1}]$ where $\tilde{C}$ is a constant and so

$$\mathbf{E}\left|y_r^\varepsilon - y_r^{x_{t_i}^\varepsilon}\right|^2 \le C\tilde{C}\Delta t_i(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2})e^{C(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2})}.$$

If we take Δt_i to be of the order $\varepsilon|\ln\varepsilon|^a$ for a suitable $a>0$, then the above qunatity converges to zero uniformly in r as $\varepsilon\to 0$. See. e.g. [47, 33, 35, 90].

In the next lemma we give the statement and the details of the computation under our standard assumptions. In particular we assume that the sectional curvature of G is bounded. Let C, c, c' denote constants.

Lemma 5.3 *Let $0=t_0<t_1<\dots<t_N=T$ and $\varepsilon\in(0,1]$. Let*

$$\alpha_i^\varepsilon(C) := C\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)e^{C(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2})}\sup_{s\in[t_i,t_{i+1}]}\mathbf{E}\rho^2\left(x_s^\varepsilon, x_{t_i}^\varepsilon\right).$$

1. *Suppose Assumption 2. Then there exist constants c and C such that:*

$$\mathbf{E}\rho^2\left(y_r^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right) \le \alpha_i^\varepsilon(C) + c\sqrt{K}\left(\alpha_i^\varepsilon(C)\right)^{\frac{1}{2}}\frac{\Delta t_i}{\varepsilon}e^{c\frac{\Delta t_i}{\varepsilon}}$$

   where K is the bound on the sectional curvature of G.
2. *Suppose furthermore that there exists a constant c' such that*

$$\sup_{i=0,1,\dots N-1}\sup_{s,t\in[t_i,t_{i+1}]}\sup_{\varepsilon\in(0,1]}\mathbf{E}\rho^2(x_s^\varepsilon, x_t^\varepsilon) \le c'|t-s|.$$

   Then there exists a constant $C>0$ such that for every $\varepsilon\in(0,1]$,

$$\mathbf{E}\rho^2\left(y_r^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right) \le C\sqrt{\Delta t_i}\left(\frac{(\Delta t_i)^2}{\varepsilon^2}+\frac{(\Delta t_i)^3}{\varepsilon^3}\right)^{\frac{1}{2}}e^{C(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2})}, \quad \forall r\in[t_i,t_{i+1}], \forall i.$$

   In particular, if Δt_i is of the order $\varepsilon|\ln\varepsilon|^a$ where $a>0$, then $\sup_i\sup_{r\in[t_i,t_{i+1}]}\mathbf{E}\rho^2\left(y_r^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right)$ is of order ε^δ where $\delta\in(0,\frac{1}{2})$.

Proof. Since the sectional curvature of G is bounded above by K, its conjugate radius is bounded from below by $\frac{\pi}{\sqrt{K}}$. Let us consider a distance function on N that agrees with the Riemannian distance, which we denote by ρ, on the tubular neighbourhood of the diagonal of $N\times N$ with radius $\frac{\pi}{2\sqrt{K}}$. More precisely let $\tau:=\tau^\varepsilon$ be the first exit time when the distance between y_r^ε and $y_r^{x_{t_i}^\varepsilon}$ is greater than or equal to $A=\frac{\pi}{2\sqrt{K}}$. We use the identity

$$\mathbf{E}\rho^2\left(y^\varepsilon_{r\wedge\tau}, y^{x^\varepsilon_{t_i}}_{r\wedge\tau}\right) = \mathbf{E}\left[\rho^2\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right)\chi_{r<\tau}\right] + A^2 P(\tau \le r),$$

to obtain that

$$P(\tau \le r) \le \frac{1}{A^2}\mathbf{E}\left[\rho^2\left(y^\varepsilon_{r\wedge\tau}, y^{x^\varepsilon_{t_i}}_{r\wedge\tau}\right)\right].$$

Thus,

$$\mathbf{E}\rho^2\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right) \le \mathbf{E}\left[\rho^2\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right)\chi_{r<\tau}\right] + \mathbf{E}\left[\rho^4\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right)\chi_{r\ge\tau}\right]^{\frac{1}{2}}\sqrt{P(\tau \le r)}.$$

By the earlier argument, it is sufficient to estimate $\mathbf{E}\rho^2\left(y^\varepsilon_{r\wedge\tau}, y^{x^\varepsilon_{t_i}}_{r\wedge\tau}\right)$, and we will show that $\mathbf{E}\rho^2\left(y^\varepsilon_{r\wedge\tau}, y^{x^\varepsilon_{t_i}}_{r\wedge\tau}\right)$ converges to zero sufficiently fast as $\varepsilon \to 0$ to compensate with the possible divergence from the factor $\left(\mathbf{E}\rho^4\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right)\right)^{\frac{1}{2}}$.

On $\{r<\tau\}$, x, y are not on each other's cut locus, we may apply Itô's formula to the pair of stochastic processes $(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r)$ and obtain

$$\begin{aligned}\left[\rho(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r)\right]^2 &= \int_{t_i}^r d\rho^2\left(\frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2} Y_k(x^\varepsilon_s, y^\varepsilon_s)\circ dW^k_s + \frac{1}{\varepsilon}Y_0(x^\varepsilon_s, y^\varepsilon_s)\,ds\right)\\ &\quad + \int_{t_i}^r d\rho^2\left(\frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2} Y_k(x^\varepsilon_{t_i}, y^{x^\varepsilon_{t_i}}_s)\circ dW^k_s + \frac{1}{\varepsilon}Y_0(x^\varepsilon_{t_i}, y^{x^\varepsilon_{t_i}}_s)\,ds\right).\end{aligned}$$

Here the notation d in the first $d\rho^2$ refers to differentiation w.r.t. the first variable, as a gradient we use $\nabla^{(1)}(\rho^2)$, and the d in the second $d\rho^2$ is with respect to the second variable whose gradient is denoted by $\nabla^{(2)}(\rho^2)$. However $\nabla^{(1)}(\rho^2)(x,y) = -/\!/\nabla^{(2)}(\rho^2)(x,y)$, where $/\!/$ denotes the parallel translation of the relevant gradient vector along the geodesic from y to x. In the following let us denote by $d\rho^2$ the differential of ρ^2 w.r.t to the first variable. Using the assumption that each Y_k, $k = 1, \dots, m_2$, has bounded first order derivative, and the fact that $\nabla\rho$ and $\nabla^2\rho$ are bounded, the latter follows from the assumption that the sectional curvature is bounded, we see:

$$\left|d\rho^2(Y_k)(x^\varepsilon_s, y^\varepsilon_s) - d\rho^2(/\!/Y_k)\left(x^\varepsilon_{t_i}, y^{x^\varepsilon_{t_i}}_s\right)\right| \le 2\rho(y^\varepsilon_s, y^{x^\varepsilon_{t_i}}_s)\left(\rho(x^\varepsilon_s, x^\varepsilon_{t_i}) + \rho(y^\varepsilon_s, y^{x^\varepsilon_{t_i}}_s)\right).$$

It is useful to observe that Y_i is a vector field on G depending on $x \in N$, so the (product) distance function on $N \times G$ is needed for the estimate. On the other hand we only need to control the Hessian of the Riemannian distance on G and the assumption on the boundedness of the sectional curvature of G suffices.

A similar estimate applies to the first order differential involving $\tilde{Y}_0$, the sum of the Stratnovich correction for the stochastic integrals and Y_0. Again we use the assumption that each Y_k where k ranges from 1 to m_2 is bounded, and $\tilde{Y}_0$ has bounded first order covariant derivative. To summing up, for a constant C independent of ε and i, we have

$$\mathbf{E}\rho^2\left(y^\varepsilon_{r\wedge\tau}, y^{x^\varepsilon_{t_i}}_{r\wedge\tau}\right) \le C\left(\frac{1}{\varepsilon}+\frac{\Delta t_i}{\varepsilon^2}\right)\left(\mathbf{E}\int_{t_i}^{r\wedge\tau}\rho^2\left(x^\varepsilon_s, x^\varepsilon_{t_i}\right)ds + \mathbf{E}\int_{t_i}^{r\wedge\tau}\rho^2\left(y^\varepsilon_s, y^\varepsilon_s\right)ds\right).$$

Use Gronwall's inequality we obtain that,

$$\mathbf{E}\rho^2\left(y^\varepsilon_{r\wedge\tau}, y^{x^\varepsilon_{t_i}}_{r\wedge\tau}\right) \le C\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)\sup_{s\in[t_i,t_{i+1}]}\mathbf{E}\rho^2\left(x^\varepsilon_s, x^\varepsilon_{t_i}\right)e^{C\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)}. \tag{5}$$

We can now plug in the uniform estimates that $\mathbf{E}\rho^2\left(x^\varepsilon_s, x^\varepsilon_{t_i}\right) \le C|t_i - s|$ we see that

$$\mathbf{E}\rho^2\left(y^\varepsilon_{r\wedge\tau}, y^{x^\varepsilon_{t_i}}_{r\wedge\tau}\right) \le C\Delta t_i\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)e^{C\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)}.$$

Observe that the constant here is independent of ε, i and independent of $r \in [t_i, t_{i+1}]$. A similar estimates hold for $\mathbf{E}\rho^2\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right)\chi_{\tau>r}$.

On $\{r > \tau\}$ we use a more crude estimate, which we obtain without using estimates on the slow variables at time s and time t_i. It is sufficient to estimate $\mathbf{E}\rho^4\left(y^\varepsilon_r, y^\varepsilon_{t_i}\right)$ and $\mathbf{E}\rho^4\left(y^{x^\varepsilon_{t_i}}_r, y^\varepsilon_{t_i}\right)$. Observing that on $[t_i, t_{i+1}]$, the processes begin with the same initial point and the driving vector fields of the SDEs to which they are solutions are $\frac{1}{\varepsilon}Y_i(x^\varepsilon_r, \cdot)$ and $\frac{1}{\varepsilon}Y_i(x^\varepsilon_{t_i}, \cdot)$ respectively. We have assumed that $\sum_{k=1}^m |Y_k|$ and $\tilde{Y}_0$ are bounded. We then apply Lemma 4.2 to these SDEs. In Lemma 4.2 we take $K = \frac{c}{\varepsilon}$ where c is a constant. Then we have

$$\mathbf{E}\rho^4\left(y^{x^\varepsilon_{t_i}}_r, y^\varepsilon_{t_i}\right) + \mathbf{E}\rho^4\left(y^\varepsilon_r, y^\varepsilon_{t_i}\right) \le c\left(\Delta t_i + \frac{\Delta t_i}{\varepsilon}\right)e^{c\frac{\Delta t_i}{\varepsilon}}. \tag{6}$$

Again, the constant is independent of ε, i and independent of $r \in [t_i, t_{i+1}]$. We put the two estimates together to see that

$$\begin{aligned}\mathbf{E}\rho^2\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right) \le& C\Delta t_i\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)e^{C\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)}\\ &+\frac{2\sqrt{K}}{\pi}\left(C\Delta t_i\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)e^{C\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)}\right)^{\frac{1}{2}}\sqrt{c}\left(\Delta t_i+\frac{\Delta t_i}{\varepsilon}\right)^{\frac{1}{2}}e^{\frac{1}{2}c\frac{\Delta t_i}{\varepsilon}}.\end{aligned}$$

For ε small the first term is small. The second factor in the second term on the right hand side is large. We conclude that for another constant $\tilde{C}$,

$$\mathbf{E}\rho^2\left(y^\varepsilon_r, y^{x^\varepsilon_{t_i}}_r\right) \le \tilde{C}\sqrt{\Delta t_i}(1+\varepsilon)^{\frac{1}{2}}\left(\frac{(\Delta t_i)^2}{\varepsilon^2}+\frac{(\Delta t_i)^3}{\varepsilon^3}\right)^{\frac{1}{2}}e^{\tilde{C}\left(\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\right)}.$$

Let us suppose that $\Delta t_i \sim \varepsilon|\ln\varepsilon|^a$. Then the exponent $\frac{\Delta t_i}{\varepsilon}+\frac{(\Delta t_i)^2}{\varepsilon^2}\sim|\ln\varepsilon|^{2a}$. So for a constant C',

$$\mathbf{E}\rho^2\left(y_r^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right)\le C'\sqrt{\varepsilon}\,|\ln\varepsilon|^{2a}\,e^{\tilde C|\ln\varepsilon|^{2a}}.$$

The right hand side is of order ε^δ for $\delta<\frac{1}{2}$. We conclude the proof. □

The next lemma is on the convergence of Riemannian sums in the stochastic averaging procedure and the continuity of stochastic averages of a function with respect to a family of measures μ_x.

Lemma 5.4 *Suppose that for a sequence of numbers $\varepsilon_n\downarrow 0$, $x_\cdot^{\varepsilon_n}$ converges almost surely in $C([0,T];N)$ to a stochastic process $x_\cdot$. Suppose that there exists a constant $p\ge 1$ s.t. for $|s-t|$ sufficiently small,*

$$\mathbf{E}\left[\sup_{0\le r\le T}\rho^{2p}(x_r^\varepsilon,O)\right]<\infty,\quad \mathbf{E}\rho(x_s^\varepsilon,x_t^\varepsilon)^2\le C|t-s|,\quad \forall\varepsilon\in(0,1].$$

Let μ_x be a family of probability measures on G, continuous in x in the total variation norm. Let $f: N\times G\to\mathbb{R}$ be a BC^1 function. Let $0=t_0<t_1<\cdots<t_N=T$ and let $C_1=|f|_\infty K_2+|\nabla f|_\infty$. Then, the following statements hold:

(i)

$$\sup_{t\in[0,T]}\mathbf{E}\left|\int_G f(x_t^{\varepsilon_n},z)\,\mu^{x_t^{\varepsilon_n}}(dz)-\int_G f(\bar x_t,z)\,\mu^{\bar x_t}(dz)\right|\to 0.$$

In particular, the following converges in L^1,

$$\left|\int_0^t\int_G f(x_s^{\varepsilon_n},z)\,\mu^{x_s^{\varepsilon_n}}(dz)\,ds-\int_0^t\int_G f(\bar x_s,z)\,\mu^{\bar x_s}(dz)\,ds\right|\to 0.$$

(ii) The following convergence is uniform in ε:

$$\mathbf{E}\left|\sum_{i=0}^{N-1}\Delta t_i\int_G f(x_{t_i}^\varepsilon,z)\,\mu^{x_{t_i}^\varepsilon}(dz)-\int_0^T\int_G f(x_s^\varepsilon)\mu^{x_s^\varepsilon}(dz)ds\right|\to 0.$$

Consequently, the Riemannian sum $\sum_{i=0}^{N-1}\Delta t_i\int_G f(\bar x_{t_i},z)\,\mu^{\bar x_{t_i}}(dz)$ converges in L^1 to $\int_0^T\int_G f(\bar x_s,z)\mu^{\bar x_s}(dz)ds$.

Proof. Suppose that $x_\cdot^{\varepsilon_n}$ converges to $\bar x_\cdot$. We simplify the notation by assuming that $x^\varepsilon\to x$ almost surely. We may assume that N is not compact, the compact case is easier. Let D_n be a family of relatively compact open set such that $D_n\subset B_{a_n}\subset B_{a_{n+2}}\subset D_{n+1}$ where B_{a_n} is the geodesic ball centred at O of radius a_n where $a_n\to\infty$. This exists by a theorem of Greene and Wu. For any $t\in[0,T]$ and for any $\varepsilon\in(0,1]$,

$$\begin{aligned}
&\left|\int_G f(x_t^\varepsilon,z)\,\mu^{x_t^\varepsilon}(dz)-\int_G f(\bar x_t,z)\,\mu^{\bar x_t}(dz)\right|\\
&\le \int_G |f(x_t^\varepsilon,z)-f(\bar x_t,z)|\;\mu^{\bar x_t}(dz)+\left|\int_G f(\bar x_t,z)\mu^{\bar x_t}(dz)-\int_G f(\bar x_t,z)\mu^{x_t^\varepsilon}(dz)\right|\\
&\le |\nabla f|_\infty\,\rho(x_t^\varepsilon,\bar x_t)+|f|_\infty|\mu^{\bar x_t}-\mu^{x_t^\varepsilon}|_{TV}.
\end{aligned}$$

We have control over $\rho(x_t^\varepsilon,\bar x_t)$, it is bounded by $\rho(x_t^\varepsilon,O)$ and $\rho(\bar x_t,O)$. By the assumption, they are bounded in L^p, uniformly in $\varepsilon\in(0,1]$ and in $t\in[0,T]$. Similarly we also have uniform control over $P(\bar x_t\not\in D_n)$ and $P(x_t^\varepsilon\not\in D_n)$, they are bounded by $c\frac{1}{n}$ where c is a constant. We observe that

$$|\mu^{\bar x_t}-\mu^{x_t^\varepsilon}|_{TV}\le|\mu^{\bar x_t}-\mu^{x_t^\varepsilon}|_{TV}\chi_{\bar x_t\in D_n}\chi_{\bar x_t^\varepsilon\in D_n}+2(\chi_{\bar x_t\not\in D_n}+\chi_{x_t^\varepsilon\not\in D_n})$$

and there exists c_n such that $|\mu^{\bar x_t}-\mu^{x_t^\varepsilon}|_{TV}\chi_{\bar x_t\in D_n}\chi_{\bar x_t^\varepsilon\in D_n}\le c_n\rho(x_t^{\varepsilon_n},\bar x_t)$. We take n large, so that $P(\bar x_t\not\in D_n)$ and $P(x_t^\varepsilon\not\in D_n)$ are as small as we want. Then for n fixed we see that the $c_n\rho(x_t^{\varepsilon_n},\bar x_t)$ converges, as $\varepsilon\to0$, in L^1. Thus, $\sup_{0\le t\le T}\mathbf{E}|\mu^{\bar x_t}-\mu^{x_t^\varepsilon}|\to0$ and

$$\sup_{0\le t\le T}\mathbf{E}\left|\int_G f(x_t^{\varepsilon_n},z)\,\mu^{x_t^{\varepsilon_n}}(dz)-\int_G f(\bar x_t,z)\,\mu^{\bar x_t}(dz)\right|$$

converges to zero. This proves part (i). Since $\mathbf{E}\rho(x_s^\varepsilon,x_{t_i}^\varepsilon)^2\le C|t_{i+1}-t_i|$,

$$\begin{aligned}
&\mathbf{E}\left|\sum_{i=0}^{N-1}\Delta t_i\int_G f(x_{t_i}^\varepsilon,z)\,\mu^{x_{t_i}^\varepsilon}(dz)-\int_0^T\int_G f(x_s^\varepsilon)\mu^{x_s^\varepsilon}(dz)ds\right|\\
&\le T\,|\nabla f|_\infty\sup_{s\in[t_i,t_{i+1})}\mathbf{E}[\rho(x_s^\varepsilon,x_{t_i}^\varepsilon)]+|f|_\infty T\sup_{s\in[t_i,t_{i+1})}\mathbf{E}\left[|\mu^{x_s^\varepsilon}-\mu^{x_{t_i}^\varepsilon}|_{TV}\right]\to0.
\end{aligned}$$

The convergence can be proved, again by breaking the total variation norm into two parts, in one part the processes are in D_n, and in the other part they are not. Since x_t^ε converges to x_t as a stochastic process on $[0,T]$, we also have that $\mathbf{E}\rho(x_s,x_{t_i})\le C|t_{i+1}-t_i|$. We apply the same argument to $\bar x_t$ to obtain that

$$\mathbf{E}\left|\sum_{i=0}^{N-1}\Delta t_i\int_G f(\bar x_{t_i},z)\,\mu^{\bar x_{t_i}}(dz)-\int_0^T\int_G f(\bar x_s)\mu^{\bar x_s}(dz)ds\right|\to0.$$

This concludes the proof. □

Suppose we assume furthermore that there exists a constant K such that

$$|\mu^{x_1}-\mu^{x_2}|_{TV}\le K(1+\rho(x_1,O)+\rho(x_2,O))\rho(x_1,x_2).$$

Then explicit estimates can be made for the convergence in Lemma 5.4, e.g.

$$\left| \int_G f(x_t^\varepsilon, z)\, \mu^{x_t^\varepsilon}(dz) - \int_G f(\bar{x}_t, z)\, \mu^{\bar{x}_t}(dz) \right|$$
$$\leq |\nabla f|_\infty\, \rho(x_t^\varepsilon, \bar{x}_t) + |f|_\infty K(1 + \rho^p(\bar{x}_t, O) + \rho^p(x_t^\varepsilon, O))\rho(x_t^\varepsilon, \bar{x}_t).$$

To this we may apply Hölder's inequality and obtain:

$$\mathbf{E}\left| \int_0^T \int_G f(x_t^{\varepsilon_n}, z)\, \mu^{x_t^{\varepsilon_n}}(dz)\, dt - \int_0^T \int_G f(\bar{x}_t, z)\, \mu^{\bar{x}_t}(dz)\, dt \right|$$
$$\leq |\nabla f|_\infty\, \mathbf{E} \int_0^T \rho(x_t^\varepsilon, \bar{x}_t)\, dt + |f|_\infty K \mathbf{E} \left| \int_0^T (1 + \rho^p(\bar{x}_t, O) + \rho^p(x_t^\varepsilon, O))\rho(x_t^\varepsilon, \bar{x}_t)\, dt \right|$$
$$\leq |\nabla f|_\infty\, T \mathbf{E} \sup_{s \leq t} \rho(x_t^\varepsilon, \bar{x}_t) + |f|_\infty K T \sqrt{\mathbf{E} \sup_{t \leq T} (1 + \rho^p(\bar{x}_t, O) + \rho^p(x_t^\varepsilon, O))^2} \sqrt{\mathbf{E} \sup_{t \leq T} \rho^2(x_t^\varepsilon, \bar{x}_t)}.$$

In the proposition below we are interested in the time average concerning a product function $f_1 f_2$, where $f_1 : N \to \mathbb{R}$ is C^∞ with has compact support and $f_2 : G \to \mathbb{R}$ is smooth.

Proposition 5.5 *Suppose the following conditions.*

(1) μ_x is a family of probability measures on G for which the locally uniform LLN assumption (Assumption 3) holds.
(2) Assumption 2.
(3) There exist constants $p \geq 1$ and c such that for $s, t \in [r_1, r_2]$ where $r_2 - r_1$ is sufficiently small,

$$\sup_{\varepsilon \in (0,1]} \sup_{s,t \in [r_1, r_2]} \mathbf{E}\rho^2(x_s^\varepsilon, x_t^\varepsilon) \leq c|t - s|, \qquad \sup_{0 \leq s \leq T} \sup_{\varepsilon \in (0,1]} \mathbf{E}\rho^{2p}(x_s^\varepsilon, O) < \infty.$$

(4) ε_n is a sequence of numbers converging to 0 with $\sup_{t \leq T} \rho(x_t^{\varepsilon_n}, \bar{x}_t)$ converges to zero almost surely.
(5) Let $f : N \times G \to \mathbb{R}$ be a smooth and globally Lipschitz continuous function. Suppose that either f is independent of the first variable or for each $y \in G$, the support of $f(\cdot, y)$ is contained in a compact set D.

Then the following random variables converge to zero in L^1:

$$\int_0^T f(x_s^\varepsilon, y_s^\varepsilon) ds - \int_0^T \int_G f(\bar{x}_s, z)\, \mu^{\bar{x}_s}(dz)\, ds.$$

Proof. Let $0 = t_0 < t_1 < \cdots < t_N = T$ and let $\Delta t_i = t_{i+1} - t_i$. Then, recalling the notation given in (4),

$$\int_0^T f(x_s^\varepsilon, y_s^\varepsilon)ds = \sum_{n=0}^{N-1}\int_{t_i}^{t_{i+1}} f(x_s^\varepsilon, y_s^\varepsilon)ds$$
$$= \sum_{n=0}^{N-1}\int_{t_i}^{t_{i+1}} \left[f(x_s^\varepsilon, y_s^\varepsilon) - f(x_{t_i}^\varepsilon, y_s^\varepsilon)\right] ds + \sum_{n=0}^{N-1}\int_{t_i}^{t_{i+1}} \left[f(x_{t_i}^\varepsilon, y_s^\varepsilon) - f\left(x_{t_i}^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right)\right] ds$$
$$+ \sum_{n=0}^{N-1}\left[\int_{t_i}^{t_{i+1}} f\left(x_{t_i}^\varepsilon, y_r^{x_{t_i}^\varepsilon}\right) ds - \Delta t_i \int_G f(x_{t_i}^\varepsilon, z)\mu^{x_{t_i}^\varepsilon}(dz)\right]$$
$$+ \left[\sum_{n=0}^{N-1}\Delta t_i \int_G f(x_{t_i}^\varepsilon, z)\mu^{x_{t_i}^\varepsilon}(dz) - \int_0^T \int_G f(x_s^\varepsilon, z)\mu^{x_s^\varepsilon}(dz)\,ds\right]$$
$$+ \left[\int_0^T \int_G f(x_s^\varepsilon, z)\mu^{x_s^\varepsilon}(dz)\,ds - \int_0^T \int_G f(\bar{x}_s, z)\mu^{\bar{x}_s}(dz)\,ds\right] + \int_0^T \int_G f(\bar{x}_s, z)\mu^{\bar{x}_s}(dz)\,ds.$$

Using the fact that f is Lipschitz continuous in the first variable and the assumptions on the moments of $\rho(x_t^\varepsilon, x_s^\varepsilon)$ we see that for a constant K,

$$\sum_{i=0}^{N-1}\mathbf{E}\int_{t_i}^{t_{i+1}} \left|f(x_r^\varepsilon, y_r^\varepsilon) - f(x_{t_i}^\varepsilon, y_r^\varepsilon)\right| dr \le K\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \mathbf{E}\rho(x_r^\varepsilon, x_{t_i}^\varepsilon)\,dr$$
$$\le K\,T\mathbf{E}\rho(x_r^\varepsilon, x_{t_i}^\varepsilon) \le TKc\max_i \sqrt{\Delta t_i}.$$

By choosing $\Delta t_i = o(\varepsilon)$ we see that the first term on the right hand side converges to zero. The converges of the second term follows directly from Lemma 5.3 by choosing $\Delta t_i \sim \varepsilon|\ln\varepsilon|^a$ where $a > 0$ and Assumption 2. By Lemma 5.2 and Assumption 3, the third term converges if we choose $\frac{\varepsilon}{\Delta t_i} = o(\varepsilon)$. The convergence of the fourth and fifth terms follow respectively from part (i) and part (ii) of Lemma 5.4. □

We are now ready to prove the main averaging theorem, Theorem 2 in section 1.1. This proof has the advantage for being concrete, from this an estimate for thew rate of convergence is also expected.

Theorem 5.6 *Suppose the following statements hold.*

(a) Assumptions 1 and 2.
(b) There exists a family of probability measures μ_x on G for which the locally uniform LLN assumption (Assumption 3) holds.

Then, as $\varepsilon \to 0$, the family of stochastic processes $\{x_t^\varepsilon, \varepsilon > 0\}$ converges weakly on any compact time intervals to a Markov process with generator $\bar{\mathscr{L}}$.

Proof. By Prohorov's theorem, a set of probability measures is tight if and only if its relatively weakly compact, i.e. every sequence has a sub-sequence that converges weakly. It is therefore sufficient to prove that every limit process of the stochastic processes x_t^ε is a Markov process with the same Markov generator. Every sequence of weakly convergent stochastic processes on an interval $[0,T]$ can be realised on a probability space as a sequence of stochastic processes that converge almost surely on $[0,T]$ with respect to the supremum norm in time. It is sufficient to prove that if a subsequence $\{x_t^{\varepsilon_n}\}$ converges almost surely on $[0,T]$, the limit is a Markov process with generator $\bar{\mathscr{L}}$. For this we apply Stroock-Varadhan's martingale method [83, 75]. To ease notation we may assume that the whole family x_t^ε converges almost surely. Let f be a real valued smooth function on N with compact support. Let $\bar{x}_t$ be the limit Markov process. We must prove that $f(\bar{x}_t) - f(x_0) - \int_0^t \bar{\mathscr{L}} f(\bar{x}_r) dr$ is a martingale. In other words we prove that for any bounded measurable random variable $G_s \in \mathscr{F}_s$ and for any $s < t$, $\mathbf{E}\left(G_s(f(\bar{x}_t) - f(\bar{x}_s) - \int_s^t \bar{\mathscr{L}} f(x_r) dr)\right) = 0$. On the other hand, for each $\varepsilon > 0$,

$$f(x_t^\varepsilon) - f(x_s^\varepsilon) - \int_s^t \left(\frac{1}{2} \sum_{i=1}^{m_1} (X_i(\cdot, y_r^\varepsilon))^2 f + X_0(\cdot, y_r^\varepsilon) f \right)(x_r^\varepsilon)\, dr$$

is a martingale. Let us introduce the notation:

$$F(x_r^\varepsilon, y_r^\varepsilon) = \left(\frac{1}{2} \sum_{i=1}^{m_1} (X_i(\cdot, y_r^\varepsilon))^2 f + X_0(\cdot, y_r^\varepsilon) f \right)(x_r^\varepsilon).$$

Since x_t^ε converges to $\bar{x}_s$ it is sufficient to prove that as $\varepsilon \to 0$,

$$\mathbf{E}\left[G_s \left(\int_s^t F(x_r^\varepsilon, y_r^\varepsilon) dr - \int_s^t \bar{\mathscr{L}} f(x_r) dr \right) \right] \to 0.$$

Even simpler we only need to prove that $\int_s^t F(x_r^\varepsilon, y_r^\varepsilon) dr$ converges to $\int_s^t \bar{\mathscr{L}} f(x_r) dr$ in L^1. Under Assumption 1, we may apply Lemma 5.1 from which we see that conditions (3) and (4) of Proposition 5.5 hold. Since f has compact support, F has compact support in the first variable. We may apply Proposition 5.5 to the function F to complete the proof. □

We remark that the locally uniform law of large numbers hold if G is compact, if $\mathscr{L}_x$ satisfies *strong Hörmander's condition*, or if $\mathscr{L}_x$ satisfies Hörmander's condition with the additional assumption that $\mathscr{L}_x$ has a unique invariant probability measure.

We obtain the following Corollary.

Corollary 1. *Let G be compact. Suppose Assumptions 1 and 2. Suppose that $\mathscr{L}_x$ satisfies Hörmander's condition and that it has a unique invariant probability measure. Then $\{x_t^\varepsilon, \varepsilon > 0\}$ converges weakly, on any compact time intervals, to a Markov process with generator $\bar{\mathscr{L}}$.*

From the proof of Theorem 5.6, the Markov generator $\bar{\mathscr{L}}$ given below.

$$\bar{\mathscr{L}} f(x) = \int_G \left(\frac{1}{2} \sum_{i=1}^{m_1} X_i^2(\cdot, y) f + X_0(\cdot, y) f \right)(x)\, \mu_x(dy). \tag{7}$$

Appendix

It is possible to write the operator $\bar{\mathscr{L}}$ given by (7) as a sum of squares of vector fields. For this we need an auxiliary family of vector fields $\{E_1,\dots,E_{n_1}\}$ with the property that at each point x they span the tangent space T_xN. Let us write each vector field $X_i(\cdot,y)$ in this basis and denote by $X_i^k(\cdot,y)$ its coordinate functions, so $X_i(x,y)=\sum_{k=1}^{n_1}X_i^k(x,y)E_k(x)$. Set

$$a_{k,l}(x,y)=\sum_{i=1}^{m_1}X_i^k(x,y)X_i^l(x,y),$$
$$b_0^k(x,y)=\frac{1}{2}\sum_{l=1}^{n_1}\sum_{i=1}^{m_1}X_i^l(x,y)\left(\nabla X_k^l(\cdot,y))(E_l(x))\right)+X_0^k(x,y),$$

where ∇ denotes differentiation with respect to the first variable.We observe that

$$\frac{1}{2}\sum_{i=1}^{m_1}(X_i(\cdot,y)^2f)(x)+(X_0(\cdot,y)f)(x)=\frac{1}{2}\sum_{k,l=1}^{n_1}a_{k,l}(x,y)(E_kE_lf)(x)+\sum_{k=1}^{n_1}b_0^k(x,y)(E_kf)(x).$$

If μ_x is a family of probability measures on G, we set

$$\bar{\mathscr{L}}=\frac{1}{2}\sum_{k,l=1}^{n_1}\left(\int_G a_{k,l}(x,y)\mu_x(dy)\right)E_kE_l+\sum_{k=1}^{n_1}\left(\int_G b_0^k(x,y)dy\right)E_k. \tag{8}$$

The auxiliary vector fields can be easily constructed. For example, we may use the gradient vector fields coming from an isometric embedding $i:N\to\mathbb{R}^{n_1}$. Then they have the following properties. For $e\in\mathbb{R}^{n_1}$, we define $E(x)(e)=\sum_{i=1}^{n_1}E_i(x)e_i$ where $\{e_i\}$ is an orthonormal basis of $\mathbb{R}^{n_1}$. Then $\mathbb{R}^{n_1}$ has a splitting of the form $\ker[X(x)]^\perp\oplus X(x)$ and $X(e)$ has vanishing derivative for $e\in\ker[X(x)]^\perp$. We may also use a 'moving frames' instead of the gradient vector fields. This is particularly useful if N is an Euclidean space, or a compact space, or a Lie group. For such spaces and their moving frames, the assumption that $X_1,\dots,X_k$ and their two order derivatives, X_0 and ∇X_0 are bounded can be expressed by the boundedness of the functions $a_{k,l}$ and b_0^k and their derivatives.

6 Re-visit the examples

6.1 A dynamical description for hypo-elliptic diffusions

Let us consider two further generalisations to the dynamical theory for Brownian motions described in §2.1. Both cases allow degeneracy in the fast variables. One of which has the same type of reduced random ODE and is closer to Theorem 2A. We state this one first and will take M compact for simplicity.

Proposition 6.1 *Let M be compact. Suppose that in (1), we replace the orthonormal basis $\{A_1,\dots,A_N\}$ and A_0 by a vectors $\{A_1,\dots A_{m_2}\}\subset \mathfrak{so}(n)$ with the property that these vectors together with their commutators generates $\mathfrak{so}(n)$. (Take $A_0=0$ for simplicity. Then, as $\varepsilon\to 0$, the rescaled position stochastic processes, $x^{\varepsilon}_{\frac{t}{\varepsilon}}$, converges to a scaled Brownian motion. Their horizontal lifts from u_0 converge also.*

Proposition 6.2 *The scale is determined by the eigenvalues of the symmetry matrix $\sum_{i=1}^{m_2}(A_i)^2$.*

Proof. The reduced equation is as before:

$$\begin{cases} \dfrac{d}{dt}\tilde{x}^{\varepsilon}_t = H_{\tilde{x}^{\varepsilon}_t}(g_{\frac{t}{\varepsilon}}e_0), \quad \tilde{x}^{\varepsilon}_0=u_0,\\ dg_t=\displaystyle\sum_{k=1}^{m_2} g_tA_k\circ dw^k_t+g_tA_0\,dt, \quad g_0=Id. \end{cases}$$

We observe that the operator $\sum_{i=1}^{m_2}(A_i^*)^2+A_0^*$ satisfies Hörmander's condition and has a unique invariant probability measure. It is symmetric w.r.t the bi-invariant Haar measure dg, and the only invariant measure is dg. Then we apply Theorem 6.4 from [66] to conclude.

Suppose instead we consider the following SDE, in which the horizontal part involves a stochastic integral

$$\begin{cases} du^{\varepsilon}_t=H_{u^{\varepsilon}_t}(e_0)\,dt+\displaystyle\sum_{j=1}^{m_1}H(u^{\varepsilon}_t)(e_j)\circ dB^j_t+\frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^{m_2}A_k^*(u^{\varepsilon}_t)\circ dW^k_t+A_0^*(u^{\varepsilon}_t)\,dt,\\ u^{\varepsilon}_0=u_0. \end{cases} \tag{1}$$

where $e_j\in\mathbb{R}^n$.

Proposition 6.3 *Suppose that M has bounded sectional curvature. Suppose that $\{A_0,A_1,\dots,A_{m_2}\}$ and their iterated brackets (commutators) generate the vector space $\mathfrak{so}(n)$. Suppose that $\{e_1,\dots,e_{m_1}\}$ is an orthonormal set. Then as $\varepsilon\to 0$, the position component of u^{ε}_t, x^{ε}_t, converges to a rescaled Brownian motion, scaled by $\frac{m_1}{n}$ where $n=\dim(M)$. Their horizontal lifts converge also to a horizontal Brownian motion with the same scale.*

Proof. Set $x^{\varepsilon}_t=\pi(u^{\varepsilon}_t)$, where π takes an frame to its base point. Then x^{ε}_t is the position process. Then

$$dx^{\varepsilon}_t=\sum_{i=1}^{m_1}u^{\varepsilon}_t(e_i)\circ dB^i_t+u^{\varepsilon}_te_0\,dt.$$

Let $\tilde{x}^{\varepsilon}_t$ denote the stochastic horizontal lifts of x^{ε}_t. Then from the nature of the horizontal vector fields and the horizontal lifts, this procedure introduces a twist to the Euclidean vectors e_i. If g_t solves:

$$dg_t = \sum_{k=1}^{m_2} g_t A_k \circ dw_t^k + g_t A_0\, dt$$

with initial value the identity, then x_t^ε satisfies the equation

$$d\tilde{x}_t^\varepsilon = \mathfrak{h}_{\tilde{x}_t^\varepsilon} dx_t^\varepsilon = H(\tilde{x}_t^\varepsilon)(g_{\frac{t}{\varepsilon}} e_0)dt + \sum_{i=1}^{m_1} H(\tilde{x}_t^\varepsilon)(g_{\frac{t}{\varepsilon}} e_i) \circ dB_t^i.$$

Since g_t does not depend on the slow variable, the conditions of the Theorem is satisfied provided M has bounded sectional curvature.

The limiting process, in this case, will not be a fixed point. It is a Markov process on the orthonormal frame bundle with generator

$$\bar{\mathscr{L}} f(u) = \sum_{i=1}^{m_1} \int_{SO(n)} \nabla df(H(u)(ge_i), H(u)(ge_i))dg,$$

where ∇ is a flat connection, ∇H_i vanishes. Let dg denote the normalised bi-invariant Haar measure. Using this connection and an orthonormal basis $\{e_i\}$ of $\mathbb{R}^n$, extending our orthonormal set $\{e_1, \dots, e_{m_1}\}$ we see that

$$\bar{\mathscr{L}} f(u) = \sum_{k,l=1}^{n} \nabla df(H(E_k), H(E_l)) \sum_{i=1}^{m_1} \int_{SO(n)} \langle e_k, ge_i\rangle \langle ge_i, e_l\rangle dg.$$

It is easy to see that

$$\sum_{i=1}^{m_1} \int_{SO(n)} \langle e_k, ge_i\rangle \langle ge_i, e_l\rangle dg = \sum_{i=1}^{m_1} \delta_{k,l} \int_{SO(n)} \langle e_k, ge_i\rangle^2 dg = \frac{m_1}{n} \delta_{k,l}.$$

This means that

$$\bar{\mathscr{L}} f(u) = \frac{m_1}{n} \sum_{k,l=1}^{n} \nabla df(H(e_k), H(e_k)).$$

Thus the $\bar{\mathscr{L}}$ diffusion has Markov generator $\frac{1}{2}\frac{m_1}{n}\Delta^H$ where Δ^H is the horizontal diffusion and which means that $\pi(u_t^\varepsilon)$ converges to a scaled Brownian motion as we have guessed. □

Problem 2. The vertical vector fields in (1) are left invariant. Instead of left invariant vertical vector fields we may take more general vector fields and consider the following SDEs. Let $f : OM \to \mathbb{R}$ be smooth functions, $e_j \in \mathbb{R}^n$ are unit vectors. Let us consider the equation

$$\begin{cases} du_t^\varepsilon = H_{u_t^\varepsilon}(e_0)\, dt + \sum_{j=1}^{m_1} H(u_t^\varepsilon)(e_j) \circ dB_t^i + \frac{1}{\sqrt{\varepsilon}} \sum_{k=1}^{m_2} f_k(u_t^\varepsilon) A_k^*(u_t^\varepsilon) \circ dW_t^k + A_0^*(u_t^\varepsilon)\, dt, \\ u_0^\varepsilon = u_0. \end{cases} \tag{2}$$

Then the horizontal lift of its position processes will, in general, depend on the slow variables. It would be interesting to determine explicit conditions on f_k for which the averaging procedure is valid and if so what is the effective limit?

6.2 Inhomogeneous scaling of Riemannian metrics

Returning to section 2.3 we pose the following problem.

Problem 3. With Theorem 5.6, we can now study a fully coupled system:

$$dg_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}} \sum_{k=1}^{m_2} (a_k A_k)(g_t^\varepsilon) \circ dB_t^k + \frac{1}{\varepsilon}(a_0 A_0)(g_0^\varepsilon) dt + (b_0 Y_0)(g_t^\varepsilon) dt + \sum_{k=1}^{m_1} (b_k Y_k)(g_t^\varepsilon) \circ dW_t^k,$$

where a_k, b_k are smooth functions. It would be interesting to study the convergence of the slow variables, vanishing of the averaged processes, and the nature of the limits in terms of a_k and b_k.

6.3 An averaging principle on principal bundles

We return to the example in section 2.4. In the following proposition, ∇ denotes the flat connection on the principal bundle P.

Proposition 6.4 *Let G be a compact Lie group and dg its Haar measure. Assume that M has bounded sectional curvature. Suppose that $\mathscr{L}_u$ satisfies Hörmander's condition and has a unique invariant probability measure. Suppose that θ_k^j are bounded with bounded derivatives. Define*

$$a_{i,j}(u) = \int_G \sum_{l=1}^{m_1} \langle X_l(u,g), H_i(u)\rangle \langle X_l(ug), H_j(u)\rangle \, dg,$$
$$b(u) = \int_G \left(\frac{1}{2} \sum_{l=1}^{m_1} \nabla_{X_l} X_l(ug) + X_0(ug) \right) dg.$$

Then $\tilde{x}^\varepsilon_{\frac{t}{\varepsilon}}$ converges weakly to a Markov process on P with the Markov generator

$$\bar{\mathscr{L}} f(u) = df(b(u)) + \frac{1}{2} \sum_{i,j=1}^{n} a_{i,j}(u) \nabla df(H_i(u), H_j(u)).$$

Proof. The convergence is a trivial consequence of Theorem 5.6. To identify the limit let $f : P \to \mathbb{R}$ be any smooth function with compact support. Then

$$f(\tilde{x}_t^\varepsilon) = f(g_0) + \sum_{l=1}^{m_1} \int_0^t df\,(X_l((\tilde{x}_s^\varepsilon g_s^\varepsilon))\,dB_s^l + \sum_{l=1}^{m_1} \int_0^t \nabla df\,(X_l(\tilde{x}_s^\varepsilon g_s^\varepsilon), X_l(\tilde{x}_s^\varepsilon g_s^\varepsilon))\,ds$$
$$+ \sum_{l=1}^{m_1} \int_0^t df\left(\nabla_{X_l} X_l(\tilde{x}_s^\varepsilon g_s^\varepsilon) + X_0(\tilde{x}_s^\varepsilon g_s^\varepsilon)\right) ds.$$

Finally we take coordinates of X_l w.r.t the parallel vector fields H_i, c.f. the Appendix of §5, to complete the proof.

Conclusions and Other Open Questions. In conclusion, the examples we studied treat some of the simplest and yet universal models, they can be studied using the method we have just developed. Even for these simple models many questions remain to be answered, including the questions stated in §1.1, §2.1 and §2.3. For example we do not know the geometric nature of the limiting object. Concerning Theorem 5.6, we expect the conditions of the theorem improved for more specific examples of manifolds, and expect an upper bound for the rate of convergence if the resolvents of the operators $\mathscr{L}_x$ is bounded in x and if the rank of the operators and their quadratic forms are bounded, and also expect an averaging principle for slow-fast SDEs driven by Lévy processes, c.f. [49].

References

1. Sergio Albeverio, Alexei Daletskii, and Alexander Kalyuzhnyi. Random Witten Laplacians: traces of semigroups, L^2-Betti numbers and index. *J. Eur. Math. Soc. (JEMS)*, 10(3):571–599, 2008.
2. Jürgen Angst, Ismaël Bailleul, and Camille Tardif. Kinetic Brownian motion on Riemannian manifolds. *Electron. J. Probab.*, 20:no. 110, 40, 2015.
3. Marc Arnaudon. Semi-martingales dans les espaces homogènes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(2):269–288, 1993.
4. V. I. Arnolʹd. *Mathematical methods of classical mechanics*. Springer-Verlag, New York, 1989.
5. M. F. Atiyah. Algebraic topology and operators in Hilbert space. In *Lectures in Modern Analysis and Applications. I*, pages 101–121. Springer, Berlin, 1969.
6. Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
7. Vlad Bally and Lucia Caramellino. Asymptotic development for the CLT in total variation distance. *Bernoulli*, 22(4):2442–2485, 2016.
8. Florent Barret and Max von Renesse. Averaging principle for diffusion processes via Dirichlet forms. *Potential Anal.*, 41(4):1033–1063, 2014.
9. Fabrice Baudoin. *An introduction to the geometry of stochastic flows*. Imperial College Press, London, 2004.
10. Fabrice Baudoin, Martin Hairer, and Josef Teichmann. Ornstein-Uhlenbeck processes on Lie groups. *J. Funct. Anal.*, 255(4):877–890, 2008.
11. Lionel Bérard-Bergery and Jean-Pierre Bourguignon. Laplacians and Riemannian submersions with totally geodesic fibres. *Illinois J. Math.*, 26(2):181–200, 1982.
12. Nils Berglund and Barbara Gentz. *Noise-induced phenomena in slow-fast dynamical systems*. Probability and its Applications (New York). Springer-Verlag London, Ltd., London, 2006. A sample-paths approach.

13. Nicole Berline, Ezra Getzler, and Michèle Vergne. *Heat kernels and Dirac operators*. Springer-Verlag, New York, 1992.
14. Jeremiah Birrell, Scott Hottovy, Giovanni Volpe, and Jan Wehr. Small mass limit of a Langevin equation on a manifold. *Ann. Henri Poincaré*, 18(2):707–755, 2017.
15. Jean-Michel Bismut. The hypoelliptic Laplacian on a compact Lie group. *J. Funct. Anal.*, 255(9):2190–2232, 2008.
16. Jean-Michel Bismut and Gilles Lebeau. Laplacien hypoelliptique et torsion analytique. *C. R. Math. Acad. Sci. Paris*, 341(2):113–118, 2005.
17. A. N. Borodin. A limit theorem for the solutions of differential equations with a random right-hand side. *Teor. Verojatnost. i Primenen.*, 22(3):498–512, 1977.
18. A. N. Borodin and M. I. Freidlin. Fast oscillating random perturbations of dynamical systems with conservation laws. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(3):485–525, 1995.
19. Thomas Cass and Peter Friz. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)*, 171(3):2115–2141, 2010.
20. R. Catellier and M. Gubinelli. Averaging along irregular curves and regularisation of ODEs. *Stochastic Process. Appl.*, 126(8):2323–2366, 2016.
21. Bennett Chow, Sun-Chin Chu, David Glickenstein, Christine Guenther, James Isenberg, Tom Ivey, Dan Knopf, Peng Lu, Feng Luo, and Lei Ni. *The Ricci flow: techniques and applications. Part III. Geometric-analytic aspects*, volume 163 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2010.
22. D. Crisan and M. Ottobre. Pointwise gradient bounds for degenerate semigroups (of UFG type). *Proc. A.*, 472(2195):20160442, 23, 2016.
23. Jean Dolbeault, Clément Mouhot, and Christian Schmeiser. Hypocoercivity for linear kinetic equations conserving mass. *Trans. Amer. Math. Soc.*, 367(6):3807–3828, 2015.
24. Dmitry Dolgopyat, Vadim Kaloshin, and Leonid Koralov. Sample path properties of the stochastic flows. *Ann. Probab.*, 32(1A):1–27, 2004.
25. R. M. Dowell. *Differentiable Approximations to Brownian Motion on Manifolds*. PhD thesis, University of Warwick, 1980.
26. Manh Hong Duong, Agnes Lamacz, Mark A. Peletier, and Upanshu Sharma. Variational approach to coarse-graining of generalized gradient flows. *Calc. Var.*, 56:100, 2017. Published first online.
27. Weinan E. *Principles of multiscale modeling*. Cambridge University Press, Cambridge, 2011.
28. J.-P. Eckmann and M. Hairer. Uniqueness of the invariant measure for a stochastic PDE driven by degenerate noise. *Comm. Math. Phys.*, 219(3):523–565, 2001.
29. K. D. Elworthy. *Stochastic differential equations on manifolds*, volume 70 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1982.
30. K. D. Elworthy, Y. Le Jan, and Xue-Mei Li. *On the geometry of diffusion operators and stochastic flows*, volume 1720 of *Lecture Notes in Mathematics*. Springer-Verlag, 1999.
31. K. David Elworthy, Yves Le Jan, and Xue-Mei Li. *The geometry of filtering*. Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2010.
32. Michel Émery. *Stochastic calculus in manifolds*. Universitext. Springer-Verlag, Berlin, 1989. With an appendix by P.-A. Meyer.
33. M. I. Freidlin. The averaging principle and theorems on large deviations. *Uspekhi Mat. Nauk*, 33(5(203)):107–160, 238, 1978.
34. M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984. Translated from the Russian by Joseph Szücs.
35. Mark I. Freidlin and Alexander D. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, third edition, 2012. Translated from the 1979 Russian original by Joseph Szücs.
36. Peter Friz, Paul Gassiat, and Terry Lyons. Physical Brownian motion in a magnetic field as a rough path. *Trans. Amer. Math. Soc.*, 367(11):7939–7955, 2015.
37. Hongbo Fu and Jinqiao Duan. An averaging principle for two-scale stochastic partial differential equations. *Stoch. Dyn.*, 11(2-3):353–367, 2011.

38. Hongbo Fu and Jicheng Liu. Strong convergence in stochastic averaging principle for two time-scales stochastic partial differential equations. *J. Math. Anal. Appl.*, 384(1):70–86, 2011.
39. Ivan I. Gonzales-Gargate and Paulo R. Ruffino. An averaging principle for diffusions in foliated spaces. *Ann. Probab.*, 44(1):567–588, 2016.
40. R. E. Greene and H. Wu. C^∞ approximations of convex, subharmonic, and plurisubharmonic functions. *Ann. Sci. École Norm. Sup. (4)*, 12(1):47–84, 1979.
41. Yu Gu and Jean-Christophe Mourrat. Pointwise two-scale expansion for parabolic equations with random coefficients. *Probab. Theory Related Fields*, 166(1-2):585–618, 2016.
42. M. Hairer, J. C. Mattingly, and M. Scheutzow. Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, 149(1-2):223–259, 2011.
43. M. Hairer and G. A. Pavliotis. Periodic homogenization for hypoelliptic diffusions. *J. Statist. Phys.*, 117(1-2):261–279, 2004.
44. M. Hairer and N. S. Pillai. Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 47(2):601–628, 2011.
45. Martin Hairer and Jonathan C. Mattingly. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Ann. of Math. (2)*, 164(3):993–1032, 2006.
46. Martin Hairer and Etienne Pardoux. Homogenization of periodic linear degenerate PDEs. *J. Funct. Anal.*, 255(9):2462–2487, 2008.
47. R. Z. Has′minskii. On the principle of averaging the Itô's stochastic differential equations. *Kybernetika (Prague)*, 4:260–279, 1968.
48. Inge S. Helland. Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.*, 9(2):79–94, 1982.
49. Michael Högele and Paulo Ruffino. Averaging along foliated Lévy diffusions. *Nonlinear Anal.*, 112:1–14, 2015.
50. Lars Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.
51. Nobuyuki Ikeda and Yukio Ogura. A degenerating sequence of Riemannian metrics on a manifold and their Brownian motions. In *Diffusion processes and related problems in analysis, Vol. I*, volume 22 of *Progr. Probab.*, pages 293–312. Birkhäuser Boston, 1990.
52. David Kelly and Ian Melbourne. Deterministic homogenization for fast-slow systems with chaotic noise. *J. Funct. Anal.*, 272(10):4063–4102, 2017.
53. Yuri Kifer. *Random perturbations of dynamical systems*, volume 16 of *Progress in Probability and Statistics*. Birkhäuser Boston, Inc., Boston, MA, 1988.
54. C. Kipnis and S. R. S. Varadhan. Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, 104(1):1–19, 1986.
55. Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of differential geometry. Vol I*. Interscience Publishers, a division of John Wiley & Sons, New York-Lond on, 1963.
56. Tomasz Komorowski, Claudio Landim, and Stefano Olla. *Fluctuations in Markov processes*, volume 345 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, 2012. Time symmetry and martingale approximation.
57. A. Korepanov, Z. Kosloff, and I. Melbourne. Martingale-coboundary decomposition for families of dynamical systems. *Annales I'Institut Henri Poincaré, Analyse non-linéaire*, pages 859–885, 2018.
58. H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7:284–304, 1940.
59. Christian Kuehn. *Multiple time scale dynamics*, volume 191 of *Applied Mathematical Sciences*. Springer, Cham, 2015.
60. Thomas G. Kurtz. A general theorem on the convergence of operator semigroups. *Trans. Amer. Math. Soc.*, 148:23–32, 1970.
61. S. Kusuoka and D. Stroock. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 32(1):1–76, 1985.

62. P Langevin. Sur la thèorie du mouvement brownien. *C. R. Acad. Sci. (Paris)*, 146, 1908.
63. Xue-Mei Li. An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity*, 21(4):803–822, 2008.
64. Xue-Mei Li. Effective diffusions with intertwined structures. arxiv:1204.3250, 2012.
65. Xue-Mei Li. Random perturbation to the geodesic equation. *Annals of Probability*, 44(1):544–566, 2015.
66. Xue-Mei Li. Limits of random differential equations on manifolds. *Probab. Theory Relat. Fields*, 166(3-4):659–712, 2016. DOI 10.1007/s00440-015-0669-x.
67. Xue-Mei Li. Homogenisation on homogeneous spaces. *J. Math. Soc. Japan*, 70(2):519–572, 2018.
68. Carlangelo Liverani and Stefano Olla. Toward the Fourier law for a weakly interacting anharmonic crystal. *J. Amer. Math. Soc.*, 25(2):555–583, 2012.
69. Rafe R. Mazzeo and Richard B. Melrose. The adiabatic limit, Hodge cohomology and Leray's spectral sequence for a fibration. *J. Differential Geom.*, 31(1):185–213, 1990.
70. S. Mischler and C. Mouhot. Exponential stability of slowly decaying solutions to the kinetic-Fokker-Planck equation. *Arch. Ration. Mech. Anal.*, 221(2):677–723, 2016.
71. S. B. Myers and N. E. Steenrod. The group of isometries of a Riemannian manifold. *Ann. of Math. (2)*, 40(2):400–416, 1939.
72. Edward Nelson. *Dynamical theories of Brownian motion*. Princeton University Press, Princeton, N.J., 1967.
73. Yukio Ogura and Setsuo Taniguchi. A probabilistic scheme for collapse of metrics. *J. Math. Kyoto Univ.*, 36(1):73–92, 1996.
74. G. C. Papanicolaou and W. Kohler. Asymptotic theory of mixing stochastic ordinary differential equations. *Comm. Pure Appl. Math.*, 27:641–668, 1974.
75. G. C. Papanicolaou, D. Stroock, and S. R. S. Varadhan. Martingale approach to some limit theorems. In *Papers from the Duke Turbulence Conference (Duke Univ.,1976)*, pages ii+120 pp. Duke Univ., Durham, N.C., 1977.
76. G. C. Papanicolaou and S. R. S. Varadhan. A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations. *Comm. Pure Appl. Math.*, 26:497–524, 1973.
77. Grigorios A. Pavliotis and Andrew M. Stuart. *Multiscale methods*, volume 53 of *Texts in Applied Mathematics*. Springer, New York, 2008. Averaging and homogenization.
78. Paulo R. Ruffino. Application of an averaging principle on foliated diffusions: topology of the leaves. *Electron. Commun. Probab.*, 20:no. 28, 5, 2015.
79. R. Schoen and S.-T. Yau. *Lectures on differential geometry*. Conference Proceedings and Lecture Notes in Geometry and Topology, I. International Press, Cambridge, MA, 1994. Lecture notes prepared by Wei Yue Ding, Kung Ching Chang [Gong Qing Zhang], Jia Qing Zhong and Yi Chao Xu, Translated from the Chinese by Ding and S. Y. Cheng, Preface translated from the Chinese by Kaising Tso.
80. Anatoli V. Skorokhod, Frank C. Hoppensteadt, and Habib Salehi. *Random perturbation methods with applications in science and engineering*, volume 150 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2002.
81. R. L. Stratonovich. Selected problems in the theory of fluctuations in radio engineering. *Sov. Radio, Moscow*, 1961. In Russian.
82. R. L. Stratonovich. *Topics in the theory of random noise. Vol. I: General theory of random processes. Nonlinear transformations of signals and noise*. Revised English edition. Translated from the Russian by Richard A. Silverman. Gordon and Breach Science Publishers, New York-London, 1963.
83. D. Stroock and S. R. S. Varadhan. Theory of diffusion processes. In *Stochastic differential equations*, volume 77 of *C.I.M.E. Summer Sch.*, pages 149–191. Springer, Heidelberg, 2010.
84. Luen-Fai Tam. Exhaustion functions on complete manifolds. In *Recent advances in geometric analysis*, volume 11 of *Adv. Lect. Math. (ALM)*, pages 211–215. Int. Press, Somerville, MA, 2010.
85. Shûkichi Tanno. The first eigenvalue of the Laplacian on spheres. *Tôhoku Math. J. (2)*, 31(2):179–185, 1979.

86. G. E. Uhlenbeck and L. S. Ornstein. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physical Review*, 36:823–841, 1930.
87. Hajime Urakawa. The first eigenvalue of the Laplacian for a positively curved homogeneous Riemannian manifold. *Compositio Math.*, 59(1):57–71, 1986.
88. Erik van Erp. The Atiyah-Singer index formula for subelliptic operators on contact manifolds. Part I. *Ann. of Math. (2)*, 171(3):1647–1681, 2010.
89. Erik van Erp. The index of hypoelliptic operators on foliated manifolds. *J. Noncommut. Geom.*, 5(1):107–124, 2011.
90. A. Yu. Veretennikov. On an averaging principle for systems of stochastic differential equations. *Mat. Sb.*, 181(2):256–268, 1990.
91. Cédric Villani. Hypocoercive diffusion operators. In *International Congress of Mathematicians. Vol. III*, pages 473–498. Eur. Math. Soc., Zürich, 2006.
92. Kôsaku Yosida. *Functional analysis*. Die Grundlehren der Mathematischen Wissenschaften, Band 123. Academic Press, Inc., New York; Springer-Verlag, Berlin, 1965.