

BROWNIAN PROCESSES FOR MONTE CARLO INTEGRATION ON COMPACT LIE GROUPS

S. SAID,* *The University of Melbourne*

J. H. MANTON,** *The University of Melbourne*

Abstract

This paper proposes a Monte Carlo approach for the evaluation of integrals of smooth functions defined on compact Lie groups. The approach is based on the ergodic property of Brownian processes in compact Lie groups. The paper provides an elementary proof of this property and obtains the following results. It gives the rate of almost sure convergence of time averages along with a “large deviations” type upper bound and a central limit theorem. It derives probability of error bounds for uniform approximation of the paths of Brownian processes using two numerical schemes. Finally, it describes generalisation to compact Riemannian manifolds.

Keywords: Monte Carlo; Compact Lie group; Brownian process; Ergodic property

2010 Mathematics Subject Classification: Primary 82C80
Secondary 60J60

1. Introduction

1.1. Exact and approximate algorithms

The problem treated in the present paper is to evaluate efficiently the integral, with respect to the Haar measure, of a smooth function f defined on a given compact Lie group G . Integrals of this kind arise in the study of random matrices, harmonic analysis and Bayesian statistics. Clearly, these lead to a wide variety of applications in engineering and physics.

It is well known the Haar measure μ of the group G is a probability measure on G which is invariant under all group operations [1, 2]. Let Y be a random variable with values in G . When the distribution of Y is given by μ , it is usual to say that Y is uniformly distributed in G . In fact, for all $g \in G$, the random variables Y , gY and Yg all have the same distribution. A first example is when $G = S^1$, the unit circle group of complex numbers z with modulus $|z| = 1$. In this case, it is enough to let $Y = e^{i\theta}$ where θ is uniformly distributed in $(0, \pi)$.

In general, the statement that Y is uniformly distributed can also be given a geometric sense [3]. A compact Lie group G can always be equipped with a Riemannian metric invariant under group operations. To this metric, there corresponds a Riemannian volume form which is also invariant under group operations. The resulting volume measure, when normalised so that the total volume of G is equal to 1, is the same as μ .

* Postal address: Department of Electrical and Electronic Engineering, Victoria 3010 Australia

** Postal address: Department of Electrical and Electronic Engineering, Victoria 3010 Australia

Thus, Y has a uniform, *i.e.* constant, probability density with respect to Riemannian volume.

With some knowledge of compact Lie groups, it is clear from the above that μ should have a tractable local expression whenever the structure of G is sufficiently well known. However, analytic calculation or direct numerical evaluation based on such an expression is bound to run into two kinds of difficulties. First, the variety of functions for which analytic calculation is possible is restricted. Second, numerical evaluation suffers from the curse of dimensionality.

The current problem is closely related to the problem of generating independent samples with values in G and distributed according to μ . Indeed, if it possible to obtain such samples, then the integral of f with respect to μ can be evaluated using the strong law of large numbers. The latter problem has an interesting history and has generated exciting results [4].

A compact Lie group typically encountered in applications is the special orthogonal group $SO(n)$. This is the group of real orthogonal $n \times n$ matrices with unit determinant. The discussion of this group exemplifies similar results obtained for other classical compact Lie groups. Methods for generating independent samples from the Haar measure of $SO(n)$ can be divided into exact and approximate methods. Exact methods are based on the following observation [4]. Let M be a real $n \times n$ matrix whose elements are independent and have a standard Gaussian distribution. Let $M = QR$ be the Gram-Schmidt decompositions of M , where Q is the orthogonal factor. After a “change of sign” to ensure that Q has unit determinant, it can be shown that Q has its values in $SO(n)$ and is distributed according to the Haar measure. Now, Q can be computed using Householder transforms or Givens rotations and this leads to streamlined algorithms.

Approximate methods have relied on random walks taking place on $SO(n)$ [5, 6]. Given *i.i.d* random variables $(Z_n)_{n \geq 1}$ with values in $SO(n)$, a random walk process is defined by $S_n = S_{n-1}Z_n$ where Z_0 has a fixed value in $SO(n)$. Everything hinges on the distribution of the Z_n . This should allow easy generation of individual steps Z_1, Z_2, \dots and also guarantee a mixing property so that the distribution of S_n quickly approaches a uniform distribution. To obtain independent samples from the Haar measure, it is then possible to start with independent samples from a well chosen initial distribution and have these samples follow (independent) random walks.

Drawbacks of the methods just discussed consist in that they may require generating a significantly large amount of random numbers and in the complexity of carrying out consecutive matrix products. Unfortunately, a detailed count of operations and a comparison to the approach proposed in the current paper are too complex to be attempted in the following.

For the evaluation of the integral with respect to μ of a smooth function f defined on G , generation of independent samples is an overkill. It is clear that a Monte Carlo approach can be based on some ergodic process W with values in G such that time averages taken along individual paths of W converge almost surely to the required integral. Of course, any Markov process W which has a unique invariant probability μ and which is, in a general sense, irreducible realises this ergodic property [7]. Here, the process W will be chosen as a Brownian process in G . This has two advantages. First, convergence of time averages to integrals can be treated rather directly and proven to satisfy explicit bounds. Second, although they are continuous time processes, Brownian

processes are easy to approximate using numerical schemes with known probability of error bounds. As a result, it becomes possible to design a Monte Carlo simulation which yields the integral of f to within any chosen precision. For the two aspects just mentioned, see Sections 3 and 4. The fact of generating only one instead of a set of independent paths is seen as a motivation for the approach based on Brownian processes.

Given the interpretation of the Haar measure μ on a compact group G as a Riemannian volume measure, it may be asked whether the current approach based on Brownian processes may be generalised to arbitrary compact Riemannian manifolds. The answer is positive and this generalisation is described in Section 5. Basically, the ergodic property of Brownian processes only depends on their local behavior, so that the global group structure can be ignored.

Usually, Monte Carlo methods are introduced to avoid high dimensional integrals which suffer from curse of dimensionality. Functions defined on a compact Lie group G with high dimension may have specific symmetries which allow an analytic or numerical approach. For instance, functions which are central, *i.e.* conjugate invariant, can be integrated using Weyl's integration formula [3, 8]. Concretely, a central function f defined on $SO(n)$ is a function which "only depends on eigenvalues" of its argument. The dimension of $SO(n)$ is $n(n-1)/2$ but Weyl's integral formula transforms the integral of a continuous central function to a Riemann integral in n variables. Another simplification holds for functions f on a classical (matrix) compact Lie group which are polynomials in the matrix elements of their argument [9]. The integrals of these functions can be computed analytically, by decomposing corresponding tensor representations of G into irreducible representations.

The current approach is essentially the same as that of [13]. Indeed, both propose to use time averages, taken along approximate paths of Brownian processes, in order to evaluate integrals. Note that [13] focuses on convergence in square mean while the following studies explicit bounds for almost sure convergence. Also, the approximation of (theoretical) paths of Brownian processes using numerical schemes is here considered in the stronger sense of local uniform approximation in probability. In either case, results can be given a full expression in terms of almost sure weak convergence of empirical measures, see [21].

1.2. Brownian processes in compact Lie groups

The current paragraph gives some basic facts concerning Brownian processes and discusses conditions under which these processes are ergodic. A more complete introduction may be found in [12]. The use of irreducible representations and of the Peter-Weyl theorem follows [10].

First, some notation needs to be specified. Let G be a compact connected Lie group of dimension d with identity e , Lie algebra \mathfrak{g} and normalised Haar measure μ . Let $\text{Irr}(G)$ denote the set of equivalence classes of irreducible representations of G . With δ_0 denoting the unit representation, let $\text{Irr}_+(G) = \text{Irr}(G) - \{\delta_0\}$. For $\delta \in \text{Irr}(G)$ let d_δ be its dimension and $U^\delta : G \rightarrow SU(d_\delta)$ be a smooth unitary representative. The normalised matrix elements $\sqrt{d_\delta}U_{ab}^\delta$, for $\delta \in \text{Irr}(G)$ and $1 \leq a, b \leq d_\delta$, form a complete orthonormal system in $L^2(\mu, \mathbb{C})$. Moreover, recall the Peter-Weyl theorem.

For continuous $f : G \rightarrow \mathbb{C}$, this states the uniform convergence

$$f(g) = \bar{f} + \sum_{\delta \in \text{Irr}_+(G)} d_\delta \text{tr}[f(\delta)U^\delta(g)] \quad (1)$$

where $\bar{f} = \int_G f d\mu$. The following inversion formula holds

$$f(\delta) = \int_G f U^{\delta^\dagger} d\mu \quad (2)$$

where integration of a matrix function is carried out elementwise and \dagger indicates the Hermitian transpose. For a probability measure ν on G , its characteristic function ϕ_ν is the sequence of matrices

$$\phi_\nu(\delta) = \int_G U^\delta d\nu \quad \delta \in \text{Irr}(G) \quad (3)$$

Clearly, $\phi_\nu(\delta_0) = 1$. A probability measure on G is uniquely determined by its characteristic function. In particular, $\phi_\mu(\delta) = 0$ for $\delta \in \text{Irr}_+(G)$.

In the following, a Brownian process W in G is the solution of a left invariant stochastic differential equation, driven by a Brownian motion in \mathfrak{g} . To write this down, let $X_1, \dots, X_q \in \mathfrak{g}$ where $q \leq d$. Let B be a Brownian motion in \mathbb{R}^q with covariance matrix $C = (C_{ij}, 1 \leq i, j \leq q)$. The components of B will be denoted B^i for $1 \leq i \leq q$. The stochastic differential equation satisfied by W is the following Stratonovich equation

$$dW_t = \sum_{i=1}^q X_i^L(W_t) \circ dB_t^i \quad W_0 = g \quad (4)$$

The initial condition is deterministic, $g \in G$. By changing to Itô form, it follows for all smooth (in fact C^2) function h on G ,

$$dh(W_t) = \sum_{i=1}^q [X_i^L h](W_t) dB_t^i + [\mathcal{D}h](W_t) dt \quad (5)$$

The differential operator \mathcal{D} is the generator of the process W .

$$\mathcal{D}h = \frac{1}{2} \sum_{i,j=1}^q C_{ij} X_i^L X_j^L h \quad (6)$$

The Brownian process W will be called ergodic if, for any initial condition g ,

$$\frac{1}{T} \int_0^T f(W_t) dt \xrightarrow{a.s.} \bar{f} \quad (7)$$

where the notation \bar{f} is as in (1).

It will be shown below, see Proposition 1, that μ is an invariant probability of W . It is known, from the theory of Harris processes [7], that the following condition **(E1)** implies (7).

(E1) Let $P_t(g, \cdot)$ be the probability distribution of W_t under initial condition g . For $t > 0$, $P_t(g, \cdot)$ has density $p_t(g, \cdot)$ with respect to μ and $p_t(g, k)$ is strictly positive and continuous in (t, g, k) .

In the following, an equivalent but more specific condition is considered, which also implies that W is ergodic. Condition **(E2)** below is a local Hörmander condition equivalent to condition **(E1)**, see [10, 11].

(E2) C is strictly positive definite and $\text{Lie}\{X_1, \dots, X_q\} = \mathfrak{g}$. Here $\text{Lie}\{\dots\}$ denotes the generated Lie algebra.

This condition will be discussed in relation to Proposition 1. For now, consider another important fact for the following.

Let f be a smooth function on G , consider the potential equation

$$\mathcal{D}h = f - \bar{f} \quad (8)$$

If condition **(E2)** is verified, then this has a smooth solution h , for all smooth f . In fact, if **(E2)** holds then \mathcal{D} is elliptic self adjoint. In particular, if f is smooth and h is a solution of (8) then h is smooth. The construction of a solution h can be carried out using a spectral representation of \mathcal{D} . At least formally, this can be deduced from the general construction in [7]. The function h is said to be a recurrent potential of f . Under the assumption $\bar{h} = 0$, which is made throughout the following, it is the unique smooth solution of (8).

Proposition 1 gives a first taste of the relation between conditions **(E1)** and **(E2)**. Recall a probability measure ν on G is called an invariant probability of W if for all smooth function h on G

$$\int_G [\mathcal{D}h] d\nu = 0 \quad (9)$$

For $\delta \in \text{Irr}_+(G)$, let $D_\delta = [\mathcal{D}U^\delta](e)$.

Proposition 1. Condition **(E2)** is equivalent to either of the following (i) D_δ is (Hermitian) strictly negative definite, for $\delta \in \text{Irr}_+(G)$ and (ii) μ is the unique invariant probability of W .

A formal proof would be too long. Rather, a few remarks are made which indicate a standard approach. First, it is shown that (i) and (ii) are equivalent. Note, by the Peter-Weyl theorem, that when f is smooth \mathcal{D} can be applied to (1) term by term. A short calculation then shows that (9) holds for all smooth h iff it holds for $h = U_{ab}^\delta$. Recall,

$$\mathcal{D}U^\delta = U^\delta D_\delta \quad (10)$$

It follows that (9) holds iff $\phi_\nu(\delta)D_\delta = 0$ for $\delta \in \text{Irr}_+(G)$. Now, (ii) is equivalent to $\phi_\nu(\delta) = 0$ being the unique solution of this equation. This clearly follows from (i), so that (i) implies (ii).

To see that (ii) implies (i), note that (ii) implies the statement that $\mathcal{D}h = 0$ only if h is a constant. Indeed, if $h \neq \bar{h}$ and $\mathcal{D}h = 0$ then the smooth function $p_h = \exp(-|h|^2)$ can be normalised to obtain an invariant probability absolutely continuous with respect

to μ . But then μ is not the unique invariant probability of W . Let $\langle h, f \rangle = \int_G h f^* d\mu$ where $*$ denotes conjugation. Using integration by parts and the fact that, $\int_G X f d\mu = 0$ for smooth f , it can be shown

$$\langle h, \mathcal{D}h \rangle = -\frac{1}{2} \sum_{i,j=1}^q C_{ij} \langle X_i^L h, X_j^L h \rangle \quad (11)$$

Replacing h by an arbitrary linear combination of U_{ab}^δ , it follows that (ii) implies D_δ is Hermitian strictly negative definite. Incidentally, even without assuming (ii), the equality (11) implies D_δ is Hermitian negative definite.

To complete the proof, it is enough to show condition **(E2)** is equivalent to (i). For this, the key observation is that **(E2)** is equivalent to the same statement, $\mathcal{D}h = 0$ *only if* h is a constant. Assume **(E2)** holds and h verifies $\mathcal{D}h = 0$. Then, (11) implies $X_i^L h = 0$ for $1 \leq i \leq q$, so that h is a constant. Conversely, if condition **(E2)** does not hold, then working in normal coordinates near the identity e , it is possible to construct a smooth function $h \neq \bar{h}$ and $\mathcal{D}h = 0$ — compare to Page 11 of [10]. The fact that **(E2)** implies (i) now follows by a repetition of the above arguments showing that (ii) implies (i). On the other hand, since the U_{ab}^δ form a complete orthonormal system in $L^2(\mu, \mathbb{C})$ it is straightforward to use (11) in order to show

$$\langle h, \mathcal{D}h \rangle = \sum_{\delta \in \text{Irr}_+(G)} d_\delta \text{tr}[h(\delta)^\dagger D_\delta h(\delta)] \quad (12)$$

Now, if (i) holds then this is strictly negative unless $h = \bar{h}$, so that $\mathcal{D}h = 0$ implies $h = \bar{h}$.

2. The ergodic property

The ergodic property of Brownian processes can be stated from general results. However, it is interesting to note a direct and elementary proof is possible. This is mainly based on the spectral properties of Brownian processes. Proposition 2 shows (7) holds when f is a matrix element U_{ab}^δ . The desired statement is then Proposition 3, which follows from the Peter-Weyl theorem.

Things are greatly simplified by the Lévy property of Brownian processes. A Brownian process W as in (4) is a left Lévy process. Precisely, the following two properties hold.

(L1) For $0 \leq s \leq t$ the left increment $W_t^{-1}W_s$ is independent of past values W_u of W , where $u \leq s$.

(L2) The distribution of left increments is stationary, $W_t^{-1}W_s \stackrel{d}{=} W_{t-s}$; here, $\stackrel{d}{=}$ denotes equality in distribution.

In preparation for Proposition 2 note that W has the following covariance structure. Recall (10); by applying (5) to each U_{ab}^δ and taking expectations, it is found

$$\mathbb{E}_g[U^\delta(W_t)] = U^\delta(g) \exp[tD_\delta]$$

where \mathbb{E}_g indicates the initial condition g and \exp is the matrix exponential. Now using property **(L1)** and the fact that U^δ has unitary values

$$\mathbb{E}_g \text{tr}[U^\delta(W_s)U^\delta(W_t)^\dagger] = \text{tr} \{ \exp[(t-s)D_\delta] \} \quad (13)$$

The right hand side will be denoted as the value $R_\delta(|t-s|)$ of the covariance function R_δ . This is independent of the initial condition g , but of course depends on $\delta \in \text{Irr}_+(G)$.

Proposition 2. *If condition (E2) is verified, then for $\delta \in \text{Irr}_+(G)$ and any initial condition g in (4)*

$$\frac{1}{T} \int_0^T U^\delta(W_t) dt \xrightarrow{a.s.} 0 \quad (14)$$

Proof. Trivially, (14) holds for δ_0 ; assume $\delta \in \text{Irr}_+(G)$, arbitrarily chosen. For $T \geq 0$, let \hat{U}_T^δ be the time average, on the left hand side of (14). Let $|\cdot|$ denote the Euclidean matrix norm. By Fubini's theorem,

$$\mathbb{E}_g \left| \hat{U}_T^\delta \right|^2 = \frac{1}{T^2} \int_0^T \int_0^T R_\delta(|t_1 - t_2|) dt_1 dt_2$$

for any initial condition g .

If condition (E2) is verified then D_δ is strictly negative definite, by Proposition 1. Then, if $-\lambda_\delta < 0$ be the largest eigenvalue of D_δ , $R_\delta(|t_1 - t_2|) \leq d_\delta \exp[-\lambda_\delta |t_1 - t_2|]$. Carrying out the previous double integral,

$$\mathbb{E}_g \left| \hat{U}_T^\delta \right|^2 \leq \frac{4d_\delta}{\lambda_\delta T} \quad (15)$$

The following is a standard argument.

Let $W_n = n^2$. By (15), $\mathbb{E}_g |\hat{U}_{W_n}^\delta|^2 = O(n^{-2})$. Thus, from the Borel-Cantelli lemma,

$$\hat{U}_{W_n}^\delta \xrightarrow{a.s.} 0 \quad (16)$$

Clearly, $W_n \uparrow \infty$. For $W_n \leq T \leq W_{n+1}$, by a direct calculation

$$\left| \hat{U}_T^\delta - \hat{U}_{W_n}^\delta \right| \leq \frac{W_{n+1} - W_n}{W_n^2} \int_0^{W_n} |U^\delta(W_t)| dt + \frac{1}{W_n} \int_{W_n}^{W_{n+1}} |U^\delta(W_t)| dt$$

The right hand side does not depend on T . Replacing $|U^\delta(W_t)| = \sqrt{d_\delta}$, it follows by an immediate simplification

$$\sup_{W_n \leq T \leq W_{n+1}} \left| \hat{U}_T^\delta - \hat{U}_{W_n}^\delta \right| \xrightarrow{a.s.} 0 \quad (17)$$

The proposition holds by (16) and (17).

Proposition 3. *If condition (E2) is verified, then for any initial condition g in (4), (7) holds.*

Proof. The proof follows from Proposition 2 and the Peter-Weyl theorem. The following holds for any initial condition g .

Let f be a continuous function on G . Replacing f by $f - \bar{f}$, note that the uniformly convergent series (1) only involves $\delta \in \text{Irr}_+(G)$.

Let Λ be a finite subset of $\text{Irr}_+(G)$ and f_Λ the corresponding partial sum in (1). For $T > 0$,

$$\frac{1}{T} \int_0^T f(W_t) dt = \frac{1}{T} \int_0^T f_\Lambda(W_t) dt + \frac{1}{T} \int_0^T [f(W_t) - f_\Lambda(W_t)] dt$$

Independently of Λ , Proposition 2 implies the first term on the right hand side converges to zero almost surely. In absolute value, the second term is less than $\|f - f_\Lambda\|_\infty$, the supremum norm of $f - f_\Lambda$. By choosing Λ large enough, always finite, the Peter-Weyl theorem implies this is arbitrarily small.

3. Almost sure convergence and large deviations

3.1. Rates of almost sure convergence

In its above form, Proposition 3 gives no statement as to the rate of convergence in (7). Here, the problem of determining these rates is studied. In view of practical application, it is interesting to speak separately of “bounds” and “rates”. Even when the rate of convergence in (7) is shown to be quite fast, the corresponding bound may depend on constants which are unknown or suboptimal. It will be argued in 6 that the bound obtained here is reasonably easy to determine and thus adapted to applications.

Here, Proposition 4 gives the rate of almost sure convergence in (7), showing that it is faster than $T^{-\alpha}$ for arbitrary $0 < \alpha < 1/2$. Proposition 5 gives an exponentially decreasing upper bound for large deviation probabilities in (7). For a given function f , the rate in the exponential is given in terms of the corresponding smooth potential h of (8).

For Propositions 4 and 5, consider the Itô equation (5) where h is the smooth potential corresponding to f . If condition **(E2)** is verified, it follows from (8),

$$\int_0^T [f(W_t) - \bar{f}] dt = h(W_T) - h(g) - \sum_{i=1}^q \int_0^T [X_i^L h](W_t) dB_t^i \quad (18)$$

where g in the initial condition in (4). Note that the sum of Itô integrals on the right hand side is a continuous square integrable martingale; this will be denoted M .

Proposition 4. *Assume condition **(E2)** holds. Let \hat{f}_T be the left hand side of (7). For $0 < \alpha < 1/2$ and any initial condition g in (4),*

$$T^\alpha [\hat{f}_T - \bar{f}] \xrightarrow{a.s.} 0 \quad (19)$$

Proof. From (7) and (18), note that the left hand side of (19) is

$$T^{-\beta} [h(W_t) - h(g) - M_T] \quad \beta = 1 - \alpha > 1/2$$

Since h is smooth, it is bounded. So, (19) will follow if it can be shown $T^{-\beta} M_T \xrightarrow{a.s.} 0$. This is done in a similar way to the proof of Proposition 2. The following reasoning does not depend on the initial condition g .

From (18), the quadratic variation $[M]$ of M verifies $[M]_T \leq KT$ where $K \geq 0$ will be discussed later on. Let $\hat{M}_T = T^{-\beta} M_T$ and $W_n = n^\gamma$ where $\gamma(2\beta - 1) > 1$. Note

$$\mathbb{E}_g \left| \hat{M}_{W_n} \right|^2 = W_n^{-2\beta} \mathbb{E}_g [M]_{W_n} \leq \frac{K}{n^{\gamma(2\beta-1)}}$$

By the Borel-Cantelli lemma

$$\hat{M}_{W_n} \xrightarrow{a.s.} 0 \quad (20)$$

Note $W_n \uparrow \infty$. Similar to the proof of Proposition 2, it is possible to write

$$\mathbb{E}_g \sup_{W_n \leq T \leq W_{n+1}} |\hat{M}_T - \hat{M}_{W_n}|^2 \leq 2 \frac{(W_{n+1}^\beta - W_n^\beta)^2}{W_n^{4\beta}} \mathbb{E}_g |M_{W_n}|^2 + \frac{2}{W_n^{2\beta}} \mathbb{E}_g \sup_{W_n \leq T \leq W_{n+1}} |M_T - M_{W_n}|^2$$

By $[M]_T \leq KT$ and Doob's martingale inequality, it then follows

$$\mathbb{E}_g \sup_{W_n \leq T \leq W_{n+1}} |\hat{M}_T - \hat{M}_{W_n}|^2 \leq 2K \frac{(W_{n+1}^\beta - W_n^\beta)^2}{W_n^{4\beta-1}} + 8K \frac{W_{n+1} - W_n}{W_n^{2\beta}}$$

which is $O(n^{-2})$. This shows the Borel-Cantelli lemma applies again. That is,

$$\sup_{W_n \leq T \leq W_{n+1}} |\hat{M}_T - \hat{M}_{W_n}| \xrightarrow{a.s.} 0 \quad (21)$$

Now, (20) and (21) imply (19).

In preparation for Proposition 5, consider again the quadratic variation $[M]$. From (18), it follows

$$[M]_T = \sum_{i,j=1}^q C_{ij} \int_0^T [X_i^L h](W_t) [X_j^L h](W_t) dt \equiv \int_0^T \tilde{h}(W_t) dt \quad (22)$$

It becomes clear that, as stated in the proof of Proposition 4, there exists $K \geq 0$ such that $[M]_T \leq KT$. For instance, $K = \|\tilde{h}\|_\infty$ is a possible value of K . Moreover, it is clear K can be chosen independent of the initial condition g .

In view of the proposition, an important question concerns the optimal choice of $K \geq 0$.

Proposition 5. *Assume condition (E2) holds. Let $K \geq 0$ be such that $[M]_T \leq KT$. For $\epsilon > 0$ and $T \geq 4\|h\|_\infty/\epsilon$,*

$$\mathbb{P}_g \left(\left| \hat{f}_T - \bar{f} \right| > \epsilon \right) \leq 2 \exp \left(\frac{-\epsilon^2 T}{8K} \right) \quad (23)$$

where \mathbb{P}_g is the law of W with initial condition g .

Proof. The starting point is (18). If $T \geq 4\|h\|_\infty/\epsilon$, then $T^{-1}|h(W_T) - h(g)| \leq \epsilon/2$. Thus, to evaluate the left hand side of (23), it is enough to consider $\mathbb{P}_g(|M_T| \geq T\epsilon/2)$.

Let $\lambda \geq 0$ and Y be the stochastic exponential $Y_T = \exp[\lambda M_T - \lambda^2 [M]_T/2]$. Because M is a Itô integral, it can be shown that Y is a square integrable martingale. This is not detailed here, but follows from the usual stochastic differential equation $dY_T = \lambda Y_T dM_T$.

Applying Proposition 3 to the function \tilde{h} of (22), it follows

$$T^{-1}[M]_T \xrightarrow{a.s.} -\langle h, f \rangle \quad (24)$$

which follows from (27) below. Assume f is not constant — the case of constant f is trivial. Then, since the right hand side of (24) is strictly positive, $[M]_T \uparrow \infty$ almost

surely as $T \uparrow \infty$. In particular, for $A \geq 0$ the stopping time $\tau_A = \inf\{T \geq 0, [M]_T \geq A\}$ is almost surely finite. In particular, it holds that $[M]_{\tau_A} = A$ almost surely.

Let B be the standard Brownian motion process $B_A = M_{\tau_A}$. Then,

$$\mathbb{E}_g[Y_{\tau_A}] = \mathbb{E}_g\{\exp[\lambda B_A - \lambda^2 A/2]\} = 1 \quad (25)$$

Recall $[M]_T \leq KT$ so that $\tau_A \geq A/K$. In particular, $\tau_{KT} \geq T$. Using the fact that $\exp(M_T)$ is a submartingale,

$$\exp[\lambda M_T] \mathbb{P}_g(M_T \geq T\epsilon/2) \leq \mathbb{E}_g\{\exp[\lambda M_T]\} \leq \mathbb{E}_g\{\exp[\lambda M_{\tau_{KT}}]\}$$

Replacing from (25) it follows,

$$\mathbb{P}_g(M_T \geq T\epsilon/2) \leq \exp\left[\frac{\lambda^2}{2}KT - \lambda T\frac{\epsilon}{2}\right]$$

Putting $\lambda = \epsilon/2K$ and performing a similar reasoning for $-M_T$, it follows

$$\mathbb{P}_g(|M_T| \geq T\epsilon/2) \leq 2 \exp\left(\frac{-\epsilon^2 T}{8K}\right) \quad (26)$$

so that the proposition follows.

3.2. Central limit theorem for time averages

In view of Proposition 4, the question may be asked of whether almost sure convergence also holds for $\alpha = 1/2$. Unsurprisingly, the answer is negative on account of the existence of a nondegenerate limit distribution. Precisely, this is a Gaussian distribution whose mean is zero and whose variance is $-\langle h, f \rangle$. From Potential theory, this quantity is the Dirichlet form associated to \mathcal{D} .

The central limit theorem can be stated and proved quite directly. Again, it is enough to consider (18). The following result will be needed. Let \tilde{h} be as in (22). Then,

$$\int_G \tilde{h} d\mu = -\langle h, f \rangle \quad (27)$$

This is just a rewriting of (11).

Proposition 6. *In the setting of Proposition 4, for any initial condition g in (4),*

$$\mathcal{L}\left\{T^{\frac{1}{2}}\left[\hat{f}_T - \bar{f}\right]\right\} \Rightarrow \mathcal{N}(0, -\langle h, f \rangle) \quad (28)$$

where $\mathcal{L}\{\cdot\}$ is the probability distribution of the quantity at hand and \Rightarrow denotes weak convergences.

Proof. Referring to (18), the left hand side of (28) is the law of

$$T^{-\frac{1}{2}}[h(W_T) - h(g) - M_T]$$

Since h is bounded, it is enough to prove $\mathcal{L}\{T^{-\frac{1}{2}}M_T\}$ converges to the required Gaussian distribution.

Let $\ell_T = T^{-1}[M]_T$ and $\sigma^2 = -(h, f)$ and recall from Proposition 3,

$$\ell_T \xrightarrow{a.s.} \sigma^2 \quad (29)$$

For $A \geq 0$ let M^A be the square integrable martingale $M_T^A = A^{-\frac{1}{2}} M_T$. This has quadratic variation $[M^A]_T = A^{-1}[M]_T$.

For $A \geq 0$, there exists a standard Brownian motion B^A such that $M^A = B^A \circ [M^A]$. By definition,

$$T^{-\frac{1}{2}} M_T = M_T^T = B_{\ell_T}^T \quad (30)$$

Of course, $\mathcal{L}\{B_{\sigma^2}^T\} = \mathcal{N}(0, \sigma^2)$ for all $T \geq 0$. From (30), it is seen the proposition will follow by showing

$$B_{\ell_T}^T \xrightarrow{\mathbb{P}_g} B_{\sigma^2}^T \text{ as } T \uparrow \infty \quad (31)$$

Here $\xrightarrow{\mathbb{P}_g}$ indicates convergence in probability, where the initial condition is g .

Fixing T , let $\tau_1 = \ell_T \wedge \sigma^2$ and $\tau_2 = \ell_T \vee \sigma^2$, where \wedge and \vee denote minimum and maximum. It is clear from (29) that $|\tau_2 - \tau_1| \xrightarrow{a.s.} 0$. Thus, it is possible to assume that τ_1, τ_2 are bounded. It follows, by optional sampling,

$$\mathbb{E}_g |B_{\ell_T}^T - B_{\sigma^2}^T|^2 = \mathbb{E} |B_{\tau_1}^T - B_{\tau_2}^T|^2 = \mathbb{E}_g [\tau_2 - \tau_1]$$

This converges to zero as $T \uparrow \infty$, so that (31) follows from (29).

Further application of Proposition 6 may be based on Esseen's inequality, see for instance [14]. In the notation of the proof, let F_T, φ_T and F, φ be the cumulative distribution function and characteristic function of, respectively, $B_{\ell_T}^T$ and $B_{\sigma^2}^T$. Esseen's inequality states for $u \geq 0$,

$$\sup_x |F_T(x) - F(x)| \leq \frac{2}{\pi} \int_0^u \left| \frac{\varphi_T(v) - \varphi(v)}{v} \right| dv + \frac{24}{\pi u} \frac{1}{\sqrt{2\pi\sigma^2}} \quad (32)$$

An elementary estimate for the integrand in the first term is

$$|\varphi_T(v) - \varphi(v)| \leq |v| \left(\mathbb{E}_g |B_{\ell_T}^T - B_{\sigma^2}^T|^2 \right)^{\frac{1}{2}} \leq |v| \left(\mathbb{E}_g |\ell_T - \sigma^2|^2 \right)^{\frac{1}{4}}$$

The quantity in parentheses can be identified as,

$$\Sigma_T^2 \equiv \mathbb{E}_g \left| \frac{1}{T} \int_0^T \tilde{h}(W_t) dt - \int_G \tilde{h} d\mu \right|^2$$

An obvious choice of u finally yields

$$\sup_x |F_T(x) - F(x)| \leq \left[\frac{2}{\pi} + \frac{24}{\pi\sqrt{2\pi\sigma^2}} \right] \Sigma_T^{\frac{1}{4}} \quad (33)$$

Note that, assuming σ^2 is known *a priori*, the values of F can be found in usual statistical tables. Thus, bounds on the deviation of \hat{f}_T away from \bar{f} can be evaluated from the study of convergence in the square mean, *i.e.* of Σ_T . This gives practical value to inequality (33), as convergence in the square mean may be relatively straightforward to establish — for example, as in the proof of Proposition 2.

4. The issue of approximation

Recall the main problem of evaluating \bar{f} for a smooth function f defined on G . Proposition 3 asserts this can be done using (7). Moreover, if sufficient knowledge of f is available, Proposition 5 can be used to find T such that \hat{f}_T is within a precision ϵ of \bar{f} , with any preassigned probability. Once such T is found, there remains the issue of computing \hat{f}_T . The main difficulty here is in obtaining a valid approximation of a path of W , over the interval $[0, T]$. Since \hat{f}_T is given by a Riemann integral, it is clear a uniform approximation over the required interval allows \hat{f} to be computed with known precision. With this motivation, the current section concentrates on local uniform approximation of the paths of W . Its main results are Propositions 7 and 8 which provide explicit probability of error bounds.

Concretely, it will be assumed G is a matrix Lie group. Precisely, the assumption is that for some $v \geq 1$, a Lie group isomorphism $U : G \rightarrow GL(v)$ is given and $g \in G$ is identified with $U(g) \in GL(v)$. Note that the existence of an isomorphism U is guaranteed whatever the underlying group G — see [8], Theorem 4.1 on Page 136. A sequence of processes $(W^n)_{n \geq 1}$ with values in $GL(v)$ is said to converge locally uniformly in probability to W if for all $T \geq 0$ and $\epsilon \geq 0$,

$$\mathbb{P} \left(\sup_{t \leq T} |W_t^n - W_t| > \epsilon \right) \rightarrow 0 \quad (34)$$

where all processes are defined on the same probability space with probability measure \mathbb{P} — the corresponding expectation is denoted \mathbb{E} . The processes W^n used to approximate W follow from numerical schemes for solving the stochastic differential equation (4). With the current assumption that G is a matrix Lie group, this can be written in the following Itô form

$$dW_t = W_t [d\tilde{B}_t + \tilde{D}dt] \equiv W_t d\tilde{X}_t \quad (35)$$

Here, with the notation $\tilde{X}_i = X_i U$, \tilde{B} is the matrix process $\tilde{B}_t = \sum_{i=1}^q B_t^i \tilde{X}_i$, \tilde{D} is the matrix $\tilde{D} = \frac{1}{2} \sum_{i,j=1}^q C_{ij} \tilde{X}_i \tilde{X}_j$ and \tilde{X} is the matrix process $\tilde{X}_t = \tilde{B}_t + \tilde{D}t$.

4.1. The first order approximation

In the current paragraph, the processes W^n are constructed from a first order Euler scheme for equation (35). The convergence of this kind of scheme is well known, in the general context of locally Lipschitz stochastic differential equations [15]. Proposition 7 below gives a direct treatment of the current matrix case. It is clear from their construction, see equation (36), that the processes W^n do not have their values in G . For many applications, this should present no difficulty, as the function f is in the first place a matrix function; that is, the restriction to G of a known smooth function $F : GL(v) \rightarrow \mathbb{R}$. Thus, \hat{f}_T is approximated by a similar Riemann integral, $T^{-1} \int_0^T F(W_t^n) dt$.

Throughout the following, let $(t_k^n)_{k \geq 0}$ form a decreasing sequence of subdivisions of \mathbb{R}_+ . For $n \geq 1$, the subdivision points $0 = t_0^n < t_k^n \uparrow \infty$ and $|t_n| = \sup_{k \geq 1} |t_k^n - t_{k-1}^n| \downarrow \infty$. The processes W^n are here given by

$$W_t^n = \prod_{k \geq 1} I_v + \tilde{X}_{t_k^n \wedge t} - \tilde{X}_{t_{k-1}^n \wedge t} \quad (36)$$

where \wedge denotes the minimum, I_v is the $v \times v$ identity matrix and the product is ordered with k increasing from left to right. Without loss of generality, it is assumed $W_0 = I_v$.

Proposition 7. *Let W be the solution of (35) and $(W^n)_{n \geq 1}$ be given by (36). For $T \geq 0$ and $n \geq 1$*

$$\mathbb{E} \left(\sup_{t \leq T} |W_t^n - W_t|^2 \right) \leq [2ve^{4\rho(T)} \mathbb{E} |\tilde{X}_1|^2] |t_n| \quad (37)$$

where ρ is the increasing function $t \mapsto 12v^2 \mathbb{E} |\tilde{X}_1|^2 (t + \frac{t^3}{3})$. In particular, $(W^n)_{n \geq 1}$ converges locally uniformly in probability to W .

Proof. Only a sketch of the main steps is given. Let V^n be the process $V_t^n = W_{t_k^n}^n$ for $t_k^n \leq t < t_{k+1}^n$. It is possible to show

$$W_t^n = I_v + \int_0^t V_{s-}^n d\tilde{X}_s \quad (38)$$

where V_{s-}^n denotes the left limit. Comparing to (35) gives,

$$W_t - W_t^n = \int_0^t (W_{s-} - V_{s-}^n) d\tilde{X}_s + \int_0^t (W_{s-} - W_{s-}^n) d\tilde{X}_s \quad (39)$$

In order to bound the first term, note $W_t^n - V_t^n = V_t^n (\tilde{X}_t - \tilde{X}_{t_k^n})$ for $t_k^n \leq t < t_{k+1}^n$. Now, $\mathbb{E} |\tilde{X}_t - \tilde{X}_{t_k^n}|^2 = \mathbb{E} |\tilde{X}_1|^2 |t_n|$. Also, it can be shown from (38) using standard techniques from stochastic calculus,

$$\mathbb{E} |V_t^n|^2 \leq 2v + 2 \int_0^t \mathbb{E} |V_s^n|^2 d\rho(s) \quad (40)$$

so that it is possible to apply Gronwall's lemma, which gives

$$\mathbb{E} |V_t^n|^2 \leq [2ve^{2\rho(t)} \mathbb{E} |\tilde{X}_1|^2] |t_n|$$

Now, $\mathbb{E} |W_t^n - V_t^n|^2$ can be bounded using the fact that \tilde{X} has independent increments. Returning to (39) and applying a similar transformation, it follows

$$\mathbb{E} |D_t^n|^2 \leq [2ve^{2\rho(T)} \mathbb{E} |\tilde{X}_1|^2] |t_n| + 2 \int_0^T \mathbb{E} |D_t^n|^2 d\rho(t) \quad (41)$$

where $D_t^n = \sup_{t \leq T} |W_t^n - W_t|$. The proposition follows from Gronwall's lemma applied to (41).

4.2. The exponential approximation

In the current paragraph, the processes W^n are constructed using so called multiplicative integrals, involving the matrix exponential. This construction, unlike the previous one in 4.1, has an intrinsic sense independent of the isomorphism U . In particular, the processes W^n have their values in G so that it is possible to consider directly any smooth function f defined on G . The convergence of multiplicative integral

schemes is known for diffusion processes in general Riemannian manifolds [16]. The specific case of compact Riemannian manifolds is treated in [17]. This gives to the order of magnitude stated in Proposition 8. Explicit bounds for the current case can be collected from the proof, which is modelled on that of [17].

Given a decreasing sequence of subdivisions $(t_k^n)_{k \geq 0}$ of \mathbb{R}_+ , let

$$b_k^n = (\tilde{B}_{t_k^n} - \tilde{B}_{t_{k-1}^n}) / (t_k^n - t_{k-1}^n)$$

and consider the processes W^n given by

$$W_t^n = \prod_{k \geq 1} \exp [b_k^n (t_k^n \wedge t - t_{k-1}^n \wedge t)] \quad (42)$$

where the product is ordered as in (36). By construction of the matrix process \tilde{B} and the fact that G is a matrix Lie group, it is clear W^n has its values in G .

Proposition 8. *Let W be the solution of (35) and $(W^n)_{n \geq 1}$ be given by (42). For $T \geq 0$ and $n \geq 1$*

$$\mathbb{P} \left(\sup_{t \leq T} |W_t^n - W_t| > \epsilon \right) = O(|t_n|) \quad (43)$$

In particular, $(W^n)_{n \geq 1}$ converges locally uniformly in probability to W .

Proof. For simplicity, consider only regular subdivisions, $t_k^n = k|t_n|$. Calculations will be carried out with a fixed subdivision $(t_k^n)_{k \geq 1}$. Accordingly, there is no confusion in writing t_k instead of t_k^n . Fix $T \geq 0$ and let $k_0 = \lfloor T/|t_n| \rfloor + 1$. Let \bar{W} and \bar{W}^n be the processes where $\bar{W}_t = W_{t_k}$ and $\bar{W}_t^n = W_{t_k}^n$ for $t_k \leq t < t_{k+1}$. Consider the following triangle inequality

$$|W_t - W_t^n| \leq |W_t - \bar{W}_t| + |\bar{W}_t - \bar{W}_t^n| + |W_t^n - \bar{W}_t^n| \quad (44)$$

For the third term, it is possible to use (42) in order to obtain

$$\sup_{t \leq T} |W_t^n - \bar{W}_t^n| \leq K_1 \max_{k \leq k_0} |\tilde{B}_{t_k} - \tilde{B}_{t_{k-1}}| \quad (45)$$

where $K_1 = \sup_{g \in G} |g|$. Following [17], this can be evaluated from

$$\mathbb{P} \left(\max_{k \leq k_0} |\tilde{B}_{t_k} - \tilde{B}_{t_{k-1}}| > \epsilon \right) \leq \sum_{k=1}^{k_0} \mathbb{P} \left(|\tilde{B}_{t_k} - \tilde{B}_{t_{k-1}}| > \epsilon \right)$$

where $\epsilon > 0$ is arbitrary. Note moreover that,

$$\mathbb{P} \left(|\tilde{B}_{t_k} - \tilde{B}_{t_{k-1}}| > \epsilon \right) \leq \mathbb{P} \left(\sup_{s \leq |t_n|} |\tilde{B}_{t_{k-1}+s} - \tilde{B}_{t_{k-1}}| > \epsilon \right)$$

and by elementary properties of Brownian motion, this is dominated by

$$\sqrt{\frac{2v^2 K_2^2 \epsilon^2}{\pi |t_n|}} \exp \left(-\frac{K_2^2 \epsilon^2}{2v^2 |t_n|} \right) \equiv R(\epsilon, |t_n|)$$

where $K_2^2 = v^4 q^2 \sum_{i,j=1}^q C_{ij} \text{tr}[\tilde{X}_i \tilde{X}_j^T]$. Replacing in (45) shows that

$$\mathbb{P}\left(\sup_{t \leq T} |W_t^n - \bar{W}_t^n| > \epsilon\right) \leq \frac{T}{|t_n|} R\left(\frac{\epsilon}{K_1}, |t_n|\right) \quad (46)$$

which is $O(|t_n|)$.

For the first term in (44), it is possible to write from (35)

$$W_t - \bar{W}_t = \int_{t_{k-1}}^t [W_s - \bar{W}_s] d\tilde{X}_s + \bar{W}_{t_{k-1}} [\tilde{X}_t - \tilde{X}_{t_{k-1}}]$$

for $t_{k-1} \leq t < t_k$. This is dominated thanks to the two following bounds. First, using the same technique as in the proof of Proposition 7

$$\mathbb{E} \max_{k \leq k_0} \sup_{t_{k-1} \leq t < t_k} \left| \int_{t_{k-1}}^{t \wedge T} [W_s - \bar{W}_s] d\tilde{X}_s \right|^2 \leq \int_0^T \mathbb{E} |W_s - \bar{W}_s|^2 d\rho(s) = K_1^2 \rho(T) \rho(|t_n|)$$

Using the same property as for (46), it is possible to find

$$\mathbb{P}\left(\max_{k \leq k_0} \sup_{t_{k-1} \leq t < t_k} |\tilde{X}_t - \tilde{X}_{t_{k-1}}| > \epsilon\right) \leq \frac{T}{|t_n|} R\left(\frac{\epsilon}{2K_1}, |t_n|\right)$$

whenever $K_1 |\tilde{D}| |t_n| \leq \epsilon/2$. Combining the last two bounds, it follows

$$\mathbb{P}\left(\sup_{t \leq T} |W_t - \bar{W}_t| > \epsilon\right) = O(|t_n|) \quad (47)$$

It remains to consider the second term in (44). Using the Taylor development of the matrix exponential in (42), it follows for $1 \leq k < k_0$,

$$W_{t_k}^n - I_v = \sum_{l=1}^k W_{t_{l-1}}^n \left[\Delta_l \tilde{B} + \frac{1}{2} \Delta_l \tilde{B}^2 \right] + A_k \quad (48)$$

where $\Delta_l \tilde{B} = \tilde{B}_{t_l} - \tilde{B}_{t_{l-1}}$ and $|A_k| \leq K_1 \sum_{l=1}^k |\Delta_l \tilde{B}|^3$.

The difference $W_{t_k} - I_v$ can be expressed from (35). Subtracting from (48), it is possible to check

$$W_{t_k} - W_{t_k}^n = A_k + S_k + J_k + T_k \quad (49)$$

where

$$S_k = \int_0^{t_k} [\bar{W}_s - W_s] d\tilde{X}_s \quad J_k = \int_0^{t_k} [\bar{W}_s^n - \bar{W}_s] d\tilde{X}_s$$

and, using the matrix integration by parts formula of [18],

$$T_k = \sum_{l=1}^k W_{t_{l-1}}^n \left[\frac{1}{2} \Delta_l B^2 - \tilde{D} |t_n| \right] = \sum_{l=1}^k \bar{W}_{t_{l-1}}^n \int_0^{|t_n|} \tilde{B}_s^l d\tilde{B}_s^l + \int_0^{|t_n|} d\tilde{B}_s^l B_s^l$$

where \tilde{B}^l is the process $\tilde{B}_t^l = \tilde{B}_{t_{l-1}+t} - \tilde{B}_{t_{l-1}}$. Now, it is clear that A_k can be bounded as follows,

$$\mathbb{E} \max_{k < k_0} |A_k|^2 \leq K_1 \frac{T}{|t_n|} \sum_{l=1}^{k_0} \mathbb{E} |\Delta_l \tilde{B}|^6 = K_1 T^2 \mathbb{E} |\tilde{B}_1|^6 |t_n|$$

For the remaining terms, note that

$$\mathbb{E} \max_{k < k_0} |S_k|^2 \leq \mathbb{E} \sup_{t \leq T} \left| \int_0^t [\bar{W}_s - W_s] d\tilde{X}_s \right|^2 \leq K_1^2 \rho(T) \rho(|t_n|)$$

$$\mathbb{E} \max_{k < k_0} |J_k|^2 \leq \mathbb{E} \sup_{t \leq T} \left| \int_0^t [\bar{W}_s^n - \bar{W}_s] d\tilde{X}_s \right|^2 \leq \int_0^T \mathbb{E} \sup_{s \leq t} |\bar{W}_s^n - \bar{W}_s|^2 \rho(t)$$

and similarly,

$$\mathbb{E} \max_{k < k_0} |T_k|^2 \leq 2K_1^2 T \mathbb{E} |\tilde{B}_1|^4 |t_n|$$

By adding the bounds obtained for A_k, S_k, J_k and T_k , it follows

$$\mathbb{E} \sup_{t \leq T} |\bar{W}_t^n - \bar{W}_t|^2 \leq O(|t_n|) + \int_0^T \mathbb{E} \sup_{s \leq t} |\bar{W}_s^n - \bar{W}_s|^2 \rho(t)$$

This allows a final conclusion, after using Gronwall's lemma,

$$\mathbb{P} \left(\sup_{t \leq T} |\bar{W}_t^n - \bar{W}_t| > \epsilon \right) = O(|t_n|) \quad (50)$$

The proposition follows from (46), (47) and (50).

5. Generalisation to compact Riemannian manifolds

The current section discusses the generalisation of previous results to arbitrary compact Riemannian manifolds. Let S be a compact connected and oriented Riemannian manifold and denote μ its Riemannian volume measure. For a smooth function f defined on S , the aim is to evaluate $\bar{f} = \int_M f d\mu$. Generalising the previous approach, consider W to be a Brownian motion process corresponding to the Riemannian structure of S . Proposition 9 below will show the same property as (7) holds for the Brownian motion process W and is controlled by a similar bound to that of Proposition 5. Importantly, the approximation of Proposition 8 also generalises immediately to the current Riemannian setting. Knowledge of the Riemannian structure of S , precisely knowledge of its exponential mapping, is then sufficient to carry out practically the evaluation of \bar{f} .

Brownian processes on a compact Lie group G , as introduced in 1.2, have a straightforward relationship to Brownian motion in G [10]. Assume given an Ad-invariant scalar product $\langle \cdot, \cdot \rangle$ on \mathfrak{g} . Such a scalar product can be constructed without knowledge of the Haar measure of G , after decomposition of \mathfrak{g} into a direct sum of a semisimple and Abelian ideals [8]. If X_1, \dots, X_d is an orthonormal basis of \mathfrak{g} , with respect to $\langle \cdot, \cdot \rangle$, then the process W given by the stochastic differential equation (4), with $q = d$, is precisely a Brownian motion process corresponding to the Riemannian structure of

G given by $\langle \cdot, \cdot \rangle$. In particular, the operator \mathcal{D} of (6) is the Laplace operator for this Riemannian structure.

Let Δ be the Laplace operator corresponding to the Riemannian structure of S . A Brownian motion process W in S is a diffusion process whose generator is Δ . That is [16], W is a pathwise continuous Markov process with values in S such that for all smooth function h defined on S and $T \geq 0$,

$$h(W_T) - h(W_0) = \int_0^T [\Delta h](W_t) dt + M_t \quad (51)$$

where M is a continuous square integrable martingale depending on the function h . For the current aim, the initial condition is taken to be deterministic, $W_0 = o \in S$. The existence of Brownian motion processes in S follows from the fact that Δ is the generator of the heat semigroup of S [19].

Proposition 9 will mainly be obtained from (51). This will require modifying (51) into a form similar to (18). Note first that for all smooth function f defined on S the Potential equation

$$\Delta h = f - \bar{f} \quad (52)$$

has a unique smooth solution under the condition that $\bar{h} = 0$. Again, this can be constructed as a recurrent potential following [7]. Second, an explicit expression of the Laplace operator Δ in the form of a “sum of squares of vector fields” as in (6) can be obtained as follows [20].

Let $O(S)$ be the orthonormal frame bundle of S . As a set, this has elements (s, u) where $s \in S$ and $u : \mathbb{R}^d \rightarrow T_s S$ is an isometry between the Euclidean space \mathbb{R}^d and the tangent space $T_s S$ of S at s . Note that $O(S)$ has a natural manifold structure making it locally diffeomorphic to $S \times O(d)$. Let e_i , where $1 \leq i \leq d$, be an orthonormal basis of \mathbb{R}^d . For $(s, u) \in O(S)$, consider the orthonormal vectors $ue_i \in T_s S$. If γ_i is a differentiable curve with $\gamma_i(0) = s$ and $\gamma_i'(0) = ue_i$, then the horizontal lift of γ_i is the curve $\tilde{\gamma}_i(t) = (\gamma(t), u(t))$ in $O(S)$ where for $1 \leq j \leq d$ the vector $u(t)e_j \in T_{\gamma(t)} S$ is the parallel transport of $ue_j \in T_s S$ along γ_i . This is well defined since parallel transport is understood with respect to the Riemannian connection of S . The lift of the vector ue_i tangent to γ_i at $t = 0$ is the vector H_i tangent to $\tilde{\gamma}_i$ at $t = 0$. This gives vector fields H_1, \dots, H_d on $O(S)$ which are linear independent at each point.

The Laplace operator Δ is expressed in the following way. Let $\pi : O(S) \rightarrow S$ be the projection $(s, u) \rightarrow s$. For a smooth function h defined on S , the following relation holds,

$$\Delta h = \sum_{i=1}^d H_i H_i (h \circ \pi) \quad (53)$$

In view of (53), an explicit representation of a Brownian motion process in S can be obtained. Let \tilde{W} be a process in $O(S)$ which satisfies

$$d\tilde{W}_t = \sum_{i=1}^d H_i(\tilde{W}_t) \circ dB_t^i \quad (54)$$

where $B = (B^1, \dots, B^d)$ is a standard Brownian motion in \mathbb{R}^d . By changing into Itô form and taking into account (53), it follows that $W = \pi(\tilde{W})$ verifies (51) so that it

is a Brownian motion process in S . It is now possible to state and prove Proposition 9. As in previous sections, for a function f defined on S , the notation \hat{f}_T stands for $T^{-1} \int_0^T f(W_t) dt$. Also $\|h\|_\infty$ denotes the supremum norm.

Proposition 9. *Let h and f be as in (52). There exists $K \geq 0$ such that, in the notation of (51), $[M]_T \leq KT$. Moreover, for $\epsilon > 0$ and $T \geq 4\|h\|_\infty/\epsilon$,*

$$\mathbb{P}\left(\left|\hat{f}_T - \bar{f}\right| > \epsilon\right) \leq 2 \exp\left(\frac{-\epsilon^2 T}{8K}\right) \quad (55)$$

where \mathbb{P} is the law of W . In particular, $\hat{f}_T \xrightarrow{a.s.} \bar{f}$.

Proof. The affirmation $\hat{f}_T \xrightarrow{a.s.} \bar{f}$ follows from (55) by an application of the Borel-Cantelli lemma. The existence of K follows from

$$M_T = \sum_{i=1}^d \int_0^T [H_i(h \circ \pi)](\tilde{W}_t) dB_t^i$$

which implies

$$[M]_T = \sum_{i=1}^d \int_0^T [H_i(h \circ \pi)]^2(\tilde{W}_t) dt$$

Indeed, since S is compact, $O(S)$ is compact and the integrand is a bounded.

The proof of (55) can be carried out exactly like the proof of (23) in Proposition 5. From (51) it follows

$$\hat{f}_T - \bar{f} = T^{-1}[h(W_T) - h(o) - M_T] \quad (56)$$

Indeed, if $T \geq 4\|h\|_\infty/\epsilon$ then $T^{-1}|h(W_T) - h(o)| \leq \epsilon/2$. Thus, it is enough to consider $\mathbb{P}_g(|M_T| \geq T\epsilon/2)$. This can be treated exactly like (23), using exponential submartingales.

The key step in generalising the approximation of 4.2 to the current Riemannian setting is the replacement of the Lie group exponential by the Riemannian exponential [16]. Recall that for $s \in S$ and $\xi \in T_s S$ there exists a unique geodesic curve $c_\xi : \mathbb{R} \rightarrow S$ with $c_\xi(0) = s$ and $c'_\xi(0) = \xi$. The exponential map $\exp_s : T_s S \rightarrow S$ associates to ξ the point $c_\xi(1)$ and more generally to $t\xi$ the point $c_\xi(t)$, where the geodesic curve c_ξ has been affinely parameterised.

Let $(t_k^n)_{k \geq 0}$ form a decreasing sequence of subdivisions of \mathbb{R}_+ . The processes W^n given by the following construction converge locally uniformly in probability to W ; note here, local uniform convergence uses the distance function corresponding to the Riemannian metric. Dropping the superscript n , let $b_k = (B_{t_k} - B_{t_{k-1}})/(t_k - t_{k-1})$. On each interval $t_{k-1} < t \leq t_k$, let \tilde{W}^n be the horizontal lift of W^n and write $\tilde{W}_t^n = (W_t^n, U_t^n)$. The processes W^n are given by,

$$W_t^n = \exp_{W_{t_{k-1}}^n} \left[(t - t_{k-1}) U_{t_{k-1}}^n b_k \right] \quad t_{k-1} < t \leq t_k \quad (57)$$

This formula has the following explanation. Fix an initial condition $(o, u) \in O(S)$. For $0 < t \leq t_1$ let W_t^n be the geodesic curve $c(t) = \exp_o(tub_1)$; here, $b_1 \in \mathbb{R}^d$ and $ub_1 \in T_o S$. Then, U_t^n is such that $U_t^n e_i$ is the parallel transport of ue_i along W^n . At $t = t_1$, this construction is repeated using $(W_{t_1}^n, U_{t_1}^n)$ instead of (o, u) and b_2 instead of b_1 , etc.

6. Discussion

This final discussion will be concerned with the practical applicability of the results of previous sections, as well as with some additional problems raised by these results and which were not pursued in the above.

The practical motivation for the results of Sections 3 and 4 is their application to the design of a Monte Carlo simulation which yields the integral \bar{f} of a given smooth function f defined on G . The objective, in this case, would be to minimise the number of required operations while respecting a preassigned precision on \bar{f} . For a given simulation, the precision on \bar{f} is controlled by Proposition 5, while the number of operations (or equivalently, the required step size) can be controlled by Proposition 7.

It is clear from (5) that the simulation is completely determined by the covariance matrix C and $X_1, \dots, X_q \in \mathfrak{g}$. For simplicity of argument, it is possible to assume $C = \sigma^2 I_d$, where I_d is the $d \times d$ identity matrix, and that σ^2 completely determines the simulation. The design can be realised by adjusting σ^2 to minimise a cost function inferred from Propositions 5 and 7. In fact, increasing σ^2 improves the precision in Proposition 5 but requires a smaller step size in Proposition 7.

Application of Proposition 5 requires knowledge of the involved constant K , in (23). In general, it is too difficult to evaluate K from $K = \|\tilde{h}\|_\infty$ given after (22). Assume on the other hand that f is of the elementary form,

$$f = \Re \operatorname{tr} [A^T U^\delta] \quad (58)$$

where A is a $d_\delta \times d_\delta$ complex matrix and \Re denotes the real part. A direct calculation shows it is possible to take

$$K = -d|A|\lambda_*^{-1} \quad (59)$$

where $\lambda_* < 0$ is the largest nonzero eigenvalue of \mathcal{D} , *i.e.* its spectral gap. This is well defined as can be seen from the spectral representation of \mathcal{D} .

Just like in the stationary case, where the initial distribution of W is given by μ , it follows from (59) that there exists a uniform rate of convergence over $\delta \in \operatorname{Irr}(G)$. For a general smooth function f , formula (59) allows K to be evaluated based on the rate of uniform convergence in (1). This can be characterised from estimates of the derivatives of f .

It has been claimed that Proposition 5 provides a large deviations type upper bound. This is misleading for two reasons. First, large deviations results are asymptotic while (23) holds for finite T . Second, large deviations provides an optimal rate of convergence, which is different from the bound in (23). This is described in [22], but it seems there is no straightforward way of computing it.

The results obtained in the current paper indicate several interesting problems for further study. The approximation of 4.2 is closely connected to the exponential random walk on G . This is similar to, but more complicated than, Kac's random walk considered in [6] and [13]. Studying the asymptotic properties of this walk using harmonic analysis, transportation costs, or some other method which does not use the diffusion limit indicated by 4.2, is interesting at least from a technical point of view.

The convergence to Riemannian volume measure considered in Section 5 was only briefly treated. A very interesting problem would be to study this convergence using the analytic method of [13].

References

- [1] GRENANDER, U. (1963). *Probabilities on algebraic structures*, Dover Editions.
- [2] DIACONIS, P. (1988). *Group representations in probability and statistics*, Institute of Mathematical Statistics.
- [3] HELGASSON, S. (1962). *Differential geometry and symmetric spaces*, Academic press.
- [4] DIACONIS, P. AND SHAHSHAHANI, M. (1987). The subgroup algorithm for generating uniform random variables. *Probability in the Engineering and Information Sciences* **1**, 15–32.
- [5] POROD, U. (1996). The cut-off phenomenon for random reflections. *Ann. Prob.* **24**, 74–96.
- [6] OLIVEIRA, R. I. (2009). On the convergence to equilibrium of Kac’s random walk on matrices. *The Annals of Applied Probability* **19**, 1200–1231.
- [7] KUNITA, H. (1990). *Stochastic flows and stochastic differential equations*, Cambridge University Press.
- [8] BRÖCKER, T. AND TOM DIECK, T. (1985). *Representations of compact Lie groups*, Springer.
- [9] COLLINS, B. AND SNIADY, P. (2006). Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Commun. Math. Phys.* **264**, 773–795.
- [10] LIAO, M. (2004). *Lévy processes on Lie groups*, Cambridge University Press.
- [11] AZENCOTT, R. (1981). Géodésiques et diffusions en temps petits. *Astérisque* **84-85**.
- [12] APPLEBAUM, D. (2001). Lévy processes in stochastic differential geometry. *Lévy processes. Theory and applications*, Editors Barndorff-Nielsen, O.E. and Mikosch, T. and Resnick, S.I., Birkhäuser.
- [13] PICCIONI, M. AND SCARLATTI, S. (1994). An iterative Monte Carlo scheme for generating Lie group valued random variables. *Adv. Appl. Prob.* **26**, 616–628.
- [14] Shirayev, A.N. (1995). *Probability*, 2nd Edition, Springer (Graduate Texts in Mathematics).
- [15] Kloeden, P.E. and Platen, E. (1995). *Numerical solution of stochastic differential equation*, Springer.
- [16] Elworthy, J.M. (1982). *Stochastic differential equations on manifolds*, Cambridge University Press.
- [17] PONTIER, M. AND SZPIRGLAS, J. (1987). Convergence des approximations de McShane d’une diffusion sur une variété compacte. *Séminaire de Probabilités (Strasbourg)* **21**, 534–543.
- [18] KARANDIKAR, R.L. (1982). Multiplicative decomposition of nonsingular matrix-valued continuous semimartingales. *Ann. Prob.* **10**, 1088–1091.
- [19] Rosebner, S. (1997). *The Laplacian on a Riemannian manifold*, Cambridge University Press.
- [20] Emery, M. (1989). *Stochastic calculus in manifolds*, Springer (Universitext).
- [21] LAMBERTON, D. AND PAGÈS, G. (2002). Recursive computation of the invariant distribution of a diffusion. *Bernoulli* **8**, 367–405.
- [22] Deuschel, J.D. and Stroock D.W. (1984). *Large deviations*, Academic Press.