Limits of Random Differential Equations on Manifolds

Xue-Mei Li

Mathematics Institute, The University of Warwick, Coventry CV4 7AL, U.K.
Email address: xue-mei.li@warwick.ac.uk

Abstract

Consider a family of random ordinary differential equations on a manifold driven by vector fields of the form $\sum_k Y_k \alpha_k(z_t^\epsilon(\omega))$ where Y_k are vector fields, ϵ is a positive number, z_t^ϵ is a $\frac{1}{\epsilon}\mathcal{L}_0$ diffusion process taking values in possibly a different manifold, α_k are annihilators of $\ker(\mathcal{L}_0^*)$. Under Hörmander type conditions on $\mathcal{L}_0$ we prove that, as ϵ approaches zero, the stochastic processes $y^\epsilon_{\frac{t}{\epsilon}}$ converge weakly and in the Wasserstein topologies. We describe this limit and give an upper bound for the rate of the convergence.

AMS classification: 60H, 60J, 60F, 60D.

1 Introduction

Let M and G be finite dimensional smooth manifolds. Let Y_k, $k = 1, \dots, m$, be C^6 vector fields on M, α_k real valued C^r functions on G, ϵ a positive number, and (z_t^ϵ) diffusions on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ with values in G and infinitesimal generator $\mathcal{L}_0^\epsilon = \frac{1}{\epsilon}\mathcal{L}_0$ which will be made precise later. The aim of this paper is to study limit theorems associated to the system of ordinary differential equations on M,

$$\dot{y}_t^\epsilon(\omega) = \sum_{k=1}^m Y_k\left(y_t^\epsilon(\omega)\right)\alpha_k(z_t^\epsilon(\omega)) \tag{1.1}$$

under the assumption that α_k 'averages' to zero. The 'average' is with respect to the unique invariant probability measure of $\mathcal{L}_0$, in case $\mathcal{L}_0$ satisfies strong Hörmander's condition, and more generally the 'average' is the projection to a suitable function space. We prove that $y^\epsilon_{\frac{t}{\epsilon}}$ converges as $\epsilon \to 0$ to a Markov process whose Markov generator has an explicit expression.

This study is motivated by problems arising from stochastic homogenization. It turned out that in the study of randomly perturbed systems with a conserved quantity, which does not necessarily take value in a linear space, the reduced equations for the slow variables can sometimes be transformed into (1.5). Below, in section 2 we illustrate this by 4 examples including one on the orthonormal frame bundle over a Riemannian manifold. Of these examples, the first is from [25] where we did not know how to obtain a rate of convergence, and the last three from [26] where a family of interpolation equations on homogeneous manifolds are introduced. An additional example can be found in [24].

1.1 Outline of the Paper

In all the examples, which we described in §2 below, the scalar functions average to 0 with respect to a suitable probability measure on G. Bearing in mind that if a Hamiltonian system approximates a physical system with error ϵ on a compact time interval, over a time interval of size $\frac{1}{\epsilon}$ the physical orbits deviate visibly from that of the Hamiltonian system unless the error is reduced by oscillations, it is natural and a classical problem to study ODEs whose right hand sides are random and whose averages in time are zero.

The objectives of the present article are: (1) to prove that, as ϵ tends to zero, the law of $(y^\epsilon_{\frac{s}{\epsilon}}, s \le t)$ converges weakly to a probability measure $\bar{\mu}$ on the path space over M and to describe the properties of the limiting Markov semigroups; (2) to estimate the rate of convergence, especially in the Wasserstein distance. For simplicity we assume that all the equations are complete. In sections 4, 5, 6 and 8 we assume that $\mathcal{L}_0$ is a regularity improving Fredholm operator on a compact manifold G, see Definition 4.1. In Theorem 6.4 we assume, in addition, that $\mathcal{L}_0$ has Fredholm index 0. But strong Hörmander's condition can be used to replace the condition 'regularity improving Fredholm operator of index 0'.

For simplicity, throughout the introduction, α_k are bounded and belong to $N^\perp$ where N is the kernel of $\mathcal{L}_0^*$, the adjoint of the unbounded operator $\mathcal{L}_0$ in $L^2(G)$ with respect to the volume measure. In case $\mathcal{L}_0$ is not elliptic we assume in addition that $r \ge 3$ or $r \ge \max\{3, \frac{n}{2}+1\}$, depending on the result. The growth conditions on Y_k are in terms of a control function V and a controlled function space $B_{V,r}$ where r indicates the order of the derivatives to be controlled, see (5.1). For simplicity we assume both M and G are compact.

In Section 3 we present two elementary lemmas, Lemma 3.4 and Lemma 3.5, assuming $\mathcal{L}_0$ mixes exponentially in a weighted total variation norm with weight $W : G \to \mathbb{R}$. In Section 4, for $\mathcal{L}_0$ a regularity improving Fredholm operator and f a C^2 function, we deduce a formula for $f(y^\epsilon_{\frac{t}{\epsilon}})$ where the transmission of the randomness from the fast motion (z^ϵ_t) to the slow motion (y^ϵ_t) is manifested in a martingale. This provides a platform for the uniform estimates over large

time intervals, weak convergences, and the study of rate of convergence in later sections.

In Section 5, we obtain uniform estimates in ϵ for functionals of y_t^ϵ over $[0, \frac{1}{\epsilon}]$. Let $\mathcal{L}_0$ be a regularity improving Fredholm operator, $y_0^\epsilon = y_0$, and V a C^2 function such that $\sum_{j=1}^m |L_{Y_j} V| \leq c + KV$, $\sum_{i,j=1}^m |L_{Y_i} L_{Y_j} V| \leq c + KV$ for some numbers c and K. Then, Theorem 5.2, for every numbers $p \geq 1$ there exists a positive number ϵ_0 such that $\sup_{0<\epsilon\leq\epsilon_0} \mathbf{E} \sup_{0\leq u\leq t} V^p(y^\epsilon_{\frac{u}{\epsilon}})$ is finite and belongs to $B_{V,0}$ as a function of y_0. This leads to convergence in the Wasserstein distance and will be used later to prove a key lemma on averaging functions along the paths of $(y_t^\epsilon, z_t^\epsilon)$.

In Section 6, $\mathcal{L}_0$ is an operator on a compact manifold G satisfying Hörmander's condition and with Fredholm index 0; M has positive injectivity radius and other geometric restrictions. In particular we do not make any assumption on the ergodicity of $\mathcal{L}_0$. Let $\overline{\alpha_i \beta_j}$ denote $\sum_l u_l \langle \alpha_i \beta_j, \pi_l \rangle$ where $\{u_l\}$ is a basis of the kernel of $\mathcal{L}_0$ and $\{\pi_l\}$ the dual basis in the kernel of $\mathcal{L}_0^*$. Theorem 6.4 states that, given bounds on Y_k and its derivatives and for $\alpha_k \in C^r$ where $r \geq \max\{3, \frac{n}{2} + 1\}$, $(y^\epsilon_{\frac{s}{\epsilon}}, s \leq t)$ converges weakly, as $\epsilon \to 0$, to the Markov process with Markov generator $\bar{\mathcal{L}} = \sum_{i,j=1}^m \overline{\alpha_i \beta_j} L_{Y_i} L_{Y_j}$. This follows from a tightness result, Proposition 6.1 where no assumption on the Fredholm index of $\mathcal{L}_0$ is made, and a law of large numbers for sub-elliptic operators on compact manifolds, Lemma 6.2. Convergences of $\{(y^\epsilon_{\frac{t}{\epsilon}}, 0 \leq t \leq T)\}$ in the Wasserstein p-distance are also obtained.

In Section 7 we study the solution flows of SDEs and their associated Kolmogorov equations, to be applied to the limiting operator $\bar{\mathcal{L}}$ in Section 8. Otherwise this section is independent of the rest of the paper. Let Y_k, Y_0 be C^6 and C^5 vector fields respectively. If M is compact, or more generally if Y_k are BC^5 vector fields, the conclusions in this section holds, trivially. Denote $B_{V,4}$ the set of functions whose derivatives up to order r are controlled by a function V, c.f.(5.1). Let $\Phi_t(y)$ be the solution flow to

$$dy_t = \sum_{k=1}^m Y_k(y_t) \circ dB_t^k + Y_0(y_t) dt.$$

Let $P_t f(y) = \mathbf{E} f(\Phi_t(y))$ and $Z = \frac{1}{2} \sum_{k=1}^m \nabla_{Y_k} Y_k + Y_0$. Let $V \in C^2(M, \mathbb{R}_+)$ and $\sup_{s\leq t} \mathbf{E} V^q(\phi_s(y)) \in B_{V,0}$ for every $q \geq 1$. This assumption on V is implied by the following conditions: $|L_{Y_i} L_{Y_j} V| \leq c + KV$, $|L_{Y_j} V| \leq c + KV$, where C, K are constants. Let $\tilde{V} = 1 + \ln(1 + |V|)$. We assume, in addition, for some number

c the following hold:

$$\begin{aligned} &\sum_{k=1}^{m}\sum_{\alpha=0}^{5}|\nabla^{(\alpha)}Y_k| \in B_{V,0}, \quad \sum_{\alpha=0}^{4}|\nabla^{(\alpha)}Y_0| \in B_{V,0}, \\ &\sum_{k=1}^{m}|\nabla Y_k|^2 \le c\tilde{V}, \qquad\qquad \sup_{|u|=1}\langle \nabla_u Z, u\rangle \le c\tilde{V}. \end{aligned} \tag{1.2}$$

Then there is a global smooth solution flow $\Phi_t(y)$, Theorem 7.2. Furthermore for $f \in B_{V,4}$, $\mathcal{L}f \in B_{V,2}$, $\mathcal{L}^2 f \in B_{V,0}$, and $P_t f \in B_{V,4}$.

For $M = \mathbb{R}^n$, an example of the control pair is: $V(x) = C(1+|x|^2)$ and $\tilde{V}(x) = \ln(1+|x|^2)$. Our conditions are weaker than those commonly used in the probability literature for $d(P_t f)$, in two ways. Firstly we allow non-bounded first order derivative, secondly we allow one sided conditions on the drift and its first order derivatives. In this regard, we extend a theorem of W. Kohler, G. C. Papanicolaou [32] where they used estimations from O. A. Oleinik- E. V. Radkevič [31]. The estimates on the derivative flows, obtained in this section, are often assumptions in applications of Malliavin calculus to the study of stochastic differential equations. Results in this section might be of independent interests.

Let P_t be the Markov semigroup generated by $\bar{\mathcal{L}}$. In Section 8, we prove the following estimate: $|\mathbf{E}f(\Phi_t^\epsilon(y_0)) - P_t f(y_0)| \le C(t)\gamma(y_0)\epsilon\sqrt{|\log\epsilon|}$ where $C(t)$ is a constant, γ is a function in $B_{V,0}$ and $\Phi_t^\epsilon(y_0)$ the solution to (1.5) with initial value y_0. The conditions on the vector fields Y_k are similar to (1.2), we also assume the conditions of Theorem 5.2 and that $\mathcal{L}_0$ satisfies strong Hörmander's condition. We incorporated traditional techniques on time averaging with techniques from homogenization. The homogenization techniques was developed from [23] which was inspired by the study in M. Hairer and G. Pavliotis [12]. For the rate of convergence we were particularly influenced by the following papers: W. Kohler and G. C. Papanicolaou [32, 36] and G. C. Papanicolaou and S.R.S. Varadhan [34]. Denote $\hat{P}_{y_{\frac{t}{\epsilon}}^\epsilon}$ the probability distributions of the random variables $y_{\frac{t}{\epsilon}}^\epsilon$ and $\bar{\mu}_t$ the probability measure determined by P_t. The under suitable conditions, $W_1(\hat{P}_{y_{\frac{t}{\epsilon}}^\epsilon}, \bar{\mu}_t) \le C\epsilon^r$, where r is any positive number less or equal to $\frac{1}{4}$ and W_1 denotes the Wasserstein 1-distance, see §9.

1.2 Main Theorems

We contrive to impose as little as possible on the vector fields $\{Y_k\}$, hence a few sets of assumptions are used. For the examples we have in mind, G is a compact Lie group acting on a manifold M, and so for simplicity G is assumed to be compact throughout the article, with few exceptions. In a future study, it would be nice to provide some interesting examples in which G is not compact.

If M is also compact, only the following two conditions are needed: (a) $\mathcal{L}_0$ satisfies strong Hörmander's condition; (b) $\{\alpha_k\} \subset C^r \cap N^\perp$ where N is the annihilator of the kernel of $\mathcal{L}_0^*$ and r is a sufficiently large number. If $\mathcal{L}_0$ is elliptic, 'C^r' can be replaced by 'bounded measurable'. For the convergence condition (a) can be replaced by '$\mathcal{L}_0$ satisfies Hörmander's condition and has Fredholm index 0'. If $\mathcal{L}_0$ has a unique invariant probability measure, no condition is needed on the Fredhom index of $\mathcal{L}_0$.

Theorem 6.4 and Corollary 6.5. Under the conditions of Proposition 6.1 and Assumption 6.1, $(y^\epsilon_{\frac{t}{\epsilon}})$ converges weakly to the Markov process determined by

$$\bar{\mathcal{L}} = -\sum_{i,j=1}^{m} \overline{\alpha_i \mathcal{L}_0^{-1} \alpha_j} L_{Y_i} L_{Y_j}, \quad \overline{\alpha_i \mathcal{L}_0^{-1} \alpha_j} = \sum_{b=1}^{n_0} u_b \langle \alpha_i \mathcal{L}_0^{-1} \alpha_j, \pi_b \rangle,$$

where n_0 is the dimension of the kernel of $\mathcal{L}_0$ which, by the assumption that $\mathcal{L}_0$ has Fredholm index 0, equals the dimension of the kernel of $\mathcal{L}_0^*$. The set of functions $\{u_b\}$ is a basis of ker($\mathcal{L}_0$) and $\{\pi_b\} \subset$ ker($\mathcal{L}_0^*$) its dual basis. In case $\mathcal{L}_0$ satisfies strong Hörmander's condition, then there is a unique invariant measure and $\overline{\alpha_i \mathcal{L}_0^{-1} \alpha_j}$ is simply the average of $\alpha_i \mathcal{L}_0^{-1} \alpha_j$ with respect to the unique invariant measure. Let $p \geq 1$ be a number and V a Lyapunov type function such that $\rho^p \in B_{V,0}$, a function space controlled by V. If furthermore Assumption 5.1 holds, $(y^\epsilon_{\frac{\cdot}{\epsilon}})$ converges, on $[0, t]$, in the Wasserstein p-distance.

Theorem 8.2. Denote $\Phi^\epsilon_t(\cdot)$ the solution flow to (1.5) and P_t the semigroup for $\bar{\mathcal{L}}$. If Assumption 8.1 holds then for $f \in B_{V,4}$,

$$\left| \mathbf{E} f\left(\Phi^\epsilon_{\frac{T}{\epsilon}}(y_0)\right) - P_T f(y_0) \right| \leq \epsilon |\log \epsilon|^{\frac{1}{2}} C(T) \gamma_1(y_0),$$

where $\gamma_1 \in B_{V,0}$ and $C(T)$ is a constant increasing in T. Similarly, if $f \in BC^4$,

$$\left| \mathbf{E} f\left(\Phi^\epsilon_{\frac{T}{\epsilon}}(y_0)\right) - P_T f(y_0) \right| \leq \epsilon |\log \epsilon|^{\frac{1}{2}} C(T) \gamma_2(y_0) \left(1 + |f|_{4,\infty}\right). \tag{1.3}$$

where γ_2 is a function in $B_{V,0}$ independent of f and C are increasing functions.

A complete connected Riemannian manifold is said to have bounded geometry if it has strictly positive injectivity radius, and if the Riemannian curvature tensor and its covariant derivatives are bounded.

Proposition 9.2. Suppose that M has bounded geometry, $\rho_o^2 \in B_{V,0}$, and Assumption 8.1 holds. Let $\bar{\mu}$ be the limit measure and $\bar{\mu}_t = (ev_t)_* \bar{\mu}$. Then for every $r < \frac{1}{4}$ there exists $C(T) \in B_{V,0}$ and $\epsilon_0 > 0$ s.t. for all $\epsilon \leq \epsilon_0$ and $t \leq T$,

$$d_W(\text{Law}(y^\epsilon_{\frac{t}{\epsilon}}), \bar{\mu}_t) \leq C(T) \epsilon^r.$$

Besides the fact that we work on manifolds, where there is the inherited nonlinearity and the problem with cut locus, the following aspects of the paper are perhaps new. (a) We do not assume there exists a unique invariant probability measure on the noise and the effective processes are obtained by a suitable projection, accommodating one type of degeneracy. Furthermore the noise takes value in another manifold, accommodating 'removable' degeneracy. For example the stochastic processes in question lives in a Lie group, while the noise are entirely in the directions of a sub-group. (b) We used Lyapunov functions to control the growth of the vector fields and their derivatives, leading to estimates uniform in ϵ and to the conclusion on the convergence in the Wasserstein topologies. A key step for the convergence is a law of large numbers, with rates, for sub-elliptic operators (i.e. operators satisfying Hörmander's sub-elliptic estimates). (c) Instead of working with iterated time averages we use a solution to Poisson equations to reveal the effective operator. Functionals of the processes $y^\epsilon_{\frac{t}{\epsilon}}$ splits naturally into the sum of a fast martingale, a finite variation term involving a second order differential operator in Hörmander form, and a term of order ϵ. From this we obtain the effective diffusion, in explicit Hörmander form. It is perhaps also new to have an estimate for the rate of the convergence in the Wasserstein distance. Finally we improved known theorems on the existence of global smooth solutions for SDEs in [22], c.f. Theorem 7.2 below where a criterion is given in terms of a pair of Lyapunov functions. New estimates on the moments of higher order covariant derivatives of the derivative flows are also given.

1.3 Classical Theorems

We review, briefly, basic ideas from existing literature on random ordinary differential equations with fast oscillating vector fields. Let $F(x,t,\omega,\epsilon) := F^{(0)}(x,t,\omega) + \epsilon F^{(1)}(x,t,\omega)$, where $F^{(i)}(x,t,\cdot)$ are measurable functions, for which a Birkhoff ergodic theorem holds whose limit is denoted by $\bar{F}$. The solutions to the equations $\dot{y}^\epsilon_t = F(y^\epsilon_t, \frac{t}{\epsilon}, \omega, \epsilon)$ over a time interval $[0,t]$, can be approximated by the solution to the averaged equation driven by $\bar{F}$. If $\bar{F}^{(0)} = 0$, we should observe the solutions in the next time scale and study $\dot{x}^\epsilon_t = \frac{1}{\epsilon}F(x^\epsilon_t, \frac{t}{\epsilon^2}, \omega, \epsilon)$. See R. L. Stratonovich [42, 43]. Suppose for some functions $\bar{a}_{j,k}$ and $\bar{b}_j$ the following estimates hold uniformly:

$$\left|\frac{1}{\epsilon^3}\int_s^{s+\epsilon}\int_s^{r_1} \mathbf{E}F^{(0)}_j(x,\frac{r_2}{\epsilon^2})F^{(0)}_k(x,\frac{r_1}{\epsilon^2})\,dr_2\,dr_1 - \bar{a}_{j,k}(s,x)\right|\,dr_2\,dr_1 \le o(\epsilon),$$

$$\left|\frac{1}{\epsilon^3}\int_s^{s+\epsilon}\int_s^{r_1}\sum_{k=1}^d \mathbf{E}\frac{\partial F^{(0)}_j}{\partial x_k}(x,\frac{r_2}{\epsilon^2})F^{(0)}_k(x,\frac{r_1}{\epsilon^2})\,dr_2\,dr_1 + \frac{1}{\epsilon}\int_s^{s+\epsilon}\mathbf{E}F^{(1)}_j(x,\frac{r}{\epsilon^2})\,dr - \bar{b}_j(x,s)\right|$$
$$\le o(\epsilon). \tag{1.4}$$

Then under a 'strong mixing' condition with suitable mixing rate, the solutions of the equations $\dot{x}_t^\epsilon = \frac{1}{\epsilon}F(x_t^\epsilon, \frac{t}{\epsilon^2}, \omega, \epsilon)$ converge weakly on any compact interval to a Markov process. This is a theorem of R. L. Stratonovich [43] and R. Z. Khasminskii[14], further refined and explored in Khasminskii [15] and A. N. Borodin [3]. These theorems lay foundation for investigation beyond ordinary differential equations with a fast oscillating right hand side.

In our case, noise comes into the system via a $\mathcal{L}_0$-diffusion satisfying Hörmander's conditions, and hence we could by pass these assumptions and also obtain convergences in the Wasserstein distances. For manifold valued stochastic processes, some difficulties are caused by the inherited non-linearity. For example, integrating a vector field along a path makes sense only after they are parallel translated back. Parallel transports of a vector field along a path, from time t to time 0, involves the whole path up to time t and introduces extra difficulties; this is still an unexplored territory wanting further investigations. For the proof of tightness, the non-linearity causes particular difficulty if the Riemannian distance function is not smooth. The advantage of working on a manifold setting is that for some specific physical models, the noise can be untwisted and becomes easy to deal with.

Our estimates for the rate of convergence, section 8 and 9, can be considered as an extension to that in W. Kohler and G. C. Papanicolaou [32, 36], which were in turn developed from the following sequence of remarkable papers: R. Coghurn and R. Hersh [6], J.B. Keller and G. C. Papanicolaou [35], R. Hersh and M. Pinsky [17], R. Hersh and G. C. Papanicolaou [16] and G. C. Papanicolaou and S.R.S. Varadhan [34]. See also T. Kurtz [21] and [33] by D. Stroock and S. R. S. Varadhan.

The condition $\bar{F} = 0$ needs not hold for this type of scaling and convergence. If $F(x, t, \omega, \epsilon) = F^{(0)}(x, \zeta_t(\omega))$, where ζ_t is a stationary process with values in $\mathbb{R}^m$, and $\bar{F}^{(0)} = X_H$, the Hamiltonian vector field associated to a function $H \in BC^3(\mathbb{R}^2; \mathbb{R})$ whose level sets are closed connected curves without intersections, then $H(y_{\frac{t}{\epsilon}}^\epsilon)$ converge to a Markov process, under suitable mixing and technical assumptions. See A. N. Borodin and M. Freidlin [4], also M. Freidlin and M. Weber [8] where a first integral replaces the Hamiltonian, and also X.-M. Li [25] where the value of a map from a manifold to another is preserved by the unperturbed system.

In M. Freidlin and A. D.Wentzell [9], the following type of central limit theorem is proved: $\frac{1}{\sqrt{\epsilon}}\left(H(x_s^\epsilon) - H(\bar{x}_s)\right)$ converges to a Markov diffusion. This formulation is not suitable when the conserved quantity takes value in a non-linear space.

For the interested reader, we also refer to the following articles on limit theorems, averaging and Homogenization for stochastic equations on manifolds: N.

Enriquez, J. Franchi, Y. LeJan [7], I. Gargate, P. Ruffino [11], N. Ikeda, Y. Ochi [19], Y. Kifer [20], M. Liao and L. Wang [27], S. Manade, Y. Ochi [29], Y. Ogura [30], M. Pinsky [37], and R. Sowers [41].

1.4 Further Question.

(1) I am grateful to the associate editor for pointing out the paper by C. Liverani and S. Olla [28], where random perturbed Harmiltonian system, in the context of weak interacting particle systems, is studied. Their system is somewhat related to the completely integrable equation studied in [23] leading to a new problem which we now state. Denote X_f the Hamiltonian vector field on a symplectic manifold corresponding to a function f. If the symplectic manifold is $\mathbb{R}^{2n}$ with the canonical symplectic form, X_f is the skew gradient of f. Suppose that $\{H_1, \dots, H_n\}$ is a completely integrable system, i.e. they are poisson commuting at every point and their Hamiltonian vector fields are linearly independent at almost all points. Following [23] we consider a completely integrable SDE perturbed by a transversal Hamiltonian vector field:

$$dy_t^\epsilon = \sum_{i=1}^{n} X_{H_i}(y_t^\epsilon) \circ dW_t^i + X_{H_0}(y_t^\epsilon)dt + \epsilon X_K(y_t^\epsilon)dt.$$

Suppose that X_{H_0} commutes with X_{H_k} for $k = 1, \dots, n$, then each H_i is a first integral of the unperturbed system. Then, [23, Th 4.1], within the action angle coordinates of a regular value of the energy function $H = (H_1, \dots, H_n)$, the energies $\{H_1(y^\epsilon_{\frac{t}{\epsilon^2}}), \dots, H_n(y^\epsilon_{\frac{t}{\epsilon^2}})\}$ converge weakly to a Markov process. When restricted to the level sets of the energies, the fast motions are ellipitic. It would be desirable to remove the 'complete integrability' in favour of Hormander's type conditions. There is a non-standard symplectic form on $(\mathbb{R}^4)^N$ with respect to which the vector fields in [28] are Hamiltonian vector fields and when restricted to level sets of the energies the unperturbed system in [28] satisfies Hörmander's condition, see [28, section 5], and therefore provides a motivating example for further studies. Finally note that the driving vector fields in (1.5) are in a special form, results here would not apply to the systems in [23] nor that in [28], and hence it would be interesting to formulate and develop limit theorems for more general random ODEs to include these two cases.

(2) It should be interesting to develop a theory for the ODEs below

$$\dot{y}_t^\epsilon(\omega) = \sum_{k=1}^{m} Y_k\left(y_t^\epsilon(\omega)\right) \alpha_k(z_t^\epsilon(\omega), y_t^\epsilon)) \tag{1.5}$$

where α_k depends also on the y^ϵ process.

(3) It would be nice to extend the theory to allow the noise to live in a non-compact manifold, in which $\mathcal{L}_0$ should be an Ornstein-Uhlenbeck type operator whose drift term would provide for a deformed volume measure.

Notation. Throughout this paper $\mathcal{B}_b(M;N)$, $C^r_K(M;N)$, and $BC^r(M;N)$ denote the set of functions from M to N that are respectively bounded measurable, C^r with compact supports, and bounded C^r with bounded first r derivatives. If $N = \mathbb{R}$ the letter N will be suppressed. Also $\mathbb{L}(V_1;V_2)$ denotes the space of bounded linear maps; $C^r(\Gamma TM)$ denotes C^r vector fields on a manifold M.

2 Examples

Let $\{W^k_t\}$ be independent real valued Brownian motions on a given filtered probability space, $\circ$ denote Stratonovich integrals. In the following H_0 and A_k are smooth vector fields, and $\{A_1,\dots,A_k\}$ is an orthonormal basis at each point of the vertical tangent spaces. To be brief, we do not specify the properties of the vector fields, instead refer the interested reader to [25] for details. For any $\epsilon > 0$, the stochastic differential equations

$$du^\epsilon_t = H_0(u^\epsilon_t)dt + \frac{1}{\sqrt{\epsilon}}\sum_{k=1}^{\frac{n(n-1)}{2}} A_k(u^\epsilon_t)\circ dW^k_t$$

are degenerate and they interpolate between the geodesic equation ($\epsilon=\infty$) and Brownian motions on the fibres ($\epsilon = 0$). The fast random motion is transmitted to the horizontal direction by the action of the Lie bracket $[H_0, A_k]$. If $H_0 = 0$, there is a conserved quantity to the system which is the projection from the orthonormal bundle to the base manifold. This allows us to separate the slow variable (y^ϵ_t) and the fast variable (z^ϵ_t). The reduced equation for (y^ϵ_t), once suitable 'coordinate maps' are chosen, can be written in the form of (1.5). In [25] we proved that ($y^\epsilon_{\frac{t}{\epsilon}}$) converges weakly to a rescaled horizontal Brownian motion. Recently J. Angst, I. Bailleul and C. Tardif gave this a beautiful treatment, [1], using rough path analysis.

By theorems in this article, the above model can be generalised to include random perturbation by hypoelliptic diffusions, i.e. $\{A_1,\dots,A_k\}$ generates all vertical directions. In [25] we did not know how to obtain a rate for the convergence. Theorem 8.2, in this article, will apply and indeed we have an upper bound for the rate of convergence.

As a second example, we consider, on the special orthogonal group $SO(n)$, the following equations:

$$dg^\epsilon_t = \frac{1}{\sqrt{\epsilon}}\sum_{k=1}^{\frac{n(n-1)}{2}} g^\epsilon_t E_k \circ dW^k_t + g^\epsilon_t Y_0 dt, \tag{2.1}$$

where $\{E_k\}$ is an orthonormal basis of $\mathfrak{so}(n-1)$, as a subspace of $\mathfrak{so}(n)$, and Y_0 is a skew symmetric matrix orthogonal to $\mathfrak{so}(n-1)$. The above equation is closely related to the following set of equations:

$$dg_t = \gamma \sum_{k=1}^{\frac{n(n-1)}{2}} g_t E_k \circ dW_t^k + \delta g_t Y_0 dt,$$

where γ, δ are two positive numbers. If $\delta = 0$ and $\gamma = 1$, the solutions are Brownian motions on $SO(n-1)$. If $\delta = \frac{1}{|Y_0|}$ and $\gamma = 0$, the solutions are unit speed geodesics on $SO(n)$. These equations interpolate between a Brownian motion on the sub-group $SO(n-1)$ and a one parameter family of subgroup on $SO(n)$. See [26]. Take $\delta = 1$ and let $\gamma = \frac{1}{\sqrt{\epsilon}} \to \infty$, what could be the 'effective limit' if it exists? The slow components of the solutions, which we denote by (u_t^ϵ), satisfy equations of the form (1.5). They are 'horizontal lifts' of the projections of the solutions to S^n. If $\mathfrak{m}$ is the orthogonal complement of $\mathfrak{so}(n-1)$ in $\mathfrak{so}(n)$ then $\mathfrak{m}$ is Ad_H-irreducible and Ad_H-invariant, noise is transmitted from $\mathfrak{h}$ to every direction in $\mathfrak{m}$, and this in the uniform way. It is therefore plausible that $u^\epsilon_{\frac{t}{\epsilon}}$ can be approximated by a diffusion $\bar{u}_t$ of constant rank. The projection of u_t to S^n is a scaled Brownian motion with scale λ. The scale λ is a function of the dimension n, but is independent of Y_0 and is associated to an eigenvalue of the Laplacian on $SO(n-1)$, indicating the speed of propagation.

As a third example we consider the Hopf fibration $\pi : S^3 \to S^2$. Let $\{X_1, X_2, X_3\}$ be the Pauli matrices, they form an orthonormal basis of $\mathfrak{su}(2)$ with respect to the canonical bi-invariant Riemannian metric.

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

Denote X^* the left invariant vector field generated by $X \in \mathfrak{su}(2)$. By declaring $\{\frac{1}{\sqrt{\epsilon}} X_1^*, X_2^*, X_3^*\}$ an orthonormal frame, we obtain a family of left invariant Riemannian metrics m^ϵ on S^3. The Berger's spheres, (S^3, m^ϵ), converge in measured Gromov-Hausdorff topology to the lower dimensional sphere $S^2(\frac{1}{2})$. For further discussions see K. Fukaya [10] and J. Cheeger and M. Gromov [5]. Let W_t be a one dimensional Brownian motion and take Y from $\mathfrak{m} := \langle X_2, X_3 \rangle$. The infinitesimal generator of the equation $dg_t^\epsilon = \frac{1}{\sqrt{\epsilon}} X_1^*(g_t^\epsilon) \circ dW_t + Y^*(g_t^\epsilon)\, dt$ satisfies weak Hörmander's conditions. The 'slow motions', suitably sacled, converge to a 'horizontal' Brownian motion whose generator is $\frac{1}{2} c \,\mathrm{trace}_{\mathfrak{m}} \nabla d$, where the trace is taken in $\mathfrak{m}$. A slightly different, ad hoc, example on the Hopf fibration is discussed in [24]. An analogous equations can be considered on $SU(n)$ where the diffusion coefficients come from a maximal torus.

Finally we give an example where the noise (z_t^ϵ) in the reduced equation is not elliptic. Let $M = SO(4)$ and let $E_{i,j}$ be the elementary 4×4 matrices and

$A_{i,j} = \frac{1}{\sqrt{2}}(E_{ij} - E_{ji})$. For $k = 1, 2$ and 3, we consider the equations

$$dg_t^\epsilon = \frac{1}{\sqrt{\epsilon}} A_{1,2}^*(g_t^\epsilon) \circ db_t^1 + \frac{1}{\sqrt{\epsilon}} A_{1,3}^*(g_t^\epsilon) \circ db_t^2 + A_{k4}^*(g_t^\epsilon) dt.$$

The slow components of the solutions of these equations again satisfy an equation of the form (1.5).

3 Preliminary Estimates

Let $\mathcal{L}_0$ be a diffusion operator on a manifold G and Q_t its transition semigroup and transition probabilities. Let $\|\cdot\|_{TV}$ denote the total variation norm of a measure, normalized so that the total variation norm between two probability measures is less or equal to 2. By the duality formulation for the total variation norm,

$$\|\mu\|_{TV} = \sup_{|f|\le 1, f\in\mathcal{B}_b(G;\mathbb{R})} \left| \int_G f d\mu \right|.$$

For $W \in \mathcal{B}(G; [1, \infty))$ denote $\|f\|_W$ the weighted supremum norm and $\|\mu\|_{TV,W}$ the weighted total variation norm:

$$\|f\|_W = \sup_{x\in G} \frac{|f(x)|}{W(x)}, \quad \|\mu\|_{TV,W} = \sup_{\{\|f\|_W \le 1\}} \left| \int_G f d\mu \right|.$$

Assumption 3.1 *There is an invariant probability measure π for $\mathcal{L}_0$, a real valued function $W \in L^1(G, \pi)$ with $W \ge 1$, numbers $\delta > 0$ and $a > 0$ such that*

$$\sup_{x\in G} \frac{\|Q_t(x, \cdot) - \pi\|_{TV,W}}{W(x)} \le a e^{-\delta t}.$$

If G is compact we take $W \equiv 1$.

In the following lemma we collect some elementary estimates, which will be used to prove Lemma 3.4 and 3.5, for completeness their proofs are given in the appendix. Write $\bar{W} = \int_G W d\pi$.

Lemma 3.1 *Assume Assumption 3.1. Let $f, g : G \to \mathbb{R}$ be bounded measurable functions and let $c_\infty = |f|_\infty \|g\|_W$. Then the following statements hold for all $s, t \ge 0$.*

(1) Let (z_t) be an $\mathcal{L}_0$ diffusion. If $\int_G g d\pi = 0$,

$$\left| \frac{1}{t-s} \int_s^t \int_s^{s_1} \left(\mathbf{E}\left\{ f(z_{s_2}) g(z_{s_1}) \middle| \mathcal{F}_s \right\} - \int_G f Q_{s_1 - s_2} g d\pi \right) ds_2 ds_1 \right|$$
$$\le \frac{a^2 c_\infty}{(t-s)\delta^2} W(z_s).$$

(2) Let (z_t) be an $\mathcal{L}_0$ diffusion. If $\int_G g d\pi = 0$ then

$$\left| \frac{1}{t-s} \int_s^t \int_s^{s_1} \mathbf{E} \left\{ f(z_{s_2}) g(z_{s_1}) \middle| \mathcal{F}_s \right\} \, ds_2 \, ds_1 - \int_G \int_0^\infty f Q_r g \, dr \, d\pi \right|$$
$$\leq \frac{c_\infty}{(t-s)\delta^2} (a^2 W(z_s) + a\bar{W}) + \frac{c_\infty a}{\delta} \bar{W}.$$

(3) Suppose that either $\int_G f \, d\pi = 0$ or $\int_G g \, d\pi = 0$. Let

$$C_1 = \frac{a}{\delta^2}(aW + \bar{W})|f|_\infty \|g\|_W, \quad C_2 = \frac{2a}{\delta}|f|_\infty \|g\|_W \bar{W} + \frac{a}{\delta}|\bar{g}| \, \|f\|_W W.$$

Let (z_t^ϵ) be an $\mathcal{L}_0^\epsilon$ diffusion. Then for every $\epsilon > 0$,

$$\left| \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E} \left\{ f(z_{s_2}^\epsilon) g(z_{s_1}^\epsilon) \middle| \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 \right| \leq C_1(z_{\frac{s}{\epsilon}}^\epsilon)\epsilon^2 + C_2(z_{\frac{s}{\epsilon}}^\epsilon)(t-s).$$

To put Assumption 3.1 into context, we consider Hörmander type operators. Let L_X denote Lie differentiation in the direction of a vector field X and $[X, Y]$ the Lie bracket of two vector fields X and Y. Let $\{X_i, i = 0, 1, \ldots, m'\}$ be a family of smooth vector fields on a compact smooth manifold G and $\mathcal{L}_0 = \frac{1}{2}\sum_{i=1}^{m'} L_{X_i} L_{X_i} + L_{X_0}$. If $\{X_i, i = 1, \ldots, m'\}$ and their Lie brackets generate the tangent space $T_x G$ at each point x we say that the operator $\mathcal{L}_0$ satisfies the strong Hörmander's condition.

Lemma 3.2 *Suppose that $\mathcal{L}_0$ satisfies the strong Hörmander condition on a compact manifold G and let $Q_t(x, \cdot)$ be its family of transition probabilities. Then Assumption 3.1 holds with W identically 1. Furthermore the invariant probability measure π has a strictly positive smooth density w.r.t. the Lebesgue measure and*

$$\|Q_t(x, \cdot) - \pi(\cdot)\|_{TV} \leq Ce^{-\delta t}$$

for all x in G and for all $t > 0$.

Proof By Hörmander's theorem there are smooth functions $q_t(x, y)$ such that $Q_t(x, dy) = q_t(x, y)dy$. Furthermore $q_t(x, y)$ is strictly positive, see J.-M. Bony [2] and A. Sanchez-Calle [39]. Let $a = \inf_{x,y \in M} q_t(x, y) > 0$. Thus Döeblin's condition holds: if vol(A) denotes the volume of a Borel set A, $Q_t(x, A) \geq a \, \text{vol}(A)$. □

We say that W is a C^3 Lyapunov function for the ergodicity problem if there are constant $c \neq 0$ and $C > 0$ s.t. $\mathcal{L}_0 W \leq C - c^2 W$. If such a function exists, the $\mathcal{L}_0^\epsilon$ diffusions are conservative. Suppose that the Lyapunov function V satisfies in

addition the following conditions: there exists a number $\alpha \in (0,1)$ and $t_0 > 0$ s.t. for every $R > 0$,

$$\sup_{\{(x,y):V(x)+V(y)\le R\}} \|Q_{t_0}(x,\cdot) - Q_{t_0}(y,\cdot)\|_{TV} \le 2(1-\alpha),$$

Then there exists a unique invariant measure π such that Assumption 3.1 holds, see e.g. M. Hairer and J. Mattingly [13]. We mention the following standard estimates which helps to understand the estimates in Lemma 3.3.

Lemma 3.3 *Let W be a C^3 Lyapunov function for the ergodicity problem of $\mathcal{L}_0$, $\mathbf{E}W(z_0^\epsilon)$ is uniformly bounded in ϵ for ϵ sufficiently small. Then there exist numbers $\epsilon_0 > 0$ and c s.t. for all $t > 0$,*

$$\sup_{s\le t}\sup_{\epsilon\le\epsilon_0} \mathbf{E}W(z^\epsilon_{\frac{s}{\epsilon}}) \le c.$$

Proof By localizing (z_t^ϵ) if necessary, we see that $W(z_t^\epsilon) - W(z_0^\epsilon) - \frac{1}{\epsilon}\int_0^t \mathcal{L}_0 W(z_r^\epsilon)dr$ is a martingale. Let $c \neq 0$ and $C > 0$ be constant s.t. $\mathcal{L}_0 W \le C - c^2 W$. Then $\mathbf{E}W(z^\epsilon_{\frac{s}{\epsilon}}) \le \left(\mathbf{E}W(z_0^\epsilon) + \frac{1}{\epsilon}Ct\right) e^{-\frac{c^2}{\epsilon}t}$. □

As an application we see that, under the assumption of Lemma 3.3, the functions C_i in part (3) of Lemma 3.1 satisfy that $\sup_{\epsilon\le\epsilon_0} \mathbf{E}C_i(z^\epsilon_{\frac{s}{\epsilon}}) < \infty$.

Definition 3.1 *We say that a stochastic differential equation (SDE) on M is complete or conservative if for each initial point $y \in M$ any solution with initial value y exists for all $t \ge 0$. Let $\Phi_t(x)$ be its solution starting from x. The SDE is strongly complete if it has a unique strong solution and that $(t,x) \mapsto \Phi_t(x,\omega)$ is continuous for a.s. ω.*

From now on, by a solution we always mean a globally defined solution. For $\epsilon \in (0,1)$ we define $\mathcal{L}_0^\epsilon = \frac{1}{\epsilon}\mathcal{L}_0$. Let Q_t^ϵ denote their transition semigroups and transition probabilities. For each $\epsilon > 0$, let (z_t^ϵ) be an $\mathcal{L}_0^\epsilon$ diffusion. Let $\alpha_k \in \mathcal{B}_b(G;\mathbb{R})$ and (y_t^ϵ) solutions to the equations

$$\dot{y}_t^\epsilon = \sum_{k=1}^m Y_k(y_t^\epsilon)\alpha_k(z_t^\epsilon). \tag{3.1}$$

Let $\Phi^\epsilon_{s,t}$ be the solution flow to (3.1) with $\Phi^\epsilon_{s,s}(y) = y$. We denote by $\bar{g}$ the average of an integrable function $g : G \to \mathbb{R}$ with respect to π. Let

$$c_0(a,\delta) = \frac{a^2+a}{\delta^2} + \frac{3a}{\delta}, \quad c_W = c(a,\delta)(W + \bar{W}). \tag{3.2}$$

Lemma 3.4 *Suppose that Assumption 3.1 holds. Let $f, g \in \mathcal{B}_b(G;\mathbb{R})$ and $\bar{g} = 0$. Suppose that α_k are bounded. Then for any $F \in C^1(M;\mathbb{R})$, $0 \le s \le t$ and $0 < \epsilon < 1$,*

$$\left| \epsilon \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ F(y^\epsilon_{s_2}) g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 \right| \le 2\gamma_\epsilon |g|_\infty |f|_\infty (\epsilon^2 + (t-s)^2).$$

Here

$$\gamma_\epsilon = \left(|F(y^\epsilon_{\frac{s}{\epsilon}})| \, c_W(z^\epsilon_{\frac{s}{\epsilon}}) + \sum_{l=1}^m |\alpha_l|_\infty \frac{\epsilon}{t-s} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{ |(L_{Y_l} F)(y^\epsilon_r)| \, c_W(z^\epsilon_r) \mid \mathcal{F}_{\frac{s}{\epsilon}} \right\} dr \right).$$

Proof We first expand $F(y^\epsilon_{s_2})$ at $\frac{s}{\epsilon}$:

$$\epsilon \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ F(y^\epsilon_{s_2}) g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 = \epsilon F(y^\epsilon_{\frac{s}{\epsilon}}) \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1$$
$$+ \sum_{l=1}^m \epsilon \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \int_{\frac{s}{\epsilon}}^{s_2} \mathbf{E}\left\{ (dF)(Y_l(y^\epsilon_{s_3})) \alpha_l(z^\epsilon_{s_3}) g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_3 \, ds_2 \, ds_1$$

By part (3) of lemma 3.1

$$\left| \epsilon F(y^\epsilon_{\frac{s}{\epsilon}}) \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 \right| \le |F(y^\epsilon_{\frac{s}{\epsilon}})| |f|_\infty |g|_\infty c_W(z^\epsilon_{\frac{s}{\epsilon}}) \left(\epsilon^3 + (t-s)\epsilon \right).$$

It remain to estimate the summands in the second term, whose absolute value is bounded by the following

$$A_l := \left| \epsilon \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \int_{\frac{s}{\epsilon}}^{s_2} \mathbf{E}\left\{ (dF)(Y_l(y^\epsilon_{s_3})) \alpha_l(z^\epsilon_{s_3}) g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_3 \, ds_2 \, ds_1 \right|$$
$$= \left| \epsilon \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{ (dF)(Y_l(y^\epsilon_{s_3})) \alpha_l(z^\epsilon_{s_3}) \int_{s_3}^{\frac{t}{\epsilon}} \int_{s_2}^{\frac{t}{\epsilon}} \mathbf{E}\left\{ g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{s_3} \right\} \, ds_1 \, ds_2 \middle| \mathcal{F}_{\frac{s}{\epsilon}} \right\} ds_3 \right|.$$

For $s_3 \in [\frac{s}{\epsilon}, \frac{t}{\epsilon}]$, we apply part (3) of lemma 3.1 to bound the inner iterated integral,

$$\left| \int_{s_3}^{\frac{t}{\epsilon}} \int_{s_2}^{\frac{t}{\epsilon}} \mathbf{E}\left\{ g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{s_3} \right\} \, ds_1 \, ds_2 \right| = \left| \int_{s_3}^{\frac{t}{\epsilon}} \int_{s_3}^{s_1} \mathbf{E}\left\{ g(z^\epsilon_{s_2}) f(z^\epsilon_{s_1}) | \mathcal{F}_{s_3} \right\} \, ds_2 \, ds_1 \right|$$
$$\le \left(\epsilon^2 + t - \epsilon s_3 \right) c_W(z^\epsilon_{s_3}) |f|_\infty |g|_\infty.$$

We bring this back to the previous line, the notation $L_{Y_l}F = dF(Y_l)$ will be used,

$$A_l \leq \epsilon \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{\left|(dF)(Y_l(y^\epsilon_{s_3}))c_W(z^\epsilon_{s_3})\alpha_l(z^\epsilon_{s_3})\right|\mathcal{F}_{\frac{s}{\epsilon}}\right\}\left|\left(\epsilon^2+(t-\epsilon s_3)\right)|f|_\infty|g|_\infty\, ds_3\right.$$

$$\leq |f|_\infty|g|_\infty|\alpha_l|_\infty(t-s)(\epsilon^2+(t-s))\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{\left|(L_{Y_l}F)(y^\epsilon_{s_3})\right| c_W(z^\epsilon_{s_3})\middle|\mathcal{F}_{\frac{s}{\epsilon}}\right\} ds_3.$$

Putting everything together we see that, for γ_ϵ given in the Lemma, $\epsilon < 1$,

$$\left|\epsilon\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{F(y^\epsilon_{s_2})g(z^\epsilon_{s_2})f(z^\epsilon_{s_1})|\mathcal{F}_{\frac{s}{\epsilon}}\right\}\, ds_2\, ds_1\right| \leq 2\gamma_\epsilon|g|_\infty|f|_\infty\left(\epsilon^2+(t-s)^2\right).$$

The proof is complete. □

In Section 5.2 we will estimate γ_∞ and give uniform, in ϵ, moment estimates of functionals of (y^ϵ_t) on $[0, \frac{T}{\epsilon}]$.

Lemma 3.5 *Assume that (z^ϵ_t) satisfies Assumption 3.1 and α_j are bounded. If $F \in C^2(M;\mathbb{R})$ and $f \in \mathcal{B}_b(G;\mathbb{R})$, then for all $s \leq t$,*

$$\left|\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{F(y^\epsilon_r)f(z^\epsilon_r)|\mathcal{F}_{\frac{s}{\epsilon}}\right\} dr - \bar{f}\, F(y^\epsilon_{\frac{s}{\epsilon}})\right|$$
$$\leq \frac{2a}{\delta}|f|_\infty\left(W(z^\epsilon_{\frac{s}{\epsilon}})|F|(y^\epsilon_{\frac{s}{\epsilon}}) + \sum_{j=1}^{\mathfrak{m}}\gamma^j_\epsilon|\alpha_j|_\infty\right)\left(\frac{\epsilon^2}{t-s}+(t-s)\right)$$

where

$$\gamma^j_\epsilon(y) = c_W(z^\epsilon_{\frac{s}{\epsilon}})\,|L_{Y_j}F(y^\epsilon_{\frac{s}{\epsilon}})| + \sum_{l=1}^{m}|\alpha_l|_\infty\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{\left|L_{Y_l}L_{Y_j}F(y^\epsilon_r)\right| c_W(z^\epsilon_r)\mid\mathcal{F}_{\frac{s}{\epsilon}}\right\} dr.$$

Proof We note that,

$$\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} F(y^\epsilon_r)f(z^\epsilon_r)dr = F(y^\epsilon_{\frac{s}{\epsilon}})\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} f(z^\epsilon_r)dr$$
$$+\sum_{j=1}^{m}\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\int_{\frac{s}{\epsilon}}^{s_1} dF(Y_j(y^\epsilon_{s_2}))\alpha_j(z^\epsilon_{s_2})f(z^\epsilon_{s_1})ds_2ds_1.$$

Letting $\psi(r) = ae^{-\delta r}$, it is clear that for $k \geq 2$,

$$\left|\mathbf{E}\left\{\left(F(y^\epsilon_{\frac{s}{\epsilon}})\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} f(z^\epsilon_r)dr - \bar{f}\, F(y_{\frac{s}{\epsilon}})\right)\mid\mathcal{F}_{\frac{s}{\epsilon}}\right\}\right|$$
$$\leq \|f\|_W W(z^\epsilon_{\frac{s}{\epsilon}})\left|F(y^\epsilon_{\frac{s}{\epsilon}})\right|\frac{\epsilon^2}{t-s}\int_{\frac{s}{\epsilon^2}}^{\frac{t}{\epsilon^2}}\psi\left(r-\frac{s}{\epsilon^2}\right)dr \leq \frac{a}{\delta}\|f\|_W W(z^\epsilon_{\frac{s}{\epsilon}})\left|F(y^\epsilon_{\frac{s}{\epsilon}})\right|\frac{\epsilon^2}{t-s}.$$

To the second term we apply Lemma 3.4 and obtain the bound

$$\left|\mathbf{E}\left\{\sum_{j=1}^{m}\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\int_{\frac{s}{\epsilon}}^{s_1} dF(Y_j(y^\epsilon_{s_2}))\alpha_j(z^\epsilon_{s_2})f(z^\epsilon_{s_1})ds_2ds_1|\ \mathcal{F}_{\frac{s}{\epsilon}}\right\}\right|$$
$$\leq 2\sum_{j=1}^{m}\tilde{\gamma}^j_\epsilon|\alpha_j|_\infty|f|_\infty\left(\frac{\epsilon^2}{t-s}+(t-s)\right)$$

where

$$\gamma^j_\epsilon = |L_{Y_j}F|(y^\epsilon_{\frac{s}{\epsilon}})|\ c_W(z^\epsilon_{\frac{s}{\epsilon}})+\sum_{l=1}^{m}|\alpha_l|_\infty\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\mathbf{E}\left\{\left|(L_{Y_l}L_{Y_j}F)(y^\epsilon_r)\right|\ c_W(z^\epsilon_r)\mid\mathcal{F}_{\frac{s}{\epsilon}}\right\}dr.$$

Adding the error estimates together we conclude the proof. □

It is worth noticing that if $\phi:\mathbb{R}\to\mathbb{R}$ is a concave function $\phi(W)$ is again a Lyapunov function. Thus by using $\log W$ if necessary, we may assume uniform bounds on $\mathbf{E}W^p(z^\epsilon_{\frac{s}{\epsilon}})$ and further estimates on the conditional expectation in the error term are expected from Cauchy-Schwartz inequality. If G is compact then c_W is bounded. In Corollary 5.3, we will give uniform estimates on moments of γ^j_ϵ.

4 A Reduction

Let G be a smooth manifold of dimension n with volume measure dx. Let $H^s\equiv H^s(G)$ denote the Sobolev space of real valued functions over a manifold G and $\|-\|_s$ the Sobolev norm. The norm $(\|u\|_s)^2:=(2\pi)^{-n}\int|\hat{u}(\xi)|^2(1+|\xi|^2)^s d\xi$ extends from domains in $\mathbb{R}^n$ to compact manifolds, e.g. by taking supremum over $\|u\|_s$ on charts. If $s\in\mathcal{N}$, H^s is the completion of $C^\infty(M)$ with the norm $\|u\|_s=\sum_{j=0}^{s}\int(|\nabla^j u|)^2dx)^{\frac{1}{2}}$ where ∇ is usually taken as the Levi-Civita connection; when the manifold is compact this is independent of the Riemannian metric. And $u\in H^s$ if and only if for any function $\phi\in C^\infty_K$, ϕu in any chart belongs to H^s.

Let $\{X_i, i=0,1,\ldots,m'\}$ be a family of smooth vector fields on G and let us consider the Hörmander form operator $\mathcal{L}_0=\frac{1}{2}\sum_{i=1}^{m'}L_{X_i}L_{X_i}+L_{X_0}$. Let

$$\Lambda:=\{X_{i_1},[X_{i_1},X_{i_2}],[X_{i_1},[X_{i_2},X_{i_3}]],i_j=0,1,\ldots,m'\}.$$

If the vector fields in Λ generate T_xG at each $x\in G$, we say that Hörmander's condition is satisfied. By the proof in a theorem of Hörmander[18, Theorem1.1], if $\mathcal{L}_0$ satisfies the Hörmander condition then u is a C^∞ function in every open set where $\mathcal{L}_0u$ is a C^∞ function. There is a number $\delta>0$ such that there is an δ

improvement in the Sobolev regularity: if u is a distribution such that $\mathcal{L}_0 u \in H^s_{\text{loc}}$, then $u \in H^{s+\delta}_{\text{loc}}$.

Suppose that G is compact. Then $\|u\|_\delta \leq C(\|u\|_{L^2} + \|\mathcal{L}_0 u\|_{L^2})$, the resolvents $(\mathcal{L}_0 + \lambda I)^{-1}$ as operators from $L^2(G; dx)$ to $L^2(G; dx)$ are compact, and $\mathcal{L}_0$ is Fredholm on $L^2(dx)$, by which we mean that $\mathcal{L}_0$ is a bounded linear operator from Dom($\mathcal{L}_0$) to $L^2(dx)$ and has the Fredholm property : its range is closed and of finite co-dimension, the dimension of its kernel, ker($\mathcal{L}_0$) is finite. The domain of $\mathcal{L}_0$ is endowed with the norm $|u|_{\text{Dom}(\mathcal{L}_0)} = |u|_{L_2} + |\mathcal{L}_0 u|_{L_2}$. Let $\mathcal{L}_0^*$ denote the adjoint of $\mathcal{L}_0$. Then the kernel N of $\mathcal{L}_0^*$ is finite dimensional and its elements are measures on M with smooth densities in $L^2(dx)$. Denote $N^\perp$ the annihilator of N, $g \in L^2(dx)$ is in $N^\perp$ if and only if $\langle g, \pi \rangle = 0$ for all $\pi \in \ker(\mathcal{L}_0^*)$. Since $\mathcal{L}_0$ has closed range, $(\ker(\mathcal{L}_0^*))^\perp$ can be identified with the range of $\mathcal{L}_0$, and the set of g such that the Poisson equation $\mathcal{L}_0 u = g$ is solvable is exactly $N^\perp$. We denote by $\mathcal{L}_0^{-1} g$ a solution. Furthermore $\mathcal{L}_0^{-1} g$ is C^r whenever g is C^r. Denote by index($\mathcal{L}_0$), $\dim \ker \mathcal{L}_0 - \dim \text{Coker} \mathcal{L}_0$, the index of a Fredholm operator $\mathcal{L}_0$, where Coker $= L^2(dx)/\text{range}(\mathcal{L}_0)$. If $\mathcal{L}_0$ is self-adjoint, index($\mathcal{L}_0$) $= 0$.

Definition 4.1 *We say that $\mathcal{L}_0$ is a regularity improving Fredholm operator, if it is a Fredholm operator and $\mathcal{L}_0^{-1}\alpha$ is C^r whenever $\alpha \in C^r \cap N^\perp$.*

Let $\{W_t^k, k = 1, \ldots, m'\}$ be a family of independent real valued Brownian motions. We may and will often represent $\mathcal{L}_0^\epsilon$-diffusions (z_t^ϵ) as solutions to the following stochastic differential equations, in Stratonovich form,

$$dz_t^\epsilon = \frac{1}{\sqrt{\epsilon}} \sum_{k=1}^{m'} X_k(z_t^\epsilon) \circ dW_t^k + \frac{1}{\epsilon} X_0(z_t^\epsilon) dt.$$

Lemma 4.1 *Let $\mathcal{L}_0$ be a regularity improving Fredholm operator on a compact manifold G, $\alpha_k \in C^3 \cap N^\perp$, and $\beta_j = \mathcal{L}_0^{-1}\alpha_j$. Let (y_r^ϵ) be global solutions of (3.1) on M. Then for all $0 \leq s < t$, $\epsilon > 0$ and $f \in C^2(M; \mathbb{R})$,*

$$\begin{aligned} f(y^\epsilon_{\frac{t}{\epsilon}}) =& f(y^\epsilon_{\frac{s}{\epsilon}}) + \epsilon \sum_{j=1}^{m} \left(df(Y_j(y^\epsilon_{\frac{t}{\epsilon}}))\beta_j(z^\epsilon_{\frac{t}{\epsilon}}) - df(Y_j(y^\epsilon_{\frac{s}{\epsilon}}))\beta_j(z^\epsilon_{\frac{s}{\epsilon}}) \right) \\ & - \epsilon \sum_{i,j=1}^{m} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_i} L_{Y_j} f(y_r^\epsilon)) \alpha_i(z_r^\epsilon)\ \beta_j(z_r^\epsilon)\ dr \\ & - \sqrt{\epsilon} \sum_{j=1}^{m} \sum_{k=1}^{m'} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} df(Y_j(y_r^\epsilon))\ d\beta_j(X_k(z_r^\epsilon))\ dW_r^k. \end{aligned} \tag{4.1}$$

Suppose that, furthermore, for each $\epsilon > 0$, $j, k = 1, \ldots, m$, $\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}|df(Y_j(y_r^\epsilon))|^2|(d\beta_j(X_k)(z_r^\epsilon)|^2\ dr$ *is finite. Then,*

$$\mathbf{E}\left\{f(y^\epsilon_{\frac{t}{\epsilon}})|\mathcal{F}_{\frac{s}{\epsilon}}\right\} - f(y^\epsilon_{\frac{s}{\epsilon}}) = \epsilon\sum_{j=1}^m \left(\mathbf{E}\left\{df(Y_j(y^\epsilon_{\frac{t}{\epsilon}}))\beta_j(z^\epsilon_{\frac{t}{\epsilon}})|\mathcal{F}_{\frac{s}{\epsilon}}\right\} - df(Y_j(y^\epsilon_{\frac{s}{\epsilon}}))\beta_j(z^\epsilon_{\frac{s}{\epsilon}})\right)$$
$$-\epsilon\sum_{i,j=1}^m \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{L_{Y_i}L_{Y_j}f(y_r^\epsilon))\alpha_i(z_r^\epsilon)\ \beta_j(z_r^\epsilon)|\mathcal{F}_{\frac{s}{\epsilon}}\right\}\ dr. \tag{4.2}$$

Proof Firstly, for any C^2 function $f : M \to R$,

$$f(y^\epsilon_{\frac{t}{\epsilon}}) - f(y^\epsilon_{\frac{s}{\epsilon}}) = \sum_{j=1}^m \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} df(Y_j(y^\epsilon_{s_1}))\alpha_j(z_{s_1})ds_1.$$

Since the α_j's are C^2 so are β_j, following from the regularity improving property of $\mathcal{L}_0$. We apply Itô's formula to the functions $(L_{Y_j}f)\beta_j : M \times G \to \mathbb{R}$. To avoid extra regularity conditions, we apply Itô's formula to the function $df(Y_j)$, which is C^1, and the C^3 functions β_j separately and follow it with the product rule. This gives:

$$df(Y_j(y^\epsilon_{\frac{t}{\epsilon}}))\beta_j(z^\epsilon_{\frac{t}{\epsilon}}) = df(Y_j(y^\epsilon_{\frac{s}{\epsilon}}))\beta_j(z^\epsilon_{\frac{s}{\epsilon}}) + \sum_{j=1}^m \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_i}L_{Y_j}f(y_r^\epsilon)\,\alpha_i(z_r^\epsilon)\ \beta_j(z_r^\epsilon)\ \ dr$$
$$+\frac{1}{\sqrt{\epsilon}}\sum_{k=1}^{m'}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_j}f(y_r^\epsilon)\,d\beta_j\,(X_k(z_r^\epsilon))\,dW_r^k + \frac{1}{\epsilon}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_j}f(y_r^\epsilon)\,\mathcal{L}_0\beta_j(z_r^\epsilon)dr.$$

Substitute this into the earlier equation, we obtain (4.1). Part (4.2) is obvious, as we note that

$$\mathbf{E}\left(\sum_{k=1}^{m'}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} df(Y_j(y_r^\epsilon))(d\beta_j)\,(X_k(z_r^\epsilon))\ \ dW_r^k\right)^2 \le \sum_{k=1}^{m'}\mathbf{E}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} df(Y_j(y_r^\epsilon))|^2|d\beta(X_k(z_r^\epsilon))|^2|\ dr < \infty$$

and the stochastic integrals are L^2-martingales, so (4.2) follows. □

When G is compact, $d\beta(X_k)$ is bounded and the condition becomes: $\mathbf{E}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} df(Y_j(y_r^\epsilon))|^2\ dr$ is finite, which we discuss below. Otherwise, assumptions on $\mathbf{E}|d\beta(X_k(z_r^\epsilon))|^{2+}$ is needed.

5 Uniform Estimates

If $V : M \to \mathbb{R}_+$ is a locally bounded function such that $\lim_{y\to\infty} V(y) = \infty$ we say that V is a pre-Lyapunov function. Let $\alpha_k \in \mathcal{B}_b(G;\mathbb{R})$. Let $\{Y_k\}$ be C^1 smooth vector fields on M such that: either (a) each Y_k grows at most linearly; or (b) there exist a pre-Lyapunov function $V \in C^1(M;\mathbb{R}_+)$, positive constants c and K such that $\sum_{k=1}^m |L_{Y_k}V| \le c + KV$ then the equations (3.1) are complete. In case (a) let $o \in M$ and a be a constant such that $|Y_k(x)| \le a(1 + \rho(o,x))$ where ρ denotes the Riemannian distance function on M. For x fixed, denote $\rho_x = \rho(x,\cdot)$. By the definition of the Riemannian distance function,

$$\rho(y_t^\epsilon, y_0) \le \int_0^t |\dot{y}_s^\epsilon| ds = \sum_{k=1}^m \int_0^t |Y_k(y_s^\epsilon)\alpha_k(z_s^\epsilon)| ds \le \sum_{k=1}^m |\alpha_k|_\infty \int_0^t |Y_k(y_s^\epsilon)| ds.$$

This together with the inequality $\rho(y_t^\epsilon, o) \le \rho(y_t^\epsilon, y_0) + \rho(o, y_0^\epsilon)$ implies that for all $p \ge 1$, there exist constants C_1, C_2 depending on p such that

$$\sup_{s\le t} \rho^p(y_s^\epsilon, o) \le (C_1\rho^p(o, y_0^\epsilon) + C_2 t)\, e^{C_2 t^p}$$

where $C_2 = a^p C_1^2 (\sum_{k=1}^m |\alpha_k|_\infty)^p$. When restricted to $\{t < \tau^\epsilon\}$, the first time y_t^ϵ reaches the cut locus, the bounded is simple Ce^{Ct}. In case (b), for any $q \ge 1$,

$$\sup_{s\le t} (V(y_s^\epsilon))^q \le \left(V^q(y_0^\epsilon) + ctq \sum_{k=1}^m |\alpha_k|_\infty \right) \exp\left(q \sum_{k=1}^m |\alpha_k|_\infty (K + c) t \right),$$

which followed easily from the bound

$$|dV^q(\alpha_k Y_k)| = |qV^{q-1} dV(\alpha_k Y_k)| \le q|\alpha_k|_\infty (c + (c + K)V^q).$$

For the convenience of comparing the above estimates, which are standard and expected, with the uniform estimates of (y_t^ϵ) in Theorem 5.2 below in the time scale $\frac{1}{\epsilon}$, we record this in the following Lemma.

Lemma 5.1 *Let $\alpha_k \in \mathcal{B}_b(G;\mathbb{R})$. Let $\epsilon \in (0,1)$, $0 \le s \le t$, $\omega \in \Omega$.*

1. *If $\{Y_k\}$ grow at most linearly then (3.1) is complete and there exists $C, C(t)$ s.t.*
$$\sup_{0\le s\le t} \rho^p(y_s^\epsilon(\omega), o) \le (C\rho^p(o, y_0^\epsilon(\omega)) + C(t))\, e^{C(t)}.$$

2. *If there exist a pre-Lyapunov function $V \in C^1(M;\mathbb{R}_+)$, positive constants c and K such that $\sum_{j=1}^m |L_{Y_j}V| \le c + KV$, then (3.1) is complete.*

3. *If (3.1) is complete and there exists $V \in C^1(M;\mathbb{R}_+)$ such that $\sum_{j=1}^m |L_{Y_j}V| \le c + KV$ then there exists a constant C, s.t.*

$$\sup_{0\le s\le t} (V(y_s^\epsilon(\omega)))^q \le ((V(y_0^\epsilon(\omega)))^q + Ct)\, e^{Ct}.$$

If $V \in \mathcal{B}(M;\mathbb{R})$ is a positive function, denote by $B_{V,r}$ the following classes of functions:

$$B_{V,r} = \left\{ f \in C^r(M;\mathbb{R}) : \sum_{j=0}^r |d^j f| \le c + cV^q \text{ for some numbers } c,\, q \right\}. \quad (5.1)$$

In particular, $B_{V,0}$ is the class of continuous functions bounded by a function of the form $c + cV^q$. In $\mathbb{R}^n$, the constant functions and the function $V(x) = 1 + |x|^2$ are potential 'control' functions.

Assumption 5.1 *Assume that (i) (3.1) are complete for every $\epsilon \in (0,1)$, (ii) $\sup_\epsilon \mathbf{E}\,(V(y_0^\epsilon))^q$ is finite for every $q \ge 1$; and (iii) there exist a function $V \in C^2(M;\mathbb{R}_+)$, positive constants c and K such that*

$$\sum_{j=1}^m |L_{Y_j}V| \le c + KV, \quad \sum_{i,j=1}^m |L_{Y_i}L_{Y_j}V| \le c + KV.$$

Below we assume that $\mathcal{L}_0$ satisfies Hörmander's condition. We do not make any assumption on the mixing rate. Let $\beta_j = \mathcal{L}_0^{-1}\alpha_j$, $a_1 = \sum_{j=1}^m |\beta_j|_\infty$, $a_2 = \sum_{i,j=1}^m |\alpha_i|_\infty |\beta_j|_\infty$, $a_3 = \sum_{j=1}^m |d\beta_j|_\infty$, and $a_4 = \sum_{k=1}^m |X_k|_\infty^2$. We recall that if α_k and $\mathcal{L}_0$ satisfy Assumption 6.1 then $\mathcal{L}_0$ is a regularity improving Fredholm operator.

Theorem 5.2 *Let $\mathcal{L}_0$ be a regularity improving Fredholm operator on a compact manifold G, and $\alpha_k \in C^3(G;\mathbb{R}) \cap N^\perp$. Assume that Y_k satisfy Assumption 5.1. Then for all $p \ge 1$, there exists a constant $C = C(c, K, a_i, p)$ s.t. for any $0 \le s \le t$ and all $\epsilon \le \epsilon_0$,*

$$\mathbf{E}\left\{ \sup_{s\le u\le t} \left(V(y^\epsilon_{\frac{u}{\epsilon}})\right)^{2p} \mid \mathcal{F}_{\frac{s}{\epsilon}} \right\} \le \left(4V^{2p}(y^\epsilon_{\frac{s}{\epsilon}}) + C(t-s)^2 + C\right) e^{C(t-s+1)t}. \quad (5.2)$$

Here $\epsilon_0 \le 1$ depends on c, K, p, a_1 and V, Y_i, Y_j.

Proof Let $0 \leq s \leq t$. We apply (4.1) to $f = V^p$:

$$\begin{aligned} V^p(y^\epsilon_{\frac{t}{\epsilon}}) =& V^p(y^\epsilon_{\frac{s}{\epsilon}}) + \epsilon \sum_{j=1}^m dV^p\left(Y_j(y^\epsilon_{\frac{t}{\epsilon}})\right)\beta_j(z^\epsilon_{\frac{t}{\epsilon}}) - \epsilon\sum_{j=1}^m dV^p\left(Y_j(y^\epsilon_{\frac{s}{\epsilon}})\right)\beta_j(z^\epsilon_{\frac{s}{\epsilon}}) \\ &- \epsilon \sum_{i,j=1}^m \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_i}L_{Y_j}V^p\left(y^\epsilon_r\right)\alpha_i(z^\epsilon_r)\,\beta_j(z^\epsilon_r)\,dr \\ &- \sqrt{\epsilon}\sum_{k=1}^p \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \sum_{j=1}^m dV^p(Y_j(y^\epsilon_r))(d\beta_j)(X_k(z^\epsilon_r))\,dW^k_r. \end{aligned}$$

In the following estimates, we may first assume that $\sum_{j=1}^m |L_{Y_j}V|$ and $\sum_{i,j=1}^m |L_{Y_j}L_{Y_i}V|$ are bounded. We may then replace t by $t \wedge \tau_n$ where τ_n is the first time that either quantity is greater or equal to n. We take this point of view for proofs of inequalities and may not repeat it each time.

We take the supremum over $[s,t]$ followed by conditional expectation of both sides of the inequality:

$$\begin{aligned} \mathbf{E}\left\{\sup_{s\leq u\leq t} V^p(y^\epsilon_{\frac{u}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \leq& V^p(y^\epsilon_{\frac{s}{\epsilon}}) + \epsilon\mathbf{E}\left\{\sup_{s\leq u\leq t}\sum_{j=1}^m dV^p\left(Y_j(y^\epsilon_{\frac{u}{\epsilon}})\right)\beta_j(z^\epsilon_{\frac{u}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \\ &- \sum_{j=1}^m dV^p\left(Y_j(y^\epsilon_{\frac{s}{\epsilon}})\right)\beta_j(z^\epsilon_{\frac{s}{\epsilon}}) \\ &+ \epsilon\mathbf{E}\left\{\sup_{s\leq u\leq t}\left|\int_{\frac{s}{\epsilon}}^{\frac{u}{\epsilon}}\sum_{i,j=1}^m L_{Y_i}L_{Y_j}V^p\left(y^\epsilon_r\right)\alpha_i(z^\epsilon_r)\,\beta_j(z^\epsilon_r)\,dr\right| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \\ &+ \sqrt{\epsilon}\mathbf{E}\left\{\sup_{s\leq u\leq t}\left|\sum_{k=1}^{m'}\int_{\frac{s}{\epsilon}}^{\frac{u}{\epsilon}}\sum_{j=1}^m dV^p(Y_j(y^\epsilon_r))(d\beta_j)(X_k(z^\epsilon_r))dW^k_r\right| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}. \end{aligned}$$

By the conditional Jensen inequality and the conditional Doob's inequality, there exists a universal constant $\tilde{C}$ depending only on p s.t.,

$$
\begin{aligned}
&\mathbf{E}\left\{\sup_{s\le u\le t} V^{2p}(y^\epsilon_{\frac{u}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\\
&\le 4V^{2p}(y^\epsilon_{\frac{s}{\epsilon}}) + 4\epsilon^2\mathbf{E}\left(\left\{\sum_{j=1}^m |\beta_j|_\infty \sup_{s\le u\le t}\left|dV^p(Y_j(y^\epsilon_{\frac{u}{\epsilon}}))\right| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\right)^2\\
&+ 4\epsilon^2\left(\sum_{j=1}^m |\beta_j|_\infty \left|dV^p(Y_j(y^\epsilon_{\frac{s}{\epsilon}}))\right|\right)^2\\
&+ 8\epsilon(t-s)\mathbf{E}\left\{\left(\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\sum_{i,j=1}^m |\alpha_i|_\infty|\beta_j|_\infty \left|L_{Y_i}L_{Y_j}V^p\left(y^\epsilon_r\right)\right| dr\right)^2 \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\\
&+ \tilde{C}\sum_{k=1}^p \mathbf{E}\left\{\epsilon\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\left|\sum_{j=1}^m dV^p(Y_j(y^\epsilon_r))(d\beta_j)\left(X_k(z^\epsilon_r)\right)\right|^2 dr \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}.
\end{aligned}
$$

Since $\sum_j |L_{Y_j}V| \le c + KV$ and $\sum_{i,j=1}^p |L_{Y_i}L_{Y_j}V| \le c + KV$, there are constants c_1 and K_1 such that $\max_{j=1,\dots,m}|L_{Y_j}V^p| \le c_1 + K_1V^p$ and $\max_{i,j=1,\dots,m}|L_{Y_i}L_{Y_j}V^p| \le c_1 + K_1V^p$. This leads to the following estimate:

$$
\begin{aligned}
&\mathbf{E}\left\{\sup_{s\le u\le t} V^{2p}(y^\epsilon_{\frac{u}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\\
\le&4V^{2p}(y^\epsilon_{\frac{s}{\epsilon}}) + 8\epsilon^2(a_1)^2\left(2(c_1)^2 + (K_1)^2\mathbf{E}\left\{\sup_{s\le u\le t} V^{2p}(y^\epsilon_{\frac{u}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} + (K_1)^2V^{2p}(y^\epsilon_{\frac{s}{\epsilon}})\right)\\
&+ 16(a_2)^2(t-s)\epsilon\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\left((c_1)^2 + (K_1)^2\mathbf{E}\left\{V^{2p}(y^\epsilon_r) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\right)dr\\
&+ \tilde{C}(a_3a_4)^2\epsilon\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\mathbf{E}\left\{\left(c_1 + K_1V^p((y^\epsilon_r))\right)^2 \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} dr.
\end{aligned}
$$

Let $\epsilon_0 = \min\{\frac{1}{8a_1K_1}, 1\}$. For $\epsilon \le \epsilon_0$,

$$
\begin{aligned}
&\frac{1}{2}\mathbf{E}\left\{\sup_{s\le u\le t} V^{2p}(y^\epsilon_{\frac{u}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\\
\le&4V^{2p}(y^\epsilon_{\frac{s}{\epsilon}}) + 16\epsilon^2(a_1c_1)^2 + 16(t-s)^2(a_2c_1)^2 + 4\tilde{C}(a_3a_4c_1)^2(t-s)\\
&+ \left(16(a_2K_1)^2(t-s) + 4\tilde{C}(a_3a_4K_1)^2\right)\epsilon\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\mathbf{E}\left\{V^{2p}(y^\epsilon_r) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} dr.
\end{aligned}
$$

It follows that there exists a constant C such that for $\epsilon \leq \epsilon_0$,

$$\mathbf{E}\left\{\sup_{s\leq u\leq t} V^{2p}(y^\epsilon_{\frac{u}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \leq \left(4V^{2p}(y^\epsilon_{\frac{s}{\epsilon}}) + C(t-s)^2 + C\right) e^{C(t-s+1)t}.$$

□

Remark. If $M = \mathbb{R}^n$, Y_i are vector fields with bounded first order derivatives, then ρ_0^2 is a pre-Lyapunov function satisfying the conditions of Theorem 5.2, hence Theorem 5.2 holds. Let us recall that $B_{V,r}$ is defined in (5.1).

We return to Lemma 3.5 in Section 3 to obtain a key estimation for the estimation in Section 8. Let us recall that $B_{V,r}$ is defined in (5.1).

Corollary 5.3 *Assume (3.1) is complete, for every $\epsilon \in (0,1)$, and conditions of Assumption 3.1. Let $V \in \mathcal{B}(M;\mathbb{R}_+)$ be a locally bounded function and ϵ_0 a positive number s.t. for all $q \geq 1$ and $T > 0$, there exists a locally bounded function $C_q : \mathbb{R}_+ \to \mathbb{R}_+$, a real valued polynomial λ_q such that for $0 \leq s \leq t \leq T$ and for all $\epsilon \leq \epsilon_0$*

$$\sup_{s\leq u\leq t} \mathbf{E}\left\{V^q(y^\epsilon_{\frac{u}{\epsilon}}) \,|\mathcal{F}_{\frac{s}{\epsilon}}\right\} \leq C_q(t) + C_q(t)\lambda_q(V(y^\epsilon_{\frac{s}{\epsilon}})), \quad \sup_{0<\epsilon\leq\epsilon_0} \mathbf{E}(V^q(y^\epsilon_0)) < \infty. \tag{5.3}$$

Let $h \in \mathcal{B}_b(G;\mathbb{R})$. If $f \in B_{V,0}$ is a function s.t. $L_{Y_j} f \in B_{V,0}$ and $L_{Y_l}L_{Y_j} f \in B_{V,0}$ for all $j,l = 1,\ldots,m$, then for all $0 \leq s \leq t$,

$$\left|\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{f(y^\epsilon_r)h(z^\epsilon_r)|\mathcal{F}_{\frac{s}{\epsilon}}\right\} dr - \bar{h}\, f(y^\epsilon_{\frac{s}{\epsilon}})\right| \leq \tilde{c}|h|_\infty \gamma_\epsilon(y^\epsilon_{\frac{s}{\epsilon}})\left(\frac{\epsilon^2}{t-s} + (t-s)\right).$$

Here $\tilde{c}$ is a constant, see (5.4) below, and

$$\gamma_\epsilon = |f| + \sum_{j=1}^m |L_{Y_j} f| + \sum_{j,l=1}^m \frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{\left|L_{Y_l}L_{Y_j} f(y^\epsilon_r)\right| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} dr.$$

For all $s < t$ and $p \geq 1$,

$$\sup_{s\leq u\leq t}\sup_{\epsilon\leq\epsilon_0} \mathbf{E}\left(\gamma_\epsilon(y^\epsilon_{\frac{u}{\epsilon}})\right)^p < \infty.$$

More explicitly, if $\sum_{j=1}^m \sum_{l=1}^m |L_{Y_l}L_{Y_j} f| \leq K + KV^q$ where K, q are constants, then there exists a constant $C(t)$ depending only on the differential equation (3.1) s.t.

$$\gamma_\epsilon \leq |f| + \sum_{j=1}^m |L_{Y_j} f| + K + C(t)V^q.$$

Proof By Lemma 3.5,

$$\left|\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\mathbf{E}\left\{f(y_r^\epsilon)h(z_r^\epsilon)|\mathcal{F}_{\frac{s}{\epsilon}}\right\}dr-\bar{h}\ f(y_{\frac{s}{\epsilon}}^\epsilon)\right|$$
$$\leq\frac{2a}{\delta}|h|_\infty\left(W(z_{\frac{s}{\epsilon}}^\epsilon)|f(y_{\frac{s}{\epsilon}}^\epsilon)|+\sum_{j=1}^{\mathfrak{m}}\gamma_\epsilon^j|\alpha_j|_\infty\right)\left(\frac{\epsilon^2}{t-s}+(t-s)\right),$$

where $\gamma_\epsilon^j(y)=c_W(z_{\frac{s}{\epsilon}}^\epsilon)\ |L_{Y_j}f(y_{\frac{s}{\epsilon}}^\epsilon)|+\sum_{l=1}^m|\alpha_l|_\infty\dfrac{\epsilon}{t-s}\displaystyle\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\mathbf{E}\left\{\left|L_{Y_l}L_{Y_j}f(y_r^\epsilon)\right|\ c_W(z_r^\epsilon)\mid\mathcal{F}_{\frac{s}{\epsilon}}\right\}dr.$

Since W is bounded so is c_W, which is bounded by $2c(a,\delta)|W|_\infty$. Furthermore

$$\mathbf{E}\left\{\left|L_{Y_l}L_{Y_j}f(y_r^\epsilon)\right|\ c_W(z_r^\epsilon)\mid\mathcal{F}_{\frac{s}{\epsilon}}\right\}dr\leq 2c(a,\delta)|W|_\infty\mathbf{E}\left\{\left|L_{Y_l}L_{Y_j}f(y_r^\epsilon)\right|\ \mid\mathcal{F}_{\frac{s}{\epsilon}}\right\}dr.$$

We gather all constant together,

$$\tilde{c}=\frac{2a}{\delta}|W|_\infty+2c(a,\delta)|W|_\infty\sum_{j,l=1}^m|\alpha_j|_\infty+2\left(\sum_{j=1}^m|\alpha_j|_\infty\right)^2. \tag{5.4}$$

It is clear that,

$$\left|\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\mathbf{E}\left\{f(y_r^\epsilon)h(z_r^\epsilon)|\mathcal{F}_{\frac{s}{\epsilon}}\right\}dr-\bar{h}\ f(y_{\frac{s}{\epsilon}}^\epsilon)\right|\leq\tilde{c}\,\gamma_\epsilon|h|_\infty\left(\frac{\epsilon^2}{t-s}+(t-s)\right).$$

Since f, L_{Y_j} and $L_{Y_l}L_{Y_j}f\in B_{V,0}$, by (5.3), the following quantities are finite for all $p\geq 1$:

$$\sup_{\epsilon\leq\epsilon_0}\sup_{s\leq u\leq t}\mathbf{E}\left|(L_{Y_l}L_{Y_j}f)(y_{\frac{u}{\epsilon}}^\epsilon)\right|^p,\quad\sup_{\epsilon\leq\epsilon_0}\sup_{s\leq u\leq t}\mathbf{E}\left|L_{Y_j}f(y_{\frac{u}{\epsilon}}^\epsilon)\right|^p,\quad\sup_{\epsilon\leq\epsilon_0}\sup_{s\leq u\leq t}\mathbf{E}\left|f(y_{\frac{u}{\epsilon}}^\epsilon)\right|^p.$$

Furthermore since $\sum_{j=1}^m\sum_{l=1}^m|L_{Y_l}L_{Y_j}f|\leq K+KV^q$,

$$\sum_{j=1}^m\sum_{l=1}^m\frac{\epsilon}{t-s}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}}\mathbf{E}\left\{\left|L_{Y_l}L_{Y_j}f(y_r^\epsilon)\right|\ \mid\mathcal{F}_{\frac{s}{\epsilon}}\right\}dr\leq K+C(t)V^q(y_{\frac{s}{\epsilon}}^\epsilon).$$

Consequently, $\gamma_\epsilon\leq|f|+\sum_{j=1}^m|L_{Y_j}f|+K+C(t)V^q$, completing the proof. □

6 Convergence under Hörmander's Conditions

Below inj(M) denotes the injectivity radius of M and $\rho_y = \rho(y, \cdot)$ is the Riemannian distance function on M from a point y. Let o denote a point in M. The following proposition applies to an operator $\mathcal{L}_0$, on a compact manifold, satisfying Hörmander's condition.

Proposition 6.1 *Let M be a manifold with positive injectivity radius and $\epsilon_0 > 0$. Suppose conditions (1-5) below or conditions (1-3), (4') and (5).*

(1) $\mathcal{L}_0$ is a regularity improving Fredholm operator on $L^2(G)$ for a compact manifold G;

(2) $\{\alpha_k\} \subset C^3 \cap N^\perp$;

(3) Suppose that for $\epsilon \in (0, \epsilon_0)$, (3.1) is complete and $\sup_{\epsilon \le \epsilon_0} \mathbf{E}\rho(y_0^\epsilon, o) < \infty$;

(4) Suppose that there exists a locally bounded function V s.t. for all $\epsilon \le \epsilon_0$ and for any $0 \le s \le u \le t$, and for all $p \ge 1$,

$$\mathbf{E}V^p(y_0^\epsilon) \le c_0, \quad \sup_{s \le u \le t} \mathbf{E}\left\{ \left(V(y^\epsilon_{\frac{u}{\epsilon}}) \right)^p \mid \mathcal{F}_{\frac{s}{\epsilon}} \right\} \le K + KV^{p'}(y^\epsilon_{\frac{s}{\epsilon}})$$

where $c_0 = c_0(p)$, $K = K(p,t)$, and $p' = p'(p,t)$ is a natural number; K, p' are locally bounded in t.

(4') There exist a function $V \in C^2(M; \mathbb{R}_+)$, positive constants c and K such that

$$\sum_{j=1}^m |L_{Y_j} V| \le c + KV, \quad \sum_{i,j=1}^m |L_{Y_i} L_{Y_j} V| \le c + KV.$$

(5) For V in part (4) or in part (4'), suppose that for some number $\delta > 0$,

$$|Y_j| \in B_{V,0} \quad \sup_{\rho(y,\cdot) \le \delta} |L_{Y_i} L_{Y_j} \rho_y(\cdot)| \in B_{V,0}.$$

Then there exists a distance function $\tilde{\rho}$ on M that is compatible with the topology of M and there exists a number $\alpha > 0$ such that

$$\sup_{\epsilon \le \epsilon_0} \mathbf{E} \sup_{s \ne t} \left(\frac{\tilde{\rho}\left(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}} \right)}{|t-s|^\alpha} \right) < \infty,$$

and for any $T > 0$, $\{(y^\epsilon_{\frac{t}{\epsilon}}, t \le T), 0 < \epsilon \le 1\}$ is tight.

Proof By Theorem 5.2, conditions (1-3) and (4') imply condition (4). (a) Let $\delta < \min(1, \frac{1}{2}\text{inj}(M))$. Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be a smooth convex function such that $f(r) = r$ when $r \le \frac{\delta}{2}$ and $f(r) = 1$ when $r \ge \delta$. Then $\tilde{\rho}(x, y) = f \circ \rho$ is a distance function with $\tilde{\rho} \le 1$. Its open sets generate the same topology on M as that by ρ. Let β_j be a solution to $\mathcal{L}_0\beta_j = \alpha_j$. For any $y_0 \in M$, $|L_{Y_j}\tilde{\rho}^2(y_0, \cdot)| \le 2|Y_j(\cdot)|$. Since $|Y_j| \in B_{V,0}$, $\int_0^{\frac{t}{\epsilon}} \mathbf{E}|L_{Y_j}\tilde{\rho}|(y_r^\epsilon)|^2 dr < \infty$. We may apply (4.2) in Lemma 4.1,

$$\begin{aligned}
&\mathbf{E}\left\{\tilde{\rho}^2\left(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}}\right) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \\
=&\epsilon\sum_{j=1}^m \left(\mathbf{E}\left\{\left(L_{Y_j}\tilde{\rho}^2(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}})\right)\ \beta_j(z^\epsilon_{\frac{t}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} - \left(L_{Y_j}\tilde{\rho}^2(y^\epsilon_{\frac{s}{\epsilon}}, \cdot)\right)(y^\epsilon_{\frac{s}{\epsilon}})\ \beta_j(z^\epsilon_{\frac{s}{\epsilon}})\right) \\
&-\epsilon\sum_{i,j=1}^m \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{\left(L_{Y_i}L_{Y_j}\tilde{\rho}^2(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_r)\right)\ \alpha_i(z^\epsilon_r)\ \beta_j(z^\epsilon_r) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\ dr.
\end{aligned}$$

In the above equation, differentiation of $(\tilde{\rho})^2$ is w.r.t. to the second variable. By construction $\tilde{\rho}$ is bounded by 1 and $|\nabla\tilde{\rho}| \le |\nabla\rho| \le 1$. Furthermore since α_j are C^3 functions on a compact manifold, so β_j and $|\beta_j|$ are bounded. For any $y_0 \in M$, $L_{Y_j}\tilde{\rho}(y_0, \cdot) = \gamma'(\rho_{y_0})L_{Y_j}\rho_{y_0}$. Thus

$$\left|\mathbf{E}\left\{\left(L_{Y_j}\tilde{\rho}^2(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}})\right)\ \beta_j(z^\epsilon_{\frac{t}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\right| \le |\beta_j|_\infty \mathbf{E}\left\{\tilde{\rho}(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}})|Y_j(y^\epsilon_{\frac{t}{\epsilon}})| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}.$$

Recall $\tilde{\rho} \le 1$ and there are numbers K_1 and p_1 s.t. $|Y_j| \le K_1 + K_1 V^{p_1}$, so

$$\mathbf{E}\left\{|Y_j(y^\epsilon_{\frac{t}{\epsilon}})| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \le K_1 + K_1\mathbf{E}\left\{V^{p_1}(y^\epsilon_{\frac{t}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \le K_1 + K_1K(p_1, t)V^{p'(p_1,t)}(y^\epsilon_{\frac{s}{\epsilon}}).$$

Let $g_1 = K_1 + K_1K(p_1)V^{p'(p_1,t)}$, it is clear that $g_1 \in B_{V,0}$. We remark that,

$$L_{Y_i}L_{Y_j}(\tilde{\rho}^2) = (f^2)''(\rho)(L_{Y_i}\rho)(L_{Y_j}\rho) + (f^2)'(\rho)L_{Y_i}L_{Y_j}\rho.$$

By the assumption, there exists a function $g_2 \in B_{V,0}$ s.t.

$$\mathbf{E}\left\{\tilde{\rho}^2\left(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}}\right)|\mathcal{F}_{\frac{s}{\epsilon}}\right\} \le g_2(y_{\frac{s}{\epsilon}})\epsilon + g_2(y_{\frac{s}{\epsilon}})(t - s).$$

For $\epsilon \ge \sqrt{t-s}$, it is better to estimate directly from (3.1):

$$\begin{aligned}
\mathbf{E}\left\{\tilde{\rho}^2\left(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}}\right) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} &= \sum_{k=1}^m \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{2\tilde{\rho}\left(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}}\right) L_{Y_k}\tilde{\rho}\left(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}}\right)\alpha_k(z^\epsilon_r) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \\
&\le 2|\alpha_k|_\infty \sum_{k=1}^m \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{|Y_k(y^\epsilon_r)| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}\ dr \le g_3(y^\epsilon_{\frac{s}{\epsilon}})\left(\frac{t-s}{\epsilon}\right)
\end{aligned}$$

where $g_3 \in B_{V,0}$. We interpolate these estimates and conclude that for some function $g_4 \in B_{V,0}$ and a constant c the following holds: $\mathbf{E}\left\{\tilde{\rho}^2\left(y^\epsilon_{\frac{t}{\epsilon}}, y^\epsilon_{\frac{s}{\epsilon}}\right) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \leq (t-s)g_4(y^\epsilon_{\frac{s}{\epsilon}})$. There is a function $g_5 \in B_{V,0}$ s.t.

$$\mathbf{E}\tilde{\rho}^2\left(y^\epsilon_{\frac{t}{\epsilon}}, y^\epsilon_{\frac{s}{\epsilon}}\right) \leq \mathbf{E}g_5(y^\epsilon_0)(t-s) \leq c(t-s).$$

In the last step we use Assumption (4) on the initial value. By Kolmogorov's criterion, there exists $\alpha > 0$ such that

$$\sup_\epsilon \mathbf{E} \sup_{s \neq t}\left(\frac{\tilde{\rho}^2(y^\epsilon_{\frac{s}{\epsilon}}, y^\epsilon_{\frac{t}{\epsilon}})}{|t-s|^\alpha}\right) < \infty,$$

and the processes $(y^\epsilon_{\frac{s}{\epsilon}})$ are equi uniformly Hölder continuous on any compact time interval. Consequently the family of stochastic processes $\{y^\epsilon_{\frac{t}{\epsilon}}, 0 < \epsilon \leq 1\}$ is tight. □

If $\mathcal{L}_0$ is the Laplace-Beltrami operator on a compact Riemannian manifold and π its invariant probability measure then for any Lipschitz continuous function $f : G \to \mathbb{R}$,

$$\sqrt{\mathbf{E}\left(\frac{1}{t}\int_0^t f(z_s)ds - \int f\, d\pi\right)^2} \leq C(\|f\|_{Osc})\frac{1}{\sqrt{t}}. \tag{6.1}$$

where $\|f\|_{Osc}$ denotes the oscillation of f. If $\mathcal{L}_0$ is not elliptic we suppose it satisfies Hörmander's conditions and has index 0. The dimension of the kernel of $\mathcal{L}_0^*$ equals the dimension of the kernel of $\mathcal{L}_0$. Let $\{u_i, i = 1, \ldots, n_0\}$ be a basis in ker($\mathcal{L}_0$) and $\{\pi_i\, i = 1, \ldots, n_0\}$ the dual basis for the null space of $\mathcal{L}_0^*$. For $f \in L^2(G;\mathbb{R})$ we define $\bar{f} = \sum_{i=1}^{n_0} u_i\langle f, \pi_i\rangle$ where the bracket denotes the dual pairing between L^2 and $(L^2)^*$.

Lemma 6.2 *Suppose that (z_t) is a Markov process on a compact manifold G with generator $\mathcal{L}_0$ satisfying Hörmander's condition and having Fredholm index 0. Then for any function $f \in C^r(G;\mathbb{R})$, where $r \geq \max\{3, \frac{n}{2}+1\}$, there is a constant C depending on $|f|_{\frac{n}{2}+1}$, s.t.*

$$\sqrt{\mathbf{E}\left(\frac{1}{t-s}\int_s^t f(z_r)dr - \bar{f}\right)^2} \leq C(\|f-\bar{f}\|_{\frac{n}{2}+1})\frac{1}{\sqrt{t-s}}. \tag{6.2}$$

Proof Since $\langle f, \pi_j\rangle = \langle f, \pi_j\rangle$, $f - \bar{f} \in N^\perp$. By working with $f - \bar{f}$ we may assume that $f \in N^\perp$ and let g be a solution to $\mathcal{L}_0 g = f$. By Hörmander's theorem, [18], there is a positive number δ, such that for all $u \in C^\infty(M)$,

$$\|u\|_{s+\delta} \leq C(\|\mathcal{L}_0 u\|_s + \|u\|_{L_2}).$$

The number $\delta = 2^{1-k}$ where $k \in \mathcal{N}$ is related to the number of brackets needed to generate the tangent spaces.

Furthermore every u such that $\|\mathcal{L}_0 u\|_s < \infty$ must be in H^s. If $s > \frac{n}{2} + 1$, H^s is embedded in C^1 and for some constant c_i,

$$|g|_{C^1(M)} \leq c_1 \|g\|_{\frac{n}{2}+1+\epsilon} \leq c_2 \left(\|f\|_{\frac{n}{2}+1} + |g|_{L_2}\right) \leq c_3 \|f\|_{\frac{n}{2}+1}.$$

Recall that $\mathcal{L}_0 = \sum_{i=1}^{m'} L_{X_i} L_{X_i} + L_{X_0}$. Let $\{W_t^j, j = 1, \ldots, m'\}$ be independent one dimensional Brownian motions. Let (z_t) be solutions of $dz_t = \sum_{j=1}^{m'} X_j(z_t) \circ dW_t^j$. Since f is C^2,

$$\frac{1}{t-s}\int_s^t f(z_r)dr = \frac{1}{t-s}\left(g(z_t) - g(z_s)\right) - \frac{1}{t-s}\left(\sum_{j=1}^{m'}\int_s^t (dg(X_j))(z_r)dW_r^j\right).$$

We apply the Sobolev estimates to g and use Doob's L^2 inequality to see that for $t \geq 1$ there is a constant C such that,

$$\begin{aligned}\mathbf{E}\left(\frac{1}{t-s}\int_s^t f(z_r)dr\right)^2 &\leq \frac{4}{t^2}|g|_\infty^2 + \frac{8}{(t-s)^2}\sum_{j=1}^{m'}\int_s^t \left(\mathbf{E}|dg(z_r)|^2|X_j(z_r)|^2\right)dr \\ &\leq \frac{4}{(t-s)^2}(|g|_\infty)^2 + \frac{8m'}{t-s}(|dg|)_\infty^2 \sum_{j=1}^{m'}|X_j|_\infty^2 \leq C(\|f\|_{\frac{n}{2}+1})^2\frac{1}{t-s}.\end{aligned}$$

□

We remark that a self-adjoint operator satisfying Hörmander's condition has index zero.

Lemma 6.3 *Suppose that $\mathcal{L}_0$ satisfies Hörmander's condition. In addition it has Fredholm index 0 or it has a unique invariant probability measure. Let $r \geq \max\{3, \frac{n}{2}+1\}$. Let $h : M \times G \to \mathbb{R}$ be such that $h(y,\cdot) \in C^r$ for each y and that $|h|_\infty + \sup_z |h(\cdot,z)|_{\text{Lip}} + \sup_y |h(y,\cdot)|_{C^r} < \infty$. Let $s \leq t$ be a pair of positive numbers, and $F \in BC(C([0,s];M) \to \mathbb{R})$. For any equi -uniformly continuous subsequence, $\tilde{y}_t^n := (y_{\frac{t}{\epsilon_n}}^{\epsilon_n})$, of $(y_{\frac{t}{\epsilon}}^{\epsilon})$ that converges weakly to a continuous process $\bar{y}_\cdot$ as $n \to \infty$, the following convergence holds weakly:*

$$F(y_{\frac{\cdot}{\epsilon_n}}^{\epsilon_n})\int_s^t h(y_{\frac{u}{\epsilon_n}}^{\epsilon_n}, z_{\frac{u}{\epsilon_n}}^{\epsilon_n})du \to F(\bar{y}_\cdot)\int_s^t \overline{h(\bar{y}_u,\cdot)}du$$

where $\overline{h(y,\cdot)} = \sum_{i=1}^{n_0} u_i\langle h(y,\cdot), \pi_i\rangle$.

Proof For simplicity we omit the subscript n. The required convergence follows from Lemma 4.3 in [25] where it was assumed that (6.1) holds and $\mathcal{L}_0$ has a unique invariant measure for μ. It is easy to check that the proof there is valid. We take care to replace $\int_G h(y,z)d\mu(z)$ in Lemma 4.3 there by $\sum_{i=1}^{n_0} u_i\langle h(y,\cdot),\pi_i\rangle$. We remark that by the regularity improving property each u_i is smooth and therefore bounded. In the first part of the proof, we divide $[s,t]$ into sub-intervals of size ϵ, freeze the slow variable $(y^\epsilon_{\frac{u}{\epsilon}})$ on $[t_k,t_{k+1}]$, and approximate $h(y^\epsilon_{\frac{u}{\epsilon}},z^\epsilon_{\frac{u}{\epsilon}})$ by $h(y^\epsilon_{\frac{t_k}{\epsilon}},z^\epsilon_{\frac{u}{\epsilon}})$ on each sub-interval $[t_k,t_{k+1}]$. This approximation is clear: the computation is exactly as in Lemma 4.3 of [25] and we use the uniform continuity of (y^ϵ_t), the fact that $|h|_\infty$ and $\sup_z |h(\cdot,z)|_{\text{Lip}}$ are finite. The convergence of

$$\int_{\frac{t_{k-1}}{\epsilon}}^{\frac{t_{k-1}}{\epsilon}} h(y^\epsilon_{\frac{t_k}{\epsilon}},z^\epsilon_{\frac{u}{\epsilon}})du \to \Delta t_k \sum_{i=1}^{n_0} u_i\langle h(y^\epsilon_{\frac{t_{k-1}}{\epsilon}},\cdot),\pi_i\rangle$$

follows from the law of large numbers in Lemma 6.2. The convergence of

$$\sum_k \Delta t_k \sum_{i=1}^{n_0} u_i\langle h(y^\epsilon_{\frac{t_{k-1}}{\epsilon}},\cdot),\pi_i\rangle \to \sum_{i=1}^{n_0} u_i \int_s^t \langle h(y^\epsilon_{\frac{u}{\epsilon}},\cdot),\pi_i\rangle du$$

is also clear and follows from the Lipschitz continuity of h in the first variable and the equi continuity of the y^ϵ path. Finally denote by $y^\epsilon_{[0,s]}$ the restriction of the path $y^\epsilon_\cdot$ to the interval $[0,s]$, the weak convergence of $\sum_{i=1}^{n_0} u_i F(y^\epsilon_{[0,s]}) \int_s^t \langle h(y^\epsilon_{\frac{u}{\epsilon}},\cdot),\pi_i\rangle du$ to the required limit is trivial, as explained in Lemma 4.3, [25]. □

Assumption 6.1 *The generator $\mathcal{L}_0$ satisfies Hörmander's condition and has Fredholm index 0 (or has a unique invariant probability measure). For $k=1,\dots,m$, $\alpha_k \in C^r(G;\mathbb{R}) \cap N^\perp$ for some $r \geq \max\{3,\frac{n}{2}+1\}$.*

If $\mathcal{L}_0$ is elliptic, it is sufficient to assume $\alpha_k \in \mathcal{B}_b(G;\mathbb{R})$, instead of $\alpha_k \in C^r$.

Theorem 6.4 *If $\mathcal{L}_0$, α_k, (y^ϵ_0) and $|Y_j|$ satisfy the conditions of Proposition 6.1 and Assumption 6.1, then $(y^\epsilon_{\frac{t}{\epsilon}})$ converge weakly to the Markov process determined by the Markov generator*

$$\bar{\mathcal{L}} = -\sum_{i,j=1}^{m} \overline{\alpha_i\beta_j} L_{Y_i}L_{Y_j}, \quad \overline{\alpha_i\beta_j} = \sum_{b=1}^{n_0} u_b\langle \alpha_i\beta_j,\pi_b\rangle.$$

Proof By Proposition 6.1, $\{(y^\epsilon_{\frac{t}{\epsilon}}, t \geq 0)\}$ is tight. We prove that any convergent sub-sequence converges to the same limit. Let $\epsilon_n \to 0$ be a a monotone

sequence converging to zero such that the probability distributions of $(y^{\epsilon_n}_{\frac{t}{\epsilon_n}})$ converge weakly, on $[0,T]$, to a measure $\bar{\mu}$. For notational simplicity we may assume that $\{(y^{\epsilon}_{\frac{t}{\epsilon}}, t \geq 0)\}$ converges to $\bar{\mu}$.

Let $s < t$, $\{\mathcal{B}_s\}$ the canonical filtration, (Y_s) the canonical process, and $Y_{[0,s]}$ its restriction to $[0,s]$. By the Stroock-Varadhan martingale method, it is sufficient to prove $f(Y_t) - f(Y_s) - \int_s^t \bar{\mathcal{L}} f(Y_r)\, dr$ is a local martingale for any $f \in C^\infty_K(M)$. By (4.1), the following is a local martingale,

$$f(y^{\epsilon}_{\frac{t}{\epsilon}}) - f(y^{\epsilon}_{\frac{s}{\epsilon}}) - \epsilon \sum_{j=1}^{m} \Big(df(Y_j(y^{\epsilon}_{\frac{t}{\epsilon}}))\beta_j(z^{\epsilon}_{\frac{t}{\epsilon}}) + df(Y_j(y^{\epsilon}_{\frac{s}{\epsilon}}))\beta_j(z^{\epsilon}_{\frac{s}{\epsilon}}) \Big)$$
$$+ \epsilon \sum_{i,j=1}^{m} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_i} L_{Y_j} f(y^{\epsilon}_r))\alpha_i(z^{\epsilon}_r)\ \beta_j(z^{\epsilon}_r)\ dr.$$

Since the third term converges to zero as ϵ tends to zero, it is sufficient to prove

$$\lim_{\epsilon \to 0} \mathbf{E} \left\{ \epsilon \sum_{i,j=1}^{m} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_i} L_{Y_j} f(y^{\epsilon}_r))\alpha_i(z^{\epsilon}_r)\ \beta_j(z^{\epsilon}_r)\ dr - \int_s^t \bar{\mathcal{L}} f(y^{\epsilon}_{\frac{r}{\epsilon}})\ dr\ |\mathcal{F}_{\frac{s}{\epsilon}} \right\} = 0.$$

This follows from Lemma 6.3, completing the proof. □

Corollary 6.5 *Let $p \geq 1$ be a number and suppose that $\rho^p \in B_{V,0}$. Then, under the conditions of Theorem 6.4 and Assumption 5.1, $(y^{\epsilon}_{\cdot})$ converges in the Wasserstein p-distance on $C([0,t]; M)$.*

Proof By Theorem 5.2, $\sup_{\epsilon \leq \epsilon_0} \mathbf{E} \sup_{s \leq t} \rho^p(o, y^{\epsilon}_{\frac{s}{\epsilon}}) < \infty$. Let W_p denote the Wasserstein p distance:

$$W_p(\mu_1, \mu_2) = \left(\inf \int_{M \times M} \sup_{s \leq t} \rho(\sigma_1(s), \sigma_2(s)) d\mu(\sigma_1, \sigma_2) \right)^{\frac{1}{p}}.$$

Here the infimum is taken over all probability measures on the path spaces $C([0,t]; M)$ with marginals μ_1 and μ_2. Note that $C([0,t]; M)$ is a Banach space, a family of probability measures μ_n converges to μ in W_p, if and only if the following holds: (1) it converges weakly and (2) $\sup_n \int \sup_{s \leq t} \rho^p(o, \sigma_2(s)) d\mu_n(\sigma_2) < \infty$. The conclusion follows. □

7 A study of the semigroups

The primary aim of the section is to study the properties of $P_t f$ for $f \in B_{V,r}$ where P_t is the semigroup for a generic stochastic differential equation. These results

will be applied to the limit equation, to provide the necessary a priori estimates. Theorem 7.2 should be of independent interest, it also lead to Lemma 7.5, which will be used in Section 8.

Throughout this section M is a complete smooth Riemannian manifold. Let Y_0 be C^5 and $\{Y_k, k = 1, \dots, m\}$ be C^6 smooth vector fields on M, $\{B_t^k\}$ independent real valued Brownian motions. Let $(\Phi_t(y), t < \zeta(y))$ be the maximal solution to the following equation

$$dy_t = \sum_{k=1}^m Y_k(y_t) \circ dB_t^k + Y_0(y_t)dt \tag{7.1}$$

with initial value y. Its Markov generator is $\mathcal{L}f = \frac{1}{2}\sum_{k=1}^m L_{Y_k}L_{Y_k}f + L_{Y_0}f$. Let $Z = \frac{1}{2}\sum_{k=1}^m \nabla_{Y_k}Y_k + Y_0$ be the drift vector field, so

$$\mathcal{L}f = \frac{1}{2}\sum_{k=1}^m \nabla df(Y_k, Y_k) + df(Z). \tag{7.2}$$

If there exists a C^3 pre-Lyapunov function V, constants c and K such that $\mathcal{L}V \le c + KV$ then (7.1) is complete. However we do not limit ourselves to Lyapunov test for the completeness of the SDE. Let us denote $|f|_r = \sum_{k=1}^r |\nabla^{(k-1)}df|$ and $|f|_{r,\infty} = \sum_{k=1}^r |\nabla^{(k-1)}df|_\infty$. The following observation is useful.

Lemma 7.1 *Let $V \in \mathcal{B}(M;\mathbb{R})$ be locally bounded.*

- *Suppose that $\sum_{j=1}^m |Y_j| \in B_{V,0}$ and $|Z| \in B_{V,0}$. Then if $f \in B_{V,2}$, $\mathcal{L}f \in B_{V,0}$. If $f \in BC^2$, $|\mathcal{L}f| \le |f|_{2,\infty}F_1$ where $F_1 \in B_{V,0}$, not depending on f.*
- *Suppose that*

  $$\sum_{j=1}^m (|Y_j| + |\nabla Y_j| + |\nabla^{(2)}Y_j|) \in B_{V,0}, \quad |Z| + |\nabla Z| + |\nabla^{(2)}Z| \in B_{V,0}.$$

  If $f \in B_{V,4}$, $\mathcal{L}^2 f \in B_{V,0}$. If $f \in BC^4$, $|\mathcal{L}^2 f| \le |f|_{4,\infty}F_2$ where F_2 is a function in $B_{V,0}$, not dependent of f.

Proof That $\mathcal{L}f$ belongs to $B_{V,0}$ follows from (7.2). If $f \in BC^2$, $|\mathcal{L}f| \le (|f|_2)_\infty(\sum_{k=1}^m |Y_k|^2 + |Z|)$. For the second part we observe that $\mathcal{L}^2 f$ involves at most four derivatives of f and two derivatives of Y_j and Z where $j = 1, \dots, m$. □

Let $d\Phi_t(v)$ denote the derivative flow in the direction of $v \in T_yM$. It is the derivative of the function $y \mapsto \Phi_t(y, \omega)$, in probability. Moreover, it solves the following stochastic covariant differential equation along the solutions $y_t := \Phi_t(y_0)$,

$$Dv_t = \sum_{k=1}^m \nabla_{v_t}Y_k \circ dB_t^k + \nabla_{v_t}Y_0 dt.$$

Here $DV_t := /\!\!/_t(y_\cdot)d(/\!\!/_t^{-1}(y_\cdot)V_t)$ where $/\!\!/_t(y_\cdot) : T_{y_0}M \to T_{y_t}M$ is the stochastic parallel transport map along the path $y_\cdot$. Denote $|d\Phi_t|_{y_0}$ the norm of $d\Phi_t(y_0) : T_{y_0}M \to T_{y_t}M$. For $p > 0$, $y \in M$ and $v \in T_yM$, we define $H_p(y) \in \mathbb{L}(T_yM \times T_yM; \mathbb{R})$ by

$$H_p(y)(v,v) = \sum_{k=1}^m |\nabla Y_k(v)|^2 + (p-2)\sum_{k=1}^m \frac{\langle \nabla Y_k(v), v\rangle^2}{|v|^2} + 2\langle \nabla Z(v), v\rangle.$$

Let $\underline{h}_p(y) = \sup_{|v|=1\}} H_p(y)(v,v)$. Its upper bound will be used to control $|d\Phi_t|_y$.

Assumption 7.1 *The equation (7.1) is complete. Conditions (i) and (ii), or (i') and (ii), below hold.*

(i) There exists a locally bounded function $V \in \mathcal{B}(M;\mathbb{R}_+)$, s.t. for all $q \geq 1$ and $t \leq T$, there exists a number $C_q(t)$ and a polynomial λ_q such that

$$\sup_{s\leq t} \mathbf{E}(|V(\Phi_s(y))|^q) \leq C_q(t) + C_q(t)\lambda_q(V(y)). \tag{7.3}$$

(i') There exists $V \in C^3(M;\mathbb{R}_+)$ and constants c and K such that

$$\mathcal{L}V \leq c + KV, \quad |L_{Y_j}V| \leq c + KV, \quad j = 1,\ldots,m,$$

(ii) Let $\tilde{V} = 1 + \ln(1+|V|)$. For some constant c,

$$\sum_{k=1}^m |\nabla Y_k|^2 \leq c\tilde{V}, \quad \sup_{|v|=1} \langle \nabla Z(v), v\rangle \leq c\tilde{V}. \tag{7.4}$$

Remark. Suppose that (7.1) is complete. Since $\mathcal{L}V^q = qV^{q-1}\mathcal{L}V + q(q-1)V^{q-2}|L_{Y_j}V|^2$, (i') implies (i). In fact, $\mathbf{E}\sup_{s\leq t}(V(y_s))^q \leq (\mathbf{E}V(y_0)^q + cq^2t)\, e^{(c+K)q^2t}$.

Recall that (7.1) is strongly complete if $(t,y) \mapsto \Phi_t(y)$ is continuous almost surely on $[0,t]\times M$ for ant $t > 0$.

Theorem 7.2 *Under Assumption 7.1, the following statements hold.*

1. *The SDE (7.1) is strongly complete and for every $t \leq T$, $\Phi_t(\cdot)$ is C^4. Furthermore for all $p \geq 1$, there exists a positive number $C(t,p)$ such that*

$$\mathbf{E}\left(\sup_{s\leq t}|d\Phi_s(y)|^p\right) \leq C(t,p) + C(t,p)V^{C(t,p)}(y). \tag{7.5}$$

2. *Let $f \in B_{V,1}$. Define $\delta P_t(df)) = \mathbf{E} df(d\Phi_t(\cdot))$. Then $d(P_t f) = \delta P_t(df)$ and $|d(P_t f)| \in B_{V,0}$. Furthermore for a constant $C(t,p)$ independent of f,*

$$|d(P_t f)| \leq \sqrt{\mathbf{E}\left(|df|_{\Phi_t^x(y)}\right)^2}\sqrt{C(t,p)(1+V^{C(t,p)}(y))}.$$

3. *Suppose furthermore that*

$$\sum_{j=1}^{m}\sum_{\alpha=0}^{3}|\nabla^{(\alpha)}Y_j| \in B_{V,0}, \qquad \sum_{\alpha=0}^{2}|\nabla^{(\alpha)}Y_0| \in B_{V,0}.$$

Then, (a) $\mathbf{E}\sup_{s\leq t}|\nabla d\Phi_s|^2(y) \in B_{V,0}$; (b) If $f \in B_{V,2}$, then $P_t f \in B_{V,2}$, and

$$(\nabla dP_t f)(u_1,u_2) = \mathbf{E}\nabla df(d\Phi_t(u_1), d\Phi_t(u_2)) + \mathbf{E} df(\nabla_{u_1} d\Phi_t(u_2)).$$

Furthermore, (c) $\frac{dP_t f}{dt} = P_t\mathcal{L}f$, and $\mathcal{L}(P_t f) = P_t(\mathcal{L}f)$.

4. *Let $r \geq 2$. Suppose furthermore that*

$$\sum_{\alpha=0}^{r}|\nabla^{(\alpha)}Y_0| \in B_{V,0}, \quad \sum_{\alpha=0}^{r+1}\sum_{k=1}^{m}|\nabla^{(\alpha)}Y_k| \in B_{V,0}.$$

Then $\mathbf{E}\sup_{s\leq t}(|\nabla^{(r-1)}d\Phi_s|_y)^2$ belongs to $B_{V,0}$. If $f \in B_{V,r}$, then $P_t f \in B_{V,r}$.

Proof The statement on strong completeness follows from the following theorem, see Thm. 5.1 in [22]. Suppose that (7.1) is complete. If $\tilde{V}$ is a function and c_0 a number such that for all $t>0$, K compact, and all constants λ,

$$\sup_{y\in K}\mathbf{E}\exp\left(\lambda\int_0^t \tilde{V}(\Phi_s(y))ds\right) < \infty, \quad \sum_{k=1}^{m}|\nabla Y_k|^2 \leq c_0\tilde{V}, \quad \underline{h}_p \leq 6pc_0\tilde{V}, \tag{7.6}$$

then (7.1) is strongly complete. Furthermore for every $p \geq 1$ there exists a constant $c(p)$ such that

$$\mathbf{E}\left(\sup_{s\leq t}|d\Phi_s(y)|^p\right) \leq c(p)\mathbf{E}\left(\exp\left(6p^2\int_0^t \tilde{V}(\Phi_s(y))ds\right)\right). \tag{7.7}$$

Since Y_j are C^6, then for every t, $\Phi_t(\cdot)$ is C^4. It is easy to verify that condition (7.6) is satisfied. In fact, by the assumption $\underline{h}_p \leq 6pc\tilde{V}$. Take $\tilde{V} = 1+\ln(1+|V|)$ then for $p \geq 1$,

$$\mathbf{E}\left(\exp\left(6p^2\int_0^t \tilde{V}(\Phi_s(y))ds\right)\right) \leq C(t,p) + C(t,p)\left(V^{C(t,p)}(y)\right) < \infty.$$

This proves part (1).

For part (2) let $f \in C^1$. Then $y \mapsto f(\Phi_t(y,\omega))$ is differentiable for almost every ω. Let $\sigma : [0,t_0] \to M$ be a geodesic segment with $\sigma(0) = y$. Then

$$\frac{f(\Phi_t(\sigma_s,\omega)) - f(\Phi_t(y,\omega))}{s} = \frac{1}{s}\int_0^s \frac{d}{dr} f\left(\Phi_t(\sigma_r,\omega)\right) dr.$$

Since $\mathbf{E}|d\Phi_t(y)|^2$ is locally bounded in y, $r \mapsto \mathbf{E}|d\Phi_t(\sigma_r,\omega)|$ is continuous and the expectation of the right hand side converges to $\mathbf{E}df(d\Phi_t(\dot\sigma(0))$. The left hand side clearly converges almost surely. Since $\mathbf{E}|df(d\Phi_t(y))|^2$ is locally bounded the convergence is in L^1. We proved that $d(P_tf) = \delta P_t(df)$. Furthermore, suppose that $|df| \le K + KV^q$,

$$\begin{aligned}|d(P_tf)|_y &\le \sqrt{\mathbf{E}\left(|df|_{\Phi_t^\epsilon(y)}\right)^2}\sqrt{\mathbf{E}|d\Phi_t^\epsilon|_y^2} \\ &\le \sqrt{2K^2 + 2K^2\mathbf{E}V^{2q}(\Phi_t^\epsilon(y))}\sqrt{c(p)C(t,p) + c(p)C(t,p)\left(V^{C(t,p)}(y)\right)}.\end{aligned}$$

The latter, as a function of y, belongs to $B_{V,0}$.

We proceed to part (3a). Let $v, w \in T_yM$ and $U_t := \nabla d\Phi_t(w,v)$. Then U_t satisfies the following equation:

$$\begin{aligned}DU_t =& \sum_{k=1}^m \nabla^{(2)}Y_k(d\Phi_t(v), d\Phi_t(w)) \circ dB_t^k + \sum_{k=1}^m \nabla Y_k(U_t)\circ dB_t^k \\ &+ \nabla^{(2)}Y_0(d\Phi_t(v), d\Phi_t(w))dt + \nabla Y_0(U_t)dt.\end{aligned}$$

It follows that,

$$\begin{aligned}d|U_t|^2 =& 2\sum_{k=1}^m \left\langle \nabla^{(2)}Y_k(d\Phi_t(v), d\Phi_t(w)) \circ dB_t^k + \nabla^{(2)}Y_0(d\Phi_t(v), d\Phi_t(w))dt, U_t\right\rangle \\ &+ \left\langle \sum_{k=1}^m \nabla Y_k(U_t)\circ dB_t^k + \nabla Y_0(U_t)dt, U_t\right\rangle.\end{aligned}$$

To the first term on the right hand side we apply Cauchy Schwartz inequality to split the first term in the inner product and the second term in the inner product. This gives: $C|U_t|^2$ and other terms that does not involve U_t. The Stratonovich corrections will throw out the extra derivative $\nabla^{(3)}Y_k$ which does not involve U_t. The second term on the right hand side is a sum of the form $\sum_{k=1}^m \langle \nabla Y_k(U_t), U_t\rangle dB_t^k$ for which only bound on $|\nabla Y_k|$ is required, and

$$\left\langle \sum_{k=1}^m \nabla^{(2)}Y_k(Y_k, U_t) + \nabla Y_0(U_t), U_t\right\rangle = \langle \nabla Z(U_t), U_t\rangle - \left\langle \sum_{k=1}^m \nabla Y_k(\nabla_{U_t}Y_k), U_t\right\rangle.$$

The second term is bounded by

$$\left|\sum_{k=1}^{m}\langle\nabla Y_k(\nabla_{U_t}Y_k),U_t\rangle\right|\leq\sum_{k=1}^{m}|\nabla Y_k|^2|U_t|^2.$$

By the assumption, there exist $c>0,q\geq 1$ such that, for every $k=1,\ldots,m$,

$$|\nabla Y_k|\leq\tilde{V},|\nabla^2 Y_j|\leq c+cV^q,|\nabla^{(3)}Y_k|\leq c+cV^q,\langle\nabla_u Z,u\rangle\leq(c+KV)|u|^2.$$

There is a stochastic process I_s, which does not involve U_t, and constants C,q such that

$$\mathbf{E}|U_t|^2\leq\mathbf{E}|U_0|^2+\int_0^t\mathbf{E}I_r dr+\int_0^t C\mathbf{E}\tilde{V}^q(y_r^\epsilon)|U_r|^2dr.$$

By induction I_r has moments of all order which are bounded on compact intervals. By Gronwall's inequality, for $t\leq T$,

$$\mathbf{E}|U_t|^2\leq\left(\mathbf{E}|U_0|^2+\int_0^T\mathbf{E}I_r dr\right)\exp\left(C\int_0^t\tilde{V}^q(y_r^\epsilon)dr\right).$$

To obtain the supremum inside the expectation, we simply use Doob's L^p inequality before taking expectations. With the argument in the proof of part (1) we conclude that $\mathbf{E}\sup_{s\leq t}|\nabla d\Phi_s|^2(y)$ is finite and belongs to $B_{V,0}$.

Part (3b). Let $f\in B_{V,2}$. By part (1), $d(P_tf)=\mathbf{E}df(d\Phi_t(y))$. Let $u_1,u_2\in T_yM$. By an argument analogous to part (3), we may differentiate the right hand side under the expectation to obtain that

$$(\nabla dP_tf)(u_1,u_2)=\mathbf{E}\nabla df(d\Phi_t(u_1),d\Phi_t(u_2))+\mathbf{E}df(\nabla_{u_1}d\Phi_t(u_2)).$$

Hence $P_tf\in B_{V,2}$. This procedure can be iterated.

Part (3c). By Itô's formula,

$$f(y_t)=f(y_s)+\sum_{k=1}^{m}\int_s^t df(Y_k(y_r))dB_r^k+\int_s^t\mathcal{L}f(y_r)dr.$$

Since $df(Y_k)\in B_{V,0}$, the expectations of the stochastic integrals with respect to the Brownian motions vanish. Since $\mathcal{L}f\in B_{V,0}$ by part (3), $\mathcal{L}f(y_r)$ is bounded in L^2. It follows that the function $r\mapsto\mathbf{E}\mathcal{L}f(y_r)$ is continuous,

$$\lim_{t\to s}\frac{\mathbf{E}f(y_t)-\mathbf{E}f(y_s)}{t-s}=\mathbf{E}\mathcal{L}f(y_s)$$

and we obtain Kolmogorov's backward equation, $\frac{\partial}{\partial s}P_sf=P_s(\mathcal{L}f)$. Since $P_sf\in B_{V,2}$, we apply the above argument to P_sf, and take t to zero in $\frac{P_t(P_sf)-P_sf}{t}$ and obtain that $\frac{\partial}{\partial s}P_sf=\mathcal{L}(P_sf)$. This leads to the required statement $\mathcal{L}P_sf=P_s\mathcal{L}f$.

Part (4). For higher order derivatives of Φ_t we simply iterate the above procedure and note that the linear terms in the equation for $\frac{d}{dt}|\nabla^{k-1}d\Phi_t(u_1,\ldots,u_k)|^2$ are always of the same form. □

Remark 7.3 *With the assumption of part (3), we can show that for all integer p,* $\mathbf{E}\sup_{s\le t}|\nabla d\Phi_s|_y^p \in B_{V,0}$.

If we assume the additional conditions that

$$|\nabla Y_0| \le c\tilde{V}, \quad \sum_{k=1}^m |\nabla^{(2)}Y_k||Y_k| \le c\tilde{V},$$

the conclusion of the remark follows more easily. With the assumptions of part (5) we need to work a bit more which we illustrate below. Let $U_t = \nabla d\Phi_t(w,v)$. Instead of writing down all term in $|U_t|^p$ we classify the terms in $|U_t|^p$ into two classes: those involving U_t and those not. For the first class we must assume that they are bounded by $c\tilde{V}$ for some c. For the second class we may use induction and hence it is sufficient to assume that they belong to $B_{V,0}$. The terms that involving U_t are:

$$\nabla Y_k(U_t), \quad \sum_{k=1}^m \nabla^{(2)}Y_k(Y_k,U_t) + \nabla Y_0(U_t).$$

The essential identity to use is:

$$\sum_{k=1}^m \nabla^{(2)}Y_k(Y_k,U_t) + \nabla Y_0(U_t) = \nabla Z(U_t) - \sum_{k=1}^m \nabla Y_k(\nabla Y_k(U_t)).$$

We do not need to assume that the second order derivatives $|\nabla^{(2)}Y_k||Y_k| \le c\tilde{V}$, it is sufficient to assume that for $|\nabla Y_k|^2$ and ∇Z for all $k = 1,\ldots,m$. With a bit of care, we check that only one sided derivatives of Z are involved.

For example we can convert it to the $p = 2$ case,

$$d|U_t|^p = \frac{p}{2}(|U_t|^{p-2}) \circ d|U_t|^2 = \frac{p}{2}|U_t|^{p-2}d|U_t|^p + \frac{1}{4}p(p-1)|U_t|^{p-4}\langle d|U_t|^2\rangle.$$

By the first term $\frac{p}{2}|U_t|^{p-2}d|U_t|^p$ we mean that in place of $d|U_t|^p$ plug in all terms on the right hand side of the equation for $d|U_t|^2$, after formally converting the integrals to Itô form. By $\langle d|U_t|^2\rangle$ we mean the bracket of the martingale term on the right hand side of $d|U_t|^2$. It is now easy to check that in all the terms that involving U_t, higher order derivatives of Y_k does not appear, except in the form of $|U_t|^{p-2}\langle \nabla_{U_t}Z, U_t\rangle$.

Remark 7.4 *Assume the SDE is complete. Suppose that for some positive number C,*

$$\sum_{k=1}^{m}\sum_{k=0}^{5}|\nabla^{(k)}Y_k| \le C, \quad \sum_{k'=0}^{4}|\nabla^{(k')}Y_0| \le C.$$

Then for all $p \ge 1$, there exists a constant $C(t,p)$ such that

$$\mathbf{E}\left(\sup_{s\le t}|d\Phi_s(x)|^p\right) \le C(t,p).$$

Furthermore the statements in Theorem 7.2 hold for $r \le 4$.

Recall that $|f|_r = \sum_{k=1}^{r}|\nabla^{(k-1)}df|$ and $|f|_{r,\infty} = \sum_{k=1}^{r}|\nabla^{(k-1)}df|_\infty$.

Lemma 7.5 *Assume Assumption 7.1 and*

$$\sum_{\alpha=0}^{4}|\nabla^{(\alpha)}Y_0| \in B_{V,0}, \quad \sum_{\alpha=0}^{5}\sum_{k=1}^{m}|\nabla^{(\alpha)}Y_k| \in B_{V,0}.$$

Then there exist constants $q_1, q_2 \ge 1$, c_1 and c_2 depending on t and f and locally bounded in t, also functions $\gamma_i \in B_{V,0}$, λ_{q_i} polynomials, such that for $s \le t$,

$$\begin{aligned}
|P_tf(y_0) - P_sf(y_0)| &\le (t-s)c_1\left(1+\lambda_{q_1}(V(y_0))\right), \quad f \in B_{V,2}\\
|P_tf(y_0) - P_sf(y_0) - (t-s)P_s(\mathcal{L}f)(y_0)| &\le (t-s)^2c_2\left(1+\lambda_{q_2}(V(y_0))\right), \quad f \in B_{V,4}\\
|P_tf(y_0) - P_sf(y_0)| &\le (t-s)\left(1+|f|_{2,\infty}\right)\gamma_1(y_0), \quad \forall f \in BC^2\\
|P_tf(y_0) - P_sf(y_0) - (t-s)P_s(\mathcal{L}f)(y_0)| &\le (t-s)^2\left(1+|f|_{4,\infty}\right)\gamma_2(y_0), \quad \forall f \in BC^4.
\end{aligned}$$

Proof Denote $y_t = \Phi_t(y_0)$, the solution to (7.1). Then for $f \in C^2$,

$$P_tf(y_0) = P_sf(y_0) + \int_s^t P_r(\mathcal{L}f)(y_0)dr + \sum_{k=1}^{m}\mathbf{E}\left(\int_s^t df(Y_k(y_r))dB_r^k\right).$$

Since $|L_{Y_k}f| \le |df|_\infty|Y_k|$ and $|df|$, Y_k belong to $B_{V,0}$, by Assumption 7.1(i), $\int_0^t \mathbf{E}|L_{Y_k}f|^2_{y_r}dr$ is finite and the last term vanishes. Hence $|P_tf(y_0) - P_sf(y_0)| \le \int_s^t P_{s_2}(\mathcal{L}f)(y_0)ds_2$. By Lemma 7.1, $\mathcal{L}f \in B_{V,0}$ if $f \in B_{V,2}$. Let K, q_1 be s.t. $|\mathcal{L}f| \le K + KV^{q_1}$.

$$\int_s^r |P_{s_2}(\mathcal{L}f)(y_0)|ds_2 \le \int_0^r \left(K + K\mathbf{E}V^{q_1}(\Phi_{s_2}(y_0))\right)ds_2.$$

By the assumption, we see easily that $\sum_{k=0}^{3}|\nabla^{(\alpha)}Z| \in B_{V,0}$. By Assumption 7.1, $\sup_{s\le t}\mathbf{E}(|V(\Phi_s(y_0))|^{q_1}) \le C_{q_1}(t) + C_{q_1}(t)\lambda_{q_1}(V(y_0))$ and the first conclusion holds. We repeat this procedure for $f \in C^4$ to obtain:

$$P_tf(y_0) - P_sf(y_0)$$
$$= \int_s^t \left(P_s(\mathcal{L}f)(y_0) + \int_s^r P_{s_2}(\mathcal{L}^2f)(y_0)ds_2 + \sum_{k=1}^m \mathbf{E}\int_s^t \left(L_{Y_k}(\mathcal{L}f)\right)(y_{s_2}))dB^k_{s_2}\right) ds_1.$$

The last term also vanishes, as every term in $L_{Y_k}\mathcal{L}f$ belongs to $B_{V,0}$. Indeed

$$L_{Y_k}\mathcal{L}f = \sum_i \nabla^{(2)}df(Y_k, Y_i, Y_i) + 2\sum_i \nabla df\left(\nabla_{Y_k}Y_i, Y_i\right) + \nabla df\left(Y_k, Z\right)$$
$$+ \sum_i df(\nabla^{(2)}Y_i(Y_k, Y_i) + \nabla Y_i\left(\nabla_{Y_k}Y_i + \nabla_{Y_k}Y_0\right).$$

This gives, for all $f \in B_{V,4}$,

$$|P_tf(y_0) - P_sf(y_0) - (t-s)P_s(\mathcal{L}f)(y_0)| \le \left|\int_s^t\int_s^{s_1} P_{s_2}(\mathcal{L}^2f)(y_0)ds_2ds_1\right|. \tag{7.8}$$

Let q_2, K be numbers such that $|\mathcal{L}^2f| \le K + KV^{q_2}$. Then,

$$\sup_{s\le t} P_s(\mathcal{L}^2f)(y_0) \le K + K\mathbf{E}\left(V(y_s)\right)^{q_2} \le K + C_{q_2}(t) + KC_{q_2}(t)\tilde{\lambda}_{q_2}(V(y_0)).$$

Consequently, there exist a constant $c_2(t)$ s.t.

$$|P_tf(y_0) - P_sf(y_0) - (t-s)P_s(\mathcal{L}f)(y_0)| \le (t-s)^2c_2(t, K, q_2)(1 + \lambda_{q_2}(V(y_0))).$$

completing the proof for $f \in B_{V,2}$ and $B_{V,4}$. Next suppose that $f \in BC^2$. By Lemma 7.1, $|\mathcal{L}f| \le |f|_{2,\infty}F_1$, and $|\mathcal{L}^2f| \le |f|_{4,\infty}F_2$ if $f \in BC^4$. Here $F_1, F_2 \in B_{V,0}$ and do not depend on f. We iterate the argument above to complete the proof for $f \in BC^4$. □

8 Rate of Convergence

If $\mathcal{L}_0$ has a unique invariant probability measure π and $f \in L^1(G, d\pi)$ denote $\bar{f} = \int_G f d\pi$. Let $\bar{\mathcal{L}} = -\sum_{i,j=1}^m \overline{\alpha_i\beta_j}L_{Y_i}L_{Y_j}$. Let $\{\sigma^i_k, i, k = 1, \dots, m\}$ be the entries in a square root of the matrix $(\overline{-\alpha_i\beta_j})$. They satisfy $\sum_{k=1}^m \sigma^i_k\sigma^j_k = (\overline{-\alpha_i\beta_j})$ and are constants. Let us consider the SDE:

$$dy_t = \sum_{k=1}^m \left(\sum_{i=1}^m \sigma^i_k Y_i(y_t)\right) \circ dB^k_t, \tag{8.1}$$

where $\{B_t^k\}$ are independent one dimensional Brownian motions. Let

$$\tilde{Y}_k = \sum_{i=1}^m \sigma_k^i Y_i(y_t), \quad \tilde{Z} = \sum_{i,j=1}^m \overline{-\alpha_i \beta_j} \nabla_{Y_i} Y_j.$$

The results from section 7 apply. Recall that $\mathcal{L}_0 = \frac{1}{2}\sum_{i=1}^p L_{X_i} L_{X_i} + L_{X_0}$ and (z_t^ϵ) are $\mathcal{L}^\epsilon = \frac{1}{\epsilon}\mathcal{L}_0$ diffusions. Let $\Phi_t^\epsilon(y)$ be the solution to the SDE (1.5): $\dot{y}_t^\epsilon = \sum_{k=1}^m \alpha_k(z_t^\epsilon) Y_k(y_t^\epsilon)$ with initial value y.

Assumption 8.1 *G is compact, $Y_0 \in C^5(\Gamma TM)$, and $Y_k \in C^6(\Gamma TM)$ for $k = 1, \ldots, m$. Conditions (1)-(5) below hold or Conditions (1), (2') and (3-5) hold.*

(1) The SDEs (8.1) and (3.1) are complete.

(2) $V \in \mathcal{B}(M; \mathbb{R}_+)$ is a locally bounded function and ϵ_0 a positive number s.t. for all $q \geq 1$ and $T > 0$, there exists a locally bounded function $C_q : \mathbb{R}_+ \to \mathbb{R}_+$, a real valued polynomial λ_q such that for $0 \leq s \leq t \leq T$ and for all $\epsilon \leq \epsilon_0$

$$\sup_{s \leq u \leq t} \mathbf{E}\left\{ V^q(\Phi^\epsilon_{\frac{u}{\epsilon}}(y)) \,|\mathcal{F}_{\frac{s}{\epsilon}}\right\} \leq C_q(t) + C_q(t)\lambda_q\left(V(\Phi^\epsilon_{\frac{s}{\epsilon}}(y)\right). \tag{8.2}$$

(2') There exists a function $V \in C^3(M; \mathbb{R}_+)$ s.t. for all $i, j \in \{1, \ldots, m\}$, $|L_{Y_i} L_{Y_j} V| \leq c + KV$ and $|L_{Y_j} V| \leq c + KV$.

(3) For V defined above, let $\tilde{V} = 1 + \ln(1 + |V|)$. Suppose that

$$\sum_{\alpha=0}^4 |\nabla^{(\alpha)} Y_0| \in B_{V,0}, \quad \sum_{\alpha=0}^5 \sum_{k=1}^m |\nabla^{(\alpha)} Y_k| \in B_{V,0},$$
$$\sum_{j=1}^m |\nabla Y_j|^2 \leq c\tilde{V}, \quad \sup_{|u|=1} \langle \nabla \tilde{Z}(u), u\rangle \leq c\tilde{V}$$

(4) $\mathcal{L}_0$ satisfies Hörmander's conditions and has a unique invariant measure π satisfying Assumption 3.1.

(5) $\alpha_k \in C^3(G; \mathbb{R}) \cap N^\perp$.

We emphasize the following:

Remark 8.1 *(a) If V in (2') is a pre-Lyapunov function, then (3.1) is complete. Furthermore $|\bar{\mathcal{L}} V| \leq c + KV$ and so (8.1) is complete.*

(b) Under conditions (1), (2') and (4-5), (2) holds. See Theorem 5.2. Also Corollary 5.3 holds. Conditions (1-5) implies the conclusions of Theorem 7.2.

(c) If $\mathcal{L}_0$ satisfies strong Hörmander's condition, condition (4) is satisfied.

Let P_t^ϵ be the probability semigroup associated with (y_t^ϵ) and P_t the Markov semigroup for $\bar{\mathcal{L}}$. Recall that $|f|_{r,\infty} = \sum_{j=1}^r |\nabla^{(j-1)} df|_\infty$. We recall that operator $\mathcal{L}_0$ on a compact manifold G satisfying strong Hörmander's condition has an exponential mixing rate, so $\mathcal{L}_0$ satisfy Assumption 3.1.

Theorem 8.2 *Assume that Y_k, α_k and $\mathcal{L}_0$ satisfy Assumption 8.1. For every $f \in B_{V,4}$,*

$$\left|\mathbf{E} f\left(\Phi^\epsilon_{\frac{T}{\epsilon}}(y_0)\right) - P_T f(y_0)\right| \leq \epsilon |\log \epsilon|^{\frac{1}{2}} C(T)\gamma_1(y_0),$$

where $\gamma_1 \in B_{V,0}$ and $C(T)$ are constant increasing in T. Similarly, if $f \in BC^4$,

$$\left|\mathbf{E} f\left(\Phi^\epsilon_{\frac{T}{\epsilon}}(y_0)\right) - P_T f(y_0)\right| \leq \epsilon |\log \epsilon|^{\frac{1}{2}} C(T)\gamma_2(y_0)\left(1 + |f|_{4,\infty}\right).$$

where γ_2 is a function in $B_{V,0}$ that does not depend on f and $C(T)$ are constants increasing in T.

Proof *Step 1.* To obtain optimal estimates we work on intervals of order ϵ, c.f. Lemma 3.4. Let $t_0 = 0 < t_1 < \cdots < t_N = T$ be a partition of $[0, T]$ with $\Delta t_k = t_k - t_{k-1} = \epsilon$ for $k < N$ and $t_1 \leq \epsilon$. Write $y_t^\epsilon = \Phi_t^\epsilon(y_0)$. Then,

$$\begin{aligned}
&f\left(y^\epsilon_{\frac{T}{\epsilon}}\right) - P_T f(y_0) = \sum_{k=1}^N \left(P_{T-t_k} f(y^\epsilon_{\frac{t_k}{\epsilon}}) - P_{T-t_{k-1}} f(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\right) \\
&= \sum_{k=1}^N \left(P_{T-t_k} f(y^\epsilon_{\frac{t_k}{\epsilon}}) - P_{T-t_k} f(y^\epsilon_{\frac{t_{k-1}}{\epsilon}}) + \Delta t_k \left(P_{T-t_{k-1}} \bar{\mathcal{L}} f(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\right)\right) \\
&+ \sum_{k=1}^N \left(P_{T-t_k} f(y^\epsilon_{\frac{t_{k-1}}{\epsilon}}) - P_{T-t_{k-1}} f(y^\epsilon_{\frac{t_{k-1}}{\epsilon}}) - \Delta t_k \left(P_{T-t_{k-1}} \bar{\mathcal{L}} f\right)(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\right).
\end{aligned}$$

Define

$$\begin{aligned}
I_k^\epsilon &= P_{T-t_k} f(y^\epsilon_{\frac{t_k}{\epsilon}}) - P_{T-t_k} f(y^\epsilon_{\frac{t_{k-1}}{\epsilon}}) + \Delta t_k \left(P_{T-t_{k-1}} \bar{\mathcal{L}} f(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\right), \\
J_k^\epsilon &= P_{T-t_k} f - P_{T-t_{k-1}} f - \Delta t_k P_{T-t_{k-1}} \bar{\mathcal{L}} f.
\end{aligned}$$

Since $f \in B_{V,4}$, Lemma 7.5 applies and obtain the desired estimate on the second term:

$$\left|J_k^\epsilon(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\right| \leq (\Delta t_k)^2 \tilde{c}_2(T, f)\left(1 + \left(\lambda_{q_2}(V(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\right)\right)$$

where $\tilde{c}_2(T, f)$ is a constant and λ_{q_2} a polynomial.

Let K, q be constants such that $\lambda_{q_2}(V) \leq K + KV^q$. We apply (8.2) from Assumption 8.1 to see that for some constant $C_q(T)$ depending on $\lambda_{q_2}(V)$,

$$\mathbf{E}\left(\lambda_{q_2}(V(y^\epsilon_{t_{\frac{k-1}{\epsilon}}})\right) \leq K + KC_q(T) + KC_q(T)\lambda_q(V(y_0)).$$

Since $\Delta t_k \leq \epsilon$ and $N \sim \frac{1}{\epsilon}$,

$$\sum_{k=1}^{N} \mathbf{E}\left|J^\epsilon_k(y^\epsilon_{t_{\frac{k-1}{\epsilon}}})\right| \leq \epsilon\tilde{c}_2(T, f)(K+1)\left(1 + C_q(T) + C_q(T)\lambda_q(V(y_0))\right). \quad (8.3)$$

If f belongs to BC^4, we apply Lemma 7.5 to see that there exists a function $F \in B_{V,0}$, independent of f s.t.

$$\left|J^\epsilon_k(y^\epsilon_{t_{\frac{k-1}{\epsilon}}})\right| \leq (\Delta t_k)^2\left(1 + |f|_{4,\infty}\right)\left(F(y^\epsilon_{t_{\frac{k-1}{\epsilon}}})\right).$$

Hence

$$\sum_{k=1}^{N} \mathbf{E}\left|J^\epsilon_k(y^\epsilon_{t_{\frac{k-1}{\epsilon}}})\right| \leq \epsilon\left(1 + |f|_{4,\infty}\right)\mathbf{E}\left(F(y^\epsilon_{t_{\frac{k-1}{\epsilon}}})\right). \quad (8.4)$$

The rest of the proof is just as for the case of $f \in B_{V,4}$.

Step 2. Let $0 \leq s < t$. By part (3) of Theorem 7.2, $\bar{\mathcal{L}}P_t f = P_t\bar{\mathcal{L}}f$ for any $t > 0$ and $P_{T-t_k}\bar{\mathcal{L}}f = \bar{\mathcal{L}}P_{T-t_k}f$. We will approximate $P_{T-t_{k-1}}\bar{\mathcal{L}}f$ by $P_{T-t_k}\bar{\mathcal{L}}f$ and estimate the error

$$\sum_{k=1}^{N} \Delta t_k\left(P_{T-t_k}\bar{\mathcal{L}}f - P_{T-t_{k-1}}\bar{\mathcal{L}}f\right)(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}).$$

By Lemma 7.1, $\mathcal{L}f \in B_{V,2}$, and we may apply Lemma 7.5 to $\bar{\mathcal{L}}f$. We have,

$$|P_{T-t_k}\bar{\mathcal{L}}f(y_0) - P_{T-t_{k-1}}\bar{\mathcal{L}}f(y_0)| \leq \Delta t_k\tilde{c}_1(T)\left(1 + \lambda_{q_1}(V(y_0))\right).$$

Recall that $\lambda_{q_1}(V) \in B_{V,0}$. Summing over k and take the expectation of the above inequality we obtain that

$$\sum_{k=1}^{N} \Delta t_k\left|P_{T-t_k}\bar{\mathcal{L}}f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) - P_{T-t_{k-1}}\bar{\mathcal{L}}f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}})\right| \leq \epsilon c_1(T)\left(1 + \lambda_{q_1}(V(y_0))\right). \quad (8.5)$$

If $f \in BC^2$, $\mathcal{L}f \in BC^2$. By Lemma 7.5 ,

$$|P_{T-t_k}\bar{\mathcal{L}}f(y_0) - P_{T-t_{k-1}}\bar{\mathcal{L}}f(y_0)| \leq \Delta t_k\tilde{c}_1(T)\left(1 + \lambda_{q_1}(V(y_0))\right).$$

there exist constant $C(T)$ and a function $\gamma_1 \in B_{V,0}$, independent of f, s.t.

$$|P_t f(y_0) - P_s f(y_0)| \le (t-s)\,(1+|f|_{2,\infty})\,\gamma_1(y_0).$$

Here $\gamma_1 \in B_{V,0}$. Thus for $f \in BC^2$,

$$\sum_{k=1}^{N} \Delta t_k \left| P_{T-t_k} \bar{\mathcal{L}} f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) - P_{T-t_{k-1}} \bar{\mathcal{L}} f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) \right| \le 2\epsilon |f|_{2,\infty}(1+\gamma_1(y_0)). \quad (8.6)$$

Finally instead of estimating I^ϵ_k, we estimate

$$D^\epsilon_k := P_{T-t_k} f(y^\epsilon_{t_{\frac{k}{\epsilon}}}) - P_{T-t_k} f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) + \Delta t_k P_{T-t_k} \bar{\mathcal{L}} f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}).$$

Step 3. If $f \in B_{V,4}$, by Theorem 7.2, $P_t f \in B_{V,4}$ for any t. Since $\alpha_k \in N^\perp \cap C^3$, we may apply Lemma 4.1 to $P_{T-t_k} f$ and obtain the following formula for D^ϵ_k.

$$\begin{aligned}
D^\epsilon_k &= P_{T-t_k} f(y^\epsilon_{t_{\frac{k}{\epsilon}}}) - P_{T-t_k} f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) + \Delta t_k P_{T-t_k} \bar{\mathcal{L}} f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) \\
&= \epsilon \sum_{j=1}^{m} \left(dP_{T-t_k} f(Y_j(y^\epsilon_{t_{\frac{k}{\epsilon}}}))\beta_j(z^\epsilon_{t_{\frac{k}{\epsilon}}}) - dP_{T-t_k} f(Y_j(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}))\beta_j(z^\epsilon_{t_{\frac{k-1}{\epsilon}}}) \right) \\
&+ \Delta t_k P_{T-t_k} \bar{\mathcal{L}} f(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) - \epsilon \sum_{i,j=1}^{m} \int_{t_{\frac{k-1}{\epsilon}}}^{t_{\frac{k}{\epsilon}}} \left(L_{Y_i} L_{Y_j} P_{T-t_k} f(y^\epsilon_r) \right) \alpha_i(z^\epsilon_r)\, \beta_j(z^\epsilon_r)\, dr \\
&- \sqrt{\epsilon} \sum_{j=1}^{m} \sum_{k=1}^{m'} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} dP_{T-t_k} f(Y_j(y^\epsilon_r))\, d\beta_j(X_k(z^\epsilon_r))\, dW^k_r.
\end{aligned}$$

Since $Y_0, Y_k \in B_{V,0}$, $L_{Y_i} L_{Y_j} P_{T-t_k} f \in B_{V,0}$, which follows the same argument as for Lemma 7.1. In particular, for each $0 < \epsilon \le \epsilon_0$,

$$\int_0^{\frac{t}{\epsilon}} \mathbf{E}\left(\left| L_{Y_i} L_{Y_j} P_{T-t_k} f(y^\epsilon_r) \right| \right)^2 dr < \infty.$$

The expectation of the martingale term in the above formula vanishes. For $j = 1, \dots, m$ and $k = 1, \dots, N$, let

$$\begin{aligned}
A^\epsilon_{jk} &= dP_{T-t_k} f\left(Y_j(y^\epsilon_{t_{\frac{k}{\epsilon}}}) \right) \beta_j(z^\epsilon_{t_{\frac{k}{\epsilon}}}) - dP_{T-t_k} f\left(Y_j(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) \right) \beta_j(z^\epsilon_{t_{\frac{k-1}{\epsilon}}}), \\
B^\epsilon_k &= \Delta t_k (P_{T-t_k} \bar{\mathcal{L}} f)(y^\epsilon_{t_{\frac{k-1}{\epsilon}}}) - \epsilon \sum_{i,j=1}^{m} \int_{t_{\frac{k-1}{\epsilon}}}^{t_{\frac{k}{\epsilon}}} \left(L_{Y_i} L_{Y_j} P_{T-t_k} f \right)(y^\epsilon_r) \alpha_i(z^\epsilon_r)\, \beta_j(z^\epsilon_r)\, dr.
\end{aligned}$$

Step 4. We recall that $\bar{\mathcal{L}}P_{T-t_k}f = \sum_{i,j=1}^m \overline{\alpha_i\beta_j} L_{Y_i}L_{Y_j}P_{T-t_k}f$. By Theorem 7.2, $L_{Y_i}L_{Y_j}P_{T-t_k}f$ is C^2. Furthermore by Assumption 3.1, the (z_t^ϵ) diffusion has exponential mixing rate. We apply Corollary 5.3 to each function of the form $L_{Y_i}L_{Y_j}P_{T-t_k}f$ and take $h = \alpha_i\beta_j$ There exist a constant $\tilde{c}$ and a function $\gamma_{i,j,,k,\epsilon} \in B_{V,0}$ such that

$$|B_k^\epsilon| \le \Delta t_k \sum_{i,j=1}^m \left| \overline{\alpha_i\beta_j}\; L_{Y_i}L_{Y_j}P_{T-t_k}f\left(y^\epsilon_{\frac{t_{k-1}}{\epsilon}}\right) - \frac{\epsilon}{\Delta t_k}\int_{\frac{t_{k-1}}{\epsilon}}^{\frac{t_k}{\epsilon}} \mathbf{E}\left\{ L_{Y_i}L_{Y_j}P_{T-t_k}f(y_r^\epsilon)(\alpha_i\beta_j)(z_r^\epsilon)|\mathcal{F}_{\frac{t_{k-1}}{\epsilon}}\right\} d$$

$$\le \sum_{i,j=1}^m \tilde{c}|\alpha_i\beta_j|_\infty \gamma_{i,j,k,\epsilon}(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\left(\epsilon^2 + (\Delta t_k)^2\right),$$

where denoting $G^k_{i,j} := L_{Y_i}L_{Y_j}P_{T-t_k}f$,

$$\gamma_{i,j,k,\epsilon} = |G^k_{i,j}| + \sum_{l'=1}^m |L_{Y_{l'}}G^k_{i,j}| + \sum_{l,l'=1}^m \frac{\epsilon}{\Delta t_k}\int_{\frac{t_{k-1}}{\epsilon}}^{\frac{t_k}{\epsilon}} \mathbf{E}\left\{ \left| L_{Y_l}L_{Y_{l'}}G^k_{i,j}(y_r^\epsilon)\right| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} dr.$$

By Theorem 7.2, $G^k_{i,j} = L_{Y_i}L_{Y_j}P_{T-t_k}f$ belong to $B_{V,2}$. Furthermore $G^k_{i,j}$ and its first two derivatives are bounded by a function in $B_{V,0}$ which depends on f only through $\sum_{k=0}^4 P_{T-t_k}(|\nabla^{(k)}df|^p)$, for some p. Thus there are numbers c, q such that for all k, $\max_{i,j}|\gamma_{i,j,k,\epsilon}| \le c + cV^q$, for some c, q. Since $\Delta t_k \le \epsilon \le 1$, $N \sim O(\frac{1}{\epsilon})$, we summing over k,

$$\sum_{k=1}^N \mathbf{E}|B_k^\epsilon| \le 2\epsilon\cdot c\cdot\tilde{c}\sum_{i,j=1}^m |\alpha_i\beta_j|_\infty C_q(T)\sup_k \mathbf{E}\left(1 + V^q(y^\epsilon_{\frac{t_{k-1}}{\epsilon}})\right) \le \epsilon C(T)\tilde{\gamma}(y_0), \tag{8.7}$$

for some constant $C(T)$ and some function $\tilde{\gamma}$ in $B_{V,0}$. If $f \in BC^4$, it is easy to see that there is a function $g \in B_{V,0}$, not depending on f, s.t. $\max_{i,j,k}\mathbf{E}\gamma_{i,j,k,\epsilon}(y^\epsilon_{\frac{t_{k-1}}{\epsilon}}) \le C(T)g(y_0)|f|_{4,\infty}$.

Step 5. Finally, by Lemma 8.4 below, for $\epsilon \le s \le t \le T$ and $f \in B_{V,3}$, there is a constant C and function $\tilde{\gamma} \in B_{V,0}$, depending on T, f s.t. for $0 \le s < t \le T$,

$$\left|\sum_{j=1}^m \mathbf{E}df(Y_j(y^\epsilon_{\frac{t}{\epsilon}}))\beta_j(z^\epsilon_{\frac{t}{\epsilon}}) - \mathbf{E}df(Y_j(y^\epsilon_{\frac{s}{\epsilon}}))\beta_j(z^\epsilon_{\frac{s}{\epsilon}})\right| \le C\gamma(y_0)\epsilon\sqrt{|\log\epsilon|} + C\gamma(y_0)(t-s). \tag{8.8}$$

For the partition $t_0 < t_1 < \cdots < t_N$, we assumed that $t_1 - t_0 \le \epsilon$ and $\Delta t_k = \epsilon$ for $k \ge 1$. Let $k \ge 2$. Since $dP_{T-t_k}f(Y_j) \in B_{V,3}$, estimate (8.8) holds also with f replaced by $dP_{T-t_k}f(Y_j)$, and we have:

$$\left|\sum_{j=1}^m \epsilon\mathbf{E}A^\epsilon_{jk}\right| \le C\tilde{\gamma}(y_0)\epsilon^2\sqrt{|\log\epsilon|}, \quad k \ge 2 \tag{8.9}$$

Since β_j are bounded and by Theorem 7.2 $dP_{T-t_k}f$ is bounded by a function in $B_{V,0}$ that does not depend on k, for $\epsilon \leq \epsilon_0$, each term $\mathbf{E}|A^\epsilon_{jk}|$ is bounded by a function in $B_{V,0}$ and $\sup_{0<\epsilon\leq\epsilon_0}|\mathbf{E}A^\epsilon_{jk}|$ is of order $\epsilon\tilde{\gamma}(y_0)$ for some function $\tilde{\gamma} \in B_{V,0}$. We ignore a finite number of terms in the summation. In particular we will not need to worry about the terms with $k = 1$. Since the sum over k involves $O(\frac{1}{\epsilon})$ terms the following bound follows from (8.9):

$$\sum_{k=1}^{N}\left|\sum_{j=1}^{m}\epsilon\mathbf{E}A^\epsilon_{jk}\right| \leq C\tilde{\gamma}(y_0)\epsilon\sqrt{|\log\epsilon|}. \tag{8.10}$$

Here $\tilde{\gamma} \in B_{V,0}$ and may depend on f. The case of $f \in BC^4$ can be treated similarly. The estimate is of the form $\tilde{\gamma}(\epsilon) = (1 + |f|_{4,\infty})\gamma_0$ where $\gamma_0 \in B_{V,0}$ does not depend on f. We putting together (8.3), (8.5), (8.7) and (8.10)to see that if $f \in B_{V,4}$,

$$\left|\mathbf{E}f\left(\Phi^\epsilon_{\frac{t}{\epsilon}}(y_0)\right) - P_tf(y_0)\right| \leq C(T)\gamma(y_0)\epsilon\sqrt{|\log\epsilon|},$$

where $\gamma \in B_{V,0}$. If $f \in BC^4$, collecting the estimates together, we see that there is a constant $C(T)$ s.t.

$$\left|\mathbf{E}f\left(\Phi^\epsilon_{\frac{t}{\epsilon}}(y_0)\right) - P_tf(y_0)\right| \leq \epsilon\sqrt{|\log\epsilon|}\,C(T)\left(1 + \sum_{k=1}^{4}|\nabla^{(k-1)}df|_\infty\right)\tilde{\gamma}(y_0)$$

where $\tilde{\gamma}$ is a function in $B_{V,0}$ that does not depend on f. By induction the finite dimensional distributions converge and hence the required weak convergence. The proof is complete. □

Lemma 8.3 *Assume that (3.1) are complete for all $\epsilon \in (0, \epsilon_0)$, some $\epsilon_0 > 0$.*

(1) $\mathcal{L}_0$ is a regularity improving Fredholm operator on a compact manifold G, $\alpha_k \in C^3 \cap N^\perp$.

(2) There exists $V \in C^2(M;\mathbb{R}_+)$, constants c, K, s.t.

$$\sum_{j=1}^{m}|L_{Y_j}V| \leq c + KV, \quad \sum_{j=1}^{m}|L_{Y_i}L_{Y_j}V| \leq c + KV.$$

(2') There exists a locally bounded $V : M \to \mathbb{R}_+$ such that for all $q \geq 2$ and $t > 0$ there are constants $C(t)$ and q', with the property that

$$\sup_{s\leq u\leq t}\mathbf{E}\left\{(V(y^\epsilon_u))^q \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} \leq CV^{q'}(y^\epsilon_{\frac{s}{\epsilon}}) + C. \tag{8.11}$$

(3) For V in part (2) or in part (2'), $\sup_\epsilon \mathbf{E}V^q(y_0^\epsilon) < \infty$ for all $q \geq 2$.

For $f \in C^2$ with the property that $L_{Y_j}f, L_{Y_i}L_{Y_j}f \in B_{V,0}$ for all i, j, there exists a number $\epsilon_0 > 0$ s.t. for every $0 < \epsilon \leq \epsilon_0$,

$$\left|\mathbf{E}\left\{f(y^\epsilon_{\frac{t}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} - f(y^\epsilon_{\frac{s}{\epsilon}})\right| \leq \gamma_1(y^\epsilon_{\frac{s}{\epsilon}}) \max_j |\beta_j|_\infty \, \epsilon + (t-s)\gamma_2(y^\epsilon_{\frac{s}{\epsilon}}) \max_i |\alpha_i|_\infty \max_j |\beta_j|_\infty.$$

Here $\gamma_1, \gamma_2 \in B_{V,0}$ and depend on $|f|$ only through $|L_{Y_j}f|$ and $|L_{Y_j}L_{Y_i}f|$. In particular there exists $\gamma \in B_{V,0}$ s.t. for all $0 < \epsilon \leq \epsilon_0$,

$$\left|\mathbf{E}f(y^\epsilon_{\frac{t}{\epsilon}}) - \mathbf{E}f(y^\epsilon_{\frac{s}{\epsilon}})\right| \leq \sup_{0<\epsilon\leq\epsilon_0} \mathbf{E}\gamma(y_0^\epsilon)(t - s + \epsilon).$$

Furthermore, $\sup_{0<\epsilon\leq\epsilon_0} \mathbf{E}\left|f(y^\epsilon_{\frac{t}{\epsilon}}) - f(y^\epsilon_{\frac{s}{\epsilon}})\right| \leq (\epsilon + \sqrt{t-s}))\mathbf{E}\gamma(y_0^\epsilon)$.

Proof Since the hypothesis of Theorem 5.2 holds, if V is as defined in (2), it satisfies (2'). Since $L_{Y_j}f \in B_{V,0}$, $\sup_{s\leq t} \mathbf{E}|L_{Y_j}f(y^\epsilon_{\frac{s}{\epsilon}})|^2$ is finite. We apply Lemma 4.1:

$$\mathbf{E}\left\{f(y^\epsilon_{\frac{t}{\epsilon}}) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\} = f(y^\epsilon_{\frac{s}{\epsilon}}) + \epsilon \sum_{j=1}^m \mathbf{E}\left\{\left(df(Y_j(y^\epsilon_{\frac{t}{\epsilon}}))\beta_j(z^\epsilon_{\frac{t}{\epsilon}}) - df(Y_j(y^\epsilon_{\frac{s}{\epsilon}}))\beta_j(z^\epsilon_{\frac{s}{\epsilon}})\right) \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}$$
$$- \epsilon \sum_{i,j=1}^m \mathbf{E}\left\{\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} L_{Y_i}L_{Y_j}f(y_r^\epsilon))\alpha_i(z_r^\epsilon)\, \beta_j(z_r^\epsilon)\, dr \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}.$$

Let

$$\gamma_1(y^\epsilon_{\frac{s}{\epsilon}}) = 2 \sup_{s\leq r\leq t} \sum_{j=1}^m \mathbf{E}\left\{|L_{Y_j}f(y^\epsilon_{\frac{r}{\epsilon}})| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}, \quad \gamma_2(y^\epsilon_{\frac{s}{\epsilon}}) = \sup_{s\leq r\leq t} \sum_{i,j=1}^m \mathbf{E}\left\{|L_{Y_i}L_{Y_j}f(y^\epsilon_{\frac{s}{\epsilon}}))| \mid \mathcal{F}_{\frac{s}{\epsilon}}\right\}.$$

Since $L_{Y_j}f$ and $L_{Y_i}L_{Y_j}f \in B_{V,0}$, $\gamma_1, \gamma_2 \in B_{V,0}$. The required conclusion follows for there conditioned inequality, and hence the estimate for $\left|\mathbf{E}f(y^\epsilon_{\frac{t}{\epsilon}}) - \mathbf{E}f(y^\epsilon_{\frac{s}{\epsilon}})\right|$. To estimate $\mathbf{E}\left|f(y^\epsilon_{\frac{t}{\epsilon}}) - f(y^\epsilon_{\frac{s}{\epsilon}})\right|$, we need to involve the diffusion term in (4.1) and hence $\sqrt{t-s}$ appears.

□

Lemma 8.4 *Assume the conditions of Lemma 8.3 and Assumption 3.1. Let $y_0^\epsilon = y_0$. If $f \in C^3$ is s.t. $|L_{Y_j}f|$, $|L_{Y_i}L_{Y_j}f|$, $|L_{Y_l}L_{Y_i}L_{Y_j}f|$ belong to $B_{V,0}$ for all i, j, k, then for some ϵ_0 and all $0 < \epsilon \leq \epsilon_0$ and for all $0 \leq \epsilon \leq s < t \leq T$ where $T > 0$,*

$$\left|\sum_{l=1}^m \mathbf{E}df(Y_l(y^\epsilon_{\frac{t}{\epsilon}}))\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) - \mathbf{E}df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right| \leq C(T)\gamma(y_0)\epsilon\sqrt{|\log\epsilon|} + C(T)\gamma(y_0)(t-s),$$

where $\gamma \in B_{V,0}$ and $C(T)$ is a constant. If the assumptions of Theorem 8.2 holds, the above estimate holds for any $f \in B_{V,3}$; if $f \in BC^3$, we may take $\gamma = (|f|_{3,\infty}+1)\tilde{\gamma}$ where $\tilde{\gamma} \in B_{V,0}$.

Proof Let $t \le T$. Since $\beta_l(z^\epsilon_{\frac{t}{\epsilon}})$ is the highly oscillating term, we expect that averaging in the oscillation in β_l gains an ϵ in the estimation. We first split the sums:

$$\begin{aligned}&\left(df(Y_l(y^\epsilon_{\frac{t}{\epsilon}}))\beta_l(z^\epsilon_{\frac{t}{\epsilon}})\right) - \left(df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right)\\ &= df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\left(\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) - \beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right) + \left(df(Y_l(y^\epsilon_{\frac{t}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\right)\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) = I_l + II_l.\end{aligned} \tag{8.12}$$

By Assumption 3.1, $\mathcal{L}_0$ has mixing rate $\psi(r) = ae^{-\delta r}$. Let $s' < s \le t$,

$$\begin{aligned}&\left|\mathbf{E}df(Y_l(y^\epsilon_{\frac{s'}{\epsilon}}))\left(\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) - \beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right)\right| \le \mathbf{E}\left(\left|df\left(Y_l(y^\epsilon_{\frac{s'}{\epsilon}})\right)\right| \cdot \left|\frac{1}{\epsilon}\int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \mathbf{E}\left\{\alpha_l(z^\epsilon_r)|\mathcal{F}_{\frac{s'}{\epsilon}}\right\} dr\right|\right)\\ &\le \mathbf{E}\left|df\left(Y_l(y^\epsilon_{\frac{s'}{\epsilon}})\right)\right| \frac{1}{\epsilon}\int_0^{\frac{t-s}{\epsilon}} \psi\left(\frac{r+\frac{s-s'}{\epsilon}}{\epsilon}\right) dr\\ &\le \frac{a^2}{\delta}e^{-\frac{\delta(s-s')}{\epsilon^2}}\,\mathbf{E}\left|df\left(Y_l(y^\epsilon_{\frac{s'}{\epsilon}})\right)\right|.\end{aligned}$$

If $s - s' = \delta_0\epsilon^2|\log\epsilon|$, $\exp\left(-\frac{\delta(s-s')}{\epsilon^2}\right) = \epsilon^{\delta\delta_0}$. We apply Theorem 5.2 to the functions $L_{Y_l}f \in B_{V,0}$. For a constant $\epsilon_0 > 0$,

$$\frac{a^2}{\delta}\sup_{0<\epsilon\le\epsilon_0}\sup_{0\le s'\le t}\mathbf{E}\left|\left(df(Y_l(y^\epsilon_{\frac{s'}{\epsilon}}))\right)\right| \le \tilde{\gamma}_l(y_0)$$

where $\tilde{\gamma}_l$ is a function in $B_{V,0}$, depending on T. Thus for $s' < s < t$,

$$\left|\mathbf{E}\left(df(Y_l(y^\epsilon_{\frac{s'}{\epsilon}}))\left(\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) - \beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right)\right)\right| \le \tilde{\gamma}_l(y_0)\frac{a^2}{\delta}\exp\left(-\frac{\delta(s-s')}{\epsilon^2}\right). \tag{8.13}$$

Let us split the first term on the right hand side of (8.12). Denoting $s' = s - \frac{1}{\delta}\epsilon^2|\log\epsilon|$,

$$\begin{aligned}I_l &= \mathbf{E}df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\left(\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) - \beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right)\\ &= \mathbf{E}df(Y_l(y^\epsilon_{\frac{s'}{\epsilon}}))\left(\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) - \beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right) + \mathbf{E}\left(\left(df(Y_l(y^\epsilon_{\frac{s}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s'}{\epsilon}}))\right)\left(\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) - \beta_l(z^\epsilon_{\frac{s}{\epsilon}})\right)\right).\end{aligned}$$

The first term on the right hand side is estimated by (8.13). To the second term we take the supremum norm of β_l and use Lemma 8.3. For some $\tilde{C}(T)$ and $\gamma \in B_{V,0}$,

$$\mathbf{E}\left|df(Y_l(y^\epsilon_{\frac{s}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s'}{\epsilon}}))\right| \le \tilde{C}(T)\gamma(y_0)\left(\epsilon + \frac{1}{\sqrt{\delta}}\epsilon|\log\epsilon|^{\frac{1}{2}}\right). \tag{8.14}$$

Then for some number $C(T)$,

$$\sum_l I_l \le \frac{1}{\sqrt{\delta}}\epsilon\sqrt{|\log\epsilon|}C(T)\gamma(y_0) \tag{8.15}$$

where $\gamma \in B_{V,0}$. Let us treat the second term on the right hand side of (8.12). Let $t' = t - \frac{1}{\delta}\epsilon^2|\log\epsilon|$. Then

$$\begin{aligned} II_l &= \mathbf{E}\left(df(Y_l(y^\epsilon_{\frac{t}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\right)\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) \\ &= \mathbf{E}\left(df(Y_l(y^\epsilon_{\frac{t}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{t'}{\epsilon}}))\right)\beta_l(z^\epsilon_{\frac{t}{\epsilon}}) + \mathbf{E}\left(df(Y_l(y^\epsilon_{\frac{t'}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\right)\beta_l(z^\epsilon_{\frac{t}{\epsilon}}). \end{aligned}$$

To the first term we apply (8.14) and obtain a rate $\frac{1}{\sqrt{\delta}}\epsilon\sqrt{|\log\epsilon|}$. We could assume that β_l averages to zero. Subtracting the term $\bar{\beta}_l$ does not change I_l. Alternatively Lemma 8.3 provides an estimate of order ϵ for $\left|\mathbf{E}\left(df(Y_l(y^\epsilon_{\frac{t}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\right)\right|$. Finally, since $\int \beta d\pi = 0$,

$$\begin{aligned} &\left|\mathbf{E}\left(df(Y_l(y^\epsilon_{\frac{t'}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\right)\beta_l(z^\epsilon_{\frac{t}{\epsilon}})\right| = \left|\mathbf{E}\left(df(Y_l(y^\epsilon_{\frac{t'}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\right)\mathbf{E}\left\{\beta_l(z^\epsilon_{\frac{t}{\epsilon}})\,|\mathcal{F}_{\frac{t'}{\epsilon}}\right\}\right| \\ &\le \mathbf{E}\left|df(Y_l(y^\epsilon_{\frac{t'}{\epsilon}})) - df(Y_l(y^\epsilon_{\frac{s}{\epsilon}}))\right||\beta_l|_\infty a e^{-\delta\frac{t-t'}{\epsilon^2}} \le \gamma_l(y_0)|\beta_l|_\infty a\epsilon. \end{aligned}$$

In the last step we used condition (2') and γ_l is a function in $B_{V,0}$. We have proved the first assertion.

If the assumptions of Theorem 8.2 holds, for any $f \in B_{V,3}$, the following functions belong to $B_{V,0}$: $|L_{Y_j}f|$, $|L_{Y_i}L_{Y_j}f|$, and $|L_{Y_l}L_{Y_i}L_{Y_j}f|$. If $f \in BC^3$, the above mentioned functions can be obviously controlled by $|f|_{3,\infty}$ multiplied by a function in $B_{V,0}$, thus completing the proof. □

9 Rate of Convergence in Wasserstein Distance

Let $\mathcal{B}(M)$ denotes the collection of Borel sets in a C^k smooth Riemannian manifold M with the Riemannian distance function ρ; let $\mathbb{P}(M)$ be the space of probability measures on M. Let $\epsilon \in (0, \epsilon_0)$ where ϵ_0 is a positive number. If $P_\epsilon \to P$ weakly, we may use either the total variation distance or the Wasserstein distance, both imply weak convergence, to measure the rate of the convergence of P_ϵ to P. Let ρ denotes the Riemannian distance function. The Wasserstein 1-distance is

$$d_W(P,Q) = \inf_{(\pi_1)^*\mu=P,(\pi_2)^*\mu=Q}\int_{M\times M}\rho(x,y)d\mu(x,y).$$

Here $\pi_i : M \times M \to M$ are projections to the first and the second factors respectively, and the infimum are taken over probability measures on $M\times M$ that couples

Q and P. If the diameter, diam(M), of M is finite, then the Wasserstein distance is controlled by the total variation distance, $d_W(P,Q) \leq \text{diam}(M)\|P-Q\|_{TV}$. See C. Villani [44].

Let us assume that the manifold has bounded geometry; i.e. it has positive injectivity radius, inj(M), the curvature tensor and the covariant derivatives of the curvature tensor are bounded. The exponential map from a ball of radius r, $r < \text{inj}(M)$, at a point x defines a chart, through a fixed orthonormal frame at x. Coordinates that consists of the above mentioned exponential charts are said to be canonical. In canonical coordinates, all transitions functions have bounded derivatives of all order. That f is bounded in C^k can be formulated as below: for any canonical coordinates and for any integer k, $|\partial^\lambda f|$ is bounded for any multi-index λ up to order k. The following types of manifolds have bounded geometry: Lie groups, homogeneous spaces with invariant metrics, Riemannian covering spaces of compact manifolds.

In the lemma below we deduce from the convergence rate of P_ϵ to P in the $(C^k)^*$ norm a rate in the Wasserstein distance. Let ρ be the Riemannian distance with reference to which we speak of Lipschitz continuity of a real valued function on M and the Wasserstein distance on $\mathbb{P}(M)$. If ξ is a random variable we denote by $\hat{P}_\xi$ its probability distribution. Denote by $|f|_{\text{Lip}}$ the Lipschitz constant of the function f. Let $p \in M$. Let $|f|_{C^k} = |f|_\infty + \sum_{j=0}^{k-1} |\nabla^j df|_\infty$.

Lemma 9.1 *Let ξ_1 and ξ_2 be random variables on a C^k manifold M, where $k \geq 1$, with bounded geometry. Suppose that for a reference point $p \in M$, $c_0 := \sum_{i=1}^2 \mathbf{E}\rho^2(\xi_i, p)$ is finite. Suppose that there exist numbers $c \geq 0, \alpha \in (0,1), \epsilon \in (0,1]$ s.t. for $g \in BC^k$,*

$$|\mathbf{E}g(\xi_1) - \mathbf{E}g(\xi_2)| \leq c\epsilon^\alpha(1 + |g|_{C^k}).$$

Then there is a constant C, depending only on the geometry of the manifold, s.t.

$$d_W(\hat{P}_{\xi_1}, \hat{P}_{\xi_2}) \leq C(c_0 + c)\epsilon^{\frac{\alpha}{k}}.$$

Proof If $k = 1$, this is clear. Let us take $k \geq 2$ and let $f : M \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant 1. Since we are concerned only with the difference of the values of f at two points, $|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)|$, we first shift f so that its value at the reference point is zero. By the Lipschitz continuity of f, $|f(x)| \leq |f|_{\text{Lip}}\ \rho(x,p)$. We may also assume that f is bounded; if not we define a family of functions $f_n = (f \wedge n) \vee (-n)$. Then f_n is Lipschitz continuous with its Lipschitz constant bounded by $|f|_{\text{Lip}}$. Let $i = 1, 2$. The correction term $(f - f_n)(\xi_i)$ can be easily controlled by the second moment of $\rho(p, \xi_i)$:

$$\mathbf{E}|(f - f_n)(\xi_i)| \leq \mathbf{E}|f(\xi_i)|\mathbf{1}_{\{|f(\xi_i)|>n\}} \leq \frac{1}{n}\mathbf{E}f(\xi_i)^2 \leq \frac{1}{n}\mathbf{E}\rho^2(p, \xi_i).$$

Let $\eta : \mathbb{R}^n \to \mathbb{R}$ be a function supported in the ball $B(x_0, 1)$ with $|\eta|_{L_1} = 1$ and $\eta_\delta = \delta^{-n}\eta(\frac{x}{\delta})$, where δ is a positive number and n is the dimension of the manifold. If $M = \mathbb{R}^n$,

$$\begin{aligned}
&|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)| \\
&\leq |\mathbf{E}(f * \eta_\delta)(\xi_1) - \mathbf{E}(f * \eta_\delta)(\xi_2)| + \sum_{i=1}^{2} |\mathbf{E}(f * \eta_\delta)(\xi_i) - \mathbf{E}f(\xi_i)| \\
&\leq c\epsilon^\alpha(1 + |f * \eta_\delta|_{C^k}) + 2\delta|f|_{\text{Lip}}.
\end{aligned}$$

In the last step we used the assumption on $\mathbf{E}|f * \eta_\delta(\xi_1) - f * \eta_\delta(\xi_2)|$ for the BC^k function $f * \eta_\delta$. By distributing the derivatives to η_δ we see that the norm of the first k derivatives of $f * \eta_\delta$ are controlled by $|f|_{\text{Lip}}$. If f is bounded,

$$c\epsilon^\alpha(1 + |f * \eta_\delta|_{C^k}) \leq c\epsilon^\alpha(1 + |f|_\infty + c_1\delta^{-k+1}|f|_{\text{Lip}}),$$

where c_1 is a combinatorial constant. To summarize, for all Lipschitz continuous f with $|f|_{\text{Lip}} = 1$,

$$\begin{aligned}
|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)| &\leq 2\delta|f|_{\text{Lip}} + c\epsilon^\alpha(1 + |f_n * \eta_\delta|_{C^k}) + \frac{c_0}{n} \\
&\leq 2\delta + c\epsilon^\alpha + c\epsilon^\alpha n + c_1 c\epsilon^\alpha\delta^{-k+1} + \frac{c_0}{n}.
\end{aligned}$$

Let $\delta = \epsilon^{\frac{\alpha}{k}}$. Since $k \geq 2$, we choose n with the property $\epsilon^{-\frac{\alpha}{k}} \leq n \leq 2\epsilon^{-\alpha+\frac{\alpha}{k}}$, then for f with $|f|_{\text{Lip}} = 1$,

$$|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)| \leq (2 + 2c + c_1 c + 2c_0)\epsilon^{\frac{\alpha}{k}}.$$

Let δ be a positive number with $4\delta < \text{inj}(M)$. Let $B_x(r)$ denotes the geodesic ball centred at x with radius r, whose Riemannian volume is denoted by $V(x, r)$. There is a countable sequence $\{x_i\}$ in M with the following property: (1) $\{B_{x_i}(\delta)\}$ covers M; (2) There is a natural number N such that any point y belongs to at most N balls from $\{\mathcal{B}_{x_i}(3\delta)\}$; i.e. the cover $\{\mathcal{B}_{x_i}(3\delta)\}$ has finite multiplicity. Moreover this number N is independent of δ. See M. A. Shubin [40]. To see the independence of N on δ, let us choose a sequence $\{x_i, i \geq 1\}$ in M with the property that $\{B_{x_i}(\delta)\}$ covers M and $\{B_{x_i}(\frac{\delta}{2})\}$ are pairwise disjoint. Since the curvature tensors and their derivatives are bounded, there is a positive number C such that

$$\frac{1}{C} \leq \frac{V(x, r)}{V(y, r)} \leq C, \quad x, y \in M, r \in (0, 4\delta).$$

Let $y \in M$ be a fixed point that belongs to N balls of the form $B_{x_i}(\frac{\delta}{2})$. Since $B_{x_i}(\frac{\delta}{2}) \subset B(y, 4\delta)$, the sum of the volume satisfies: $\sum V(x_i, \frac{\delta}{2}) \leq V(y, 4\delta)$ and

$\frac{N}{C}V(y,\frac{\delta}{2}) \leq V(y,4\delta)$. The ratio $\sup_y \frac{V(y,4\delta)}{V(y,\frac{\delta}{2})}$ depends only on the dimension of the manifold.

Let us take a C^k smooth partition of unity $\{\alpha_i, i \in \Lambda\}$ that is subordinated to $\{B_{x_i}(2\delta)\}$: $1 = \sum_{i\in\Lambda} \phi_i$, $\phi_i \geq 0$, ϕ_i is supported in $B_{x_i}(2\delta)$, and for any point x there are only a finite number of non-zero summands in $\sum_{i\in\Lambda} \alpha_i(x)$. The partition of unity satisfies the additional property: $\sup_i |\partial^\lambda \alpha_i| \leq C_\lambda$, $\alpha_i \geq 0$.

Let $(B_{x_i}(\text{inj}(M)), \phi_i)$ be the geodesic charts. Let $f_i = f\alpha_i$ and let $\tilde{g} = g \circ \phi_i$ denote the representation of a function g in a chart.

$$
\begin{aligned}
|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)| &= \left| \sum_{i\in\Lambda} \mathbf{E}\tilde{f}_i\left(\phi_i^{-1}(\xi_1)\right) - \sum_{i\in\Lambda} \mathbf{E}\tilde{f}_i\left(\phi_i^{-1}(\xi_2)\right) \right| \\
&\leq \left| \sum_{i\in\Lambda} \mathbf{E}\tilde{f}_i * \eta_\delta\left(\phi_i^{-1}(\xi_1)\right) - \sum_{i\in\Lambda} \mathbf{E}\tilde{f}_i * \eta_\delta\left(\phi_i^{-1}(\xi_2)\right) \right| \\
&+ \sum_{j=1}^{2} \left| \sum_{i\in\Lambda} \mathbf{E}\tilde{f}_i * \eta_\delta\left(\phi_i^{-1}(\xi_j)\right) - \sum_{i\in\Lambda} \mathbf{E}\tilde{f}_i\left(\phi_i^{-1}(\xi_j)\right) \right|.
\end{aligned}
$$

It is crucial to note that there are at most N non-zero terms in the summation. By the assumption, for each i,

$$
\left| \mathbf{E}\tilde{f}_i * \eta_\delta\left(\phi_i^{-1}(\xi_1)\right) - \mathbf{E}\tilde{f}_i * \eta_\delta\left(\phi_i^{-1}(\xi_2)\right) \right| \leq c\epsilon^\alpha |\tilde{f}_i * \eta_\delta \circ \phi_i^{-1}|_{C^k}.
$$

By construction, $\sup_i |\alpha_i|_{C^k}$ is bounded. There is a constant c' that depends only on the partition of unity, such that

$$
|\tilde{f}_i * \eta_\delta \circ \phi_i^{-1}|_{C^k} \leq c'|\tilde{f}_i * \eta_\delta|_{C^k} \leq c'|\tilde{f}|_\infty + c'c_1\delta^{1-k}|\tilde{f}|_{\text{Lip}}
$$

Similarly for the second summation, we work with the representatives of f_i,

$$
\left| \tilde{f}_i * \eta_\delta\left(\phi_i^{-1}(y)\right) - \tilde{f}_i\left(\phi_i^{-1}(y)\right) \right| \leq \delta|\tilde{f}_i|_{\text{Lip}} \leq c'\delta.
$$

Since we work in the geodesic charts the Lipschitz constant of $\tilde{f}_i$ are comparable to that of $|f|_{\text{Lip}}$. Let $|f|_{\text{Lip}} = 1$. If f is bounded,

$$
|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)| \leq Nc\epsilon^\alpha(1 + c'|f|_\infty + c'\delta^{1-k}) + 2c'\delta N
$$

Let $\delta = \epsilon^{\frac{\alpha}{k}}$,

$$
|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)| \leq Nc\epsilon^\alpha(c'|f|_\infty + 1) + Nc'\epsilon^{\frac{\alpha}{k}} + 2c'N\epsilon^{\frac{\alpha}{k}}.
$$

On a compact manifold, $|f|_\infty$ can be controlled by $|f|_{\text{Lip}}$; otherwise we use the cut off function f_n in place of f and the estimate $\mathbf{E}|(f - f_n)(\xi_i)| \leq \frac{c_0}{n}$. Choose

n sufficiently large, as before, to see that $|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)| \leq C\epsilon^{\frac{\alpha}{k}}$. Finally we apply the Kantorovich-Rubinstein duality theorem,

$$d_W(\hat{P}_{\xi_1}, \hat{P}_{\xi_2}) = \sup_{f:|f|_{\text{Lip}}\leq 1} \{|\mathbf{E}f(\xi_1) - \mathbf{E}f(\xi_2)|\} \leq C\epsilon^{\frac{\alpha}{k}},$$

to obtain the required estimate on the Wasserstein 1-distance and concluding the proof. □

Let $\text{ev}_t : C([0,T];M) \to M$ denote the evaluation map at time t : $\text{ev}(\sigma) = \sigma(t)$. Let $\hat{P}_\xi$ denote the probability distribution of a random variable ξ. Let $o \in M$.

Proposition 9.2 *Assume the conditions and notations of Theorem 8.2. Suppose that M has bounded geometry and $\rho_o^2 \in B_{V,0}$. Let $\bar{\mu}$ be the limit measure and $\bar{\mu}_t = (ev_t)_*\bar{\mu}$. Then for every $r < \frac{1}{4}$ there exists $C(T) \in B_{V,0}$ and $\epsilon_0 > 0$ s.t. for all $\epsilon \leq \epsilon_0$ and $t \leq T$,*

$$d_W(\hat{P}_{y^\epsilon_{\frac{t}{\epsilon}}}, \bar{\mu}_t) \leq C(T)\epsilon^r.$$

Proof By Theorem 8.2, for $f \in BC^4$,

$$\left|\mathbf{E}f(\Phi^\epsilon_{\frac{t}{\epsilon}}(y_0)) - P_t f(y_0)\right| \leq C(T)(y_0)\epsilon\sqrt{|\log\epsilon|},$$

where $C(T)(y_0) \leq \tilde{C}(T)(y_0)(1 + |f|_{C^4})$ for some function $\tilde{C}(T) \in B_{V,0}$. Since by Theorem 5.2, there exists $\epsilon_0 > 0$ such that $\sup_{\epsilon\leq\epsilon_0} \mathbf{E}\rho_o^2(\Phi^\epsilon_t(y_0))$ is finite, we take α in Lemma 9.1 to be any number less than 1 to conclude the proposition. □

10 Appendix

We began with the proof of Lemma 3.1, follow it with a discussion on conditional inequalities without assuming conditions on the σ-algebra concerned.

Proof of Lemma 3.1

Step 1. Denote $\psi(t) = ae^{-\delta t}$. Firstly, if $f \in \mathcal{B}_b(G;\mathbb{R})$ and $z \in G$,

$$|Q_t f(z) - \pi f| \leq \|f\|_W \cdot \psi(t) \cdot W(z).$$

Next, by the Markov property of (z_t) and the assumption that $\int g d\pi = 0$:

$$\begin{aligned}
&\left|\mathbf{E}\{f(z_{s_2})g(z_{s_1})|\mathcal{F}_s\} - \int_G f Q_{s_1-s_2} g d\pi\right| \\
&= \left|\mathbf{E}\left\{\left(f Q_{s_1-s_2} g\right)(z_{s_2})\middle|\mathcal{F}_s\right\} - \int_G f Q_{s_1-s_2} g d\pi\right| \\
&\leq \psi(s_2-s)\left\|f Q_{s_1-s_2} g\right\|_W W(z_s) \leq \psi(s_2-s)\sup_{z\in G}\left(\frac{|f(z)||Q_{s_1-s_2}g(z)|}{W(z)}\right) W(z_s) \\
&\leq \psi(s_2-s)\psi(s_1-s_2)|f|_\infty \|g\|_W W(z_s) \leq a\psi(s_1-s)|f|_\infty\|g\|_W W(z_s).
\end{aligned}$$

From this we see that,

$$\left|\frac{1}{t-s}\int_s^t\int_s^{s_1}\left(\mathbf{E}\left\{f(z_{s_2})g(z_{s_1})\Big|\mathcal{F}_s\right\}-\int_G fQ_{s_1-s_2}g d\pi\right)ds_2 ds_1\right|$$
$$\leq a|f|_\infty\|g\|_W W(z_s)\frac{1}{t-s}\int_s^t\int_s^{s_1}\psi\left(s_1-s\right)\,ds_2\,ds_1$$
$$\leq \frac{a^2}{\delta^2(t-s)}|f|_\infty\|g\|_W W(z_s)\int_0^{(t-s)\delta} re^{-r}\,dr \leq \frac{a^2}{\delta^2(t-s)}|f|_\infty\|g\|_W W(z_s).$$

This concludes (1). Step 2. For (2), we compute the following:

$$\frac{1}{t-s}\int_s^t\int_s^{s_1}\int_G fQ_{s_1-s_2}g\,d\pi\,ds_2\,ds_1 = \int_G\frac{1}{t-s}\int_0^{t-s} fQ_r g(t-s-r)\,drd\pi$$
$$=\int_G\int_0^\infty\left(fQ_r g\right)\,dr\,d\pi-\int_G\int_{t-s}^\infty fQ_r g\,dr\,d\pi-\frac{1}{t-s}\int_G\int_0^{t-s} rfQ_r g\,drd\pi.$$

We estimate the last two terms. Firstly,

$$\left|\int_G\int_{t-s}^\infty f(z)Q_r g(z)\,dr\,d\pi(z)\right|\leq|f|_\infty\left|\int_G\int_{t-s}^\infty|Q_r g(z)|\,dr\,d\pi(z)\right|_\infty$$
$$\leq|f|_\infty\|g\|_W\int_G W(z)\pi(dz)\int_{t-s}^\infty\psi(r)dr\leq\frac{1}{\delta}|f|_\infty\|g\|_W\bar{W}\int_{(t-s)\delta}^\infty ae^{-r}dr$$
$$\leq\frac{a}{\delta}|f|_\infty\|g\|_W\bar{W}.$$

It remains to calculate the following:

$$\left|\frac{1}{t-s}\int_G\int_0^{t-s} rfQ_r g\,drd\pi\right|\leq\frac{1}{t-s}|f|_\infty\|g\|_W\bar{W}\int_0^{t-s} r\psi(r)\,dr$$
$$\leq\frac{a}{(t-s)\delta^2}|f|_\infty\|g\|_W\bar{W}.$$

Gathering the estimates together we obtain the bound:

$$\left|\frac{1}{t-s}\int_s^t\int_s^{s_1}\int_G fQ_{s_1-s_2}g\,d\pi\,ds_2\,ds_1-\int_G\int_0^\infty\left(fQ_r g\right)\,dr\,d\pi\right|$$
$$\leq\frac{a}{\delta}|f|_\infty\|g\|_W\bar{W}+\frac{a}{(t-s)\delta^2}|f|_\infty\|g\|_W\,\bar{W}.$$

By adding this estimate to that in part (1), we conclude part (2):

$$\left|\frac{1}{t-s}\int_s^t\int_s^{s_1}\mathbf{E}\left\{f(z_{s_2})g(z_{s_1})\Big|\mathcal{F}_s\right\}-\int_G\int_0^\infty\left(fQ_r g\right)\,dr\,d\pi\right|$$
$$\leq\frac{a}{\delta}|f|_\infty\|g\|_W\bar{W}+\frac{a}{(t-s)\delta^2}|f|_\infty\|g\|_W\,\bar{W}+\frac{a^2}{\delta^2(t-s)}|f|_\infty\|g\|_W W(z_s). \tag{10.1}$$

We conclude part (2). Step 3. We first assume that $\bar{g} = 0$, then,

$$\left| \frac{\epsilon}{t-s} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ f(z^\epsilon_{s_2}) g(z^\epsilon_{s_1}) \Big| \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 \right|$$
$$\leq \left| \frac{\epsilon}{t-s} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ f(z^\epsilon_{s_2}) g(z^\epsilon_{s_1}) \Big| \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 - \int_G \int_0^\infty f Q^\epsilon_r g \, dr \, d\pi \right|$$
$$+ \left| \int_G \int_0^\infty f Q^\epsilon_r g \, dr \, d\pi \right|.$$

We note that for every $x \in G$, $\|Q^\epsilon_r(x,\cdot) - \pi\|_{TV,W} \leq \psi(\frac{r}{\epsilon}) W(x)$. In line (10.1) we replace s, t, δ by $\frac{s}{\epsilon}$, $\frac{t}{\epsilon}$, and $\frac{\delta}{\epsilon}$ respectively to see the first term on the right hand side is bounded by

$$\frac{a\epsilon^3}{\delta^2(t-s)} (aW(z^\epsilon_{\frac{s}{\epsilon}}) + \bar{W}) |f|_\infty \|g\|_W + \frac{a\epsilon}{\delta} |f|_\infty \|g\|_W \bar{W}.$$

Next we observe that

$$\int_0^\infty f(z) Q^\epsilon_s g(z) \, ds = \int_0^\infty f(z) Q_{\frac{s}{\epsilon}}(z) \, ds = \epsilon \int_0^\infty f(z) Q_s g(z) \, ds$$
$$\left| \int_G \int_0^\infty f(z) Q^\epsilon_s g(z) \, ds \, d\pi(z) \right| \leq \epsilon \, |f|_\infty \|g\|_W \bar{W} \int_0^\infty \psi(s) \, ds = \frac{a\epsilon}{\delta} |f|_\infty \|g\|_W \bar{W}.$$

This gives the estimate for the case of $\bar{g} = 0$:

$$\left| \frac{\epsilon}{t-s} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ f(z^\epsilon_{s_2}) g(z^\epsilon_{s_1}) \Big| \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 \right| \leq C_1(z^\epsilon_{\frac{s}{\epsilon}}) \frac{\epsilon^3}{t-s} + C_2'(z^\epsilon_{\frac{s}{\epsilon}}) \epsilon.$$

where

$$C_1 = \frac{a}{\delta^2}(aW(\cdot) + \bar{W}) |f|_\infty \|g\|_W, \quad C_2' = \frac{2a}{\delta} |f|_\infty \|g\|_W \bar{W}.$$

If $\int g \, d\pi \neq 0$, we split $g = g - \bar{g} + \bar{g}$ and estimate the remaining term. We use the fact that $\pi f = 0$,

$$\left| \frac{\epsilon}{t-s} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ f(z^\epsilon_{s_2}) \bar{g} | \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 \right| \leq |\bar{g}| \left| \frac{\epsilon}{t-s} \int_0^{\frac{t-s}{\epsilon}} \int_0^{s_1} \left| Q^\epsilon_{s_2} f(z_{\frac{s}{\epsilon}}) \right| \, ds_2 \, ds_1 \right|$$
$$\leq |\bar{g}| \|f\|_W W(z^\epsilon_{\frac{s}{\epsilon}}) \sup_{s_1 > 0} \left\{ \left| \int_0^{s_1} \psi(\frac{s_2}{\epsilon}) ds_2 \right| \right\} \leq |\bar{g}| \, \|f\|_W W(z^\epsilon_{\frac{s}{\epsilon}}) \epsilon \int_0^\infty \psi(r) dr$$
$$= \frac{a\epsilon}{\delta} |\bar{g}| \, \|f\|_W W(z^\epsilon_{\frac{s}{\epsilon}}).$$

Finally we obtain the required estimate in part (3):

$$\left| \frac{\epsilon}{t-s} \int_{\frac{s}{\epsilon}}^{\frac{t}{\epsilon}} \int_{\frac{s}{\epsilon}}^{s_1} \mathbf{E}\left\{ f(z^\epsilon_{s_2}) g(z^\epsilon_{s_1}) \Big| \mathcal{F}_{\frac{s}{\epsilon}} \right\} \, ds_2 \, ds_1 \right|$$
$$\leq C_1(z^\epsilon_{\frac{s}{\epsilon}}) \left(\frac{\epsilon^3}{t-s} \right) + C_2'(z^\epsilon_{\frac{s}{\epsilon}})\epsilon + \epsilon \frac{a}{\delta} |\bar{g}| \, \|f\|_W W(z^\epsilon_{\frac{s}{\epsilon}}),$$

thus concluding part (3).

The following conditional inequalities are elementary. We include a proof for a partial conditional Burkholder-Davis-Gundy inequality for completeness. We do not assume the existence of regular conditional probabilities.

Lemma 10.1 *Let (M_t) be a continuous L^2 martingale vanishing at 0. Let (H_t) be an adapted stochastic process with left continuous sample paths and right limits. If for stopping times $s < t$, $\mathbf{E} \int_s^t (H_r)^2 d\langle M \rangle_r < \infty$. Then*

$$\mathbf{E}\left\{ \left(\int_s^t H_r dM_r \right)^2 \Big| F_s \right\} = \mathbf{E}\left\{ \int_s^t (H_r)^2 d\langle M \rangle_r \Big| F_s \right\}.$$

Lemma 10.2 *Let $p > 1$ and (M_t) is a right continuous $(\mathcal{F}_t)$ martingale or a right continuous positive sub-martingale index by an interval I of $\mathbb{R}_+$. Then,*

$$\mathbf{E}\left\{ \sup_{s \in I} |M_t|^p \Big| \mathcal{F}_s \right\} \leq \left(\frac{p}{p-1} \right)^p \sup_{s \in I} \mathbf{E}\left\{ |M_s|^p \Big| \mathcal{F}_s \right\}.$$

If $(M_u, s \leq u \leq t)$ is a right continuous $(\mathcal{F}_t)$ martingale and $p \geq 2$, there exists a constant $c(p) > 0$ s.t.

$$\mathbf{E}\left\{ \sup_{s \leq u \leq t} |M_u|^p \Big| F_s \right\} \leq c_p \mathbf{E}\left\{ \langle M \rangle_t^{\frac{p}{2}} \Big| F_s \right\}.$$

This proof is the same as the proof for $\mathcal{F}_s$ the trivial σ-algebra, c.f. D. Revuz, M. Yor [38].

References

[1] J. Angst, I. Bailleul, and C. Tardif. Kinetic Brownian motion on Riemannian manifolds. Preprint, 2014.

[2] Jean-Michel Bony. Principe du maximum, inégalite de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés. *Ann. Inst. Fourier (Grenoble)*, 19(fasc. 1):277–304 xii, 1969.

[3] 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.

[4] 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.

[5] Jeff Cheeger and Mikhael Gromov. Collapsing Riemannian manifolds while keeping their curvature bounded. I. *J. Differential Geom.*, 23(3):309–346, 1986.

[6] R. Cogburn and R. Hersh. Two limit theorems for random differential equations. *Indiana Univ. Math. J.*, 22:1067–1089, 1972/73.

[7] N. Enriquez, J. Franchi, and Y. Le Jan. Central limit theorem for the geodesic flow associated with a Kleinian group, case $\delta > d/2$. *J. Math. Pures Appl. (9)*, 80(2):153–175, 2001.

[8] Mark Freidlin and Matthias Weber. On stochasticity of solutions of differential equations with a small delay. *Stoch. Dyn.*, 5(3):475–486, 2005.

[9] 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.

[10] K. Fukaya. Collapsing Riemannian manifolds to ones of lower dimensions. *J. Differential Geom.*, 25(1):139–156, 1987.

[11] Ivan I. Gonzales Gargate and Paulo R. Ruffino. An averaging principle for diffusions in foliated spaces. http://arxiv.org/abs/1212.1587, 2012.

[12] M. Hairer and G. A. Pavliotis. Periodic homogenization for hypoelliptic diffusions. *J. Statist. Phys.*, 117(1-2):261–279, 2004.

[13] Martin Hairer and Jonathan C. Mattingly. Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations. *Ann. Probab.*, 36(6):2050–2091, 2008.

[14] R. Z. Has′minskii. A limit theorem for solutions of differential equations with a random right hand part. *Teor. Verojatnost. i Primenen*, 11:444–462, 1966.

[15] R. Z. Has′minskii. Stochastic processes defined by differential equations with a small parameter. *Teor. Verojatnost. i Primenen*, 11:240–259, 1966.

[16] R. Hersh and G. Papanicolaou. Non-commuting random evolutions, and an operator-valued Feynman-Kac formula. *Comm. Pure Appl. Math.*, 25:337–367, 1972.

[17] R. Hersh and M. Pinsky. Random evolutions are asymptotically Gaussian. *Comm. Pure Appl. Math.*, 25:33–44, 1972.

[18] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Math.*, 119:147–171, 1967.

[19] Nobuyuki Ikeda and Yoko Ochi. Central limit theorems and random currents. In *Stochastic differential systems (Bad Honnef, 1985)*, volume 78 of *Lecture Notes in Control and Inform. Sci.*, pages 195–205. Springer, Berlin, 1986.

[20] Yuri Kifer. Averaging in dynamical systems and large deviations. *Invent. Math.*, 110(2):337–370, 1992.

[21] Thomas G. Kurtz. A general theorem on the convergence of operator semigroups. *Trans. Amer. Math. Soc.*, 148:23–32, 1970.

[22] Xue-Mei Li. Strong p-completeness of stochastic differential equations and the existence of smooth flows on noncompact manifolds. *Probab. Theory Related Fields*, 100(4):485–511, 1994.

[23] Xue-Mei Li. An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity*, 21(4):803–822, 2008.

[24] Xue-Mei Li. Effective diffusions with intertwined structures. http://arxiv.org/abs/1204.3250, 2012.

[25] Xue-Mei Li. Random perturbation to the geodesic equation. To appear: Annals of Probability, http://arxiv.org/abs/1402.5861, 2014.

[26] Xue-Mei Li. Homogenization on homogeneous spaces. http://arxiv.org/abs/1505.06772, 2015.

[27] Ming Liao and Longmin Wang. Motion of a rigid body under random perturbation. *Electron. Comm. Probab.*, 10:235–243 (electronic), 2005.

[28] Carlangelo Liverani and Stefano Olla. Toward the Fourier law for a weakly interacting anharmonic crystal. *J. Amer. Math. Soc.*, 25(2):555–583, 2012.

[29] Shojiro Manabe and Yoko Ochi. The central limit theorem for current-valued processes induced by geodesic flows. *Osaka J. Math.*, 26(1):191–205, 1989.

[30] Yukio Ogura. Weak convergence of laws of stochastic processes on Riemannian manifolds. *Probab. Theory Related Fields*, 119(4):529–557, 2001.

[31] O. A. Oleinik and E. V. Radkevič. *Second order equations with nonnegative characteristic form*. Plenum Press, New York-London, 1973. Translated from the Russian by Paul C. Fife.

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

[33] 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.

[34] 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.

[35] George Papanicolaou and Joseph B. Keller. Stochastic differential equations with applications to random harmonic oscillators and wave propagation in random media. *SIAM J. Appl. Math.*, 21:287–305, 1971.

[36] George C. Papanicolaou and Werner Kohler. Asymptotic analysis of deterministic and stochastic equations with rapidly varying components. *Comm. Math. Phys.*, 45(3):217–232, 1975.

[37] Mark A. Pinsky. Homogenization and stochastic parallel displacement. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, volume 851 of *Lecture Notes in Math.*, pages 271–284. Springer, Berlin, 1981.

[38] D. Revuz and M. Yor. *Continuous martingales and Brownian motion, 2nd edition*. Springer-Verlag, 1993.

[39] Antonio Sánchez-Calle. Fundamental solutions and geometry of the sum of squares of vector fields. *Invent. Math.*, 78(1):143–160, 1984.

[40] M. A. Shubin. Spectral theory of elliptic operators on noncompact manifolds. *Astérisque*, (207):5, 35–108, 1992. Méthodes semi-classiques, Vol. 1 (Nantes, 1991).

[41] Richard B. Sowers. On the tangent flow of a stochastic differential equation with fast drift. *Trans. Amer. Math. Soc.*, 353(4):1321–1334 (electronic), 2001.

[42] R. L. Stratonovich. A limit theorem for solutions of differential equations with random right-hand side. *Theory Prob. Appl.,*, 11, 1960. In Russian.

[43] R. L. Stratonovich. Selected problems in the theory of fluctuations in radio engineering. *Sov. Radio, Moscow*, 1961. In Russian.

[44] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.