AN ENGINEER'S GUIDE TO PARTICLE FILTERING ON MATRIX LIE GROUPS

Goran Marjanovic † *and Victor Solo* ‡

School of Electrical Engineering and Telecommunications,
The University of New South Wales, Sydney, AUSTRALIA

ABSTRACT

In many important engineering applications the state dynamics of a system are modelled by **S**tochastic **D**ifferential **E**quations (**SDE**s) evolving in non-Euclidean spaces such as matrix Lie groups. Due to the advances in computing power, the problem of state estimation can be efficiently addressed by the particle filtering method. This requires dealing with both the geometry and the stochastics of the problem. However, the very few papers that properly deal with either are in the mathematics literature and not accessible. The engineering literature is also small but plagued with problems. With this in mind, we give a direct accessible derivation of the particle filter algorithm for state estimation in matrix Lie groups. We do not rely on differential geometry or advanced stochastic calculus. Simulation examples are provided.

Index Terms— Particle filter, sequential Monte Carlo methods, stochastic differential equations, matrix Lie group integrators

1. INTRODUCTION

In numerous engineering applications, state estimation of dynamical systems evolving in non-Euclidean spaces such as matrix Lie groups is required. Examples include computer vision [1–4], array signal processing [5], satellite attitude and pose estimation [6–8], robotics [6, 9, 10], etc. **P**article **F**iltering (**PF**) methods have become a very popular class of algorithms to numerically solve these estimation problems in a recursive fashion as observations become available. They are flexible, and can easily be applied to nonlinear and non-Gaussian dynamic models.

What is a Matrix Lie Group? – A matrix Lie group $\mathcal{G}$ is simply a closed subset (under matrix multiplication) of all $n \times n$ invertible matrices. The set of $n \times n$ matrices $\mathbf{A}$ such that $\mathrm{e}^{\mathbf{A}} \in \mathcal{G}$ forms a vector space called the Lie algebra of $\mathcal{G}$, and is denoted by $\mathfrak{g}$. There is a also a geometric interpretation for $\mathfrak{g}$ – By multiplying $\mathrm{e}^{\mathbf{A}}$ with itself, we can interpolate the generated sequence $\mathrm{e}^{\mathbf{A}}, \mathrm{e}^{2\mathbf{A}}, \mathrm{e}^{3\mathbf{A}}, \cdots \in \mathcal{G}$ to obtain $\mathrm{e}^{t\mathbf{A}} \in \mathcal{G}$, $t \geq 0$. Since $\mathrm{e}^{0\cdot\mathbf{A}} = \mathbf{I} \in \mathcal{G}$ and $\frac{d}{dt}\mathrm{e}^{t\mathbf{A}}\big|_{t=0} = \mathbf{A}$, we can thus think of $\mathfrak{g}$ as a tangent space at the identity.

In addition to $\mathcal{G}$ being a closed set, it is also a manifold, i.e. a curved space that locally looks like $\mathbb{R}^m$ for some m. Hence the fact that $\mathfrak{g}$ is a tangent space to $\mathcal{G}$ implies that we can think of $\mathfrak{g}$ behaving like $\mathbb{R}^m$. A good exposé on the topic is given in [11, 12].

† Email: g.marjanovic@unsw.edu.au. Author information also available on Research Gate: https://www.researchgate.net/profile/Goran_Marjanovic4

‡ This work was partially supported by an ARC (Australian Research Council) grant. Email: v.solo@unsw.edu.au.

The State Space Model – In practice, the complete information on the dynamics of the state $\mathbf{X}(t)$ is not available, and so, a stochastic model is considered [13]. We use the following general SDE (in Itō form) to describe the dynamics of the state $\mathbf{X}(t) \in \mathcal{G}$:

$$d\mathbf{X}(t) = \mathbf{X}(t)\mathbf{V}_0(\mathbf{X}(t))\,dt + \sum_{i=1}^{d} \mathbf{X}(t)\mathbf{V}_i(\mathbf{X}(t))\,dW_i(t) \quad (1)$$

where $\mathbf{V}_i(\mathbf{X}(t))$'s are known matrix functions of the state $\mathbf{X}$ and time t, and satisfy (18) in Theorem 4 below. $dW_i(t)$'s are scalar i.i.d. Gaussian increments, i.e. $dW_i \sim \mathcal{N}(0, dt)$. The SDE in (1) is very common in geometric state estimation problems, see [1, 6–8, 14, 15].

Prior Work – Until very recently, PFs have solely been addressing systems in $\mathbb{R}^m$, see [16, 17]. However, such filters cannot be considered for filtering problems in other manifolds such as matrix Lie groups $\mathcal{G}$ because the update schemes tend to immediately leave the manifold, e.g. see the example in [8]. Hence, these schemes are unstable.

PFs for state estimation that take into account the geometry of the state space are given in [1–3, 6, 13, 18–22]. [13] is the first paper to explicitly address particle filtering on $\mathcal{G}$. However, it does not deal with the general SDE in (1) but rather a very special case. [3] addresses explicitly in detail the particle filtering on the special Euclidean group $\mathcal{G} = SE(3)$. Thus, the state dynamics are also a special case of (1). Similarly, [18, 19] consider the special orthogonal group $\mathcal{G} = SO(3)$. In [6], the state dynamics satisfy (1), and a general PF is given (without derivation), which is also used in [1, 2]. In [1], the emphasis is placed on the Affine group $\mathcal{G} = \mathit{Aff}(2)$, while in [2, 6], it is again on $SE(3)$. [6] also considers $SO(3)$. However, even though the PF algorithm in [6] seems to be in the form that respects the geometry, i.e. the update schemes supposedly remain in $\mathcal{G}$, it definitely does not take into account the stochastic nature of the problem (Itō's lemma). In other words, the PF state updates do not correspond to the system SDE (1).

[20, 21] propose PFs, where the state evolves on a Riemannian manifold (which includes $\mathcal{G}$). However, the given schemes are too abstract, do not explicitly deal with $\mathcal{G}$, and so, extensive knowledge of differential geometry is required in order to specialise the procedures to $\mathcal{G}$. In [20], the example provided considers state updates in $\mathbb{R}^m$, which is not extendable to updates on $\mathcal{G}$. In [21], the discrete state updates follow a multivariate affine generalised Hyperbolic distribution. It is not clear whether states generated from this distribution satisfy (1).

Lastly, existing literature for dealing with matrix Lie groups considers deterministic state dynamics [23–26]. These works do not apply in the stochastic setting, where the available literature for engineers [27–29] only deals with systems in $\mathbb{R}^m$. Dealing with geometric SDEs, such as (1), requires knowledge in both stochastic calculus and differential geometry. So far, the very little literature on this [30–32] has been oriented to mathematicians and not accessible to engineers.

Current Contribution – We give a direct accessible derivation of the general PF algorithm for state estimation on a matrix Lie group $\mathcal{G}$, where the state dynamics are give by (1). Knowledge of only basic probability concepts is assumed, and we do not rely on differential geometry or stochastic processes theory. The paper is organised as follows: In Section 2 we state the problem of state estimation. Sections 3 and 4 describe the particle filter method for solving the problem, and the method is stated in Section 5. Simulations and the conclusion are given in Section 6 and 7 respectively.

2. PROBLEM STATEMENT – STATE ESTIMATION IN $\mathcal{G}$

For the discrete observation times $t = t_1, t_2, \ldots, t_k$, define $\mathbf{X}_k = \mathbf{X}(t_k)$, also, let the noisy $l \times q$ measurement $\mathbf{Y}_k$ of the state $\mathbf{X}_k \in \mathcal{G}$ be given by $\mathbf{Y}_k = \mathbf{C}(\mathbf{X}_k) + \mathbf{E}$, where $\mathbf{C} : \mathcal{G} \to \mathbb{R}^{l\times q}$ and $\mathbf{E}_{l\times q}$ is the noise whose entries are i.i.d. zero mean Gaussian.

Given the observations $\mathbf{Y}_{1:k} = \{\mathbf{Y}_1, \ldots, \mathbf{Y}_k\}$, the aim is to find the minimum **M**ean **S**quared **E**rror (**MSE**) estimates of the states $\mathbf{X}_{1:k} = \{\mathbf{X}_1, \ldots, \mathbf{X}_k\}$, where the MSE is defined by [5]

$$\mathbb{E}\left[d(\mathbf{X}, \mathbf{X}_k)^2\right] = \int_{\mathcal{G}} d(\mathbf{X}, \mathbf{X}_k)^2\, p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k})\, d_{\mathcal{G}}\mathbf{X}_k \tag{2}$$

where $d : \mathcal{G}\times\mathcal{G} \to \mathbb{R}$ is the distance between $\mathbf{X}$ and $\mathbf{X}_k$ (see Section 6), $p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k})$ is the posterior probability density function, and $d_{\mathcal{G}}\mathbf{X}_k$ is the infinitesimal area on the curved space $\mathcal{G}$ at $\mathbf{X}_k$, see [33]. Note that if $\mathcal{G}$ is replaced by $\mathbb{R}^m$, then the infinitesimal area on $\mathbb{R}^m$ at $\mathbf{x} = [x_1, \ldots, x_m] \in \mathbb{R}^m$ simplifies to $d_{\mathbb{R}^m}\mathbf{x} = dx_1 \ldots dx_m$.

Remark 1. Here we do not explicitly calculate integrals containing probability densities on $\mathcal{G}$, rather approximate them using random samples, i.e. letting $\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(N)} \sim p(\mathbf{Z})$, where $N \gg 1$, we have the following (Monte Carlo) sample average approximation

$$\mathbb{E}\left[\phi(\mathbf{Z})\right] = \int_{\mathcal{G}} \phi(\mathbf{Z}) p(\mathbf{Z})\, d_{\mathcal{G}}\mathbf{Z} \approx \frac{1}{N}\sum_{i=1}^{N} \phi(\mathbf{Z})|_{\mathbf{Z}=\mathbf{S}^{(i)}} \tag{3}$$

3. STATE ESTIMATION VIA PARTICLE FILTERING

To obtain a state estimator $\widehat{\mathbf{X}}_k \in \mathcal{G}$ we need to minimise the MSE (2) with respect to $\mathbf{X} \in \mathcal{G}$, which might be difficult in general if the direct approach of calculating the integral is considered. Hence, we approximate (2) by its sample average using (3), which requires sampling from the posterior $p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k})$, and then minimise the approximation with respect to $\mathbf{X}$. However, direct sampling from the posterior is difficult and/or inefficient in general because the relationship between $\mathbf{X}$ and $\mathbf{Y}$ might be too complicated. Thus, a recursive procedure is needed, which takes the samples (particles) from the previous posterior $p(\mathbf{X}_{k-1} \,|\, \mathbf{Y}_{1:k-1})$ and transforms them into samples (particles) from the current posterior $p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k})$. This leads to particle filtering.

The Filtering Equations – Using samples drawn from $p(\mathbf{X}_{k-1} \,|\, \mathbf{Y}_{1:k-1})$ to obtain samples from $p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k})$ requires deriving a relationship between these two posteriors. – By the standard formula for the conditional density, we have

$$p(\mathbf{X}_k, \mathbf{X}_{k-1} \,|\, \mathbf{Y}_{1:k-1}) = p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1}, \mathbf{Y}_{1:k-1})\, p(\mathbf{X}_{k-1} \,|\, \mathbf{Y}_{1:k-1})$$

The state SDE (1) implies that $\mathbf{X}$ is Markov, i.e. $\mathbf{X}_k$ depends only on $\mathbf{X}_{k-1}$. Hence, $p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1}, \mathbf{Y}_{1:k-1}) = p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1})$. By substituting this and integrating the result over $\mathbf{X}_{k-1}$, we obtain the data dependant **C**hapman-**K**olmogorov (**CK**) equation

$$\begin{aligned} p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k-1}) &= \int_{\mathcal{G}} p(\mathbf{X}_k, \mathbf{X}_{k-1} \,|\, \mathbf{Y}_{1:k-1})\, d_{\mathcal{G}}\mathbf{X}_{k-1} \\ &= \int_{\mathcal{G}} p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1}) p(\mathbf{X}_{k-1} \,|\, \mathbf{Y}_{1:k-1})\, d_{\mathcal{G}}\mathbf{X}_{k-1} \end{aligned} \tag{4}$$

Since $p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k}) = p(\mathbf{X}_k \,|\, \mathbf{Y}_k, \mathbf{Y}_{1:k-1})$, by applying Bayes' rule to the RHS, and using $p(\mathbf{Y}_k \,|\, \mathbf{X}_k, \mathbf{Y}_{1:k-1}) = p(\mathbf{Y}_k \,|\, \mathbf{X}_k)$ due to the observation equation, we obtain the update equation

$$\begin{aligned} p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k}) &= \frac{1}{z_k} p(\mathbf{Y}_k \,|\, \mathbf{X}_k, \mathbf{Y}_{1:k-1}) p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k-1}) \\ &= \frac{1}{z_k} p(\mathbf{Y}_k \,|\, \mathbf{X}_k) p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k-1}) \end{aligned} \tag{5}$$

where $z_k = p(\mathbf{Y}_k \,|\, \mathbf{Y}_{1:k-1})$ is a normalisation factor that depends only on the data. The observation noise $\mathbf{E}_{l\times q}$ has Gaussian entries, so $p(\mathbf{Y}_k \,|\, \mathbf{X}_k)$ is known, and thus, can be calculated.

Particle Filtering – Letting $\mathbf{S}_{k-1}^{(1)}, \ldots, \mathbf{S}_{k-1}^{(N)} \in \mathcal{G}$ denote the N samples (particles) drawn from $p(\mathbf{X}_{k-1} \,|\, \mathbf{Y}_{1:k-1})$, if $\widetilde{\mathbf{S}}_k^{(s)} \sim p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1} = \mathbf{S}_{k-1}^{(s)})$ for each $s = 1, \ldots, N$, then eq. (4) implies

Theorem 1. $\widetilde{\mathbf{S}}_k^{(1)}, \ldots, \widetilde{\mathbf{S}}_k^{(N)} \sim p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k-1})$ when $N \gg 1$.

Sampling from $p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1})$ so as to ensure that the samples remain in $\mathcal{G}$ will be discussed in Section 4. Defining the probabilities

$$p_{k,s} = \frac{p(\mathbf{Y}_k \,|\, \mathbf{X}_k = \widetilde{\mathbf{S}}_k^{(s)})}{\sum_{i=1}^{N} p(\mathbf{Y}_k \,|\, \mathbf{X}_k = \widetilde{\mathbf{S}}_k^{(i)})}, \quad s = 1, \ldots, N \tag{6}$$

and letting $\mathbf{S}_k^{(1)}, \ldots, \mathbf{S}_k^{(N)}$ denote the N samples (particles) chosen from $\widetilde{\mathbf{S}}_k^{(1)}, \ldots, \widetilde{\mathbf{S}}_k^{(N)} \in \mathcal{G}$ with probability $p_{k,s}$, eq. (5) implies

Theorem 2. $\mathbf{S}_k^{(1)}, \ldots, \mathbf{S}_k^{(N)} \sim p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k})$ when $N \gg 1$.

Remark 2. Choosing a sample from $\widetilde{\mathbf{S}}_k^{(1)}, \ldots, \widetilde{\mathbf{S}}_k^{(N)}$ with probability $p_{k,s}$ is done by constructing the cumulative distribution $\mathbb{P}_N$ (a stair-case function with steps at each s with height $p_{k,s}$), choosing a uniformly random point α between $p_{k,1}$ and $\sum_{s=1}^{N} p_{k,s}$, and selecting $\widetilde{\mathbf{S}}_k^{(j)}$, where $j = \max_i\{i : \mathbb{P}_N(i) \leq \alpha\}$.

Obtaining the State Estimator $\widehat{\mathbf{X}}_k$ – We can now use Theorem 2 to obtain $\widehat{\mathbf{X}}_k \in \mathcal{G}$ since we have that

$$\begin{aligned} \widehat{\mathbf{X}}_k &= \arg\min_{\mathbf{X}\in\mathcal{G}} \mathbb{E}\left[d(\mathbf{X}, \mathbf{X}_k)^2\right] \\ &\overset{(3)}{\approx} \arg\min_{\mathbf{X}\in\mathcal{G}} \frac{1}{N}\sum_{i=1}^{N} d(\mathbf{X}, \mathbf{S}_k^{(s)})^2, \quad \mathbf{S}_k^{(s)} \sim p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k}) \end{aligned} \tag{7}$$

4. SAMPLING FROM $p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1})$ BY SOLVING (1)

Obtaining an expression for $p(\mathbf{X}_k \,|\, \mathbf{X}_{k-1})$ to achieve sampling is not needed when a discrete solution of the SDE (1) can be derived on the small time interval $[t_{k-1}, t_k]$ – Since it is assumed that $\mathbf{X}_{k-1} \in \mathcal{G}$, the solution $\mathbf{X}(t) \in \mathcal{G}$ of (1) must have the form

$$\mathbf{X}(t) = \mathbf{X}_{k-1} \mathrm{e}^{\mathbf{\Omega}(t)}, \;\; \mathbf{\Omega}(t) \in \mathfrak{g}, \;\; \mathbf{\Omega}(t_{k-1}) = \mathbf{0} \tag{8}$$

by the definition of $\mathfrak{g}$, (see Section 1), and so, $\mathbf{X}_k = \mathbf{X}(t)|_{t=t_k}$. To obtain an expression for $\boldsymbol{\Omega}(t)$, we need to derive the differential of the solution, and equate it with the differential (SDE) in (1). So, by the Taylor series of $e^{\boldsymbol{\Omega}}$, note that

$$d\mathbf{X} = \mathbf{X}_{k-1}\left\{\frac{d}{d\epsilon}e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}} + \frac{1}{2}\frac{d^2}{d\epsilon^2}e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}} + \cdots\right\}\Big|_{\epsilon=0} \tag{9}$$

where $\epsilon d\boldsymbol{\Omega} \in \mathfrak{g}$ is a very small perturbation of $\boldsymbol{\Omega} \in \mathfrak{g}$. Note that $\boldsymbol{\Omega} + \epsilon d\boldsymbol{\Omega} \in \mathfrak{g}$ since $\mathfrak{g}$ is a vector space. Then, by standard calculus

$$\frac{d}{d\epsilon}e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}} \approx e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}} d\boldsymbol{\Omega} \tag{10}$$

By differentiating (10) we obtain $\frac{d^r}{d\epsilon^r}e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}} \approx e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}}(d\boldsymbol{\Omega})^r$ for $r \geq 1$. Next, let $\boldsymbol{\Omega}(t)$ obey the following general SDE

$$d\boldsymbol{\Omega} = \mathbf{A}\, dt + \sum_{i=1}^{d} \mathbf{B}_i\, dW_i \tag{11}$$

where $\mathbf{A}, \mathbf{B}_i$ are some function of $\boldsymbol{\Omega}$ and t. Since $\boldsymbol{\Omega} \in \mathfrak{g}$, (11) has to be an equation in $\mathfrak{g}$. So, since $\mathfrak{g}$ is a vector space, by definition we must have $\mathbf{A} \in \mathfrak{g}$ and $\mathbf{B}_i \in \mathfrak{g}$. We now need to find $\mathbf{A}$ and $\mathbf{B}_i$'s.

We firstly evaluate $(d\boldsymbol{\Omega})^2$, which has matrices scaled by the terms $(dt)^2$, $dt\, dW_i$ and $dW_i\, dW_j$ for each $i = 1, \ldots, d$ and $j = 1, \ldots, d$. Now, $dW_i \sim \mathcal{N}(0, dt)$ implies that the term $dt\, dW_i$ has a mean of 0 and variance of order $(dt)^3$. When $i \neq j$, dW_i and dW_j are i.i.d., hence the term $dW_i\, dW_j$ has a mean 0 and variance of order $(dt)^2$. Lastly, when $i = j$, the term $(dW_i)^2$ has a mean of dt and variance of order $(dt)^2$. Since dt is a very small quantity, we can assume $(dt)^p \approx 0$ for any $p \geq 2$. This implies that the terms $dt\, dW_i$, $dW_i\, dW_j$ have zero variance, i.e. can be treated as deterministic quantities, which therefore must be equal to their mean. So, we can conclude that

$$(dt)^2 = 0,\ dt\, dW_i = 0,\ dW_i\, dW_j = \begin{cases} 0 & \text{if } i \neq j \\ dt & \text{otherwise} \end{cases} \tag{12}$$

which are called Itō's rules, and so

$$(d\boldsymbol{\Omega})^2 \overset{(11)}{=} \left(\mathbf{A}\, dt + \sum_{i=1}^{d} \mathbf{B}_i\, dW_i\right)^2 \overset{(12)}{=} \sum_{i=1}^{d} \mathbf{B}_i^2\, dt \tag{13}$$

Thus, by applying (12) to the product of (11) and (13) we obtain that $(d\boldsymbol{\Omega})^3 = \mathbf{0}$, and so, for any $r \geq 3$ we must have $(d\boldsymbol{\Omega})^r = (d\boldsymbol{\Omega})^3(d\boldsymbol{\Omega})^{r-3} = \mathbf{0}$. Consequently, the derivative $\frac{d^r}{d\epsilon^r}e^{\boldsymbol{\Omega}+\epsilon d\boldsymbol{\Omega}}$ is zero for all $r \geq 3$, and hence, from (9) we obtain Itō's lemma (14)

Theorem 3. (Itō's lemma) For $\mathbf{X} = \mathbf{X}_{k-1}e^{\boldsymbol{\Omega}}$, where $\boldsymbol{\Omega}$ satisfies the SDE (11), we have

$$d\mathbf{X} = \mathbf{X}_{k-1}\left\{\frac{d}{d\epsilon}e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}} + \frac{1}{2}\frac{d^2}{d\epsilon^2}e^{\boldsymbol{\Omega}+\epsilon\, d\boldsymbol{\Omega}}\right\}\Big|_{\epsilon=0} \tag{14}$$

Next, substituting into (14) the derivative $\frac{d^r}{d\epsilon^r}e^{\boldsymbol{\Omega}+\epsilon d\boldsymbol{\Omega}}|_{\epsilon=0}$ with $r = 1, 2$, as well as the expressions for $d\boldsymbol{\Omega}$ and $(d\boldsymbol{\Omega})^2$ from (11) and (13) respectively, the differential reduces to

$$\begin{aligned} d\mathbf{X} &= \mathbf{X}_{k-1}e^{\boldsymbol{\Omega}}\, d\boldsymbol{\Omega} + \frac{1}{2}\mathbf{X}_{k-1}e^{\boldsymbol{\Omega}}(d\boldsymbol{\Omega})^2 \\ &= \mathbf{X}\left\{\mathbf{A} + \frac{1}{2}\sum_{i=1}^{d}\mathbf{B}_i^2\right\}dt + \sum_{i=1}^{d}\mathbf{X}\mathbf{B}_i\, dW_i \end{aligned} \tag{15}$$

We can finally equate (15) with the state SDE (1), resulting in

$$\mathbf{V}_0 = \mathbf{A} + \frac{1}{2}\sum_{i=1}^{d}\mathbf{B}_i^2 \ \text{ and } \ \mathbf{V}_i = \mathbf{B}_i,\ i = 1, \ldots, d$$

and so, noting that $\mathbf{V}_i(\mathbf{X}(t)) \overset{(8)}{=} \mathbf{V}_i\left(\mathbf{X}_{k-1}e^{\boldsymbol{\Omega}(t)}\right)$, we have

$$\mathbf{A}(\boldsymbol{\Omega}(t)) = \mathbf{V}_0\left(\mathbf{X}_{k-1}e^{\boldsymbol{\Omega}(t)}\right) - \frac{1}{2}\sum_{i=1}^{d}\mathbf{V}_i\left(\mathbf{X}_{k-1}e^{\boldsymbol{\Omega}(t)}\right)^2 \tag{16}$$

$$\mathbf{B}_i(\boldsymbol{\Omega}(t)) = \mathbf{V}_i\left(\mathbf{X}_{k-1}e^{\boldsymbol{\Omega}(t)}\right),\ i = 1, \ldots, d \tag{17}$$

which are in $\mathfrak{g}$. Hence, we immediately obtain two results

Theorem 4. If $\mathbf{X} \in \mathcal{G}$ from (8) solves (1), then (16) and (17) imply

$$\mathbf{V}_0 - \frac{1}{2}\sum_{i=1}^{d}\mathbf{V}_i^2 \in \mathfrak{g} \ \text{ and } \ \mathbf{V}_i \in \mathfrak{g},\ i = 1, \ldots, d \tag{18}$$

Theorem 5. If $\mathbf{X}(t) \in \mathcal{G}$ from (8) solves (1), then $\boldsymbol{\Omega}(t) \in \mathfrak{g}$ solves (11), where $\mathbf{A}$ and $\mathbf{B}_i$ are given by (16) and (17) respectively.

Hence, the geometry of the problem in the curved space $\mathcal{G}$ is managed by working in $\mathfrak{g}$ (the tangent plane). Treating $\mathfrak{g}$ as Euclidean space, see Section 1, we then apply standard linear algebra to solve the problem in $\mathfrak{g}$, i.e. we obtain $\boldsymbol{\Omega}(t_k)$ by solving/discretising (11)

$$\begin{aligned} d\boldsymbol{\Omega} &\approx \boldsymbol{\Omega}(t_k) - \boldsymbol{\Omega}(t_{k-1}) \\ &\overset{(11)}{=} \mathbf{A}(\boldsymbol{\Omega}(t_{k-1}))\Delta + \sum_{i=1}^{d}\mathbf{B}_i(\boldsymbol{\Omega}(t_{k-1}))\Delta W_{i,k-1} \end{aligned} \tag{19}$$

where $\Delta = t_k - t_{k-1}$ and $\Delta W_{i,k-1} \sim \mathcal{N}(0, \Delta)$. Recalling that $\boldsymbol{\Omega}(t_{k-1}) = \mathbf{0}$, from (16), (17) and (19), we obtain the algorithm

$$\begin{aligned} \boldsymbol{\Omega}(t_k) = &\left\{\mathbf{V}_0(\mathbf{X}_{k-1}) - \frac{1}{2}\sum_{i=1}^{d}\mathbf{V}_i(\mathbf{X}_{k-1})^2\right\}\Delta \\ &+ \sum_{i=1}^{d}\mathbf{V}_i(\mathbf{X}_{k-1})\Delta W_{i,k-1} \end{aligned} \tag{20}$$

$$\mathbf{X}_k = \mathbf{X}_{k-1}e^{\boldsymbol{\Omega}(t_k)} \tag{21}$$

which is the simplest Euler method on $\mathcal{G}$, see Figure 1. For a detailed derivation see our paper [34]. Special case examples of this Euler method are found in [8, 35]. The algorithm in [1, 6] for numerically solving (1) is also of Euler type, however, it does not have the correct update. Namely, it is missing the "Itō correction term" $\frac{1}{2}\sum_{i=1}^{d}\mathbf{V}_i^2$, see eq. (4) in [1], and eq. (13) in [6].

5. THE STATE ESTIMATION (PF) ALGORITHM

Let $\mathbf{X}(0) = \mathbf{X}_0$, and assume we can sample from $p(\mathbf{X}_0\,|\,\mathbf{Y}_{1:0}) = p(\mathbf{X}_0)$ at $t_0 = 0$.
Initialisation: For $s = 1, \ldots, N$, draw $\mathbf{S}_0^{(s)} \sim p(\mathbf{X}_0)$. Compute the estimator $\widehat{\mathbf{X}}_0$ using (7). Let $t = t_k$ and $k = 1$. Then:

(a) For $s = 1, \ldots, N$, using $\mathbf{S}_{k-1}^{(s)} \sim p(\mathbf{X}_{k-1}\,|\,\mathbf{Y}_{1:k-1})$, draw $\widetilde{\mathbf{S}}_k^{(s)} \sim p(\mathbf{X}_k\,|\,\mathbf{X}_{k-1} = \mathbf{S}_{k-1}^{(s)})$:

  (i) Using (20), let $\boldsymbol{\Omega}_s(t_k) = \boldsymbol{\Omega}(t_k)|_{\mathbf{X}_{k-1}=\mathbf{S}_{k-1}^{(s)}}$.

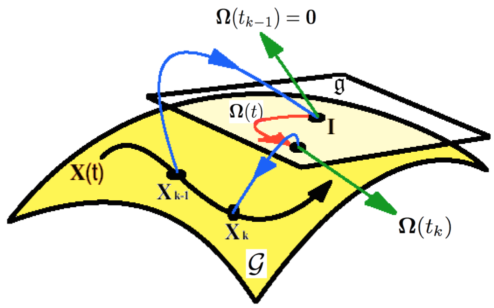

Fig. 1: Illustration of the Euler method ((20) & (21)) – The smooth function $\mathbf{X}(t) \in \mathcal{G}$ can be thought of as a flow on $\mathcal{G}$ (a smooth manifold, see Section 1). Given the current iterate $\mathbf{X}_{k-1} = \mathbf{X}(t_{k-1})$, an approximation of $\mathbf{X}(t) \in \mathcal{G}$ on the small time interval $[t_{k-1}, t_k]$ is $\mathbf{X}_k = \mathbf{X}(t_k)$. To obtain $\mathbf{X}_k$ we shift the SDE problem (1) from $\mathcal{G}$ to its Lie algebra $\mathfrak{g}$ (tangent at $\mathbf{I}$). The SDE problem in $\mathfrak{g}$, given by (11), is then solved on $[t_{k-1}, t_k]$, i.e. we obtain an approximation of $\boldsymbol{\Omega}(t) \in \mathfrak{g}$ at t_k. Using the exponential map $\mathrm{e} : \mathfrak{g} \to \mathcal{G}$ we then shift the solution $\boldsymbol{\Omega}(t_k)$ from $\mathfrak{g}$ back to $\mathcal{G}$, i.e. $\mathbf{X}_k = \mathbf{X}_{k-1}\,\mathrm{e}^{\boldsymbol{\Omega}(t_k)}$.

(ii) Using (21), compute $\widetilde{\mathbf{S}}_k^{(s)} = \mathbf{S}_{k-1}^{(s)}\,\mathrm{e}^{\boldsymbol{\Omega}_s(t_k)}$.

(b) For $s = 1, \dots, N$, draw $\mathbf{S}_k^{(s)} \sim p(\mathbf{X}_k \,|\, \mathbf{Y}_{1:k})$:

(i) Compute the probabilities $p_{k,s}$ in (6).

(ii) For each s draw a sample, denoted by $\mathbf{S}_k^{(s)}$, from the set $\{\widetilde{\mathbf{S}}_k^{(s)}\}_{s=1}^N$ with probability $p_{k,s}$ (see Remark 2).

(c) Compute the estimator $\widehat{\mathbf{X}}_k$ using (7).

(d) Let $k = k+1$ and $t = t_k$. Go to (a).

6. SIMULATION

Here we illustrate the above PF algorithm for $\mathcal{G} = SO(3)$, which is a set of 3×3 invertible matrices $\mathbf{X}$ such that $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ and $\det(\mathbf{X}) = 1$. The corresponding Lie algebra $\mathfrak{g} = \mathfrak{so}(3)$ is the set of 3×3 skew symmetric matrices, i.e. $\mathbf{A} \in \mathfrak{so}(3)$ implies $\mathbf{A}+\mathbf{A}^T = \mathbf{0}$.

Since the state $\mathbf{X} \in SO(3)$ satisfies the state dynamics SDE (1), we need to firstly use Theorem 4 to obtain the restrictions on $\mathbf{V}_i$'s in (1) for $i = 0, 1, \dots, d$. Letting $d = 3$, according to the theorem, $\mathbf{V}_i \in \mathfrak{so}(3)$ for $i = 1, 2, 3$, and so, $\mathbf{V}_i + \mathbf{V}_i^T = \mathbf{0}$ for $i = 1, 2, 3$, which implies the following structure holds

$$\mathbf{V}_i = \mathbf{S}(\mathbf{v}_i) = \begin{bmatrix} 0 & -v_{3,i} & v_{2,i} \\ v_{3,i} & 0 & -v_{1,i} \\ -v_{2,i} & v_{1,i} & 0 \end{bmatrix}, \quad \mathbf{v}_i = \begin{bmatrix} v_{1,i} \\ v_{2,i} \\ v_{3,i} \end{bmatrix} \tag{22}$$

for $i = 1, 2, 3$. Regarding $\mathbf{V}_0$, according to the theorem, $\mathbf{V}_0 - \frac{1}{2}\sum_{i=1}^3 \mathbf{V}_i^2 \in \mathfrak{so}(3)$, and so,

$$\mathbf{V}_0 - \frac{1}{2}\sum_{i=1}^{3}\mathbf{V}_i^2 + \left(\mathbf{V}_0 - \frac{1}{2}\sum_{i=1}^{3}\mathbf{V}_i^2\right)^T = \mathbf{0} \tag{23}$$

For $i = 1, 2, 3$ we have $(\mathbf{V}_i^2)^T = (\mathbf{V}_i^T)(\mathbf{V}_i^T) = (-\mathbf{V}_i)(-\mathbf{V}_i) = \mathbf{V}_i^2$, which when substituted in (23) gives $\mathbf{V}_0 + \mathbf{V}_0^T = \sum_{i=1}^3 \mathbf{V}_i^2$. Thus, $\mathbf{V}_0$ cannot be skew-symmetric, i.e. $\mathbf{V}_0 \notin \mathfrak{so}(3)$, unless $\mathbf{V}_i = \mathbf{0}$ for all $i = 1, 2, 3$. So, to satisfy the restrictions on the $\mathbf{V}_i$'s, here we simply let $\mathbf{V}_0 = -\mathbf{I}$ and $\mathbf{V}_i = \mathbf{S}(\mathbf{e}_i)$, $i = 1, 2, 3$, where $\mathbf{e}_i$ is a vector with 1 in the i-th entry and 0 in all other entries.

Next, defining $\mathbf{y}_0 = [0, 0, 1]^T$, we let the measurement function $\mathbf{C}(\mathbf{X}) = \mathbf{X}^T\mathbf{y}_0 \in \mathbb{R}^3$. In this case, the observation noise $\mathbf{E}_{3\times 1} \sim \mathcal{N}(\mathbf{0}_{3\times 1}, \boldsymbol{\Sigma}_{3\times 3})$, and so, in (6) we have

$$p(\mathbf{Y}_k|\mathbf{X}_k = \widetilde{\mathbf{S}}_k^{(s)}) \propto \mathrm{e}^{-\frac{1}{2}\left(\mathbf{Y}_k - \mathbf{C}\left(\widetilde{\mathbf{S}}_k^{(s)}\right)\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{Y}_k - \mathbf{C}\left(\widetilde{\mathbf{S}}_k^{(s)}\right)\right)}$$

We let $\boldsymbol{\Sigma} = 0.1 \times \mathbf{I}_{3\times 3}$. Note that $\mathbf{C}(\mathbf{X})^T\mathbf{C}(\mathbf{X}) = 1$ when $\mathbf{X} \in SO(3)$, and so, the observations lie on the unit sphere.

To obtain $\widehat{\mathbf{X}}_k$ in step (c) of the PF algorithm, we let the distance $d : SO(3) \times SO(3) \to \mathbb{R}$ in (7) be given by $d(\mathbf{X}_1, \mathbf{X}_2) = \|\mathbf{X}_1 - \mathbf{X}_2\|_F$, where $\|\cdot\|_F$ is the Frobenius norm. In this case, from [36] we have that $\widehat{\mathbf{X}}_k$ in (7) is given by

$$\widehat{\mathbf{X}}_k = \begin{cases} \mathbf{V}\mathbf{U}^T & \text{if } \det\left(\bar{\mathbf{R}}^T\right) > 0 \\ \mathbf{V}\,\mathrm{diag}([1, 1, -1])\,\mathbf{U}^T & \text{otherwise} \end{cases} \tag{24}$$

where $\mathbf{U}, \mathbf{V}$ are obtained form the singular decomposition of $\bar{\mathbf{R}}^T$, i.e. $\bar{\mathbf{R}}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T$, and $\bar{\mathbf{R}} = \frac{1}{N}\sum_{s=1}^N \mathbf{S}_k^{(s)}$. We let $N = 200$, and $k = 0, \dots, 200$.

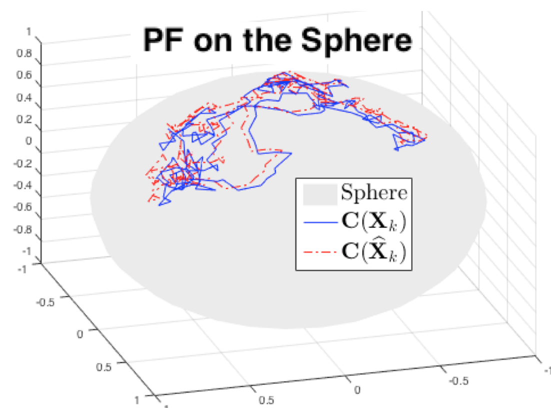

Fig. 2: Plotting $\mathbf{C}(\mathbf{X}_k) = [x_k, y_k, z_k]^T$ vs. its PF estimate $\mathbf{C}(\widehat{\mathbf{X}}_k) = [\widehat{x}_k, \widehat{y}_k, \widehat{z}_k]^T$. $\mathbf{X}_k$ is obtained by the Euler method ((20) & (21)). Since $\widehat{\mathbf{X}}_k \in SO(3)$, we see that $\mathbf{C}(\widehat{\mathbf{X}}_k)$ remain on the sphere. The initial set of particles are given by $\mathbf{X}_0\,\mathrm{e}^{\mathbf{S}(\mathbf{v}_s)}$, $s = 1, \dots, N$, where $\mathbf{v}_s \sim \mathcal{N}(\mathbf{0}_{3\times 1}, 5\times 10^{-3}\mathbf{I}_{3\times 3})$.

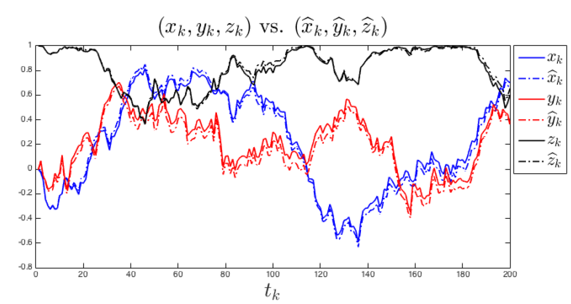

Fig. 3: Plotting $[x_k, y_k, z_k]^T$ vs. its PF estimates $[\widehat{x}_k, \widehat{y}_k, \widehat{z}_k]^T$.

7. CONCLUSION

We have demonstrated in an accessible way how the particle filters developed for state estimation in $\mathbb{R}^m$ can be extended to matrix Lie groups. A general numerical method has been derived, and an example simulation has illustrated this method for the $SO(3)$ Lie group.

8. REFERENCES

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

[2] C. Choi and H. I. Christensen, "Robust 3D visual tracking using particle filtering on the special Euclidean group: A combined approach of keypoint and edge features," *The International Journal of Robotics Research*, vol. 31, pp. 498–519, 2012.

[3] A. Srivastava, "Bayesian filtering for tracking pose and location of rigid targets," *Proc. SPIE, Signal Processing, Sensor Fusion, and Target Recognition IX*, vol. 4052, pp. 160–171, 2000.

[4] Y. M. Lui, "Advances in matrix manifolds for computer vision," *Image and Vision Computing*, vol. 30, pp. 380–388, 2012.

[5] A. Srivastava, "A Bayesian approach to geometric subspace estimation," *IEEE Trans. Signal Process.*, vol. 48, 2000.

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

[7] V. Solo, "Attitude estimation and Brownian motion on SO(3)," *Proc. IEEE CDC*, pp. 4857–4862, 2010.

[8] M. J. Piggott and V. Solo, "Stochastic numerical analysis for Brownian motion on SO(3)," *Proc. IEEE CDC*, pp. 3420–3425, 2014.

[9] W. Park, Y. Liu, Y. Zhou, M. Moses, and G. S. Chirikjian, "Kinematic state estimation and motion planning for stochastic nonholonomic systems using the exponential map," *Robotica*, vol. 26, pp. 419–434, 2008.

[10] G. S. Chirikjian and A. B. Kyatkin, *Engineering Applications of Noncommutative Harmonic Analysis*. CRC, Boca Raton, 2000.

[11] G. S. Chirikjian, *Sstochastic Models, Information Theory, and Lie Groups, Vol. 2*. Birkhäuser, Basel, 2012.

[12] B. Hall, *Lie groups, Lie algebras, and representations - An elementary introduction*. Springer, New York, 2003.

[13] A. Chiuso and S. Soatto, "Monte Carlo Filtering on Lie groups," *IEEE CDC*, pp. 304–309, 2000.

[14] R. W. Brockett, "Lie algebras and Lie groups in control theory," in *Geometric Methods in System Theory*. Riedel, 1973, pp. 43–82.

[15] M. Liao and L. Wang, "Motion of a rigid body under random perturbation," *Electron. Commun. Prob.*, vol. 10, pp. 235–243, 2005.

[16] D. Crisan and A. Doucet, "A Survey of Convergence Results on Particle Filtering Methods for Practitioners," *IEEE Trans. Signal Process.*, vol. 50, pp. 736–746, 2002.

[17] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: fifteen years later," 2011, http://www.cs.ubc.ca/~arnaud/doucet_johansen_tutorialPF.pdf.

[18] A. Srivastava and E. Klassen, "Monte Carlo extrinsic estimators of manifold-valued targets," *IEEE Trans. Signal Process.*, vol. 50, pp. 299–308, 2002.

[19] A. Srivastava, U. Grenander, G. R. Jensen, and M. I. Miller, "Jump-diffusion Marcov processes on orthogonal groups for object pose estimation," *J. Stat. Plan. Infer.*, vol. 103, pp. 15–37, 2002.

[20] H. Snoussi and C. Richard, "Monte Carlo Tracking on the Riemannian Manifold of Multivariate Normal Distributions," *IEEE DSP/SPE*, pp. 280–285, 2009.

[21] H. Snoussi and A. Mohammad-Djafari, "Particle Filtering on Riemannian Manifolds," *AIP Conf. Proc.*, vol. 872, 2006.

[22] Q. Rentmeesters, P.-A. Absil, P. Van Dooren, K. Gallivan, and A. Srivastava, "An efficient particle filtering technique on the Grassmann manifold," *IEEE ICASSP*, pp. 3838–3841, 2010.

[23] H. Munthe-Kaas, "Runge-Kutta methods on Lie groups," *BIT Numer. Math.*, vol. 38, pp. 92–111, 1998.

[24] ——, "Higher order Runge-Kutta methods on manifolds," *Appl. Numer. Math.*, vol. 29, pp. 115–127, 1999.

[25] A. Iserles, H. Z. Munthe-Kaas, S. P. Nørsett, and A. Zanna, *Lie group methods*. Acta Numerica, 2005.

[26] E. Hairer, C. Lubich, and G. Wanner, *Geometric Numerical Integration - Structure Preserving Algorithms for Ordinary Differential Equations*. Springer-Verlag, Berlin, 2002.

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

[28] S. J. A. Malham and A. Wiese, "An introduction to SDE simulation," 2010, http://arxiv.org/abs/1004.0646.

[29] D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," *SIAM Review*, vol. 43, pp. 525–546, 2001.

[30] K. Burrage and P. M. Burrage, "High Strong Order Methods for Non-commutative Stochastic Ordinary Differential Equation Systems and the Magnus Formula," *Phys. D*, vol. 133, pp. 34–48, 1999.

[31] P. M. Burrage, "Runge-Kutta methods for stochastic differential equations," 1999, PhD Thesis, http://espace.library.uq.edu.au/view/UQ:157833.

[32] S. J. A. Malham and A. Wiese, "Stochastic Lie Group Integrators," *SIAM J. Sci. Comput.*, vol. 30, pp. 597–617, 2008.

[33] X. Pennec, "Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements," *NSIP*, 1999.

[34] G. Marjanovic, M. J. Piggott, and V. Solo, "A simple approach to numerical methods for stochastic differential equations in Lie groups," *accepted in IEEE CDC*, 2015.

[35] M. J. Piggott and V. Solo, "Tikhonov-Galerkin stochastic system identification in SO(3)," *IEEE ICASSP*, 2015.

[36] M. Moakher, "Means and averaging in the group of rotations," *SIAM J. Matrix Analysis Appl.*, vol. 24, pp. 1–16, 2002.