2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Stratonovich, Ito and Numerical Analysis on Lie Groups

Victor Solo

***Abstract*— The description of stochastic mechanical systems naturally leads to Stratonovich stochastic differential equations on Lie groups. But calculating statistical properties is naturally done using Ito stochastic differential equations. All this is well known. But numerical implementations raise new issues. In particular, we show that the common practice of constructing numerical schemes for Stratonovich stochastic differential equations in Lie groups by following deterministic numerical schemes is in general flawed. Such numerical schemes do not in general converge to the correct limits. In a particular case, we give conditions showing when they do and when they don't.**

I. Introduction

Stochastic differential equations (SDEs) on manifolds and Lie Groups have a long history in mathematics [1]. And the engineering literature is growing due to e.g. applications in mechanics [2] molecular dynamics [3] and robotics [4].

The literature on numerical schemes for orinary differential equations (ODEs) evolving on Lie groups is well established, including a classic monograph [5]. The book also describes numerical schemes on Riemannian manifolds but the development there is less mature. However the situation with numerical schemes for SDEs on manifolds remains rudimentary. There is a mature development for SDEs in Euclidean spaces [6] but it remains to be fully generalised to Lie groups and manifolds.

In some earlier work [7] we pointed out the problem with some numerical schemes in the robotics literature [8],[9]. But while we gave examples, and an heuristic argument, we did not spell out any theory explaining what goes wrong.

Here we tackle that problem of finding what SDE a particular numerical scheme converges to as the step size tends to 0. This is a challenging problem and we restrict attention to multiplicative SDEs, which however arise commonly in applications [2],[4]. It turns out problems arise when the diffusion coefficients in the multiplicative SDE depend on the state. Our previous work on convergence [10],[11] did not cover that case; while [12] does, it deals with a very special case and a very different numerical scheme to that discussed here.

Here we discuss a particular first order SDE scheme from [7] and show that it agrees with an ODE scheme when the diffusion coefficients are constant matrices. A convergence analysis of the SDE scheme then shows that when the diffusion coefficients are state dependent the SDE scheme converges to the correct SDE but the ODE scheme does not.

This work was funded by an ARC Discovery Project.

V Solo is with School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, AUSTRALIA. v.solo@unsw.edu.au

Notation: SDE = stochastic differential equation; iid = independent identically distributed. We say $f(\delta)$ is $O(\delta)$ if $f(\delta)/\delta \rightarrow$ const. as $\delta \rightarrow 0$. We say $f(\delta)$ is $o(\delta)$ if $f(\delta)/\delta \rightarrow 0$ as $\delta \rightarrow 0$. Matrix Hilbert-Schmidt norm: $\| A \| = \sqrt{\sum_{kl}|A_{kl}|^2}$. Prefixes: I- Ito; S- Stratonovich.

In the sequel, 'Theorem' refers to an existing result, whereas 'Result' denotes a new one.

The remainder of the paper is organised as follows. In section II we review some features of Stratonovich and Ito SDEs in manifolds and Lie groups. In section III we introduce two numerical schemes. In section IV are the main results. We analyse convergence of one of the schemes and argue that the other only converges in special cases. Conclusions are in section V. There is one appendix containing the longer proofs.

II. Review of Ito and Stratonovich Integrals in Riemannian Manifolds and Lie Groups

We first recap vector results and then less well known matrix versions. We refer to [13],[14] for basic definitions and properties of Stratonovich and Ito SDEs. We also develop one new result in this section.

A. Vector SDEs

Consider a system whose Euclidean $p-$dimensional state $\mathbf{r} = \mathbf{r}(t)$ obeys a Stratonovich [13] SDE

$$\eth\mathbf{r} = \alpha_S(\mathbf{r})dt + \textstyle\sum_j \alpha_j(\mathbf{r})\eth b_j \tag{2.1}$$

where $b_j(t)$ are independent Brownian motions.

We can now state some results.

Theorem I. Vector Stratonovich-Ito (SI) transformation [13]. If $\mathbf{r}$ obeys the S-SDE then it also obeys the I-SDE

$$\begin{aligned} d\mathbf{r} &= \alpha_I(\mathbf{r})dt + \textstyle\sum_j \alpha_j(\mathbf{r})db_j \\ \alpha_I(\mathbf{r}) &= \alpha_S(\mathbf{r}) + \alpha_\Delta(\mathbf{r}) \\ \alpha_\Delta(\mathbf{r}) &= \frac{1}{2}\textstyle\sum_j \frac{\partial \alpha_j(\mathbf{r})}{\partial \mathbf{r}^\top}\alpha_j(\mathbf{r}) \end{aligned}$$

We now need to find properties the drift and diffusion coefficients must satisfy to ensure the state trajectory lies in a constraint space $h_{(p-d)\times 1}(\mathbf{r}) = 0$. We assume $h(\cdot)$ is twice differentiable and can then regard the constraint as defining an embedded Riemannian manifold (RM). We introduce, at the point $\mathbf{r}$, the matrix of normals

$$N_{p\times(p-d)} = \frac{\partial h^\top}{\partial \mathbf{r}} = [n^u] = [\frac{\partial h^u}{\partial \mathbf{r}}]$$

which we assume to have rank $p-d$. The $(p-d)$ columns/normals of N are linearly independent and span the normal space $\mathcal{N}$ of the embedded RM at the point $\mathbf{r}$. The

$d-$dimensional tangent space at the point $\mathbf{r}$, $\mathcal{T}$ is spanned by linearly independent vectors of dimension p that are each orthogonal to the $p-d$ normal vectors.

Theorem II. [15]. Constrained S-SDE.
The S-SDE (2.1) evolves in the embedded RM $h(\mathbf{r})=0$ iff

- $N^\top \alpha_S(\mathbf{r})=0$.
- $N^\top \alpha_j(\mathbf{r})=0$ for all j

B. Multiplicative Matrix SDEs

The vector results are first converted into general matrix versions and we then focus on multiplicative matrix SDEs. This vector to matrix conversion massively simplifies, statements of results, proofs and intuitions. Also moving to matrices will put us in a matrix Lie group setting [16].

Theorem III. Matrix SI transformation [15]
Suppose $X_{n\times p}$ obeys the matrix S-SDE

$$đX = A_S(X)dt + \sum_j A_j(X)đb_j \qquad (2.2)$$

This just means that each column of X obeys a vector S-SDE. Then X also obeys the matrix I-SDE

$$\begin{aligned} dX &= A_I(X)dt + \sum_j A_j(X)db_j \qquad (2.3)\\ A_I(X) &= A_S(X) + A_\Delta(X)\\ A_\Delta(X) &= \frac{1}{2}\sum_j \sum_{rs:(A_j)_{rs}\neq 0}\frac{\partial A_j}{\partial X_{rs}}(A_j)_{rs} \end{aligned}$$

Result I. Matrix SI transform for Multiplicative Noise SDE.
Suppose X obeys the multiplicative noise S-SDE

$$đX = A_S(X)dt + X\sum_j B_j đb_j$$

Then it also obeys the multiplicative noise I-SDE

$$\begin{aligned} dX &= A_I(X)dt + X\sum_j B_j db_j\\ A_I(X) &= A_S(X) + A_\Delta(X)\\ A_\Delta(X) &= \frac{1}{2}X\sum_j (B_j)^2 + \frac{1}{2}X\sum_j K_j\\ K_j &= \sum_{rs}\frac{\partial B_j}{\partial X_{rs}}(XB_j)_{rs} \end{aligned}$$

Proof. See the appendix.

Corollary Ia. [15].
In Result I, when the diffusion coefficients B_j are constants $B_j = B_j^o$ then $K_j = 0$ so $A_\Delta(X) = \frac{1}{2}X\sum_j (B_j^o)^2$.

Corollary Ib.
In Result I, when $B_j(X) = \sigma_j(X)B_j^o$ where $\sigma_j(X) > 0$ are state dependent scalar standard deviations, then

$$\begin{aligned} K_j &= B_j(X)\tau_j \text{ where } \tau_j = tr(\sigma_j'(X)XB_j^o)\\ \sigma_j'(X) &= [\frac{\partial \sigma_j(X)}{\partial X_{rs}}] \end{aligned}$$

Proof. We have

$$\begin{aligned} K_j &= \sum_{rs}(\sigma_j'(X))_{rs}B_j^o(XB_j^o)_{rs}\sigma_j(X)\\ &= \sigma_j(X)B_j^o tr(\sigma_j'(X)XB_j^o)\\ &= B_j(X)tr(\sigma_j'(X)XB_j^o) \end{aligned}$$

We now specialise further to fully multiplicative square SDEs.

Theorem IV. SDEs in SO(n). [7].
In order that the solution $X_{n\times n}$ of the multiplicative I-SDE

$$dX = XB_{o,I}(X)dt + X\sum_j B_j(X)db_j$$

lies in $SO(n)$ (i.e. $X^\top X = I$) it is necessary and sufficient that $B_j(X)$ are skew matrices and that

$$B_{o,I} + B_{o,I}^\top = \sum_1^d B_j^2(X)$$

Corollary IV. In order that the solution X to the multiplicative S-SDE

$$đX = XB_{o,S}(X)dt + X\sum_j B_j(X)đb_j$$

lies in $SO(n)$ (i.e. $X^\top X = I$) it is necessary and sufficient that $B_j(X)$ are skew matrices and that $B_{o,S}$ is skew.

Proof. We convert to an I-SDE and apply Theorem IV. We find via Theorem III and Result I that

$$\begin{aligned} B_{o,I} &= B_{o,S} + \frac{1}{2}\sum_j (B_j)^2 + \frac{1}{2}\sum_j K_j\\ \Rightarrow B_{o,I} + B_{o,I}^\top &= \sum_j (B_j)^2 + \frac{1}{2}\sum_j [K_j + K_j^\top]\\ &+ B_{o,S} + B_{o,S}^\top\\ &= \sum_j (B_j)^2 + B_{o,S} + B_{o,S}^\top \end{aligned}$$

where we have used the fact that B_j are skew $\Rightarrow K_j$ are skew. Then we see that the condition of Theorem IV holds iff $B_{o,S} + B_{o,S}^\top = 0$.

Further Notation. In the sequel when $X(t)$ obeys the S-SDE of corollary IV we write

$$X \sim \mathcal{S}_{SO(n)}(B_{o,S}, B_j)$$

and when $X(t)$ obeys the I-SDE of theorem IV we write

$$X \sim \mathcal{I}_{SO(n)}(B_{o,I}, B_j)$$

III. Numerical Schemes for Multiplicative SDEs in Lie Groups

Here we discuss numerical solution of SDEs. For ODEs evolving in Lie groups there is a well developed literature, for which the now classic reference is [5]. That book also has some schemes applicable to embedded RMs.

For Euclidean SDEs there is a rich literature on construction and analysis [6]. However development of numerical schemes for SDEs in Lie groups and RMs remains rudimentary: see [7] for methods and references.

One approach to finding a numerical solution to a multiplicative SDE (2.2) uses the Magnus expansion (ME) [5]

- $X(t) = X(0)e^{\Omega(t)}$

The idea is to find the SDE obeyed by $\Omega(t)$ and apply a standard Euclidean numerical scheme to that. The second advantage of the ME is that it enables easy application of constraints. Thus if we require that $X(t)$ evolve in a Stiefel manifold so that $X^\top X = I$, this is easily achieved by ensuring Ω is skew symmetric.

A. ME Numerical Schemes based on I-SDE

We consider the multiplicative I-SDE of Theorem IV, which is equivalent to the multiplicative S-SDE of corollary IV. We find the I-SDE obeyed by $\Omega(t)$ and develop a Euclidean numerical solution for that which induces a numerical scheme for the original I-SDE.

To do this we need the following result.

Theorem V. [7].
Suppose $X(t) = X(0)e^{\Omega(t)}$ obeys the multiplicative I-SDE

$$dX = XB_{o,I}(X)dt + X\sum_{j=1}^{d} B_j(X)db_j$$

where $\Omega(0) = 0$. Then $\Omega(t)$ obeys the I-SDE

$$d\Omega = \Gamma_o dt + \sum_j \Gamma_j db_j$$

where

$$\begin{aligned}\Gamma_o &= dexp_{-\Omega}^{-1}(C_o) - \frac{1}{2}\sum_1^d dexp_{-\Omega}^{-1}(C_j) \\ C_o &= B_{o,I} - \frac{1}{2}\sum_1^d B_j^2 \\ \Gamma_j &= dexp_{-\Omega}^{-1}(B_j)\end{aligned}$$

and $C_j, j \geq 1$ are doubly infinite series [15] and not given here since we will not need them. Also

$$dexp_{-\Omega}^{-1}(W) = \sum_0^\infty \frac{\pi_k}{k!} ad_{-\Omega}^k(W)$$

where π_k are the Bernoulli numbers and ad^k are defined recursively by

$$\begin{aligned}&ad_R^0(W) = W, ad_R^1(W) = [R, W] = RW - WR \\ &ad_R^k(W) = [R, ad_R^{k-1}(W)]\end{aligned}$$

Numerical schemes can be constructed by truncating the infinite series. Here we consider a single term numerical method. In that case we find

$$\begin{aligned}dexp_{-R}^{-1}(W) &\approx ad_{-R}^0(W) = W \\ \Rightarrow \Gamma_o &\approx C_o = B_{o,I} - \frac{1}{2}\sum_1^d B_j^2 \\ \Gamma_j &\approx B_j\end{aligned}$$

We now state the equispaced special case of the first order algorithm described in [7] in the table above remark 6.

First Order I-ME.
Divide the interval $[0, T]$ into M subintervals each of width δ so that $M\delta = T$. Set $\hat{X}_o = X(0)$ (so that $\hat{X}_o^\top \hat{X}_o = I$) and iterate the following three steps:

(i) At step m draw $\epsilon_{j,m}, j = 1, \cdots, d$ iid $\sim N(0, 1)$.
(ii) Given $\hat{X}_m$ compute $\hat{B}_{j,m} = B_j(\hat{X}_m)$ and

$$\Omega_{m+1}^\delta = \delta B_{o,I}(\hat{X}_m) - \frac{\delta}{2}\sum_1^d \hat{B}_{j,m}^2 + \sqrt{\delta}\sum_1^d \hat{B}_{j,m}\epsilon_{j,m}$$

(iii) Set $\hat{X}_{m+1} = \hat{X}_m e^{\Omega_{m+1}^\delta}(\Rightarrow \hat{X}_{m+1}^\top \hat{X}_{m+1} = I)$.

We need to define the numerical approximation in continuous time. A step function interpolation is sufficient. A linear interpolation can also be done but is mathematically more complicated but does not improve the convergence rate. We set

$$\hat{X}(t) = \hat{X}_{k-1}, k\delta - \delta \leq t < k\delta \text{ for } 1 \leq k \leq M$$

and we write

$$\hat{X}(t) \sim \mathcal{M}_{SO(n)}(B_{o,I}, B_j)$$

IV. Convergence of I-ME Numerical Schemes

The analysis of the behaviour of matrix SDE numerical schemes as $\delta \to 0$ is very challenging. While convergence analysis is well developed in the Euclidean case [6] it is in its early stages for SDEs on manifolds and Lie groups. We extend the approach previously developed in [17],[12]. It does not deliver optimal rates (i.e. $O(\delta)$) but is fairly straightforward. To develop optimal rates, a totally different approach will be required [11]. Note that [11] does not treat the case considered here, where B_j are functions of X.

We deal only with the case of a multiplicative SDE evolving on SO(n). In this case we have three important properties:

(i) B_j are skew.
(ii) $B_{o,I} + B_{o,I}^\top = \sum_j (B_j)^2$
(iii) Ω_m^δ are skew and so $X_m^\top X_m = I$.

To proceed we introduce some assumptions.

A1. For $X \in SO(n)$, $B_o(X), B_j(X)$ obey Lipschitz conditions with bounded Lipschitz constant.
A2. For $X \in SO(n)$, $B_o(X), B_j(X)$ are bounded.

A crucial part of the argument is to introduce the intermediate quantity $X_\delta(t)$ with $X_\delta(0) = X(0)$ defined by

$$\begin{aligned}X_\delta(t) - X_\delta(0) &= \int_0^t \hat{X}(u)B_{o,I}(\hat{X}(u))du \\ &+ \int_0^t \hat{X}(u)\sum_j B_j(\hat{X}(u))db_j(u)\end{aligned}$$

Now introduce the error processes

$$\begin{aligned}e(t) &= \hat{X}(t) - X(t) = \nu(t) + \eta(t) \\ \nu(t) &= \hat{X}(t) - X_\delta(t) \\ \eta(t) &= X_\delta(t) - X(t)\end{aligned}$$

It follows that

$$\begin{aligned}&E[\sup_{0\leq t\leq T} \parallel e(t) \parallel^2] \\ \leq\ & 2E[\sup_{0\leq t\leq T} \parallel \eta(t) \parallel^2] + 2E[\sup_{0\leq t\leq T} \parallel \nu(t) \parallel^2]\end{aligned}$$

We now bound each term separately. This requires the following result.

Theorem VI. [17].
Let $W(t)$ be a standard Brownian motion and introduce

$$U_k = \sup_{k\delta-\delta\leq t<k\delta} \parallel W(t) - W(k\delta - \delta) \parallel$$

Then by the stationary independent increments property of Brownian motion, $U_1, \cdots, U_M$ are iid. Then

$$E[\max_{1\leq k\leq M} U_k^2] \leq O(\delta^{1-\theta})$$

where $\theta > 0$ can be made arbitrarily small.

Result IIIa. Under A1, A2 for $\hat{X} \sim \mathcal{M}_{SO(n)}(B_{o,I}, B_j)$ and $X \sim \mathcal{I}_{SO(n)}(B_{o,I}, B_j)$

$$E[\sup_{0\le t\le T} \| \nu(t) \|^2] \le O(\delta^{1-\theta})$$

Proof. See the appendix.

Result IIIb. Under A1, A2 for $\hat{X} \sim \mathcal{M}_{SO(n)}(B_{o,I}, B_j)$ and $X \sim \mathcal{I}_{SO(n)}(B_{o,I}, B_j)$

$$E[\sup_{0\le t\le T} \| \eta(t) \|^2] \le O(\delta^{1-\theta})$$

Proof. See the appendix.

Putting these together gives the main result.

Result IV. Under A1, A2 for $\hat{X} \sim \mathcal{M}_{SO(n)}(B_{o,I}, B_j)$ and $X \sim \mathcal{I}_{SO(n)}(B_{o,I}, B_j)$

$$E[\sup_{0\le t\le T} \| \hat{X}(t) - X(t) \|^2] \le O(\delta^{1-\theta})$$

Thus $\hat{X}(t)$ converges to $X(t)$ uniformly in mean square as $\delta \to 0$.

V. Convergence of S-ME

We now consider convergence of Stratonovich motived numerical schemes which we denote as S-ME. In particular we consider firstly the setup in corollary Ia.

Suppose $X(t) \sim \mathcal{S}_{SO(n)}(B_{o,S}, B_j)$ then the corresponding S-ME numerical scheme is

$$\hat{X}(t) \sim \mathcal{M}_{SO(n)}(B_{o,S} + \frac{1}{2}\textstyle\sum_j B_j^2, B_j).$$

because the first two terms in the Ω_m^δ update reduce to $\delta\hat{B}_{o,S}$. Now introduce the I-SDE

$$X(t) \sim \mathcal{I}_{SO(n)}(B_{o,S} + \frac{1}{2}\textstyle\sum_j B_j^2, B_j)$$

Then from Result IV we can conclude

$$E[\sup_{0\le t\le T} \| \hat{X}(t) - X(t) \|] \le O(\delta^{1-\theta})$$

However, from corollary Ib $X(t) \sim \mathcal{S}_{SO(n)}(B_{o,S}, B_j)$ corresponds to the I-SDE

$$X(t) \sim \mathcal{I}_{SO(n)}(B_{o,S} + \frac{1}{2}\textstyle\sum_j B_j^2 + \frac{1}{2}\sum_j \tau_j B_j, B_j)$$

So the S-ME numerical scheme converges to the wrong SDE!

In the case of corollary Ia the two SDEs agree and we get convergence to the correct SDE.

We summarise this as follows.

Result V. Consider the multiplicative S-SDE $X(t) \sim \mathcal{S}_{SO(n)}(B_{o,S}, B_j)$ and S-ME numerical scheme $\hat{X}(t) \sim \mathcal{M}_{SO(n)}(B_{o,S} + \frac{1}{2}\sum_j B_j^2, B_j)$. Then

(i) If the diffusion coefficients B_j are constants then, as $\delta \to 0$ S-ME converges to the correct I-SDE $X(t) \sim \mathcal{I}_{SO(n)}(B_{o,S} + \frac{1}{2}\sum_j B_j^2, B_j)$.

(ii) If the diffusion coefficients $B_j = \sigma_j(X)B_j^o$ are state-dependent then, as $\delta \to 0$, S-ME converges to the wrong I-SDE namely $X(t) \sim \mathcal{I}_{SO(n)}(B_{o,S} + \frac{1}{2}\sum_j B_j^2 + \frac{1}{2}\sum_j \tau_j B_j, B_j)$.

Proof. (ii) is already established. (i) follows from (ii) since then $\tau_j = 0$.

VI. Conclusions

In this paper we have shown, for the first time, (uniform mean-square) convergence of a Magnus expansion based numerical scheme for simulating a multiplicative SDE, constrained to lie in $SO(n)$, with state dependent coefficients.

We then studied a second numerical scheme based on a Stratonovich/ODE formulation. We showed that with state dependent diffusion coefficients it converges to the wrong SDE. Only if the diffusion coefficients are constants does it converge to the correct SDE.

This is an alarming result because it appears that most SDE numerical simulation in applied literatures are using the wrong numerical schemes. It provides another insight into the relative properties of Ito and Stratonovich SDEs.

In future work we will study methods to obtain the optimal convergence rate.

VII. Appendix

In the proofs of IIIa, IIIb:

- we use R, L to denote generic constants.
- we denote $\hat{B}_{o,I}(u) = B_{o,I}(\hat{X}(u))$ and $\hat{B}_j(u) = B_j(\hat{X}(u))$
- we denote $\hat{X}_m = \hat{X}(m\delta)$.
- we denote $\hat{B}_{o,I,m} = B_{o,I}(\hat{X}_m), \hat{B}_{j,m} = B_j(\hat{X}_m)$.
- We repeatedly use the fact that $\| \hat{X}(u) \| \le R$.
- We repeatedly use the elementary inequalities: $(a+b)^2 \le 2a^2 + 2b^2$ and more generally $(\sum_1^d a_j)^2 \le d\sum_j a_j^2$ and also $|a+b|^3 \le 6|a|^3 + 6|b|^3$.

A. Proof of Result I

Introduce E_{rs} which is a matrix of $0s$ but with a 1 in row r, column s. Applying the chain rule and Theorem III we find

$$\begin{aligned}
\frac{\partial A_j}{\partial X_{rs}} = \frac{\partial (XB_j)}{\partial X_{rs}} &= E_{rs}B_j + X\frac{\partial B_j}{\partial X_{rs}} \\
\Rightarrow \textstyle\sum_{rs} \frac{\partial A_j}{\partial X_{rs}}(A_j)_{rs} &= \textstyle\sum_{rs}(A_j)_{rs}E_{rs}B_j + XK_j \\
&= A_jB_j + XK_j \\
&= X(B_j)^2 + XK_j \\
K_j &= \textstyle\sum_{rs}\frac{\partial B_j}{\partial X_{rs}}(A_j)_{rs} \\
&= \textstyle\sum_{rs}\frac{\partial B_j}{\partial X_{rs}}(XB_j)_{rs}
\end{aligned}$$

from which the result follows.

B. Proof of Result IIIa

We have

$$\begin{aligned}
& X_\delta(t) - X_\delta(0) \\
= & \int_0^t \hat{X}(u)\hat{B}_{o,I}(u)du + \textstyle\sum_j \int_0^t \hat{X}(u)\hat{B}_j(u)db_j(u)
\end{aligned}$$

Next for $m\delta \le t < m\delta + \delta$, since $\hat{X}(u)$ is a step function, we find

$$\begin{aligned}
& X_\delta(t) - X_\delta(0) \\
= & \hat{X}_m\hat{B}_{o,I,m}(t - m\delta) + \textstyle\sum_j \hat{X}_m\hat{B}_{j,m}(b_j(t) - b_j(m\delta)) \\
+ & \textstyle\sum_1^m \hat{X}_{k-1}(\hat{B}_{o,I,k-1}\delta + \sqrt{\delta}\hat{B}_{j,k-1}\epsilon_{j,k-1})
\end{aligned}$$

To calculate ν_t we need to calculate $\hat{X}_m$. We have

$$\begin{aligned}\hat{X}_m &= \hat{X}_{m-1}e^{\Omega^\delta_m} \\ &= \hat{X}_{m-1}(I+\Omega^\delta_m+\frac{1}{2}(\Omega^\delta_m)^2)+\hat{X}_{m-1}E_m \\ E_m &= e^{\Omega^\delta_m}-I-\Omega^\delta_m-\frac{1}{2}(\Omega^\delta_m)^2\end{aligned}$$

Summing gives

$$\begin{aligned}\hat{X}_m &= \textstyle\sum_1^m(\hat{X}_{k-1}\Omega^\delta_k+\frac{1}{2}\hat{X}_{k-1}(\Omega^\delta_k)^2) \\ &+ \textstyle\sum_1^m\hat{X}_{k-1}E_k+\hat{X}_o\end{aligned}$$

Now consider that

$$\begin{aligned}\Omega^\delta_{m+1} &= \delta\hat{B}_{e,m}+\sqrt{\delta}\textstyle\sum_j\hat{B}_{j,m}\epsilon_{j,m} \\ \hat{B}_{e,m} &= \hat{B}_{o,I,m}-\frac{1}{2}\textstyle\sum_j\hat{B}^2_{j,m} \\ &\Rightarrow \frac{1}{2}(\Omega^\delta_{m+1})^2=\frac{\delta}{2}\textstyle\sum_j\hat{B}^2_{j,m}\epsilon^2_{j,m}+E^\delta_m \\ E^\delta_m &= \frac{\delta^2}{2}\hat{B}^2_{e,m}+\frac{\delta}{2}(\textstyle\sum_{j\neq j'}\hat{B}_{j,m}\epsilon_{j,m}\hat{B}_{j',m}\epsilon_{j',m}) \\ &+ \delta^{3/2}\hat{B}_{e,m}\textstyle\sum_j\hat{B}_{j,m}\epsilon_{j,m}\end{aligned}$$

Thus we find

$$\begin{aligned}&\Omega^\delta_{m+1}+\frac{1}{2}(\Omega^\delta_{m+1})^2 \\ &= \delta\hat{B}_{e,m}+\sqrt{\delta}\textstyle\sum_j\hat{B}_{j,m}\epsilon_{j,m} \\ &+ \frac{1}{2}\delta\textstyle\sum_j\hat{B}^2_{j,m}\epsilon^2_{j,m}+E^\delta_m \\ &= \delta\hat{B}_{o,I,m}+\sqrt{\delta}\textstyle\sum_j\hat{B}_{j,m}\epsilon_{j,m}+E^\delta_m+E^b_m \\ E^b_m &= \frac{1}{2}\delta\textstyle\sum_j(\hat{B}_{j,m})^2(\epsilon^2_{j,m}-1)\end{aligned}$$

Putting this together we get

$$\begin{aligned}&\hat{X}_m-\hat{X}_o=\Delta_m+\Delta^\delta_m+\Delta^b_m \\ +&\textstyle\sum_1^m\hat{X}_{k-1}(\delta\hat{B}_{o,I,k-1}+\sqrt{\delta}\sum_j\hat{B}_{j,k-1}\epsilon_{j,k-1}) \\ &(\Delta_m,\Delta^\delta_m,\Delta^b_m)=\textstyle\sum_1^m\hat{X}_{k-1}(E_k,E^\delta_{k-1},E^b_{k-1})\end{aligned}$$

Now we can form $\nu(t)$. Noting that $X_\delta(0)=X(0)=\hat{X}_o$ we find the important terms cancel leaving

$$\begin{aligned}-\nu(t) &= X_\delta(t)-\hat{X}(t) \\ &= a(t)+b(t)+\Delta_m+\Delta^\delta_m+\Delta^b_m \\ a(t) &= \hat{X}_m\hat{B}_{o,I,m}(t-m\delta) \\ b(t) &= \textstyle\sum_j\hat{X}_m\hat{B}_{j,m}(b_j(t)-b_j(m\delta))\end{aligned}$$

Thus

$$\begin{aligned}\nu_T &= E[\max_{0\le t\le T}\parallel\nu(t)\parallel^2]\le 5a+5b+5\alpha+5\alpha^\delta+5\alpha^b \\ a &= const.\delta \\ b &= E[\sup_{0\le t\le T}b^2(t)] \\ \alpha &= E[\max_{1\le m\le M}\parallel\Delta_m\parallel^2] \\ \alpha^\delta &= E[\max_{1\le m\le M}\parallel\Delta^\delta_m\parallel^2] \\ \alpha^b &= E[\max_{1\le m\le M}\parallel\Delta^b_m\parallel^2]\end{aligned}$$

We treat each term separately.

$\boxed{b}=O(\delta^{1-\theta})$.

We have

$$\begin{aligned}b(t) &\le \textstyle\sum_j\max_{m\delta\le t<m\delta+\delta}|b_j(t)-b_j(m\delta)| \\ &= \textstyle\sum_jU_{jk}\le\sqrt{d\sum_jU^2_{jk}} \\ \Rightarrow E[\sup_{0\le t\le T}b^2(t)] &\le d\textstyle\sum_jE[\max_{1\le k\le M}U^2_{jk}]\le O(\delta^{1-\theta})\end{aligned}$$

where we have applied Theorem VI.

$\boxed{\alpha^b}=O(\delta)$.

Since the $\epsilon_{j,k}$ are iid then Δ^b_m is a martingale and we can use Doob's maximal inequality [14] to find

$$\begin{aligned}E(\max_{1\le m\le M}\parallel\Delta^b_m\parallel^2) &\le RE(\parallel\Delta^b_M\parallel^2 \\ &\le R\textstyle\sum_1^ME\parallel E^b_{k-1}\parallel^2 \\ &\le LM\delta^2=LT\delta=O(\delta)\end{aligned}$$

$\boxed{\alpha^\delta}=O(\delta)$.

Δ^δ_m has three components (see E^δ_m). The first is bounded by $\frac{1}{2}Rm\delta^2\le RT\delta=O(\delta)$. The third is a martingale and following the same argument used for α^b we get a maximal bound $\delta^3\delta=\delta^4=o(\delta)$. The second term is also a martingale and following the same argument used for α^b we get a maximal bound $\delta^2\delta=\delta^3=o(\delta)$.

$\boxed{\alpha}=O(\delta)$.

We have

$$\begin{aligned}\parallel\Delta_m\parallel &\le \textstyle\sum_1^M\parallel E_{k-1}\parallel \\ \Rightarrow\max_{1\le m\le M}\parallel\Delta_m\parallel^2 &\le M\textstyle\sum_1^M\parallel E_{k-1}\parallel^2 \\ \Rightarrow E[\max_{1\le m\le M}\parallel\Delta_m\parallel^2] &\le M\textstyle\sum_1^ME[\parallel E_{k-1}\parallel^2]\end{aligned}$$

We now use the following result [11][Lemma 11].

If A is skew then $\parallel e^A-I-A-\frac{1}{2}A^2\parallel\le\frac{1}{3}\parallel A\parallel^3$.

Then $\parallel E_m\parallel\le\parallel\Omega^\delta_m\parallel^3$. We have

$$\begin{aligned}\parallel\Omega^\delta_m\parallel &\le R\delta+\sqrt{\delta}\textstyle\sum_j|\epsilon_{j,m}| \\ \Rightarrow\parallel\Omega^\delta_m\parallel^3 &\le L\delta^3+\delta^{3/2}(\textstyle\sum_j|\epsilon_{j,m}|)^3 \\ &\le L\delta^3+L\delta^{3/2}\textstyle\sum_j|\epsilon_{j,m}|^3 \\ \Rightarrow\parallel\Omega^\delta_m\parallel^6 &\le R\delta^6+R\delta^3(\textstyle\sum_j|\epsilon_{j,m}|^3)^2 \\ &\le R\delta^6+L\delta^3\textstyle\sum_j|\epsilon_{j,m}|^6\end{aligned}$$

Thus

$$\begin{aligned}E\parallel E_m\parallel^2 &\le E\parallel\Omega^\delta_m\parallel^6 \\ &\le R\delta^6+R\delta^3\le L\delta^3 \\ \Rightarrow M\textstyle\sum_1^ME\parallel E_{k-1}\parallel^2 &\le LM^2\delta^2\delta\le LT^2\delta=O(\delta)\end{aligned}$$

We see that all terms are $O(\delta)$, except $b=O(\delta^{1-\theta})$ and the result is established.

C. Proof of Result IIIb

We have

$$\begin{aligned} \eta(t) &= A(t) + \textstyle\sum_j C_j(t) \\ A(t) &= \int_0^t (\hat{X}(u)\hat{B}_{o,I}(u) - X(u)B_{o,I}(u))du \\ C_j(t) &= \int_0^t [\hat{X}(u)\hat{B}_j(u) - X(u)B_j(u))]db_j(u) \end{aligned}$$

Since $X(u), \hat{X}(u), \in SO(p)$ then in view of A1,A2 we have

$$\begin{aligned} & \hat{X}(u)\hat{B}_{o,I}(u) - X(u)B_{o,I}(u) \\ = \quad & (\hat{X}(u) - X(u))\hat{B}_{o,I}(u) + X(u)(\hat{B}_{o,I}(u) - B_{o,I}(u)) \\ \Rightarrow \quad & \| \hat{X}(u)\hat{B}_{o,I}(u) - X(u)B_{o,I}(u) \| \\ \leq \quad & \frac{1}{2}R \| \hat{X}(u) - X(u) \| + \frac{1}{2}R \| \hat{X}(u) - X(u) \| \\ \leq \quad & R \| \nu(u) \| + R \| \eta(u) \| \end{aligned}$$

Thus

$$\begin{aligned} & \| \eta(t) \| \leq \textstyle\sum_j \| C_j(t) \| \\ + \quad & R\int_0^t \| \nu(u) \| du + R\int_0^t \| \eta(u) \| du \end{aligned}$$

Introduce $\eta_*(t) = \sup_{0 \leq s \leq t} \| \eta(t) \|$ and similarly define $\nu_*(t)$ and $c_{j,*}(t)$. Next use the Cauchy-Schwarz inequality $(\int_0^t f(u)du)^2 \leq t\int_0^t f^2(u)du \leq T\int_0^t f^2(u)du)$ to find

$$\begin{aligned} \eta_*^2(t) \quad \leq \quad & 3R^2T\int_0^t \eta_*^2(u)du \\ + \quad & 3R^2T\int_0^t \nu_*^2(u)du + 3R^2d\textstyle\sum_j c_{j,*}^2(t) \end{aligned}$$

We will use this below.

Now observe that $C_j(t)$ are independent martingales and so Doob's martingale inequality [14] gives

$$E(c_{j,*}^2(t)) \leq 4E(C_j^2(t))$$

Using properties of the Ito integral [14] and A1,A2 we find

$$\begin{aligned} & E(C_j^2(t)) \\ = \quad & \int_0^t E \| \hat{X}(u)\hat{B}_j(u) - X(u)B_j(u) \|^2 du \\ \leq \quad & R^2\int_0^t E \| \hat{X}(u) - X(u) \|^2 du \\ \leq \quad & 2R^2\int_0^t E(\| \nu(u) \|^2 du + 2R^2\int_0^t E(\| \eta(u) \|^2 du \\ \leq \quad & 2R^2\int_0^t E(\nu_*^2(u))du + 2R^2\int_0^t E(\eta_*^2(u))du \end{aligned}$$

Putting this together we get

$$\begin{aligned} & E(\eta_*^2(t)) \\ \leq \quad & 3R^2T\int_0^t E(\eta_*^2(u))du + 3R^2T\int_0^t E(\nu_*^2(u))du \\ + \quad & 6R^4d^2\int_0^t [E(\nu_*^2(u)) + E(\eta_*^2(u))du \\ = \quad & L\int_0^t E(\eta_*^2(u))du + L\int_0^t E(\nu_*^2(u))du \end{aligned}$$

Now applying Gronwall's lemma gives

$$\begin{aligned} E(\eta_*^2(t)) \quad \leq \quad & L\int_0^T E(\nu_*^2(u))due^{LT} \\ \leq \quad & Le^{LT}TE(\nu_*^2(T)) \leq R\delta^{1-\theta} \end{aligned}$$

where we applied result IIIa. Result IIIb is now established.

D. Acknowledgement

We thank the reviewers for a careful reading of the manuscript, including one in particular who spent extra time chasing down a long series of typos in the proof section. This high level of reviewing scholarship is greatly appreciated. Of course I should have done a better job of proof-reading.

References

[1] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes, 2nd Edn.*, North-Holland, Tokyo, 1989.

[2] G.S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, vol. I*, Birkhauser, Basel, 2009.

[3] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, "Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics", *ChemPhysChem*, vol. 6, pp. 1809–1814, 2005.

[4] G.S. Chirikjian, *Stochastic Models, Information Theory, and Lie Groups, vol. II*, Birkhauser, Basel, 2012.

[5] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration Structure-Preserving Algorithms for Ordinary Differential Equations, 2nd. Edn.*, Springer, Berlin, 2006.

[6] P.E. Kloeden and E. Platen, *Numerical solution of stochastic differential equations*, Springer-Verlag, Berlin, 1992.

[7] G. Marjanovic and V. Solo, "Numerical methods for stochastic differential equations in matrix lie groups made simple", *IEEE Trans. Autom. Contr.*, vol. 63, pp. 4035–4050, 2018.

[8] J. Kwon andd M. Choi, FC. Park, and C. Chun, "Particle filtering on the euclidean group: framework and applications", *Robotica*, vol. 25, pp. 725–737, 2007.

[9] J. Kwon and FC. Park, "Visual tracking via particle filtering on the affine group", *J. Robot. Res.*, vol. 29, pp. 198–217, 2010.

[10] A.J. Seneviratne and V. Solo, "Group sparsity via sure based on regression parameter mean squared error", *IEEE Sig. Proc. Lett.*, vol. 21, pp. 1125 – 1129, 2014.

[11] M. Piggott and V. Solo, "Geometric euler-maruyama schemes for stochastic differential equations in so(n) and se(n)", *SIAM. Jl. Numerical Analysis*, vol. 54, pp. 2490–2516, 2016.

[12] Z. Wang and V. Solo, "Convergence of a tangent space numerical scheme for a stochastic differential equation on a sphere", in *Proc. IEEE CDC*. IEEE, 2021, pp. 5807–5812.

[13] B. Oksendal, *Stochastic Differential Equations*, Springer-Verlag, Berlin, 1995.

[14] F. Klebaner, *Introduction to Stochastic Calculus with Applications, 2nd Edn.*, Imperial College Press, London, UK, 2005.

[15] V. Solo and G. Chirikjian, "Ito, stratonovich and geometry", in *Proc. IEEE CDC, Nice*. IEEE, 2019, pp. 3026–3032.

[16] BC. Hall, *Lie Groups, Lie Algebras, and Representations*, Springer, Berlin, 2003.

[17] M Piggott and V Solo, "Stochastic numerical analysis for brownian motion on so(3)", in *Proc. IEEE CDC*. IEEE, 2014, pp. 3420–3425.