AN ENGINEER'S GUIDE TO PARTICLE FILTERING ON THE STIEFEL MANIFOLD

Goran Marjanovic † *and Victor Solo* ‡

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

ABSTRACT

In many engineering applications the state of a dynamical system is modelled by a Stochastic Differential Equation (SDE) evolving in a "curved" (non-Euclidean) space such as the Stiefel manifold – the set of $n \times p$ real matrices with orthonormal columns, $(n \geq p)$. Due to the advances in computing power, the problem of state estimation can be efficiently addressed by the Particle Filter (PF). However, PF algorithms have to be completely reworked to handle the geometry, and the very few papers that properly deal with either the geometry or the stochastics of the problem are in the mathematics literature and are not accessible to an engineering audience. With this in mind and motivated by deterministic schemes on the Stiefel manifold, we give a direct accessible derivation of a novel PF algorithm for state estimation on the Stiefel manifold such that the resulting estimators always remain on the manifold. Our method can be applied to ANY dynamical system (SDE) on the Stiefel manifold. We do not rely on differential geometry or advanced stochastic calculus. Simulation examples are provided.

Index Terms— Sequential Monte Carlo methods, particle filter, stochastic processes, Stiefel manifold

1. INTRODUCTION

Numerous engineering applications deal with systems evolving in the Stiefel manifold $V_{p,n}$ – the set of $n \times p$ real matrices with orthonormal columns. They include signal processing [1], attitude control and filtering [2–5], image processing [6], optimization [7], robotics [8], etc. In this paper with consider the application of particle filtering (PF) to the state estimation problem on the Stiefel manifold. The state dynamics are modelled by a general stochastic differential equation (SDE) in the Stiefel manifold, while the (noisy) observations lie in Euclidean space. Lately, PF methods have become a very popular class of algorithms for state estimation, which is carried out sequentially as new observations become available. They are flexible, and can be applied to nonlinear and non-Gaussian dynamic models.

Prior Work – Until very recently PFs have only been addressing systems in Euclidean space ($\mathbb{R}^m$) – See [9, 10]. However, such filters cannot be considered for filtering problems in other manifolds ("curved" spaces) because the update schemes tend to immediately leave the manifold – See examples of this in [5, 11, 12].

The application of PF methods to state estimation in "curved" spaces is a relatively recent development. Some examples include [1, 4, 13–16]. [1] is the only paper to explicitly address state estimation on $V_{p,n}$, $(p \leq n)$, using a PF. Other papers only deal with the special case where $p = n$. Here, the Stiefel manifold is a matrix Lie group called the orthogonal group $O(n)$. In engineering, a frequently considered special case (subgroup) of $O(n)$ is the special orthogonal group $SO(n)$, and examples that deal with state dynamics evolving in $SO(n)$ are given in [4, 14, 15, 17].

Motivated by the work in [18], the state dynamics in [1] are not explicity given by a state space model but rather are implicitly described by a state transition probability density function (pdf) – the von Mises-Fischer distribution. There are two issues with this; Firstly, it is not clear how to relate a given SDE to the state transition pdf in [1]. Secondly, sampling from the von Mises-Fischer pdf can be difficult. The scheme we develop does not have these problems – The state dynamics are explicitly described by a general SDE, and it will be shown that the sampling procedures in the PF are very simple.

PFs for state estimation on arbitrary Riemannian manifold (which includes $V_{p,n}$) have also been proposed in [19] and [20]. However, these schemes are too abstract, do not explicitly deal with $V_{p,n}$, and so, extensive knowledge of differential geometry is required in order to specialise the procedures to $V_{p,n}$. In [19], the example provided considers state updates in $\mathbb{R}^m$, which is not extendable to updates on $V_{p,n}$. In [20], the state updates follow a multivariate affine generalised Hyperbolic distribution. It is not at all clear what SDE this distribution corresponds to.

Lastly, further existing literature dealing with the Stiefel manifold considers deterministic state dynamics [21, 22], i.e. the dynamical system is modelled by an ordinary differential equation (ODE). These works do not apply in the stochastic setting, where the available literature for engineers [23, 24] only deals with systems in $\mathbb{R}^m$. Dealing with geometric SDEs requires extensive knowledge in BOTH stochastic calculus and differential geometry – See [25–27].

Current Contribution – In this paper we derive a novel PF algorithm for state estimation on the Stiefel manifold $V_{p,n}$, where the state dynamics are described by a general SDE. The derivation is both direct and accessible to engineers. We do not rely on any differential geometry or stochastic processes theory. The development is in vein of our previous work [3, 4, 11, 12, 28] and Chirikjian's work [29], and is motivated by the ODE algorithms of [21, 22]. The numerical scheme we develop can be applied to any dynamical system (SDE) on the Stiefel manifold.

The paper is organised as follows: In Section 2 we state the problem of state estimation. Sections 3 and 4 describe the particle

† 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.

filtering 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 $V_{p,n}$

The State Space Model – We consider a matrix state lying in $V_{p,n}$ and evolving according to a general SDE

$$d\mathbf{X}(t) = \mathbf{F}_0(t,\mathbf{X})\,dt + \sum_{r=1}^{d} \mathbf{F}_r(t,\mathbf{X})\,dW_r(t) \tag{1}$$

where $\mathbf{F}_i$ coefficients are known matrix functions of time t and state $\mathbf{X}$, and $dW_r(t) \sim \mathcal{N}(0, dt)$ are i.i.d. Note that $\mathbf{F}_r$'s must satisfy constraints given in Theorem 1 in [12] in order for (1) to evolve in $V_{p,n}$, i.e. namely

$$\mathbf{F}_0^T\mathbf{X} + \mathbf{X}^T\mathbf{F}_0 = -\sum_{r=1}^{d} \mathbf{F}_r^T\mathbf{F}_r \tag{2}$$

$$\mathbf{F}_r^T\mathbf{X} + \mathbf{X}^T\mathbf{F}_r = \mathbf{0},\ r = 1,\ldots,d \tag{3}$$

When $p = n$, (2) and (3) reduce to the special case conditions in [4, 11].

Tangent of $V_{p,n}$ – Thinking of $V_{p,n}$ as a "curved" space allows us to envisage a tangent at a point $\mathbf{X} \in V_{p,n}$. It is given by [7]

$$\mathcal{T}_{\mathbf{X}}V_{p,n} = \{\mathbf{Z}_{n\times p} : \mathbf{Z}^T\mathbf{X} + \mathbf{X}^T\mathbf{Z} = \mathbf{0}\} \tag{4}$$

So, (3) implies that $\mathbf{F}_r \in \mathcal{T}_{\mathbf{X}}V_{p,n}$ for $r = 1,\ldots,d$. However, from (2) we see that $\mathbf{F}_0 \notin \mathcal{T}_{\mathbf{X}}V_{p,n}$, but it can be shown that the drift term $\mathbf{F}_0\,dt$ is a sum of two components: one in the tangent plane, and the other in the normal plane [12]. The latter we call the "pinning drift" because it "pins" or keeps the trajectory in $V_{p,n}$.

How does one derive (4)? Firstly, supposing that $p = 1$ implies $V_{p,n} = V_{1,n}$ is a hypersphere. So, a trajectory $\mathbf{x}(t) \in V_{1,n}$ satisfies $\mathbf{x}(t)^T\mathbf{x}(t) = 1$. Differentiating (w.r.t. time) gives $\dot{\mathbf{x}}^T\mathbf{x} = 0$, and recalling the elementary geometry that $\mathbf{x}(t)$ is normal to the hypersphere, we see that $\dot{\mathbf{x}}$ must be tangent to it. So, in a more general scenario, if $\mathbf{X}(t)$ evolves in $V_{p,n}$, then differentiating $\mathbf{X}(t)^T\mathbf{X}(t) = \mathbf{I}$ gives $\dot{\mathbf{X}}^T\mathbf{X} + \mathbf{X}^T\dot{\mathbf{X}} = \mathbf{0}$, where $\dot{\mathbf{X}}$ is an element of the tangent, which gives (4).

The Estimation Problem – For discrete times $t = t_1,\ldots,t_k$, define $\mathbf{X}_k = \mathbf{X}(t_k)$. At each time interval t_k let the noisy measurement $\mathbf{Y}_k \in \mathbb{R}^{u\times v}$ of the state $\mathbf{X}_k \in V_{p,n}$ be given by

$$\mathbf{Y}_k = \mathbf{C}(\mathbf{X}_k) + \mathbf{E}_k \tag{5}$$

where $\mathbf{C} : V_{p,n} \to \mathbb{R}^{u\times v}$ and $\mathbf{E}_k$ is the noise whose entries are i.i.d. zero mean Gaussian.

Given the observations $\{\mathbf{Y}_1,\ldots,\mathbf{Y}_k\} = \mathcal{Y}_1^k$ up to time $t = t_k$, the aim is to find the minimum Mean Square Error (MSE) estimate $\widehat{\mathbf{X}}_k$ of the state $\mathbf{X}_k$ for each $k \geq 1$. The MSE is defined by

$$\mathbb{E}\left[d(\mathbf{X},\mathbf{X}_k)^2\right] = \int_{V_{p,n}} d(\mathbf{X},\mathbf{X}_k)^2\,p(\mathbf{X}_k|\mathcal{Y}_1^k)\,d\mathbf{X}_k \tag{6}$$

where $d(\cdot,\cdot)$ is a distance between $\mathbf{X}$ and $\mathbf{X}_k$ on the "curved" space $V_{p,n}$, $p(\mathbf{X}_k|\mathcal{Y}_1^k)$ is the posterior probability density function (pdf), and $d\mathbf{X}_k$ is the infinitesimal area[1] on $V_{p,n}$ – See [18, 30].

[1]The Stiefel manifold admits a unit invariant measure whose differential form we denote by $d\mathbf{X}_k$.

Remark 1. (Monte Carlo (MC) approximation) The purpose of considering PF methods is to avoid calculating difficult integrals such as (6). So, we approximate them using random samples drawn from $p(\cdot)$ – Letting $\mathbf{S}^{(1)},\ldots,\mathbf{S}^{(N)} \sim p(\mathbf{Z})$, where $N \gg 1$, the following is an MC (sample average) approximation of the mean

$$\int_{V_{p,n}} \phi(\mathbf{Z})\,p(\mathbf{Z})\,d\mathbf{Z} = \mathbb{E}[\phi(\mathbf{Z})] \approx \frac{1}{N}\sum_{i=1}^{N}\phi(\mathbf{S}^{(i)}) \tag{7}$$

3. STATE ESTIMATION VIA PARTICLE FILTERING (PF)

Obtaining a state estimator $\widehat{\mathbf{X}}_k \in V_{p,n}$ by minimising (6) with respect to $\mathbf{X} \in V_{p,n}$ is usually not possible due to the difficulty in evaluating the integral. So, using the sample average (7) to approximate (6) leads to a much easier minimisation problem. This, however, involves directly sampling from the posterior $p(\mathbf{X}_k|\mathcal{Y}_1^k)$ and is difficult and/or inefficient in general because the relationship between the state and the observation might be very complicated. Thus, a recursive procedure is needed that takes the samples (particles) from the previous posterior $p(\mathbf{X}_{k-1}|\mathcal{Y}_1^{k-1})$ and transforms them into samples (particles) from the current posterior $p(\mathbf{X}_k|\mathcal{Y}_1^k)$. This is the essence of the PF algorithm.

Filtering Equations – Using samples from $p(\mathbf{X}_{k-1}|\mathcal{Y}_1^{k-1})$ to derive samples from $p(\mathbf{X}_k|\mathcal{Y}_1^k)$ requires obtaining a relation between the two posteriors. It is given by the standard filtering equations

$$p(\mathbf{X}_k|\mathcal{Y}_1^{k-1}) = \int_{V_{p,n}} p(\mathbf{X}_k|\mathbf{X}_{k-1})\,p(\mathbf{X}_{k-1}|\mathcal{Y}_1^{k-1})\,d\mathbf{X}_{k-1} \tag{8}$$

$$p(\mathbf{X}_k|\mathcal{Y}_1^k) \propto p(\mathbf{Y}_k|\mathbf{X}_k)\,p(\mathbf{X}_k|\mathcal{Y}_1^{k-1}) \tag{9}$$

Equation (8) is referred to as the prediction equation, and (9) is the update equation. These equations are well known (for 50 years [31]) when the system states evolve in $\mathbb{R}^m$ [9]. When dealing with systems in non-Euclidean ("curved") spaces the integration is no longer over $\mathbb{R}^m$ but rather over the "curved" space. Such examples of the filtering equations are found in [1, 4, 13–16, 32, 33], and their derivation relies on the Markovian property of the state due to (1), and the standard conditional pdf relations – See [4].

Particle Filtering – Let $\mathbf{S}_{k-1}^{(1)},\ldots,\mathbf{S}_{k-1}^{(N)} \in V_{p,n}$ denote the N particles (samples) drawn from the posterior $p(\mathbf{X}_{k-1}|\mathcal{Y}_1^{k-1})$. By approximating the prediction equation (8) with the MC approximation (7), the first stage involves obtaining particles from $p(\mathbf{X}_k|\mathcal{Y}_1^{k-1})$. So, if $\tilde{\mathbf{S}}_k^{(i)} \sim p(\mathbf{X}_k|\mathbf{X}_{k-1} = \mathbf{S}_{k-1}^{(i)})$ for $i = 1,\ldots,N$, then

Theorem 1. $\tilde{\mathbf{S}}_k^{(1)},\ldots,\tilde{\mathbf{S}}_k^{(N)} \sim p(\mathbf{X}_k|\mathcal{Y}_1^{k-1})$ when $N \gg 1$.

Sampling from the state transition pdf $p(\mathbf{X}_k|\mathbf{X}_{k-1})$ so as to ensure the resulting particles remain in $V_{p,n}$ is discussed in Section 4. The procedure is very simple and no assumptions are imposed on the state transition pdf as done in [1].

Obtaining particles from $p(\mathbf{X}_k|\mathcal{Y}_1^k)$ requires Theorem 1 and the update equation (9) – Define the "weight" quantities

$$w_k^i = \frac{p(\mathbf{Y}_k|\mathbf{X}_k = \tilde{\mathbf{S}}_k^{(i)})}{\sum_{j=1}^{N} p(\mathbf{Y}_k|\mathbf{X}_k = \tilde{\mathbf{S}}_k^{(j)})},\ i = 1,\ldots,N \tag{10}$$

and let $\mathbf{S}_k^{(1)}, \ldots, \mathbf{S}_k^{(N)}$ denote the N particles chosen (with replacement) from the set $\{\tilde{\mathbf{S}}_k^{(i)}\}_{i=1}^N$ according to probabilities $\{w_k^i\}_{i=1}^N$.[2] Then, (9) implies

Theorem 2. $\mathbf{S}_k^{(1)}, \ldots, \mathbf{S}_k^{(N)} \sim p(\mathbf{X}_k|\mathcal{Y}_1^k)$ when $N \gg 1$.

So, approximating (6) using (7), by Theorem 2 we have

$$\widehat{\mathbf{X}}_k \approx \arg\min_{\mathbf{X}\in V_{p,n}} \frac{1}{N}\sum_{i=1}^{N} d(\mathbf{X}, \mathbf{S}_k^{(i)})^2,\ \mathbf{S}_k^{(i)} \sim p(\mathbf{X}_k|\mathcal{Y}_1^k) \quad (11)$$

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

Sampling from $p(\mathbf{X}_k|\mathbf{X}_{k-1})$ can be done without actually knowing what this transitional pdf looks like. We only need to obtain $\mathbf{X}_k$ given $\mathbf{X}_{k-1}$, and this can be done by solving the SDE in (1) on the small time interval $[t_{k-1}, t_k]$. Both the geometry and the stochastics need to be taken into account to do this.

Solution of (1) – Denoting the solution of (1) by $\mathbf{X}(t)$, what form can $\mathbf{X}(t)$ have? We know that at $t = t_{k-1}$ it is given by $\mathbf{X}(t_{k-1}) = \mathbf{X}_{k-1} \in V_{p,n}$, and so, we need to ensure that $\mathbf{X}(t) \in V_{p,n}$ for $t \in (t_{k-1}, t_k]$. Letting $\mathbf{\Phi}_{n\times n}$ be an orthogonal matrix, note that

$$(\mathbf{\Phi}\mathbf{X}_{k-1})^T(\mathbf{\Phi}\mathbf{X}_{k-1}) = \mathbf{X}_{k-1}^T(\mathbf{\Phi}^T\mathbf{\Phi})\mathbf{X}_{k-1} = \mathbf{X}_{k-1}^T\mathbf{X}_{k-1} = \mathbf{I}$$

and so, setting $\mathbf{X}(t) = \mathbf{\Phi}(t)\mathbf{X}_{k-1}$ ensures $\mathbf{X}(t) \in V_{p,n}$. Then

Lemma 1. Any orthogonal $\mathbf{\Phi}_{n\times n}$ can be written as $\mathbf{\Phi} = \mathrm{e}^{\mathbf{\Omega}}$ for some skew-symmetric $\mathbf{\Omega}_{n\times n}$ (i.e. $\mathbf{\Omega} + \mathbf{\Omega}^T = \mathbf{0}$), where $\mathrm{e}^{(\cdot)}$ is the matrix exponential.

So, our "guess" solution for $t \in [t_{k-1}, t_k]$ becomes

$$\mathbf{X}(t) = \mathrm{e}^{\mathbf{\Omega}(t)}\mathbf{X}_{k-1},\ \mathbf{\Omega}(t_{k-1}) = \mathbf{0} \quad (12)$$

Solving $\mathbf{\Omega}^T + \mathbf{\Omega} = \mathbf{0}$ leads to $n(n-1)/2$ unknowns, and thus, $\mathbf{\Omega}$ is completely specified by a vector in $\mathbb{R}^{n(n-1)/2}$.

Remark 2. Consequently, we have transferred the problem of finding an $\mathbf{X}(t)$ in the Stiefel manifold (complicated geometry) to the problem of finding a skew-symmetric $\mathbf{\Omega}(t)$, or equivalently an $n(n-1)/2$-vector (simple geometry).

Now we have $\mathbf{X}_k = \mathbf{X}(t_k) = \mathrm{e}^{\mathbf{\Omega}(t_k)}\mathbf{X}_{k-1}$. We are given $\mathbf{X}_{k-1}$, hence to get $\mathbf{X}_k$ we simply need to find $\mathbf{\Omega}(t_k)$.

Finding $\mathbf{\Omega}(t_k)$ – Requires deriving the differential $d\mathbf{X}$ of the solution (12) and equating it with the differential $d\mathbf{X}$ from (1). This will give us an SDE in $\mathbf{\Omega}(t)$ (space of skew-symmetric matrices), which is easy to solve numerically,[3] allowing us to obtain $\mathbf{\Omega}(t_k)$.

So, by the Taylor series of $\mathrm{e}^{\mathbf{\Omega}}$, the differential of (12) is

$$d\mathbf{X} = \left\{\frac{d}{d\epsilon}\mathrm{e}^{\mathbf{\Omega}+\epsilon\, d\mathbf{\Omega}} + \frac{1}{2}\frac{d^2}{d\epsilon^2}\mathrm{e}^{\mathbf{\Omega}+\epsilon\, d\mathbf{\Omega}} + \cdots\right\}\mathbf{X}_{k-1}\bigg|_{\epsilon=0} \quad (13)$$

where $\epsilon\, d\mathbf{\Omega}$ is a small perturbation of the skew-symmetric $\mathbf{\Omega}$. Note that both $d\mathbf{\Omega}$ and $\mathbf{\Omega} + \epsilon\, d\mathbf{\Omega}$ are skew-symmetric since the space of skew-symmetric matrices is a vector space (equivalent to $\mathbb{R}^{n(n-1)/2}$). Then, by standard calculus

$$\frac{d^q}{d\epsilon^q}\mathrm{e}^{\mathbf{\Omega}+\epsilon\, d\mathbf{\Omega}} \approx (d\mathbf{\Omega})^q\,\mathrm{e}^{\mathbf{\Omega}+\epsilon\, d\mathbf{\Omega}},\ q \geq 1 \quad (14)$$

Now, we need to calculate $(d\mathbf{\Omega})^q$ for all $q \geq 1$, and to do this we suppose a very general structure for $d\mathbf{\Omega}$, i.e. without loss of generality we suppose that $\mathbf{\Omega}(t)$ obeys the following SDE on the interval $[t_{k-1}, t_k]$

$$d\mathbf{\Omega} = \mathbf{A}\, dt + \sum_{r=1}^{d}\mathbf{B}_r\, dW_r \quad (15)$$

where $\mathbf{A}$ and $\mathbf{B}_r$'s are matrices to be found. Since $d\mathbf{\Omega}$ is skew-symmetric, $\mathbf{A}$ and $\mathbf{B}_r$ must be skew-symmetric as well. By evaluating $(d\mathbf{\Omega})^q$ we can proceed to find $\mathbf{A}$ and $\mathbf{B}_r$.

We firstly evaluate $(d\mathbf{\Omega})^2$, and from (15) we see that it has matrices scaled by $(dt)^2$, $dt\, dW_r$ and $dW_r\, dW_s$. Now, $dW_r \sim \mathcal{N}(0, dt)$ implies that $dt\, dW_r$ has mean 0 and variance $\mathcal{O}((dt)^3)$. When $r \neq s$, dW_r and dW_s are i.i.d., hence the term $dW_r\, dW_s$ has mean 0 and variance $\mathcal{O}((dt)^2)$. Lastly, when $r = s$, the term $(dW_r)^2$ is χ^2 distributed, and hence has mean dt and variance $\mathcal{O}((dt)^2)$. Since dt is very small, we can assume $(dt)^2 = (dt)^3 = \cdots = 0$, which implies that the terms $dt\, dW_r$, $dW_r\, dW_s$ ($r \neq s$) have zero variance, i.e. can be treated as deterministic quantities, and thus, equal their respective means. So, we can conclude that

$$(dt)^2 = 0,\ dt\, dW_r = 0,\ dW_r\, dW_s = \begin{cases} 0 & \text{if } r \neq s \\ dt & \text{otherwise} \end{cases} \quad (16)$$

which are called Itō's rules. Using these

$$(d\mathbf{\Omega})^2 = \left(\mathbf{A}\, dt + \sum_{r=1}^{d}\mathbf{B}_r\, dW_r\right)^2 = \sum_{r=1}^{d}\mathbf{B}_r^2\, dt \quad (17)$$

By applying Itō's rules to the product of (15) and (17) we obtain that $(d\mathbf{\Omega})^3 = \mathbf{0}$. So, we see that $(d\mathbf{\Omega})^q = \mathbf{0}$ for any $q \geq 3$. Consequently by substituting this in (14), (13) reduces to (Itō's Lemma)

Theorem 3. $d\mathbf{X} = \left\{\frac{d}{d\epsilon}\mathrm{e}^{\mathbf{\Omega}+\epsilon\, d\mathbf{\Omega}} + \frac{1}{2}\frac{d^2}{d\epsilon^2}\mathrm{e}^{\mathbf{\Omega}+\epsilon\, d\mathbf{\Omega}}\right\}\mathbf{X}_{k-1}\Big|_{\epsilon=0}$

which is an exact expression (in the mean square sense). Substituting in Theorem 3 the 1st and 2nd derivatives of $\mathrm{e}^{\mathbf{\Omega}}$ as well as the expressions for $d\mathbf{\Omega}$ and $(d\mathbf{\Omega})^2$ respectively, the differential becomes

$$\begin{aligned} d\mathbf{X} &= (d\mathbf{\Omega})\,\mathrm{e}^{\mathbf{\Omega}}\mathbf{X}_{k-1} + \frac{1}{2}(d\mathbf{\Omega})^2\,\mathrm{e}^{\mathbf{\Omega}}\mathbf{X}_{k-1} \\ &= \left\{\mathbf{A} + \frac{1}{2}\sum_{r=1}^{d}\mathbf{B}_r^2\right\}\mathbf{X}\, dt + \sum_{r=1}^{d}\mathbf{B}_r\mathbf{X}\, dW_r \end{aligned} \quad (18)$$

To solve for $\mathbf{A}$ and $\mathbf{B}_r$, we equate (18) and (1), which gives

$$\mathbf{F}_0 = \left\{\mathbf{A} + \frac{1}{2}\sum_{r=1}^{d}\mathbf{B}_r^2\right\}\mathbf{X},\ \text{and } \mathbf{F}_r = \mathbf{B}_r\mathbf{X},\ r \geq 1 \quad (19)$$

Proceeding further requires the following simple result.

Theorem 4. Let $\mathbf{X} \in V_{p,n}$ and $\mathbf{M} \in \mathcal{T}_{\mathbf{X}}V_{p,n}$. For an $n \times n$ matrix $\mathcal{S}_{\mathbf{X}}(\mathbf{M}) = \mathbf{H}\mathbf{M}\mathbf{X}^T - \mathbf{X}\mathbf{M}^T\mathbf{H}$, where $\mathbf{H} = \left(\mathbf{I} - \frac{1}{2}\mathbf{X}\mathbf{X}^T\right)$, we have: (a) $\mathcal{S}_{\mathbf{X}}(\cdot)$ is skew-symmetric, and (b) $\mathcal{S}_{\mathbf{X}}(\mathbf{M})\mathbf{X} = \mathbf{M}$.

[2] Choosing samples from a set according to discrete probabilities is well known. For example, see Remark 2 in [4] to see how this can be done.

[3] using any numerical technique developed for SDEs in $\mathbb{R}^m$ – See [23]

Proof. Very simple – See [12]. Omitted due to lack of space.

Recalling from Section 2 that $\mathbf{F}_r \in \mathcal{T}_{\mathbf{X}} V_{p,n}$ for $r = 1, \ldots, d$, we can use Theorem 4 (b) to first obtain an expression for $\mathbf{B}_r$. We have that $\mathcal{S}_{\mathbf{X}}(\mathbf{F}_r)\mathbf{X} = \mathbf{F}_r = \mathbf{B}_r\mathbf{X}$, where the last equality is from (19). Thus, $\mathbf{B}_r = \mathcal{S}_{\mathbf{X}}(\mathbf{F}_r)$ and is indeed a skew-symmetric matrix by Theorem 4 (a). Next, since $\mathbf{A}$ is skew-symmetric note that $\mathbf{AX} \in \mathcal{T}_{\mathbf{X}} V_{p,n}$ because it satisfies the equation in (4). So, using Theorem 4 (b) we have $\mathbf{AX} = \mathcal{S}_{\mathbf{X}}(\mathbf{AX})\mathbf{X}$, which implies $\mathbf{A} = \mathcal{S}_{\mathbf{X}}(\mathbf{AX})$, indeed also skew-symmetric. But, from (19) we have

$$\mathbf{AX} = \mathbf{F}_0 - \frac{1}{2}\sum_{r=1}^{d} \mathbf{B}_r^2\mathbf{X} = \mathbf{F}_0 - \frac{1}{2}\sum_{r=1}^{d} \mathcal{S}_{\mathbf{X}}(\mathbf{F}_r)\mathbf{F}_r$$

and so, $\mathbf{A} = \mathcal{S}_{\mathbf{X}}\left(\mathbf{F}_0 - \frac{1}{2}\sum_{r=1}^{d} \mathcal{S}_{\mathbf{X}}(\mathbf{F}_r)\mathbf{F}_r\right)$.

Now that we have found expressions for $\mathbf{A}$ and $\mathbf{B}_r$ we substitute them into the SDE for $\mathbf{\Omega}$ in (15) and solve it numerically to obtain $\mathbf{\Omega}(t_k)$. Since the space of skew-symmetric matrices is equivalent to the Euclidean space we obtain $\mathbf{\Omega}(t_k)$ by standard linear algebra, i.e. $d\mathbf{\Omega} \approx \mathbf{\Omega}(t_k) - \mathbf{\Omega}(t_{k-1}) = \mathbf{A}(t_{k-1}, \mathbf{X}_{k-1})\Delta + \sum_{r=1}^{d} \mathbf{B}_r(t_{k-1}, \mathbf{X}_{k-1})\Delta W_{r,k-1}$, where $\Delta = t_k - t_{k-1}$ and $\Delta W_{r,k-1} \sim \mathcal{N}(0, \Delta)$. In general, $\mathbf{F}_r$'s are functions of $(t, \mathbf{X})$, thus, $\mathbf{A}$ and $\mathbf{B}_r$ are also functions of $(t, \mathbf{X})$. Finally, recalling that $\mathbf{\Omega}(t_{k-1}) = \mathbf{0}$ results in

(Geometric Euler) method – Obtain $\mathbf{X}_k$ given $\mathbf{X}_{k-1}$:

$$\mathbf{F}_{r,k} = \mathbf{F}_r(t_{k-1}, \mathbf{X}_{k-1}),\ r = 0, 1, \ldots, d \tag{20}$$

$$\mathbf{\Omega}(t_k) = \mathcal{S}_{\mathbf{X}_{k-1}}\left(\mathbf{F}_{0,k} - \frac{1}{2}\sum_{r=1}^{d} \mathcal{S}_{\mathbf{X}_{k-1}}(\mathbf{F}_{r,k})\mathbf{F}_{r,k}\right)\Delta + \sum_{r=1}^{d} \mathcal{S}_{\mathbf{X}_{k-1}}(\mathbf{F}_{r,k})\Delta W_{r,k-1} \tag{21}$$

$$\mathbf{X}_k = \mathrm{e}^{\mathbf{\Omega}(t_k)}\mathbf{X}_{k-1} \tag{22}$$

Clearly, since $\mathbf{\Omega}(t_k)$ is always skew-symmetric we have that the state iterates $\mathbf{X}_k$ always remain in $V_{p,n}$ (provided $\mathbf{X}_{k-1} \in V_{p,n}$).

5. STATE ESTIMATION (PF) ALGORITHM

Let $\mathbf{X}(0) = \mathbf{X}_0$, and at $t = t_0 = 0$ assume we can sample from the posterior $p(\mathbf{X}_0|\mathcal{Y}_1^0) = p(\mathbf{X}_0)$.

Initialisation: For $i = 1, \ldots, N$, draw $\mathbf{S}_0^{(i)} \sim p(\mathbf{X}_0)$. Compute the estimator $\widehat{\mathbf{X}}_0$ using (11). Let $t = t_k$ and $k = 1$. Then:

(a) For each $i = 1, \ldots, N$, using $\mathbf{S}_{k-1}^{(i)} \sim p(\mathbf{X}_{k-1}|\mathcal{Y}_1^{k-1})$, draw $\tilde{\mathbf{S}}_k^{(i)} \sim p(\mathbf{X}_k|\mathbf{X}_{k-1} = \mathbf{S}_{k-1}^{(i)})$ in the following way:

1. In (20) and (21) let $\mathbf{X}_{k-1} = \mathbf{S}_{k-1}^{(i)}$.
2. Compute $\tilde{\mathbf{S}}_k^{(i)}$ using (22), i.e. $\tilde{\mathbf{S}}_k^{(i)} = \mathrm{e}^{\mathbf{\Omega}(t_k)}\mathbf{S}_{k-1}^{(i)}$.

(b) Draw $\{\mathbf{S}_k^{(i)}\}_{i=1}^N \sim p(\mathbf{X}_k|\mathcal{Y}_1^k)$ in the following way:

1. Using $\{\tilde{\mathbf{S}}_k^{(i)}\}_{i=1}^N$ compute the weights $\{w_k^i\}_{i=1}^N$ in (10).
2. Draw N samples, denoted by $\{\mathbf{S}_k^{(i)}\}_{i=1}^N$, from $\{\tilde{\mathbf{S}}_k^{(i)}\}_{i=1}^N$ with probabilities $\{w_k^i\}_{i=1}^N$.

(c) Using $\{\mathbf{S}_k^{(i)}\}_{i=1}^N$, compute the estimator $\widehat{\mathbf{X}}_k$ in (11).

(d) Let k be $k + 1$, and go to (a).

6. SIMULATIONS

Here we illustrate the above PF algorithm. We let $\mathbf{C}(\mathbf{X}_k) = \mathrm{vec}(\mathbf{X}_k) = \mathbf{x}_k \in \mathbb{R}^{np}$ obtained by stacking columns of $\mathbf{X}_k$ on top of one another. So, $\mathbf{Y}_k = \mathbf{y}_k \in \mathbb{R}^{np}$ and $\mathbf{E}_k = \boldsymbol{\epsilon}_{np\times 1} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. Thus, in (10) we have $p(\mathbf{Y}_k|\mathbf{X}_k = \tilde{\mathbf{S}}_k^{(i)}) \propto \mathrm{e}^{-\frac{1}{2}\boldsymbol{\delta}_k^T\mathbf{\Sigma}^{-1}\boldsymbol{\delta}_k}$, where $\boldsymbol{\delta}_k = \mathbf{y}_k - \mathrm{vec}(\tilde{\mathbf{S}}_k^{(i)})$. To obtain $\widehat{\mathbf{X}}_k$ in step (c) of the PF algorithm we solve the minimisation problem in (11) using the simple Algorithm 1 from [6]. It requires the map $\mathrm{Exp} : \mathcal{T}_{(\cdot)}V_{p,n} \to V_{p,n}$ and its inverse $\mathrm{Exp}^{-1} = \mathrm{Log}$. Simple schemes (written in Matlab) to obtain these maps are Algorithm 2.1 and 3.1 respectively, in [34]. Finally, in the state SDE (1) we let $d = 1$, $\mathbf{F}_1 = \mathbf{P}_1\mathbf{X}$ and $\mathbf{F}_0 = \mathbf{XP}_0 - \frac{1}{2}\mathbf{XF}_1^T\mathbf{F}_1$, where $\mathbf{P}_0 \in \mathbb{R}^{p\times p}$ and $\mathbf{P}_1 \in \mathbb{R}^{n\times n}$ are arbitrary skew-symmetric matrices.[4]

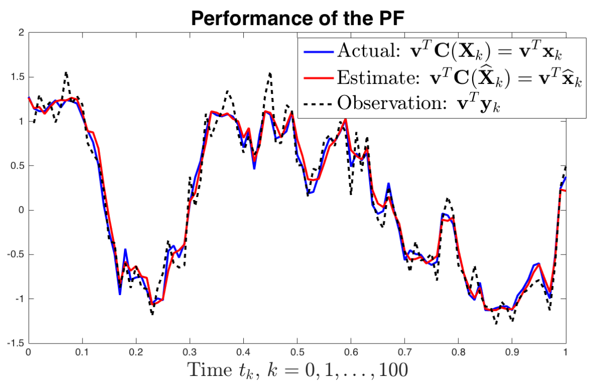

Fig. 1: PF performance, where $\mathbf{v}$ is an arbitrary np-vector. The initial particles $\mathbf{S}_0^{(i)}$'s are $\{\mathrm{e}^{\epsilon\mathbf{Q}_i}\mathbf{X}_0\}_{i=1}^N$, where $\mathbf{X}_0 = [\mathbf{I}_{p\times p}^T, \mathbf{0}^T]^T$, $\mathbf{Q}_i$'s are skew symmetric with non-zero entries $\sim \mathcal{N}(0, 1)$ and $\epsilon = 0.2$. We let $n = 3$, $p = 2$, $N = 1500$ and covariance $\mathbf{\Sigma} = 0.1 \times \mathbf{I} \in \mathbb{R}^{np\times np}$. Lastly, $\Delta = t_k - t_{k-1} = 1/100$.

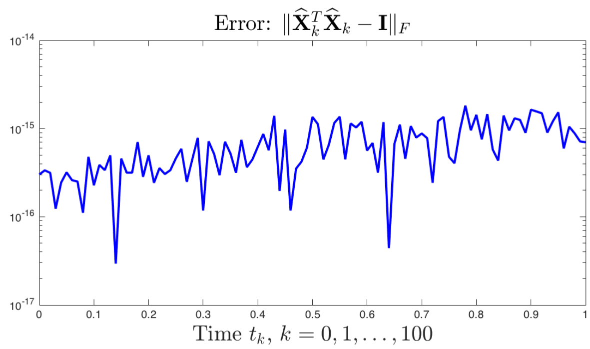

Fig. 2: Plotting the error in the PF estimator for the setup in Fig 1. As we can see the PF iterates remain in $V_{p,n}$.

7. CONCLUSION

We demonstrated in an accessible way how the particle filter developed for state estimation in $\mathbb{R}^m$ can be extended to the Stiefel manifold. A simple numerical method has been derived, and an example simulation has illustrated its usage.

[4] It can be easily verified that the chosen $\mathbf{F}_0$ and $\mathbf{F}_1$ satisfy the necessary conditions (2) and (3).

8. REFERENCES

[1] F. Tompkins and P. J. Wolfe, “Bayesian filtering on the Stiefel manifold,” *CAMPSAP*, pp. 261–264, 2007.

[2] J. Boulanger, S. Said, N. Le Bihan, and J. H. Manton, “Filtering from observations on Stiefel manifolds,” *Signal Processing*, vol. 122, pp. 52–64, 2016.

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

[4] G. Marjanovic and V. Solo, “An engineer’s guide to particle filtering on matrix Lie groups,” *IEEE ICASSP*, pp. 3969–3973, 2015.

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

[6] P. Turaga, A. Veeraraghavan, and R. Srivastava, A. Chellappa, “Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2273–2286, 2011.

[7] A. Edelman, T. A. Arias, and S. T. Smith, “The geometry of algorithms with orthogonality constraints,” *SIAM J. Matrix Anal. Appl.*, vol. 20, pp. 303–353, 1998.

[8] 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*, pp. 419–434, 2008.

[9] 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.

[10] A. Doucet and A. M. Johansen, “A tutorial on particle filtering and smoothing: fifteen years later,” 2011. [Online]. Available: http://www.stats.ox.ac.uk/~doucet/doucet_johansen_tutorialPF2011.pdf

[11] G. Marjanovic, M. Piggott, and V. Solo, “A Simple Approach to Numerical Methods for Stochastic Differential Equations in Lie Groups,” *IEEE CDC*, pp. 7143–7150, 2015.

[12] ——, “Numerical Methods for Stochastic Differential Equations in the Stiefel Manifold Made Simple,” *IEEE CDC, accepted*, 2016. [Online]. Available: https://www.researchgate.net/profile/Goran_Marjanovic4

[13] A. Srivastava, “Bayesian filtering for tracking pose and location of rigid targets,” *SPIE*, vol. 4052, pp. 160–171, 2000.

[14] 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.

[15] A. Chiuso and S. Soatto, “Monte Carlo filtering on Lie groups,” *IEEE CDC*, pp. 304–309, 2000.

[16] A. Srivastava and E. Klassen, “Bayesian and geometric subspace tracking,” *Advances in Applied Probability*, vol. 36, pp. 43–56, 2004.

[17] 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.

[18] Y. Chikuse, “State space models on special manifolds,” *Journal of Multivariate Analysis*, vol. 97, pp. 1284–1294, 2006.

[19] H. Snoussi and C. Richard, “Monte Carlo tracking on the Riemannian manifold of multivariate normal distributions,” *IEEE DSP/SPE*, pp. 280–285, 2009.

[20] H. Snoussi and A. Mohammad-Djafari, “Particle filtering on Riemannian manifolds,” *AIP Conf. Proc.*, vol. 872, 2006.

[21] E. Celledoni and B. Owren, “On the implementation of Lie group methods on the Stiefel manifold,” *Numerical Algorithms*, vol. 32, pp. 163–183, 2003.

[22] ——, “A class of intrinsic schemes for orthogonal integration,” *SIAM J. Numer. Anal.*, vol. 40, pp. 2069–2084, 2003.

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

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

[25] M. Emery, *Stochastic calculus on Manifolds*. Springer-Verlag, New-York, Berlin, 1989.

[26] P. Malliavin, *Stochastic analysis*. Springer, New-York, 1997.

[27] K. D. Elworthy, *Stochastic differential equations on manifolds*. London Math. Soc. Lecture Note Ser. 70, Cambridge University Press, London, 1982.

[28] V. Solo, “An approach to stochastic system identification in Riemannian manifolds,” *Proc. IEEE CDC*, pp. 6510–6515, 2014.

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

[30] X. Pennec, “Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements,” *IEEE DSP/SPE*, pp. 280–285, 2009.

[31] A. E. Bryson and Y. C. Ho, *Applied optimal control: optimization, estimation, and control*. Waltham, 1969.

[32] S. K. Ng and P. E. Caines, “Nonlinear filtering in Riemannian manifolds,” *IMA J. Math. Control and Inf.*, pp. 25–36, 1985.

[33] T. E. Duncan, “Some filtering results in Riemannian manifolds,” *Information and Control*, pp. 182–195, 1977.

[34] R. Zimmermann, “A matrix-algebraic algorithm for the Riemannian logarithm on the Stiefel manifold,” 2016. [Online]. Available: https://arxiv.org/pdf/1604.05054