The Annals of Probability
2016, Vol. 44, No. 1, 544–566
DOI: 10.1214/14-AOP981

RANDOM PERTURBATION TO THE GEODESIC EQUATION[1]

BY XUE-MEI LI

University of Warwick

We study random "perturbation" to the geodesic equation. The geodesic equation is identified with a canonical differential equation on the orthonormal frame bundle driven by a horizontal vector field of norm 1. We prove that the projections of the solutions to the perturbed equations, converge, after suitable rescaling, to a Brownian motion scaled by $\frac{8}{n(n-1)}$ where n is the dimension of the state space. Their horizontal lifts to the orthonormal frame bundle converge also, to a scaled horizontal Brownian motion.

1. Introduction. Let M be a complete smooth Riemannian manifold of dimension n and T_xM its tangent space at $x \in M$. Let OM denote the space of orthonormal frames on M and π the projection that takes an orthonormal frame $u : \mathbb{R}^n \to T_xM$ to the point x in M. Let $T_u\pi$ denote its differential at u. For $e \in \mathbb{R}^n$, let $H_u(e)$ be the basic horizontal vector field on OM such that $T_u\pi(H_u(e)) = u(e)$, that is, $H_u(e)$ is the horizontal lift of the tangent vector $u(e)$ through u. If $\{e_i\}$ is an orthonormal basis of $\mathbb{R}^n$, the second-order differential operator $\Delta_H = \sum_{i=1}^n L_{H(e_i)}L_{H(e_i)}$ is the Horizontal Laplacian. Let $\{w_t^i, 1 \le i \le n\}$ be a family of real valued independent Brownian motions. The solution $(u_t, t < \zeta)$, to the following semi-elliptic stochastic differential equation (SDE), $du_t = \sum_{i=1}^n H_{u_t}(e_i) \circ dw_t^i$, is a Markov process with infinitesimal generator $\frac{1}{2}\Delta_H$ and lifetime ζ. We denote by $\circ$ Stratonovich integration. The solutions are known as horizontal Brownian motions. It is well known that a horizontal Brownian motion projects to a Brownian motion on M. We recall that a Brownian motion on M is a sample continuous strong Markov process with generator $\frac{1}{2}\Delta$ where Δ is the Laplace–Beltrami operator. This construction of Brownian motions on a Riemannian manifold is canonical and has fundamental applications in analysis on path spaces.

For $e_0 \in \mathbb{R}^n$, the horizontal vector field $H(e_0)$ does not project to a vector field on M. It, however, induces a vector field X on TM which is a geodesic spray. If $(u_t^{e_0})$ is the solution to the first-order differential equation

$$\dot{u}(t) = H_{u(t)}(e_0), \qquad u(0) = u_0,$$

then $\pi(u_t^{e_0})$ is the geodesic on M with initial velocity $u_0(e_0)$ and initial value $\pi(u_0)$.

Received August 2014; revised October 2014.

[1]Supported by EPSRC Grant EP/E058124/1.

MSC2010 subject classifications. 60H10, 58J65, 37Hxx, 53B05.

Key words and phrases. Horizontal flows, horizontal Brownian motions, vertical perturbation, stochastic differential equations, homogenisation, geodesics.

Let $N = \frac{n(n-1)}{2}$ and let $\mathfrak{so}(n)$ be the space of skew symmetric matrices in dimension n. It is the Lie algebra of the orthogonal group $O(n)$. For $A \in \mathfrak{so}(n)$, we denote by A^* the fundamental vertical vector field on OM determined by right actions of the exponentials of tA; see (2.1) below. If X is a vector field, we denote by L_X Lie differentiation in the direction of X. Let us fix a time $T > 0$. Let ρ be the Riemannian distance function on M, ∇ the Levi–Civita connection and Δ the Laplace–Beltrami operator. Let ε a positive number. Our main theorems concern the convergence, as ε approaches zero, of the "horizontal part" of the solutions to a family of stochastic differential equations with parameter ε. The definitions for the horizontal and vertical vector fields and for the horizontal lift of a curve are given in Section 2. Let e_0 be a unit vector in $\mathbb{R}^n$.

THEOREM 1.1. *Let M be a complete Riemannian manifold of dimension $n > 1$ and of positive injectivity radius. Suppose that there are positive numbers C and a such that $\sup_{\rho(x,y)\le a} |\nabla d\rho|(x, y) \le C$. Let $x_0 \in M$ and $u_0 \in \pi^{-1}(x_0)$. Let $\bar{A} \in \mathfrak{so}(n)$ and $\{A_1, \ldots, A_N\}$ be an orthonormal basis of $\mathfrak{so}(n)$. Let $(u_t^\varepsilon, 0 \le t \le T)$ be the solution to the SDE*

$$\text{(1.1)} \qquad \begin{cases} du_t^\varepsilon = H_{u_t^\varepsilon}(e_0)\,dt + \dfrac{1}{\sqrt{\varepsilon}} \displaystyle\sum_{k=1}^{N} A_k^*(u_t^\varepsilon) \circ dw_t^k + \bar{A}^*(u_t^\varepsilon)\,dt, \\ u_0^\varepsilon = u_0. \end{cases}$$

Let $x_t^\varepsilon = \pi(u_t^\varepsilon)$ and let $(\tilde{x}_t^\varepsilon, 0 \le t \le T)$ be the horizontal lift of $(x_t^\varepsilon, 0 \le t \le T)$ to OM through u_0. Then the following statements hold:

(1) *The SDE does not explode.*

(2) *The processes $(x_{t/\varepsilon}^\varepsilon, 0 \le t \le T)$ and $(\tilde{x}_{t/\varepsilon}^\varepsilon, 0 \le t \le T)$ converge in law, as $\varepsilon \to 0$.*

(3) *The limiting law of $(x_{t/\varepsilon}^\varepsilon, 0 \le t \le T)$ is independent of e_0. It is a scaled Brownian motion with generator $\frac{4}{n(n-1)}\Delta$. The limiting law of $(\tilde{x}_{t/\varepsilon}^\varepsilon, 0 \le t \le T)$ is that associated to the generator $\frac{4}{n(n-1)}\Delta_H$.*

If "$\varepsilon = \infty$" and $\bar{A} = 0$, the SDE (1.1) reduces to the first-order differential equation $\dot{u}(t) = H_{u(t)}(e_0)$ whose solutions are geodesics. If "$\varepsilon = 0$", the SDE "reduces" to the "vertical SDE", $du_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}} \sum_{k=1}^N A_k^*(u_t^\varepsilon) \circ dw_t^k$. This vertical equation does not have a meaning for $\varepsilon = 0$, nevertheless the "vertical SDE" has a first integral $\pi : OM \to M$, that is, $\pi(u_t^\varepsilon) = \pi(u_0^\varepsilon)$. By a preliminary multi-scale analysis, we see that $\pi(u_t^\varepsilon)$ varies slowly with ε and there is a visible effective motion in the time interval $[0, \frac{1}{\varepsilon}]$. The first integral π is not a real-valued function. It is a function from a manifold to a manifold and the slow variables $\{(x_t^\varepsilon), \varepsilon > 0\}$ are not Markov processes. Before further discussions on conservation laws related to the SDEs, we remark the following features: (1) the slow motion solves a first-order differential equation, (2) the "fast motion" on OM is not elliptic, (3) the limiting process is semi-elliptic. Another feature of Theorem 1.1 is that the pair of the intertwined family of stochastic processes $(x^{\varepsilon}_{t/\varepsilon}, \tilde{x}^{\varepsilon}_{t/\varepsilon})$ converge. We will explore (3) in a forthcoming article on homogeneous manifolds. For now, the following observation indicates a potential application of (3): the stochastic area of two linear Brownian motions $\{w^1_t, w^2_t\}$ is the principal part of the horizontal lift of the two-dimensional Brownian motion (w^1_t, w^2_t) to the three-dimensional Heisenberg group. We remark also that the first-order horizontal geodesic equation on the orthonormal frame bundle corresponds a second-order differential equation on the manifold, which explains the unusual scaling in (1.1).

There have been many studies of limit theorems whose geometric settings or scalings or methodologies relate that in this article. For example, our philosophy agrees with that in Bismut [3] where the equation $\ddot{x} = \frac{1}{T}(-\dot{x} + \dot{w})$ interpolates between classical Brownian motion ($T \to 0$) and the geodesic flow ($T \to \infty$). In Ikeda [16] and Ikeda and Ochi [17], the authors studied limit theorems for line integrals of the form $\int_0^t \phi(dx_s)$, where ϕ is a differential form and (x_s) is a suitable process such as a Brownian motion. In Manabe and Ochi [23] the authors obtained central limit theorems for line integrals along geodesic flows. One of their tools is symbolic representations of geodesic flows. Another related work can be found in Pinsky [27], where a piecewise geodesic with a Poisson-type switching mechanism is shown to converge to the horizontal Brownian motion. We also note that geodesic flows perturbed by vertical Brownian motions were considered by Franchi and Le Jan [10], in the context of relativistic diffusions.

The conclusion of (1.1) is consistent with the following central limit theorems for geodesic flows. Let M be a manifold of constant negative curvature and of finite volume. Let $(\gamma_t(x, v))$ denote the geodesic with initial value (x, v) in the unit tangent bundle STM and let $\theta_t(v) = (\gamma_t(x, v), \dot{\gamma}_t(x, v))$, a stochastic process on STM. Let f be a bounded measurable function on STM with the property that it is centered with respect to the normalized Liouville measure $\mathfrak{m}$. Then there is a number σ with the property that

$$\lim_{t\to\infty} \mathfrak{m}\left\{\xi : \frac{\int_0^t f(\theta_s(\xi))\, ds}{\sigma\sqrt{t}} \le a\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-y^2/2}\, dy.$$

See Sinai [30], Ratner [28]; see Guivarch and Le Jan [12] and Enriquez, Franchi and Le Jan [8] for further developments. See also Helland [15] and Kipnis and Varadhan [19]. These results exploit the chaotic nature of the deterministic dynamical system on manifolds of negative curvature.

In the homogenisation literature, the following works are particularly relevant: Khasminskii [14, 18], Nelson [24], Borodin and Freidlin [4], Freidlin and Wentzell [11] and Bensoussan, Lions and Papanicolaou [1]. We note in particular Theorem 2.1 in [4] which deals with the convergence of path integrals of a suitable function along a family of ergodic Markov processes. In this article, such integrals are better understood as integrals of differential 1-forms along random paths. Finally, we mention the following work: Li [21] for averaging of integrable systems and Ruffino and Gonzales Gargate [29] for averaging on foliated manifolds. See also [22] for an earlier work on the orthonormal frame bundle. We also refer to Dowell [5] for a scaling limit of Ornstein–Uhlenbeck type.

Open question. The local uniform bound on $\nabla d\rho$ is only used in Lemma 3.2 for the proof of tightness. This bound can be weakened, for example, replaced by a local uniform control over the rate of growth of the norms of $\frac{\nabla d\rho}{\rho}$ and $\frac{\nabla \rho}{\rho}$. We remark that Brownian motion constructed in Theorem 1.1 is automatically complete. The conditions in Theorem 1.1 appear to be related to the uniform cover criterion on stochastic completeness and could be studied in connection with that in Li [20]. Also, much of the work in this article is valid for a connection ∇ with torsion, the horizontal tangent bundle and Δ_H will then be induced by this connection with torsion. The effect of the torsion will generally lead to an additional drift to the Brownian motion downstairs. In this case the geodesic completeness of the manifold M may no longer be equivalent to the metric completeness of (M, ρ).

2. Preliminaries. Given a Riemannian metric on M, an orthonormal frame $u = \{u_1, \dots, u_n\}$ is an ordered basis of T_xM that is orthonormal. We denote by OM the set of all orthonormal frames on M and π the map that takes the frame u to the point $x \in M$. Let $\pi^{-1}(x) = \{u \in OM : \pi(u) = x\}$. If (O, x) is a coordinate system on M, $u_i = \sum_j u_i^j \frac{\partial}{\partial x_j}|_x$. This gives a coordinate map on OM. The map (x, u_i^j) is a homeomorphism from $\pi^{-1}(O)$ to $(x(O), O(n))$. If we identify a frame u with the transformation $u : \mathbb{R}^n \to T_xM$, then OM is a principal bundle with fibre $O(n)$ and group G, acting on the right. We adopt the notation $ue = u(e)$. For $g \in O(n)$ let R_g denote right multiplication on $O(n)$ and the right action of $O(n)$ on OM. For $A, B \in \mathfrak{so}(n)$ let $\langle A, B\rangle = \operatorname{tr} AB^T$.

A tangent vector v in OM is vertical if $T\pi(v) = 0$ where $T\pi$ denotes the differential of π. If A belongs to the Lie algebra $\mathfrak{so}(n)$, we denote by $\exp(tA)$ the exponential map. If u is a frame, the composition $u\exp(tA)$ is again a frame in the same fibre. We define the fundamental vertical vector fields associated to A by A^*,

$$A^*(u) = \frac{d}{dt}\Big|_{t=0} u\exp(tA). \tag{2.1}$$

By a linear connection on the principal bundle OM, we mean a splitting of the tangent bundle TOM with the following properties: (1) $T_uOM = HT_uOM \oplus VT_uOM$ (2) $(R_a)_* H_uTOM = H_{ua}TOM$ for all $u \in OM$ and $a \in G$. The spaces HT_uOM and VT_uOM are, respectively, the horizontal tangent spaces and the vertical tangent spaces. We will introduce a metric on OM such that π is an isometry between H_uTOM and $T_{\pi(u)}M$ and such that H_uTOM and VT_uOM are orthogonal. The metric on $\mathfrak{so}(n)$ is the bi-invariant metric introduced earlier. We will restrict our attention to the Levi–Civita connection.

Let $\mathfrak{h}_u(v)$ denote the horizontal lift of $v \in T_xM$ through $u \in \pi^{-1}(x)$. To each $e \in \mathbb{R}^n$ we denote $H_u(e) = \mathfrak{h}_u(ue)$ the basic vector field. Later, we also use H_ue for

$H_u(e)$. If $\{e_1, \dots, e_n\}$ is an orthonormal basis of $\mathbb{R}^n$, then $\{H_u(e_1), \dots, H_u(e_n)\}$ is an orthonormal basis for the horizontal tangent space HT_uOM.

A piecewise C^1 curve γ on OM is horizontal if the one-sided derivatives $\dot\gamma(\pm)$ are horizontal for all t. If c is a C^1 curve on M, there is a horizontal curve $\tilde c$ on OM such that $\tilde c$ covers c, that is, $\pi(\tilde c(t)) = c(t)$. In fact, $\tilde c(t)$ is the family of orthonormal frames along c that are obtained by parallel transporting the frame $\tilde c(0)$. We say that $\tilde c$ is a horizontal lift of c. The map $\tilde c(t)(\tilde c(0))^{-1} : T_{c(0)}M \to T_{c(t)}M$ is the parallel translation along the curve $c(t)$. In a coordinate chart (O, x), the principal part of $\tilde c(t)$ is a $n \times n$ matrix whose column vectors $\{\tilde c_1(t), \dots, \tilde c_n(t)\}$ form a frame. In components, write $\tilde c_l(t) = (\tilde c_l^1(t), \dots, \tilde c_l^n(t))^T$. Then

$$\frac{\partial \tilde c_l^k(t)}{\partial t} + \sum_{i=1, j=1}^{n} \frac{\partial c^i(t)}{\partial t} \Gamma_{ij}^k\big(c(t)\big)\tilde c_l^j(t) = 0.$$

Take $c(t) = (0, \dots, t, \dots, 0)$, where the nonzero entry is in the ith-place. We obtain the principal part of the horizontal lift of $\frac{\partial}{\partial x_i}$ through $u = \tilde c(0) = (u_l^j)$:

$$\left(\mathfrak{h}_{\tilde c(0)}\left(\frac{\partial}{\partial x_i}\right)\right)_l = \left(\frac{\partial \tilde c}{\partial t}(0)\right)_l = -\left(\sum_j \Gamma_{ij}^1 u_l^j, \dots, \sum_{j=1}^n \Gamma_{ij}^n u_l^j\right)^T.$$

Denote by A_i the matrix whose element at the (b, l) position is $\sum_j \Gamma_{ij}^b u_l^j$. Then A_i is the principal part of $H_u(\frac{\partial}{\partial x_i})$ and the horizontal space at u is spanned by the basis $\{(\frac{\partial}{\partial x_i}, A_i)\}$.

A basic object we use in our computation is the connection 1-form ϖ on OM. A connection 1-form assigns a skew symmetric matrix to every tangent vector on OM and it satisfies the following conditions:

(1) $\varpi(A^*) = A$ for all $A \in \mathfrak{so}(n)$;

(2) for all $a \in O(n)$ and $w \in OM$, $\varpi(R_{a*}w) = Ad(a^{-1})\varpi(w)$. We recall that $R_{a*}(A^*) = (Ad(a^{-1})A)^*$ for all $a \in O(n)$. It is convenient to consider horizontal tangent vectors on OM as elements of the kernel of ϖ. If $\{A_1, \dots, A_N\}$ is a basis of $\mathfrak{so}(n)$, then the horizontal component of a vector w is $w^h = w - \sum_j \langle \varpi(w), A_j\rangle A_j^*$.

The connection 1-form ϖ is basically the set of Christoffel symbols. Let $E = \{E_1, \dots, E_n\}$ be a local frame; we define the Christoffel symbols relative to E by $\nabla E_j = \sum_{ki} \Gamma_{ij}^k \, dx_i \otimes E_k$. Let θ^i be the set of dual differential 1-forms on M to $\{E_i\}$: $\theta^i(E_j) = \delta_{ij}$. We define $\omega_k^i = \Gamma_{lk}^i \theta^l$. Then $d\theta^i = -\sum_k \omega_k^i \wedge \theta^k$. Let $\{A_i^j\}$ be a basis of $\mathfrak{g}$. To each moving frame E, we associate a 1-form, $\omega = \sum_{i,j} \omega_j^i A_i^j$, on M. If (O, x) is a chart of M and $s : O \to OM$ is a local section of OM, let us denote by ω_s the differential 1-form given above, then $\varpi(s_* v) = \omega_s(v)$. Conditions (1) and (2) are equivalent to the following: if $a : U \to G$ is a smooth function,

$$\varpi\big((s \cdot a)_* v\big) = a^{-1}(x)\, da(v) + a^{-1}(x)\varpi(s_* v)a(x).$$

This corresponds to the differentiation of $s \cdot a$ and this type of consideration will be used in the next section.

3. Some lemmas.

LEMMA 3.1. *Let M be a geodesically complete Riemannian manifold. Let (u_t^ε) be the solution to the SDE (1.1) on OM. Let $x_t^\varepsilon = \pi(u_t^\varepsilon)$, which has a unique horizontal lift, $\tilde{x}_t^\varepsilon$, through $u_0 \equiv u_0^\varepsilon$. Then*

$$\frac{d}{dt}\tilde{x}_t^\varepsilon = H_{\tilde{x}_t^\varepsilon}(g_t^\varepsilon e_0),$$

$$dg_t^\varepsilon = \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^m g_t^\varepsilon A_k \circ dw_t^k + g_t^\varepsilon \bar{A}\,dt,$$

where g_0^ε is the unit matrix. Consequently the SDE (1.1) is conservative.

PROOF. By the defining properties of the basic horizontal vector fields, $\dot{x}_t^\varepsilon = \pi_*(H_{u_t^\varepsilon}(e_0)) = u_t^\varepsilon e_0$. Let $\mathfrak{h}_u(v)$ denote the horizontal lift of a tangent vector v through $u \in OM$. Since $u_t^\varepsilon e_0$ has unit speed, the solution exists for all time if (u_t^ε) does, and

$$\frac{d}{dt}\tilde{x}_t^\varepsilon = \mathfrak{h}_{\tilde{x}_t^\varepsilon}(\dot{x}_t^\varepsilon) = \mathfrak{h}_{\tilde{x}_t^\varepsilon}(u_t^\varepsilon e_0).$$

At each time t, the horizontal lift $(\tilde{x}_t^\varepsilon)$ of the curve (x_t^ε) through u_0 and the original curve u_t^ε belong to the same fibre. Let g_t^ε be an element of G with the property that $u_t^\varepsilon = \tilde{x}_t^\varepsilon g_t^\varepsilon$. Then g_0^ε is the unit matrix and

$$\frac{d}{dt}\tilde{x}_t^\varepsilon = \mathfrak{h}_{\tilde{x}_t^\varepsilon}(\tilde{x}_t^\varepsilon g_t^\varepsilon e_0) = H_{\tilde{x}_t^\varepsilon}(g_t^\varepsilon e_0).$$

If a_t is a C^1 path with values in $O(n)$, $a_t^{-1}\dot{a}_t = \frac{d}{dr}|_{r=0}e^{ra_t^{-1}\dot{a}_t}$, its action on u gives rise to a fundamental vector field,

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

We denote by DL_g and DR_g, respectively, the differentials of the left multiplication and of the right action. By Itô's formula applied to the product $\tilde{x}_t^\varepsilon g_t^\varepsilon$,

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

Since right translation of horizontal vectors are horizontal, the connection 1-form vanishes on the first term and $\varpi(\circ du_t^\varepsilon) = DL_{(g_t^\varepsilon)^{-1}} \circ dg_t^\varepsilon$. We apply ϖ to the SDE for u_t^ε,

$$dg_t^\varepsilon = DL_{g_t^\varepsilon}\varpi(\circ du_t^\varepsilon) = DL_{g_t^\varepsilon}\varpi\left(\frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^N A_k^*(u_t^\varepsilon)\circ dw_t^k + \bar{A}^*(u_t^\varepsilon)\,dt\right)$$
$$= \frac{1}{\sqrt{\varepsilon}}\sum_{k=1}^m g_t^\varepsilon A_k \circ dw_t^k + g_t^\varepsilon \bar{A}\,dt.$$

There is a global solution to the above equation. The ODE $\frac{d}{dt}\tilde{x}_t^\varepsilon = H_{\tilde{x}_t^\varepsilon}(g_t^\varepsilon e_0)$ has bounded right-hand side and has a global solution. It follows that $u_t^\varepsilon = \tilde{x}_t^\varepsilon g_t^\varepsilon$ has a global solution. □

REMARK 3.1. Since the stochastic process (g_t^ε) is sample continuous with initial value the unit matrix, it stays in the connected component $SO(n)$ of $O(n)$.

If $\{A_k\}$ is an orthonormal basis of $\mathfrak{so}(n)$ let $\mathcal{L}_G = \frac{1}{2}\sum_{k=1}^N L_{gA_k}L_{gA_k}$. Then (g_t^ε) is a Markov process with infinitesimal generator

$$\mathcal{L}^\varepsilon = \frac{1}{\varepsilon}\mathcal{L}_G + L_{g\bar{A}}.$$

LEMMA 3.2. *Let M be a complete Riemannian manifold with positive injectivity radius. Suppose that there are numbers $C > 0$ and $a_2 > 0$ such that* $\sup_{\rho(x,y)\le a_2}|\nabla d\rho|(x,y) \le C$. *Let $T > 0$. The probability distributions of the family of stochastic processes $\{\tilde{x}_{t/\varepsilon}^\varepsilon, t \le T\}$ are tight. There is a metric $\tilde{d}$ on M such that $\{(\tilde{x}_{t/\varepsilon}^\varepsilon)\}$ is equi-Hölder continuous with exponent $\alpha < \frac{1}{2}$.*

PROOF. Let μ^ε be the probability laws of $(\tilde{x}_t^\varepsilon)$ on the path space over OM with initial value u_0, which we denote by $C([0,T]; OM)$. Since $\tilde{x}_0^\varepsilon = u_0$, it suffices to estimate the modulus of continuity and show that for all positive numbers a, η, there exists $\delta > 0$ such that for all ε sufficiently small (see Billingsley [2] and Ethier and Kurtz [9])

$$P\Big(\omega : \sup_{|s-t|<\delta} d(\tilde{x}_t^\varepsilon, \tilde{x}_s^\varepsilon) > a\Big) < \delta\eta.$$

Here, d denotes a distance function on OM. We will choose a suitable distance function. The Riemannian distance function $\tilde{\rho}(x,y)$ is not smooth in y if y is in the cut locus of x. To avoid any assumption on the cut locus of OM, we construct a new distance function that preserves the topology of OM.

Let $2a$ be the minimum of 1, a_2 and the injectivity radius of M. Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ be a smooth concave function such that $\phi(r) = r$ when $r < a$ and $\phi(r) = 1$ when $r \ge 2a$. Let ρ and $\tilde{\rho}$ be, respectively, the Riemannian distance on M and on

OM. Then $\phi \circ \rho$ and $\tilde{d} = \phi \circ \tilde{\rho}$ are distance functions on M and on OM, respectively. Then for $r < t$,

$$\phi^2 \circ \tilde{\rho}(\tilde{x}^\varepsilon_{t/\varepsilon}, \tilde{x}^\varepsilon_{r/\varepsilon}) = \int_{r/\varepsilon}^{t/\varepsilon} D(\phi^2 \circ \tilde{\rho}(\tilde{x}^\varepsilon_r, \cdot))_{\tilde{x}^\varepsilon_s}(H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0))\, ds.$$

Since $H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0)$ has unit length, from the equation above we do not observe, directly, a uniform bound in ε.

For further estimates, we work with a C^2 function $F : OM \to \mathbb{R}$ to simplify the notation. Also, the computations below and some of the identities will be used later in the proof of Theorem 1.1. Let $0 \le r < t$,

$$F(\tilde{x}^\varepsilon_{t/\varepsilon}) = F(\tilde{x}^\varepsilon_{r/\varepsilon}) + \int_{r/\varepsilon}^{t/\varepsilon} (DF)_{\tilde{x}^\varepsilon_s}(H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0))\, ds. \tag{3.1}$$

Let $\{e_i\}$ be an orthonormal basis of $\mathbb{R}^n$. We define two sets of functions $f_i : OM \to \mathbb{R}$ and $h_i : O(n) \to \mathbb{R}$:

$$f_i(u) = (DF)_u(H_u e_i), \qquad \alpha_i(g) = \langle g e_0, e_i \rangle.$$

From the linearity of H_u, we obtain the identity $H_u(g e_0) = \sum_{i=1}^n H_u(e_i)\alpha_i(u)$. Thus, the integrand in (3.1) factorizes and we have

$$F(\tilde{x}^\varepsilon_{t/\varepsilon}) = F(\tilde{x}^\varepsilon_{r/\varepsilon}) + \sum_{i=1}^n \int_{r/\varepsilon}^{t/\varepsilon} f_i(\tilde{x}^\varepsilon_s)\alpha_i(g^\varepsilon_s e_0)\, ds. \tag{3.2}$$

Since the Riemannian metric on $G = SO(n)$ is bi-invariant, the Riemannian volume measure, which locally has the form $\sqrt{\det(g_{ij})}\, dx^1 \wedge \cdots \wedge dx^N$, is the Haar measure. Let dg be the Haar measure normalized to be a probability measure on G. Let $\tilde{g}$ be a rotation such that $\tilde{g} e_0 = -e_0$. Then $\int_G g(\tilde{g} e_0)\, dg = \int_G g(e_0)\, dg$. The integral of $g e_0$ with respect to the Haar measure vanishes. In particular, $\int_G \alpha_i\, dg = 0$. On a compact Riemannian manifold the Poisson equation with a smooth function that is centered with respect to the Riemannian volume measure has a unique centered smooth solution. For each i, let $h_i : G \to \mathbb{R}$ be the smooth centred solution to the Poisson equation

$$\mathcal{L}_G h_i = \alpha_i = \langle g e_0, e_i \rangle. \tag{3.3}$$

We apply Itô's formula to the function $f_i h_i$ and $r < t$,

$$\begin{aligned} f_i(\tilde{x}^\varepsilon_{t/\varepsilon}) h_i(g^\varepsilon_{t/\varepsilon}) = {} & f_i(\tilde{x}^\varepsilon_{r/\varepsilon}) h_i(g^\varepsilon_{r/\varepsilon}) + \int_{r/\varepsilon}^{t/\varepsilon} (Df_i)_{\tilde{x}^\varepsilon_s}(H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0)) h_i(g^\varepsilon_s)\, ds \\ & + \frac{1}{\sqrt{\varepsilon}} \sum_k \int_{r/\varepsilon}^{t/\varepsilon} f_i(\tilde{x}^\varepsilon_s)(Dh_i)_{(g^\varepsilon_s)}(g^\varepsilon_s A_k)\, dw^k_s \\ & + \int_{r/\varepsilon}^{t/\varepsilon} f_i(\tilde{x}^\varepsilon_s) L_{g^\varepsilon_s \bar{A}} h_i(g^\varepsilon_s)\, ds + \frac{1}{\varepsilon} \int_{r/\varepsilon}^{t/\varepsilon} f_i(\tilde{x}^\varepsilon_s) \mathcal{L}_G h_i(g^\varepsilon_s)\, ds. \end{aligned}$$

We sum up the above equation from $i = 1$ to n. Note that

$$\sum_{i=1}^{n} f_i(u)\mathcal{L}_G h_i(g) = \sum_{i=1}^{n} f_i(u)\alpha_i(g).$$

We compare the last term in the above formula for $f_i(\tilde{x}^\varepsilon_{t/\varepsilon})h_i(g^\varepsilon_{t/\varepsilon})$ with the integral in (3.2) to obtain that

$$\begin{aligned} F(\tilde{x}^\varepsilon_{t/\varepsilon}) = F(\tilde{x}^\varepsilon_{r/\varepsilon}) &+ \varepsilon \sum_{i=1}^{n} \big(f_i(\tilde{x}^\varepsilon_{t/\varepsilon})h_i(g^\varepsilon_{t/\varepsilon}) - f_i(\tilde{x}^\varepsilon_{r/\varepsilon})h_i(g^\varepsilon_{r/\varepsilon})\big) \\ &- \varepsilon \sum_{i=1}^{n} \int_{r/\varepsilon}^{t/\varepsilon} (Df_i)_{\tilde{x}^\varepsilon_s}\big(H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0)\big)h_i(g^\varepsilon_s)\,ds \\ &- \varepsilon \sum_{i=1}^{n} \int_{r/\varepsilon}^{t/\varepsilon} f_i(\tilde{x}^\varepsilon_s)L_{g^\varepsilon_s \bar{A}} h_i(g^\varepsilon_s)\,ds \\ &- \sqrt{\varepsilon}\sum_{i=1}^{n}\sum_{k=1}^{N} \int_{r/\varepsilon}^{t/\varepsilon} f_i(\tilde{x}^\varepsilon_s)(Dh_i)_{(g^\varepsilon_s)}(g^\varepsilon_s A_k)\,dw^k_s. \end{aligned}$$

Let us compute the differential of $f_i(u) = (DF)_u(H_u e_i)$. Let ∇ be the flat connection on OM. It is determined by the parallelization $\mathbb{X}: OM \times \mathbb{R}^n \times \mathfrak{so}(n) \to TOM$ where $\mathbb{X}_u(e, A) = H_u(e) + \varpi_u^{-1}(A)$. In the calculation below, we use the fact that $\nabla H(e) = 0$.

$$\begin{aligned} &F(\tilde{x}^\varepsilon_{t/\varepsilon}) - F(\tilde{x}^\varepsilon_{r/\varepsilon}) \\ &\quad = \varepsilon \sum_{i=1}^{n} \big((DF)_{\tilde{x}^\varepsilon_{t/\varepsilon}}(H_{\tilde{x}^\varepsilon_{t/\varepsilon}} e_i)h_i(g^\varepsilon_{t/\varepsilon}) - (DF)_{\tilde{x}^\varepsilon_{r/\varepsilon}}(H_{\tilde{x}^\varepsilon_{r/\varepsilon}} e_i)h_i(g^\varepsilon_{r/\varepsilon})\big) \\ &\quad\quad - \varepsilon \sum_{i=1}^{n} \int_{r/\varepsilon}^{t/\varepsilon} (\nabla DF)_{\tilde{x}^\varepsilon_s}\big(H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0), H_{\tilde{x}^\varepsilon_s}(e_i)\big)h_i(g^\varepsilon_s)\,ds \\ &\quad\quad - \varepsilon \sum_{i=1}^{n} \int_{r/\varepsilon}^{t/\varepsilon} (DF)_{\tilde{x}^\varepsilon_s}(H_{\tilde{x}^\varepsilon_s} e_i)L_{g^\varepsilon_s \bar{A}} h_i(g^\varepsilon_s)\,ds \\ &\quad\quad - \sqrt{\varepsilon}\sum_{i=1}^{n}\sum_{k=1}^{N} \int_{r/\varepsilon}^{t/\varepsilon} (DF)_{\tilde{x}^\varepsilon_s}(H_{\tilde{x}^\varepsilon_s} e_i)(Dh_i)_{(g^\varepsilon_s)}(g^\varepsilon_s A_k)\,dw^k_s. \end{aligned} \tag{3.4}$$

We also remark that $|H_{\tilde{x}^\varepsilon_s} e_i| = 1$, $|H_{\tilde{x}^\varepsilon_s} g^\varepsilon_s e_i| = 1$, $|g^\varepsilon_s \bar{A}| = |\bar{A}|$. If F is a function that is BC^2, by the Kunita–Watanabe inequality, for any $p \geq 1$,

$$\mathbb{E}\big|F(\tilde{x}^\varepsilon_{t/\varepsilon}) - F(\tilde{x}^\varepsilon_{r/\varepsilon})\big|^p \leq C_1(T)\varepsilon^p\big(|DF|_\infty + |\nabla DF|_\infty\big) + C_1(T)|DF|_\infty |t-r|^{p/2},$$

for some constant $C_1(T)$. If $\varepsilon^2 \leq |t-r|$, there exists a constant $C_2(T)$, such that $\mathbb{E}|F(\tilde{x}^\varepsilon_{t/\varepsilon}) - F(\tilde{x}^\varepsilon_{r/\varepsilon})|^p \leq C_2(T)|t-r|^{p/2}$. If $|t-r| < \varepsilon^2$, we estimate directly from (3.1):

$$\left|F\left(\tilde{x}^{\varepsilon}_{t/\varepsilon}\right)-F\left(\tilde{x}^{\varepsilon}_{r/\varepsilon}\right)\right|\le C\frac{t-r}{\varepsilon}\le C\sqrt{t-r}.$$

Thus, for $C(T)=C_2(T)+C^p$,

$$\mathbb{E}\left|F\left(\tilde{x}^{\varepsilon}_{t/\varepsilon}\right)-F\left(\tilde{x}^{\varepsilon}_{r/\varepsilon}\right)\right|^p\le C(T)|t-r|^{p/2}.$$

We apply the above formula to $F=\phi^2\circ\tilde{\rho}(\cdot,u_0)$ where $u_0=\tilde{x}^{\varepsilon}_0$. Since ϕ is bounded so is F. Since $|\nabla\tilde{\rho}(\cdot,u_0)|\le 1$ and ϕ' is bounded, $\nabla F=2\phi\phi'\nabla\rho(\cdot,u_0)$ is bounded. The norm of its second derivative is

$$\left|2(\phi')^2\nabla\rho\otimes\nabla\rho+2(\phi\phi'')\nabla\rho\otimes\nabla\rho+2(\phi\phi')\nabla\, d\rho\right|,$$

and the tensor is evaluated at $\rho(x,y)$. We remark that $\phi'(x,y)=0$ when $\rho(x,y)\ge a$ and $|\nabla\, d\rho(\rho(x,y))|\le C$ when $\rho(x,y)\ge a$. Hence, for all u_0, there is a common number $C(T)$ s.t.

$$\mathbb{E}\left|\tilde{d}\left(\tilde{x}^{\varepsilon}_{t/\varepsilon},u_0\right)\right|^p\le C(T)t^{p/2}.$$

Conditioning on $\mathcal{F}_r$ to see that

$$\mathbb{E}\left|\tilde{d}\left(\tilde{x}^{\varepsilon}_{t/\varepsilon},\tilde{x}^{\varepsilon}_{r/\varepsilon}\right)\right|^p\le C(T)|t-r|^{p/2}.$$

The tightness of the law of $\{\tilde{x}^{\varepsilon}_{t/\varepsilon}\}$ follows. By Kolmogorov's criterion, $\{\tilde{x}^{\varepsilon}_{t/\varepsilon}\}$ is Hölder continuous with exponent α for any $\alpha<\frac{1}{2}$. The Hölder constants are independent of ε and, for any $p'<p$, Kolmogorov's criterion yields

$$\sup_{\varepsilon}\mathbb{E}\sup_{s\neq t}\left(\frac{\tilde{d}(\tilde{x}^{\varepsilon}_{t/\varepsilon},\tilde{x}^{\varepsilon}_{s/\varepsilon})}{|t-s|^{\alpha}}\right)^{p'}<\infty, \tag{3.5}$$

thus completing the proof. □

We will need the following lemma in which we make a statement on the limit of a function of two variables, one of which is ergodic and the other one varies significantly slower. The result is straightforward, but we include the proof for completeness. If $f:N\to\mathbb{R}$ is a Lipschitz continuous function on a metric space (N,d) with distance function d, we denote by $|f|_{\mathrm{Lip}}$ its Lipschitz semi-norm. If S is a subset of N, we let $\mathrm{Osc}_S(f)$ denote $|\sup_{x\in S}f(x)-\inf_{x\in S}f(x)|$, the Oscillation of f over S. Let $\mathrm{Osc}(f)=\mathrm{Osc}_N(f)$.

Let $E(N)$ be one of the following classes of real valued functions on a metric space (N,d):

$$E(N)=\left\{f:N\to\mathbb{R}:|f|_{\mathrm{Lip}}<\infty,\mathrm{Osc}(f)<\infty\right\}$$

or $E_r(N)=E(N)\cap C^r$, where $r=0,1,\dots,\infty$. Denote

$$|f|_E=|f|_{\mathrm{Lip}}+\mathrm{Osc}(f).$$

Let d be the metric with respect to which the Lipschitz property is defined. We define $\tilde{d} = d \wedge 1$ to be a new metric on N. Then $|f|_{\text{Lip}} \le C$ and $\text{Osc}(f) \le C$ is equivalent to f being Lipschitz with respect to $\tilde{d}$.

Let $p \ge 1$ and let $W_p(N)$ denote the Wasserstein p-distance between two probability measures on a metric space (N, d):

$$\big(W_p(\mu_1, \mu_2)\big)^p = \inf_{\{\nu : (\pi_1)_*\nu = \mu_1, (\pi_2)_*\nu = \mu_2\}} \int_{N \times N} \big(d(x, y)\big)^p \, d\nu(x, y).$$

Let μ^ε, μ be a family of probability measures on the metric space (N, d). Then $\mu^\varepsilon \to \mu$ in $W_p(N)$ if and only if they converge weakly and $\sup_{x \in N} \int (d(x, y))^p \, d\mu_\varepsilon(y)$ is bounded for any $x \in N$. If $\tilde{d} = d \wedge 1$, then $\tilde{d}$ and d induce the same topology on N and the concepts of weak convergence are equivalent. With respect to $\tilde{d}$, weak convergence is equivalent to Wasserstein p-convergence.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a filtered probability space. Let (Y, ρ), (Z, d) be metric spaces or C^m manifolds. Let $\{(y_t^\varepsilon, t \le T), \varepsilon > 0\}$ be a family of $\mathcal{F}_t$-adapted stochastic processes with state space Y. Let (z_t^ε) be a family of sample continuous $\mathcal{F}_t$-Markov processes on Z.

ASSUMPTION 3.3. (1) The stochastic processes $(y_{t/\varepsilon}^\varepsilon, t \le T)$ are equi-uniformly continuous and converge weakly to a continuous process $(\bar{y}_t, t \le T)$.

(2) For each ε, $(z_{t\varepsilon}^\varepsilon, t \le T)$ has an invariant measure μ_ε. There exists a function δ on $\mathbb{R}_+ \times Z \times \mathbb{R}_+$ with the property that $\delta(\cdot, z, \varepsilon)$ is nondecreasing for each pair of (z, ε) and $\lim_{\varepsilon \to 0} \sup_{z \in Z} \delta(K, z, \varepsilon) = 0$ for all K and for all $f \in E_r(Z)$ and $t > 0$,

$$\mathbb{E}\left| \frac{\varepsilon}{t} \int_0^{t/\varepsilon} f\big(z_{s\varepsilon}^\varepsilon\big) \, ds - \int_Z f(z) \, d\mu_\varepsilon(z) \right| \le \delta\left(|f|_E, z_0^\varepsilon, \frac{\varepsilon}{t} \right).$$

(3) There exists a probability measure μ on $W^1(C([0, T]; Z))$ s.t. $\lim_{\varepsilon \to 0} W_1(\mu_\varepsilon, \mu) = 0$.

(4) The processes $(y_{t/\varepsilon}^\varepsilon)$ converges to $(\bar{y}_t)$ in $W_1(Y)$, and there exists an exponent $\alpha > 0$ such that

$$\sup_\varepsilon \mathbb{E}\left(\sup_{s \ne t} \frac{\rho(y_{t/\varepsilon}^\varepsilon, y_{s/\varepsilon}^\varepsilon)}{|t - s|^\alpha} \right) < \infty.$$

We cannot assume that $(\bar{y}_t)$ is adapted to the filtration with respect to which $(z_{t/\varepsilon}^\varepsilon)$ is a Markov process. The process $(z_{t/\varepsilon}^\varepsilon)$ is usually not convergent and we do not assume that $(y_t^\varepsilon, z_t^\varepsilon)$ and $(\bar{y}_t)$ are realized in the same probability space.

We denote by $\hat{P}_\eta$ the probability distribution of a random variable η and let T be a positive real number. If r is a positive number, let $C([0, r]; Y)$ denote the space of continuous paths, $\sigma : [0, r] \to Y$, on Y. If $F : C([0, r]; Y) \to \mathbb{R}$ is a Borel measurable function, we use the shorter notation $F(y_{\cdot/\varepsilon}^\varepsilon)$ for $F((y_{u/\varepsilon}^\varepsilon, u \le r))$.

LEMMA 3.4. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a filtered probability space. Let (Y, ρ), (Z, d) be metric spaces or C^m manifolds in case $m \ge 1$. Let $\{(y_t^\varepsilon, t \le T), \varepsilon > 0\}$ be a family of $\mathcal{F}_t$-adapted stochastic processes on Y. Let (z_t^ε) be a family of sample continuous $\mathcal{F}_t$-Markov processes on Z. Let $G \in E_m(Y \times Z)$. Let $0 \le r < t$ and let $F : C([0, r]; Y) \to \mathbb{R}$ be a bounded continuous function. We define*

$$A(\varepsilon) \equiv A(\varepsilon, F, G) := F\big(y^\varepsilon_{\cdot/\varepsilon}\big) \int_r^t G\big(y^\varepsilon_{s/\varepsilon}, z^\varepsilon_{s/\varepsilon}\big)\, ds.$$

- *If* (1)–(3) *in Assumption* 3.3 *hold, then the random variables $A(\varepsilon)$ converge weakly to A as $\varepsilon \to 0$, where*

$$A \equiv A(F, G) := F(\bar y_\cdot) \int_r^t \int_Z G(\bar y_s, z)\, d\mu(z)\, ds.$$

- *Assume* (1)–(4) *in Assumption* 3.3. *Then there is a constant c, s.t. for $\varepsilon < 1$,*

$$\begin{aligned} &W_1(\hat P_{A(\varepsilon)}, \hat P_A) \\ &\quad \le c|F|_\infty \max_{z\in Z} \delta\bigg(|G|_E, z, \frac{\varepsilon}{t-r}\bigg) + 2\varepsilon |F|_\infty \min\big(|G|_\infty, |\mathrm{Osc}(G)|\big) \\ &\qquad + c(t-r)|F|_\infty |G|_{\mathrm{Lip}}\big(W_1\big(\hat P_{y^\varepsilon_{\cdot/\varepsilon}}, \hat P_{\bar y_\cdot}\big) + W_1\big(\mu^\varepsilon, \mu\big)\big) + c\varepsilon^\alpha |F|_\infty |G|_{\mathrm{Lip}}. \end{aligned}$$

PROOF. Let us fix the functions F, G, r, t and define

$$\mathcal{E}_1(r,t) = \int_r^t G\big(y^\varepsilon_{s/\varepsilon}, z^\varepsilon_{s/\varepsilon}\big)\, ds - \int_r^t \int_Z G\big(y^\varepsilon_{s/\varepsilon}, z\big)\, d\mu_\varepsilon(z)\, ds;$$

$$\mathcal{E}_2 = F\big(y^\varepsilon_{\cdot/\varepsilon}\big)\bigg(\int_r^t \int_Z G\big(y^\varepsilon_{s/\varepsilon}, z\big)\, d\mu_\varepsilon(z)\, ds - \int_r^t \int_Z G\big(y^\varepsilon_{s/\varepsilon}, z\big)\, d\mu(z)\, ds\bigg);$$

$$I(\varepsilon) = F\big(y^\varepsilon_{\cdot/\varepsilon}\big) \int_r^t \int_Z G\big(y^\varepsilon_{s/\varepsilon}, z\big)\, d\mu(z)\, ds.$$

The proof is split into three parts: (i) $F(y^\varepsilon_{\cdot/\varepsilon})\mathcal{E}_1(r,t)$ converges to zero in $L_p(\Omega)$ for any $p > 1$, (ii) $\mathcal{E}_2$ converges to zero in $L_p(\Omega)$ for any $p > 1$ and (iii) $I(\varepsilon)$ converges to A weakly.

We first prove that $F(y^\varepsilon([0, \frac{r}{\varepsilon}]))\mathcal{E}_1(r,t)$ converges to zero in $L_p(\Omega)$. Since F is bounded it is sufficient to take $r = 0$ and F a constant, and to work with $\mathcal{E}_1(0,t)$. Let us write

$$\mathcal{E}_1 := \int_0^t G\big(y^\varepsilon_{s/\varepsilon}, z^\varepsilon_{s/\varepsilon}\big)\, ds - \int_0^t \int_Z G\big(y^\varepsilon_{s/\varepsilon}, z\big)\, d\mu_\varepsilon(z)\, ds.$$

Let $0 = t_0 < t_1 < \cdots < t_M \le t$ be a partition of $[0,t]$ into pieces of size $t\varepsilon$. Let $M \equiv M_\varepsilon = [\frac{1}{\varepsilon}]$. Let $\Delta t_i = t_{i+1} - t_i$ and let $\tilde t = t\varepsilon M_\varepsilon$. Below $a \sim b$ indicates "$a - b = O(\varepsilon)$" as ε converges to 0. Since $G \in E_m(Y \times Z)$,

$$\begin{aligned} \big|\mathcal{E}_1(\tilde t, t)\big| &\le 2\min\bigg(|G|_\infty, \big|\mathrm{Osc}(G)\big|, |G|_{\mathrm{Lip}} \max_{0\le s\le t} \int_Z d\big(z^\varepsilon_{s/\varepsilon}, z\big)\mu_\varepsilon(dz)\bigg)(t - \tilde t) \\ &\le \varepsilon 2 \min\big(|G|_\infty, \big|\mathrm{Osc}(G)\big|\big) \le 2\varepsilon\big(|G|_E\big). \end{aligned}$$

By the Lipschitz continuity of G, for each $\varepsilon > 0$ the following holds:

$$\begin{aligned}\mathcal{E}_3 &:= \left|\sum_{i=0}^{M_\varepsilon-1}\int_{t_i}^{t_{i+1}} G\big(y^\varepsilon_{s/\varepsilon}, z^\varepsilon_{s\varepsilon}\big)\,ds - \sum_{i=0}^{M_\varepsilon-1}\int_{t_i}^{t_{i+1}} G\big(y^\varepsilon_{t_i/\varepsilon}, z^\varepsilon_{s\varepsilon}\big)\,ds\right| \\ &\le |G|_{\mathrm{Lip}}\sum_{i=0}^{M_\varepsilon-1}\int_{t_i}^{t_{i+1}}\rho\big(y^\varepsilon_{s/\varepsilon}, y^\varepsilon_{t_i/\varepsilon}\big)\,ds.\end{aligned}$$

By equi-uniform continuity of $(y^\varepsilon_{s/\varepsilon})$, for almost surely all ω, $\mathcal{E}_3$ converges to zero. Since $\mathcal{E}_3$ is bounded the convergence is in $L_p(\Omega)$. If $(y^\varepsilon_{s/\varepsilon})$ is assumed to be equi-Hölder continuous as in condition (4), there is a convergence rate of $\varepsilon^\alpha|G|_{\mathrm{Lip}}$ for the L^p convergence.

We prove next that $\sum_{i=0}^{M_\varepsilon-1}\int_{t_i}^{t_{i+1}} G(y^\varepsilon_{t_i/\varepsilon}, z^\varepsilon_{s/\varepsilon})\,ds$ converges. We apply the Markov property of (z^ε_t) and we use the fact that (y^ε_t) is adapted to the filtration $(\mathcal{F}_t)$, with respect to which (z^ε_t) is a Markov process:

$$\begin{aligned}&\sum_{i=1}^{M_\varepsilon-1}\mathbb{E}\left|\int_{t_i}^{t_{i+1}} G\big(y^\varepsilon_{t_i/\varepsilon}, z^\varepsilon_{s/\varepsilon}\big)\,ds - \Delta t_i\int_Z G\big(y^\varepsilon_{t_i/\varepsilon}, z\big)\,d\mu_\varepsilon(z)\right| \\ &\quad\le \sum_{i=1}^{M_\varepsilon-1}\Delta t_i\mathbb{E}\bigg(\mathbb{E}\bigg\{\bigg|\frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}} G\big(y^\varepsilon_{t_i/\varepsilon}, z^\varepsilon_{s/\varepsilon}\big)\,ds \\ &\qquad\qquad - \int_Z G\big(y^\varepsilon_{t_i/\varepsilon}, z\big)\,d\mu_\varepsilon(z)\bigg|\bigg|\mathcal{F}_{t_i/\varepsilon}\bigg\}\bigg) \\ &\quad= \sum_{i=1}^{M_\varepsilon-1}\Delta t_i\mathbb{E}\bigg(\mathbb{E}\bigg(\bigg|\frac{\varepsilon^2}{\Delta t_i}\int_{t_i/\varepsilon^2}^{t_{i+1}/\varepsilon^2} G\big(y, z^\varepsilon_{s\varepsilon}\big)\,ds \\ &\qquad\qquad - \int_Z G(y,z)\,d\mu_\varepsilon(z)\bigg|\bigg)\bigg|_{y=y^\varepsilon_{t_i/\varepsilon}}\bigg).\end{aligned}$$

Since $\frac{\varepsilon^2}{\Delta t_i} = \frac{\varepsilon}{t}$, we may now apply condition (2) and obtain

$$\begin{aligned}&\mathbb{E}\bigg(\bigg|\frac{\varepsilon^2}{\Delta t_i}\int_{t_i/\varepsilon^2}^{t_{i+1}/\varepsilon^2} G\big(y, z^\varepsilon_{s\varepsilon}\big)\,ds - \int_Z G(y,z)\,d\mu_\varepsilon(z)\bigg|\bigg) \\ &\qquad\le \delta\bigg(\big|G\big(y^\varepsilon_{t_i/\varepsilon},\cdot\big)\big|_E, z^\varepsilon_{t_i/\varepsilon}, \frac{\varepsilon}{t}\bigg) \le \delta\bigg(|G|_E, z^\varepsilon_{t_i/\varepsilon}, \frac{\varepsilon}{t}\bigg).\end{aligned}$$

We record that

$$\begin{aligned}\mathcal{E}_4 &:= \mathbb{E}\left|\sum_{i=0}^{M_\varepsilon-1}\int_{t_i}^{t_{i+1}} G\big(y^\varepsilon_{t_i/\varepsilon}, z^\varepsilon_{s/\varepsilon}\big)\,ds - \sum_{i=0}^{M_\varepsilon-1}\Delta t_i\int_Z G\big(y^\varepsilon_{t_i/\varepsilon}, z\big)\,d\mu_\varepsilon(z)\right| \\ &\le \max_{z\in Z}\delta\bigg(|G|_E, z, \frac{\varepsilon}{t}\bigg).\end{aligned} \tag{3.6}$$

Let us define

$$\mathcal{E}_5 := \sum_{i=0}^{M_\varepsilon - 1} \Delta t_i \int_Z G\big(y^\varepsilon_{t_i/\varepsilon}, z\big)\, d\mu_\varepsilon(z) - \int_0^t \int_Z G\big(y^\varepsilon_{s/\varepsilon}, z\big)\, d\mu_\varepsilon(z)\, ds.$$

By the definition of Riemann integral

$$\mathcal{E}_5 \le |G|_{\mathrm{Lip}} \sum_{i=0}^{M_\varepsilon - 1} \Delta t_i \mathrm{Osc}_{[s_i, s_{i+1}]}\big(y^\varepsilon_{s/\varepsilon}\big),$$

where $\mathrm{Osc}_{[a,b]}(f)$ denotes the oscillation of a function f in the indicated interval. Since $(y^\varepsilon_{s/\varepsilon})$ is equi-uniform continuous on $[0, T]$, $\mathcal{E}_5 \to 0$ in L_p. Given Hölder continuity of $(y^\varepsilon_{s/\varepsilon})$ from condition (4), we have the quantitative estimates: $|\mathcal{E}_5|_{L_p(\Omega)} \le C|G|_{\mathrm{Lip}}\varepsilon^\alpha$. To summarize,

$$\big|\mathcal{E}_1(0, t)\big| \le \big|\mathcal{E}_1(\tilde{t}, t)\big| + \mathcal{E}_3 + \mathcal{E}_4 + \mathcal{E}_5.$$

It follows that $F(y^\varepsilon_{r/\varepsilon})\mathcal{E}_1(r, t)$ converges to zero.

When condition (4) holds, there is a constant C such that

$$\begin{aligned} &\big|F(y^\varepsilon_{\cdot/\varepsilon})\mathcal{E}_1(r, t)\big|_{L_p(\Omega)} \\ &\quad \le |F|_\infty\big(2\varepsilon \min\big(|G|_\infty, \big|\mathrm{Osc}(G)\big|\big) + \mathcal{E}_3 + \mathcal{E}_4 + \mathcal{E}_5\big) \\ &\quad \le C|F|_\infty(\varepsilon^\alpha + \varepsilon)|G|_{\mathrm{Lip}} + 2\varepsilon|F|_\infty \min\big(|G|_\infty, \big|\mathrm{Osc}(G)\big|\big) \\ &\qquad + C|F|_\infty \max_{z \in Z} \delta\bigg(|G|_E, z, \frac{\varepsilon}{t - r}\bigg). \end{aligned} \tag{3.7}$$

For any two random variables on the same probability space and with the same state space, the L_p norm of their difference dominates their Wasserstein p-distance. The random variable

$$F\big(y^\varepsilon_{r/\varepsilon}\big) \int_r^t G\big(y^\varepsilon_{s/\varepsilon}, z^\varepsilon_{s/\varepsilon}\big)\, ds - F\big(y^\varepsilon_{r/\varepsilon}\big) \int_r^t \int_Z G\big(y^\varepsilon_{s/\varepsilon}, z\big)\, d\mu_\varepsilon(z)\, ds \overset{W_p(N)}{\to} 0,$$

with the same rate as indicated above.

We proceed to step (ii). It is clear that for almost all ω, $F(y^\varepsilon_{\cdot/\varepsilon}) \int_r^t G(y^\varepsilon_{s/\varepsilon}, z)\, ds$ is Lipschitz continuous in z. For any $z_1, z_2 \in Z$,

$$\begin{aligned} &\bigg|F\big(y^\varepsilon_{\cdot/\varepsilon}\big) \int_r^t G\big(y^\varepsilon_{s/\varepsilon}, z_1\big)\, ds - F\big(y^\varepsilon_{\cdot/\varepsilon}\big) \int_r^t G\big(y^\varepsilon_{s/\varepsilon}, z_2\big)\, ds\bigg| \\ &\qquad \le |F|_\infty d(z_1, z_2) \int_r^t \big|G\big(y^\varepsilon_{s/\varepsilon}, \cdot\big)\big|_{\mathrm{Lip}}\, ds \le (t - r) d(z_1, z_2)|F|_\infty |G|_{\mathrm{Lip}}. \end{aligned}$$

By the Kantorovich duality formula, for the distance between two probability measures μ_1 and μ_2,

$$W_1(\mu_1, \mu_2) = \sup\bigg\{\int U\, d\mu_1 - \int U\, d\mu_2 : |U|_{\mathrm{Lip}} \le 1\bigg\},$$

we have

$$|\mathcal{E}_2| \le (t-r) \cdot |F|_\infty \cdot |G|_{\mathrm{Lip}} \cdot W_1(\mu^\varepsilon, \mu).$$

For part (iii), let U be a continuous function on $C([0,T];Y)$. If $\sigma \in C([0,T];Y)$, let us denote by $\sigma([0,r])$ the restriction of the path to $[0,r]$. Since F is bounded continuous and G is Lipschitz continuous,

$$\sigma \mapsto U\biggl(F(\sigma([0,r]))\biggl(\int_r^t \int_Z G(\sigma_s, z)\, d\mu(z)\, ds\biggr)\biggr)$$

is a continuous function on $C([0,T];Y)$. By the weak convergence of $(y^\varepsilon_{\cdot/\varepsilon})$, $\mathbb{E}(U(I(\varepsilon)))$ converges to $\mathbb{E}(U(A(F,G)))$ and the random variables $I(\varepsilon)$ converge weakly to $A(F,G)$. By now, we have proved that $A(\varepsilon, F, G)$ converges to $A(F,G)$ weakly; we thus conclude the first part of the lemma.

Let us assume condition (4) from Assumption 3.3. In particular, $(y^\varepsilon_{\cdot/\varepsilon})$ converges in $W_1(C([0,T];Y))$. Let U be a Lipschitz continuous function on $C([0,T];Y)$. We define $\tilde{U}: C([0,T];Y) \to \mathbb{R}$ by

$$\tilde{U}(\sigma) = U\biggl(F(\sigma([0,r]))\biggl(\int_r^t \int_Z G(\sigma_s, z)\, d\mu(z)\, ds\biggr)\biggr).$$

Let σ^1, σ^2 are two paths on Y,

$$\begin{aligned}
&|\tilde{U}(\sigma_1) - \tilde{U}(\sigma_2)| \\
&\quad \le |U|_{\mathrm{Lip}} \cdot |F|_\infty \left| \int_r^t \int_Z G(\sigma_s^1, z)\, d\mu(z)\, ds - \int_r^t \int_Z G(\sigma_s^2, z)\, d\mu(z)\, ds \right| \\
&\quad \le (t-r)|U|_{\mathrm{Lip}} \cdot |F|_\infty \cdot |G|_{\mathrm{Lip}} \cdot \sup_{0 \le s \le T} \rho(\sigma_s^1, \sigma_s^2).
\end{aligned}$$

By the Kantorovitch duality and assumption (4),

$$W_1(\hat{P}_{I(\varepsilon)}, \hat{P}_I) \le (t-r) \cdot |F|_\infty \cdot |G|_{\mathrm{Lip}} \cdot W_1(\hat{P}_{y^\varepsilon_{\cdot/\varepsilon}}, \hat{P}_{\bar{y}_\cdot}).$$

We collect all the estimations together. Under assumptions (1)–(4), the following estimates hold:

$$\begin{aligned}
W_1(\hat{P}_{A(\varepsilon)}, \hat{P}_A) &\le C|F|_\infty |G|_{\mathrm{Lip}}(\varepsilon^\alpha + \varepsilon) + C|F|_\infty \max_{z \in Z} \delta\biggl(|G|_E, z, \frac{\varepsilon}{t-r}\biggr) \\
&\quad + C(t-r) \cdot |F|_\infty \cdot |G|_{\mathrm{Lip}} \cdot \bigl(W_1(\hat{P}_{y^\varepsilon_{\cdot/\varepsilon}}, \hat{P}_{\bar{y}_\cdot}) + W_1(\mu_\varepsilon, \mu)\bigr) \\
&\quad + 2\varepsilon |F|_\infty \min\bigl(|G|_\infty, |\mathrm{Osc}(G)|\bigr).
\end{aligned}$$

We may now limit ourselves to $\varepsilon \le 1$ and conclude part 2 of the lemma. □

REMARK 3.2. In the lemma above, we should really think that the z^ε process and process y^ε follow different clocks, the former is run at the fast time scale $\frac{1}{\varepsilon}$ and the latter at scale 1.

EXAMPLE 3.5. Let (g_s) be a Brownian motion on $G = SO(n)$, solving

$$dg_t = \sum_{k=1}^{N} L_{g_t A_k}\, dw_t^k.$$

Here, $\{A_1, \dots, A_N\}$ is an orthonormal basis of $\mathfrak{g}$. In Lemma 3.4 we take $z_t^\varepsilon = g_{t/\varepsilon}$, then condition (2) holds. If f is a Lipschitz continuous function, it is well known that the law of large numbers holds for $\int_0^t f(g_s)\, ds$, so does a central limit theorem. The remainder term in the central limit theorem is of order $\sqrt{t}$ and depends on f only through the Lipschitz constant $|f|_{\mathrm{Lip}}$.

It is easy to see that the remainder term in the law of large numbers depends only on the Lipschitz constant of the function. Without loss of generality, we assume that $\int f\, dg = 0$. Let α solve the Poisson equation: $\Delta^G \alpha = f$. Then

$$\frac{1}{t}\int_0^t f(g_s)\, ds = \frac{1}{t}\alpha(g_t) - \frac{1}{t}\alpha(g_0) - \sum_k \frac{1}{t}\int_0^t (D\alpha)(g_s A_k)\, dw_s^k.$$

Since α is bounded, we are only concerned with the martingale term. By Burkholder–Davis–Gundy inequality, its L^2 norm is bounded by

$$\frac{2}{t}\left(\sum_{k=1}^{N}\int_0^t \mathbb{E}\big((D\alpha)(g_s A_k)\big)^2\, ds\right)^{1/2} \le \frac{2}{t}\left(\int_0^t \mathbb{E}|D\alpha|_{g_s}^2\, ds\right)^{1/2}.$$

By elliptic estimates, $|D\alpha|$ is bounded by $|f|_{L_\infty}$. Since f is centered, it is bounded by $\mathrm{Osc}(f)$. In summary,

$$\mathbb{E}\left(\frac{1}{t}\int_0^t f(g_s)\, ds - \int_N f(g)\, dg\right)^2 \le C\big(\mathrm{Osc}(f) t^{-1/2}\big)^2.$$

In Theorem 1.1, we may wish to add an extra drift of the form $\frac{1}{\varepsilon}A^*$ where $A \in \mathfrak{g}$, so that $\mathcal{L}_G$ is $\frac{1}{2}\Delta^G + L_{gA}$. Translations by orthogonal matrices are isometries, so for any $A \in \mathfrak{g}$ the vector field gA is a killing field, and the Haar measure remains an invariant measure for the diffusion with infinitesimal generator $\frac{1}{2}\Delta^G + L_{gA}$. However, on a compact Lie group no left invariant vector field is the gradient of a function and $\frac{1}{2}\Delta^G + L_g A$ is no longer a symmetric operator. In this case, we do not know how to obtain the estimate in the example.

4. Proof. We are ready to prove the main theorem. In Lemma 3.2, we used a fundamental technique to split the integral

$$\int_{r/\varepsilon}^{t/\varepsilon} (DF)_{\tilde{x}_s^\varepsilon}(H_{x_s^\varepsilon})\big(g_s^\varepsilon e_0\big)\, ds$$

into the sum of a process of finite variation and a martingale. The computation in the proof of Lemma 3.2 will be used to prove the weak convergence. A similar consideration was used in Li [21], which was inspired by a paper of Hairer and Pavliotis [13]. In the above-mentioned papers, the convergence is in probability; while here we can only expect weak convergence. To prove the convergence, we apply Stroock–Varadhan's martingale method and Lemma 3.4; see also Borodin and Freidlin [4]; Papanicolaou, Stroock and Varadhan [25, 26] where the limit is given by a double integration in time. Our formulation for the limit is in terms of space averaging. Finally, we use explicit eigenfunctions of the Laplacian on $SO(n)$ to compute the limiting generator.

PROOF OF THEOREM 1.1. We define a Markov generator $\bar{\mathcal{L}}$ on OM. If $F : OM \to \mathbb{R}$ is bounded and Borel measurable and $\{e_i\}$ is an orthonormal basis of $\mathbb{R}^n$, we define

$$\begin{aligned}\bar{\mathcal{L}}F = &-\sum_{i=1}^n \int_G (\nabla DF)_u\big(H_u(ge_0), H_u(e_i)\big)h_i(g)\,dg \\ &-\sum_{i=1}^n \int_G (DF)_u(H_u e_i) L_{g\bar{A}} h_i(g)\,dg,\end{aligned} \tag{4.1}$$

where h_i is the solution to the Poisson equation (3.3). Since $(\tilde{x}^\varepsilon_{t/\varepsilon})$ is tight by Lemma 3.2, every sub-sequence of $(\tilde{x}^\varepsilon_{t/\varepsilon})$ has a sub-sequence that converges in distribution. We will prove that the probability distributions of $(\tilde{x}^\varepsilon_{t/\varepsilon})$ converge weakly to the probability measure, $\bar{P}$, determined by $\bar{\mathcal{L}}$. It is sufficient to prove that if $(\bar{y}_t)$ is a limit of $(\tilde{x}^\varepsilon_{t/\varepsilon})$, then

$$F(\bar{y}_t) - F(u_0) - \int_0^t \bar{\mathcal{L}}F(\bar{y}_s)\,ds$$

is a martingale. Since the convergence is weak, and the Markov process $(\tilde{x}^\varepsilon_t, g^\varepsilon_{t/\varepsilon})$ is not tight, we do not have a suitable filtration on Ω to work with. We formulate the above convergence on the space of continuous paths over OM on a given time interval $[0, T]$.

Let X_t be the coordinate process on the path space over OM, $\mathcal{G}_t = \sigma\{(X_s) : 0 \le s \le t\}$ and let $\hat{P}_{\tilde{x}^\varepsilon}$ be the probability distribution of $(\tilde{x}^\varepsilon_{t/\varepsilon})$ on the path space over OM. By taking a subsequence if necessary, we may assume that $\{\hat{P}_{\tilde{x}^\varepsilon}\}$ converges to $\bar{P}$.

Let $F : OM \to \mathbb{R}$ be a smooth function with compact support. We will prove that with respect to $\bar{P}$,

$$\mathbb{E}\left\{F(X_t) - F(X_r) - \int_r^t \bar{\mathcal{L}}F(X_s)\,ds \Big| \mathcal{G}_r\right\} = 0.$$

Since $\hat{P}_{\tilde{x}_\varepsilon} \to \bar{P}$ weakly, we only need to prove that for all bounded and continuous real value random variables ξ that are measurable with respect to $\mathcal{G}_r$,

$$\lim_{\varepsilon\to 0} \int \xi\big(F(X_t) - F(X_r)\big)\,d\hat{P}_{\tilde{x}^\varepsilon} = \int \left(\xi \int_r^t \bar{\mathcal{L}}F(X_s)\,ds\right) d\bar{P}. \tag{4.2}$$

By formula (3.4) in the proof of Lemma 3.2, for $t \geq r$,

$$
\begin{aligned}
&F(\tilde{x}^\varepsilon_{t/\varepsilon}) - F(\tilde{x}^\varepsilon_{r/\varepsilon}) \\
&\quad \sim -\varepsilon \sum_{i=1}^n \int_{r/\varepsilon}^{t/\varepsilon} (\nabla DF)_{\tilde{x}^\varepsilon_s}\big(H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0), H_{\tilde{x}^\varepsilon_s}(e_i)\big) h_i(g^\varepsilon_s)\, ds \\
&\qquad - \varepsilon \sum_{i=1}^n \int_{r/\varepsilon}^{t/\varepsilon} (DF)_{\tilde{x}^\varepsilon_s}(H_{\tilde{x}^\varepsilon_s} e_i) L_{g^\varepsilon_s \bar{A}} h_i(g^\varepsilon_s)\, ds \\
&\qquad - \sqrt{\varepsilon} \sum_{i=1}^n \sum_{k=1}^N \int_{r/\varepsilon}^{t/\varepsilon} (DF)_{\tilde{x}^\varepsilon_s}(H_{\tilde{x}^\varepsilon_s} e_i)(Dh_i)_{(g^\varepsilon_s)}(g^\varepsilon_s A_k)\, dw^k_s.
\end{aligned} \tag{4.3}
$$

Hence, up to a term of order ε,

$$
\begin{aligned}
&\int \xi\big(F(X_t) - F(X_r)\big)\, d\hat{P}_{\tilde{x}^\varepsilon} \\
&\quad = O(\varepsilon) - \varepsilon \sum_{i=1}^n \int \bigg(\xi \int_{r/\varepsilon}^{t/\varepsilon} (\nabla DF)_{X_s}\big(H_{X_s}(G_s e_0), H_{X_s}(e_i)\big) h_i(G_s)\, ds\bigg)\, d\hat{P}_{\tilde{x}^\varepsilon} \\
&\qquad - \varepsilon \sum_{i=1}^n \int \bigg(\xi \int_{r/\varepsilon}^{t/\varepsilon} (DF)_{X_s}(H_{X_s} e_i) L_{G_s \bar{A}} h_i(G_s)\, ds\bigg)\, d\hat{P}_{\tilde{x}^\varepsilon}.
\end{aligned}
$$

We prove this by working with the original processes. Let $(\tilde{x}^\varepsilon_t)$ denote a subsequence of the original sequence with limit $(\bar{y}_s)$. For each $i, l = 1, \dots, n$, let us define

$$
\beta_{li}(u) = (\nabla DF)_u\big(H_u(e_l), H_u(e_i)\big).
$$

By linearity of H_u and ∇DF,

$$
\begin{aligned}
&(\nabla DF)_u\big(H_u(ge_0), H_u e_i\big) h_i(g) \\
&\quad = \sum_{l=1}^n (\nabla DF)_u\big(H_u(e_l), H_u(e_i)\big)\langle ge_0, e_l\rangle h_i(g) = \sum_{l=1}^n \beta_{li}(u)\langle ge_0, e_l\rangle h_i(g),
\end{aligned}
$$

for each $i = 1, \dots, n$; and

$$
\begin{aligned}
&-\varepsilon \int_{r/\varepsilon}^{t/\varepsilon} (\nabla DF)_{\tilde{x}^\varepsilon_s}\big(H_{\tilde{x}^\varepsilon_s}(g^\varepsilon_s e_0), H_{\tilde{x}^\varepsilon(s)}(e_i)\big) h_i(g^\varepsilon_s)\, ds \\
&\qquad = -\varepsilon \sum_{l=1}^n \int_{r/\varepsilon}^{t/\varepsilon} \beta_{li}(\tilde{x}^\varepsilon_s)\langle g^\varepsilon_s e_0, e_l\rangle h_i(g^\varepsilon_s)\, ds \\
&\qquad = -\sum_{l=1}^n \int_r^t \beta_{li}(\tilde{x}^\varepsilon_{s/\varepsilon})\langle g^\varepsilon_{s/\varepsilon} e_0, e_l\rangle h_i(g^\varepsilon_{s/\varepsilon})\, ds.
\end{aligned}
$$

We observe that $(g^{\varepsilon}_{s\varepsilon})$ satisfies the equation $dg_t = \sum_k g_t A_k \circ dw^k_t$ with initial value the identity element. The solution stays in the connected component $SO(n)$. It is ergodic with the normalized Haar measure dg on $SO(n)$ as its invariant measure and it satisfies the Birkhoff ergodic theorem; see Example 3.5. By Lemma 3.2, $(\tilde{x}^{\varepsilon}_{s/\varepsilon})$ is tight, and equi-uniformly Hölder continuous on $[0, T]$. In Assumption 3.3, we take $z^{\varepsilon}_t = g^{\varepsilon}_t$, $d\mu_\varepsilon = dg$, $y^{\varepsilon}_t = \tilde{x}^{\varepsilon}_t$ and check that conditions (1)–(4) are satisfied. In Lemma 3.4, we take $G(u, g) = \sum_{l=1}^n \beta_{li}(u)\langle ge_0, e_l\rangle h_i(g)$. Since the functions $h_i : G \to \mathbb{R}$ are smooth and G is compact, also β_{li} are smooth and bounded by construction, we may apply Lemma 3.4. If ϕ is a bounded real valued continuous function on $C([0, r]; OM)$, let $\xi = \phi(\tilde{x}^{\varepsilon}_{u/\varepsilon}, 0 \le u \le r)$. Then

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \mathbb{E}\left(\xi \sum_{l=1}^n \int_r^t \beta_{li}(\tilde{x}^{\varepsilon}_{s/\varepsilon})\langle g^{\varepsilon}_{s/\varepsilon} e_0, e_l\rangle h_i(g^{\varepsilon}_{s/\varepsilon})\, ds\right) \\
&\quad = \sum_{l=1}^n \mathbb{E}\left(\xi \int_r^t \beta_{li}(\bar{y}_s)\, ds\right) \int_G \langle ge_0, e_l\rangle h_i(g)\, dg \\
&\quad = \sum_{l=1}^n \mathbb{E}\left(\xi \int_r^t \nabla DF_{\bar{y}_s}\big(H_{\bar{y}_s}(e_l), H_{\bar{y}_s}(e_i)\big)\right) \int_G \langle ge_0, e_l\rangle h_i(g)\, dg \\
&\quad = \sum_{l=1}^n \mathbb{E}\left(\xi \int_r^t \int_G \nabla DF_{\bar{y}_s}\big(H_{\bar{y}_s}(ge_0), H_{\bar{y}_s}(e_i)\big) h_i(g)\, dg\right).
\end{aligned}$$

By the same reasoning, we also have

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \varepsilon \mathbb{E}\left(\xi \int_{r/\varepsilon}^{t/\varepsilon} (DF)_{\tilde{x}^{\varepsilon}_s}(H_{\tilde{x}^{\varepsilon}_s} e_i) L_{g^{\varepsilon}_s \bar{A}} h_i\big(g^{\varepsilon}_s\big)\, ds\right) \\
&\qquad = \mathbb{E}\left(\xi \int_r^t (DF)_{\bar{y}_s}(H_{\bar{y}_s} e_i)\, ds \int_G L_{g\bar{A}} h_i(g)\, dg\right).
\end{aligned}$$

We have proved (4.2). Since every sub-sequence of $\hat{P}_{\tilde{x}^{\varepsilon}}$ has a sub-sequence that converges to the same limit, we have proved $\hat{P}_{\tilde{x}^{\varepsilon}} \to \bar{P}$ weakly.

Finally, we compute the limiting Markov generator $\bar{\mathcal{L}}$. We observe that there is a family of eigenfunctions of the Laplacian on G with eigenvalue $-\frac{n-1}{2}$. Indeed, since $\sum_{k=1}^{n(n-1)/2} (A_k)^2 = -\frac{n-1}{2} I$,

$$\begin{aligned}
\sum_{k=1}^{n(n-1)/2} L_{gA_k} L_{gA_k}\left(-\frac{4}{n-1}\langle ge_0, e_i\rangle\right) &= -\frac{4}{n-1} \sum_{k=1}^{n(n-1)/2} \big\langle g(A_k)^2 e_0, e_i\big\rangle \\
&= 2\langle ge_0, e_i\rangle.
\end{aligned}$$

Thus,

$$h_i = -\frac{4}{n-1}\langle ge_0, e_i\rangle$$

is the solution to the Poisson equation (3.3):

$$\mathcal{L}_G h_i = \langle g e_0, e_i\rangle \qquad \text{where } \mathcal{L}_G = \frac{1}{2}\sum_{k=1}^{n(n-1)/2} L_{gA_k} L_{gA_k}.$$

We compute the second integral in (4.1). Since $L_{g\bar{A}} h_i = -\frac{4}{n-1}\langle g\bar{A}e_0, e_i\rangle$, we have

$$\begin{aligned}\sum_{i=1}^n \int_G (DF)_u(H_u e_i) L_{g\bar{A}} h_i(g)\,dg &= -\frac{4}{n-1}\int_G (DF)_u(H_u g\bar{A}e_0)\,dg \\ &= -\frac{4}{n-1}(DF)_u\Big(H_u\Big(\int_G g\bar{A}e_0\,dg\Big)\Big) = 0.\end{aligned}$$

Consequently,

$$\begin{aligned}\bar{\mathcal{L}}F &= -\sum_{i=1}^n \int_G (\nabla DF)_u\big(H_u(ge_0), H_u(e_i)\big)h_i(g)\,dg \\ &= -\sum_{i,j=1}^n \int_G (\nabla DF)_u\big(H_u(e_j), H_u(e_i)\big)\langle ge_0, e_j\rangle h_i(g)\,dg.\end{aligned}$$

In the last step, we use the fact that $H_u(\cdot)$ is linear and that $\{e_i\}$ is an o.n.b. of $\mathbb{R}^n$. Let us define

$$\begin{aligned}a_{i,j}(e_0) &= -\int_G \langle ge_0, e_j\rangle h_i(g)\,dg \\ &= \frac{4}{n-1}\int_G \langle ge_0, e_j\rangle\langle ge_0, e_i\rangle\,dg.\end{aligned}$$

Then

$$\bar{\mathcal{L}}F = -\sum_{i,j=1}^n a_{i,j}(\nabla DF)_u\big(H_u(e_j), H_u(e_i)\big). \tag{4.4}$$

To further identify the limit, we first prove that $a_{i,j}(e_0)$ is independent of e_0. Recall that G acts transitively on the unit sphere of $\mathbb{R}^n$. Let $e_0' \in \mathbb{R}^n$ we take O such that $Oe_0' = e_0$. By the right invariant property of the Haar measure,

$$\int_G \langle ge_0', e_j\rangle\langle ge_0', e_i\rangle\,dg = \int_G \langle gOe_0, e_j\rangle\langle gOe_0, e_i\rangle\,dg = \int_G \langle ge_0, e_j\rangle\langle ge_0, e_i\rangle\,dg.$$

We first compute the case of $i \neq j$ and $n = 2$:

$$a_{1,2}(e_1) = \int_{SO(2)} \langle ge_1, e_1\rangle\langle ge_1, e_2\rangle\,dg = -\int_0^{2\pi}\cos(\theta)\sin(\theta)\,d\theta = 0.$$

If $n > 2$, for any $i \neq j$, there is an orientation preserving rotation matrix O such that $Oe_i = -e_i$ and $Oe_j = e_j$. For example, if $i = 1$, $j = 2$, we take $O = (-e_1, e_2, -e_3, e_4, \dots, e_n)$. So

$$\int_G \langle ge_0, e_j\rangle\langle ge_0, e_i\rangle\, dg = -\int_G \langle ge_0, Oe_j\rangle\langle ge_0, Oe_i\rangle\, dg$$
$$= -\int_G \langle ge_0, e_j\rangle\langle ge_0, e_i\rangle\, dg.$$

Thus, $a_{i,j} = 0$ if $i \neq j$. Let

$$C_i = \int_G \langle ge_0, e_i\rangle^2\, dg.$$

For $i = 1, \dots, n$, $C_i = \int_G \langle ge_0, e_i\rangle^2\, dg$ is independent of i and

$$\int_G \sum_{i=1}^n \langle ge_0, e_i\rangle^2\, dg = 1$$

and consequently $C_i = \frac{1}{n}$. The nonzero values of $(a_{i,j})$ are

$$a_{i,i} = -\int_G \langle ge_0, e_i\rangle h_i(g)\, dg = \frac{4}{n-1}\int_G \langle ge_0, e_i\rangle^2\, dg = \frac{4}{(n-1)n}.$$

By the definition, $\Delta_H F(u) = \sum_{i=1}^n L_{H(e_i)} L_{H(e_i)} F$. Since ∇ is the canonical flat connection, $\nabla_{H(e_i)} H(e_i) = 0$. See the paragraph before equation (3.4). By (4.4), we see that

$$\bar{\mathcal{L}} F(u) = -\sum_{i,j=1}^n a_{i,j} (\nabla DF)_u\big(H_u(e_j), H_u(e_i)\big)$$
$$= \frac{4}{(n-1)n}\sum_{i=1}^n (\nabla DF)_u\big(H_u(e_i), H_u(e_i)\big)$$
$$= \frac{4}{(n-1)n}\Delta_H F(u).$$

We conclude that $(\tilde{x}^\varepsilon_{t/\varepsilon})$ is a diffusion process with infinitesimal generator $\frac{4}{(n-1)n}\Delta_H$. Since $(x^\varepsilon_{t/\varepsilon})$ is the projection of $(\tilde{x}^\varepsilon_{t/\varepsilon})$ it is also convergent. The operators Δ_H and Δ are intertwined by π; for $f : M \to \mathbb{R}$ smooth, $(\Delta_H f) \circ \pi = \Delta(f \circ \pi)$. See, for example, Theorem 4C of Chapter II in Elworthy [6] and also Elworthy, Le Jan and Li [7]; Δ_H is cohesive and a horizontal operator in the terminology of [7] and is the horizontal lift of Δ. We see that $(x^\varepsilon_{t/\varepsilon})$ converges to a process with generator $\frac{4}{(n-1)n}\Delta$ where Δ is the Laplacian on the Riemannian manifold M. We have completed the proof of Theorem 1.1. □

Acknowledgements. It is a pleasure to thank D. Bakry, K. D. Elworthy, M. Hairer, M. Ledoux, Y. Maeda, J. Norris and S. Rosenberg for helpful discussions. I would also like to thank the referees for helpful comments.

REFERENCES

[1] BENSOUSSAN, A., LIONS, J.-L. and PAPANICOLAOU, G. (2011). *Asymptotic Analysis for Periodic Structures*. Amer. Math. Soc., Providence, RI. MR2839402

[2] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

[3] BISMUT, J.-M. (2011). *Hypoelliptic Laplacian and Orbital Integrals*. *Annals of Mathematics Studies* **177**. Princeton Univ. Press, Princeton, NJ. MR2828080

[4] BORODIN, A. N. and FREIDLIN, M. I. (1995). Fast oscillating random perturbations of dynamical systems with conservation laws. *Ann. Inst. Henri Poincaré Probab. Stat.* **31** 485–525. MR1338450

[5] DOWELL, R. M. (1980). Differentiable approximations to Brownian motion on manifolds. Ph.D. thesis, Univ. Warwick.

[6] ELWORTHY, D. (1988). Geometric aspects of diffusions on manifolds. In *École d'Été de Probabilités de Saint-Flour XV–XVII*, 1985–1987. *Lecture Notes in Math.* **1362** 277–425. Springer, Berlin. MR0983375

[7] ELWORTHY, K. D., LE JAN, Y. and LI, X.-M. (2010). *The Geometry of Filtering*. Birkhäuser, Basel. MR2731662

[8] ENRIQUEZ, N., FRANCHI, J. and LE JAN, Y. (2001). Central limit theorem for the geodesic flow associated with a Kleinian group, case $\delta > d/2$. *J. Math. Pures Appl.* (9) **80** 153–175. MR1815697

[9] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York. MR0838085

[10] FRANCHI, J. and LE JAN, Y. (2007). Relativistic diffusions and Schwarzschild geometry. *Comm. Pure Appl. Math.* **60** 187–251. MR2275328

[11] FREIDLIN, M. I. and WENTZELL, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* **260**. Springer, New York. MR0722136

[12] GUIVARC'H, Y. and LE JAN, Y. (1993). Asymptotic winding of the geodesic flow on modular surfaces and continued fractions. *Ann. Sci. École Norm. Sup.* (4) **26** 23–50. MR1209912

[13] HAIRER, M. and PAVLIOTIS, G. A. (2004). Periodic homogenization for hypoelliptic diffusions. *J. Stat. Phys.* **117** 261–279. MR2098568

[14] HAS'MINSKIĬ, R. Z. (1968). On the principle of averaging the Itô's stochastic differential equations. *Kybernetika* (*Prague*) **4** 260–279. MR0260052

[15] HELLAND, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Stat.* **9** 79–94. MR0668684

[16] IKEDA, N. (1987). Limit theorems for a class of random currents. In *Probabilistic Methods in Mathematical Physics* (*Katata/Kyoto*, 1985) 181–193. Academic Press, Boston, MA. MR0933823

[17] IKEDA, N. and OCHI, Y. (1986). Central limit theorems and random currents. In *Stochastic Differential Systems* (*Bad Honnef*, 1985). *Lecture Notes in Control and Inform. Sci.* **78** 195–205. Springer, Berlin. MR0903044

[18] KHAS'MINSKII, R. Z. (1963). The behavior of a self-oscillating system acted upon by slight noise. *J. Appl. Math. Mech.* **27** 1035–1044. MR0162032

[19] KIPNIS, C. and VARADHAN, S. R. S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19. MR0834478

[20] LI, X.-M. (1994). Properties at infinity of diffusion semigroups and stochastic flows via weak uniform covers. *Potential Anal*. **3** 339–357. MR1302415

[21] LI, X.-M. (2008). An averaging principle for a completely integrable stochastic Hamiltonian system. *Nonlinearity* **21** 803–822. MR2399826

[22] LI, X.-M. (2012). Effective diffusions with intertwined structures. Preprint.

[23] MANABE, S. and OCHI, Y. (1989). The central limit theorem for current-valued processes induced by geodesic flows. *Osaka J. Math*. **26** 191–205. MR0991288

[24] NELSON, E. (1967). *Dynamical Theories of Brownian Motion*. Princeton Univ. Press, Princeton, NJ. MR0214150

[25] PAPANICOLAOU, G. C., STROOCK, D. and VARADHAN, S. R. S. (1977). Martingale approach to some limit theorems. In *Papers from the Duke Turbulence Conference* (*Duke Univ*., *Durham*, *N.C*., 1976), *Paper No*. 6. Duke Univ., Durham, NC. MR0461684

[26] PAPANICOLAOU, G. C. and VARADHAN, S. R. S. (1973). A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations. *Comm*. *Pure Appl*. *Math*. **26** 497–524. MR0383530

[27] PINSKY, M. A. (1981). Homogenization and stochastic parallel displacement. In *Stochastic Integrals* (*Proc*. *Sympos*., *Univ. Durham*, *Durham*, 1980). *Lecture Notes in Math*. **851** 271–284. Springer, Berlin. MR0620994

[28] RATNER, M. (1973). The central limit theorem for geodesic flows on n-dimensional manifolds of negative curvature. *Israel J. Math*. **16** 181–197. MR0333121

[29] RUFFINO, P. R. and GONZALES GARGATE, I. (2012). An averaging principle for diffusions in foliated spaces. Available at http://arxiv.org/abs/1212.1587.

[30] SINAĬ, JA. G. (1960). The central limit theorem for geodesic flows on manifolds of constant negative curvature. *Soviet Math*. *Dokl*. **1** 983–987. MR0125607

MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY CV4 7AL
UNITED KINGDOM
E-MAIL: xue-mei.li@warwick.ac.uk