MONTE-CARLO ESTIMATION FROM OBSERVATION ON STIEFEL MANIFOLD

J. Boulanger [1,2] *, *N. Le Bihan* [1,2] †, *S. Said* [2], *and J. H. Manton* [2] *Senior Member*

Gipsa-Lab, 11 Rue des mathématiques, 38402 Saint Martin d'Hères, France[1]
The University of Melbourne, Parkville, Victoria 3010, Australia[2]

ABSTRACT

Partial observation of stochastic processes can occur for various reasons, ranging from faulty sensors to occultation issues. In this paper, we consider the problem of estimating the angular velocity of a rotating system from partial observation corrupted by noise. The system is assumed to evolve on the rotation group $SO(n)$, and only k noisy measurements with $k < n$ are available. We propose an optimal filter to track the angular velocity.

We show that, under some conditions, it is possible to recover the angular velocity of the rotating system and we propose a solution based on a Monte-Carlo method (particle filter). In particular, we show that if the angular velocity is stepwise constant, our algorithm succeed in estimating it. Simulations illustrate the proposed approach.

***Index Terms*—** Stochastic process, Stiefel manifold, Rotation group, angular velocity estimation, Partial observation, Particle filtering.

I. INTRODUCTION

Being able to estimate the angular velocity of a spacecraft is an important issue in control theory [1]. In standard cases, the angular velocity is estimated from observing the orientation of the spacecraft. The orientation can then be modeled as a process on $SO(3)$, the set of 3×3 rotation matrices. More generally, we consider here the case where the process evolves on the set of $n \times n$ matrices, *i.e.* the special orthogonal group $SO(n)$. Such processes also appear in problems linked with computer vision for example [2], [3]. In practice, the sensors used to determine the orientation can be faulty or only partial observation may be available. We consider in this article the case of partial observation from a rotation process, given by the k first columns of the matrix process. In this case, the observation can be modeled as a process on the Stiefel manifold $V_{n,k}$.

For a process on $V_{n,k}$, the additive noise model used in [4] or [5] does not hold anymore as the noise is multiplicative. That is noticeable that as opposed to [6], we do not consider the problem of estimating a process evolving on $V_{n,k}$ from a noisy $n \times k$ observation matrix. In our case, the noise acts as a control noise on the observation.

This article first presents the geometry of the Stiefel manifold $V_{n,k}$, based on the geometry of the special orthogonal group $SO(n)$. Then, the observation process is modeled by a stochastic differential equation on $V_{n,k}$. A Monte-Carlo solution is then presented and implemented in Section IV and Section V. A discussion concerning the limitations of the presented algorithm can be found in Section V.

II. GEOMETRY OF STIEFEL MANIFOLDS

In this section, we give a rapid introduction to some differential geometry concepts that will prove to be useful in the following sections. For a more detailed review on Stiefel manifolds, see for example [7][8]. The Stiefel manifold $V_{n,k}$ is the set of matrices P of dimension $n \times k$ for $k \leq n$ such that

$$P^T P = I_k.$$

In the case $k = 1$, $V_{n,k}$ is the unit sphere in $\mathbb{R}^n$, denoted S^{n-1}. In the case where $k = n$, $V_{n,k}$ is the group of orthogonal matrices, *i.e.* $O(n)$. For simplicity, we restrict ourself to the case $k \leq n-1$. Recall that $SO(n)$ is the set of orthogonal matrices with positive unit determinant (also called the *rotation group*). Let Π be the projection from $SO(n)$ into $V_{n,k}$ defined as the operation that consists in truncating the $n-k$ last columns of a rotation matrix

$$\Pi(R) = P \text{ with } P \in V_{n,k},\ R \in SO(n).$$

In the next sections, the observed process is considered as the image of a rotation process via the projection Π.

The projection is clearly surjective, *i.e* $\Pi\left(SO(n)\right) = V_{n,k}$ but is not injective (except for $k = n-1$), *i.e* for a element $P \in V_{n,k}$, we can find different matrices $R_1, R_2 \in SO(n)$ such that $\Pi(R_1) = \Pi(R_2) = P$. It is noticeable that Π is left invariant: for any $R_1, R_2 \in SO(n)$, $\Pi(R_1 R_2) = R_1 \Pi(R_2)$. Given a (matrix) element P of $V_{n,k}$, the space tangent to $V_{n,k}$ at P, denoted $T_P V_{n,k}$ can be described as the projection of the space tangent to $SO(n)$ at $R \in SO(n)$ such that $\Pi(R) = P$ as $T_P V_{n,k} = d\Pi_R\left(T_R SO(n)\right)$, with $d\Pi_R$ the differential of Π at the point R.

*J. Boulanger is supported by the Rhônes-Alpes region through the CMIRA Explora-Doc scholarship program.

†N. Le Bihan is supported by the ERA, European Union, through the International Outgoing Fellowship (IOF GeoSToSip 326176) program of the 7th PCRD.

One can show that $T_P V_{n,k}$ can also be written

$$T_P V_{n,k} = \{\sigma P | \sigma \in \mathfrak{so}(n)\}$$

where $\mathfrak{so}(n)$ is the set of $n \times n$ skew symmetric matrices, and also remarkable for being the Lie algebra associated to the Lie group $SO(n)$. This naturally leads to the definition of the map $\chi : \mathfrak{so}(n) \times V_{n,k} \to TV_{n,k}$ as

$$\chi(\sigma, P) = \sigma P.$$

For a matrix R such that $\Pi(R) = P$, let $\mathcal{V}_R$ be $\mathcal{V}_R = \ker(d\Pi_R)$ where $\ker$ denotes the kernel. In other words,

$$\mathcal{V}_R = \{\sigma | d\Pi_R(\sigma) = 0\}.$$

The space $\mathcal{V}_R$ is called vertical space. Its orthogonal complement in $T_R SO(n)$ is denoted $\mathcal{H}_R$ and is called horizontal space. We can show that $T_P V_{n,k}$ and $\mathcal{H}_R$ are isomorphic. It signifies that for each vector tangent to $V_{n,k}$ at P, we can associate a unique vector tangent to R in the horizontal subspace. The image of $v \in T_P V_{n,k}$ via this isomorphism is denoted $v^{\mathcal{H}} \in T_R SO(n)$. At the end, only the components of the angular velocity in the horizontal space will act on P and will be estimated, as the vertical space does not act on P (it is the kernel of $d\Pi_R$).

Denoting χ^{-1} the application

$$\chi^{-1}(v, P) = v^{\mathcal{H}} R^T, \text{ with } \Pi(R) = P, \tag{1}$$

we can show that χ^{-1} is well defined, *i.e* χ^{-1} does not depend on the choice of R (as Π is non-injective, P has several antecedents). This is due to the fact that $\mathcal{H}_R$, by definition, depends on the choice of R. Obviously, the horizontal space considered in (1) should be associated with the same R that is used in (1). An illustration of the horizontal and vertical spaces are presented in Figure 1. This application will be used to construct a rotation process in $SO(n)$ based on a process in $V_{n,k}$ for completing the filtering.

This leads to the definition of an inner product (and the associated norm) in $T_P V_{n,k}$ as

$$\begin{aligned} < v_1, v_2 > &= < v_1^{\mathcal{H}}, v_2^{\mathcal{H}} >_{\mathfrak{so}(n)} \\ &= \frac{1}{2} \operatorname{trace}\left((v_1^{\mathcal{H}})^T v_2^{\mathcal{H}}\right). \end{aligned}$$

This notion will be used in the determination of the likelihood in the filtering section.

III. MODEL OF OBSERVATION

Rotation processes are involved in several engineering applications such as mechanics [1] or computer vision for example [2], [3]. If the system modeled by a rotation matrix is evolving with time, determining the angular velocity of this rotation process can be useful to determine the dynamics of the system.

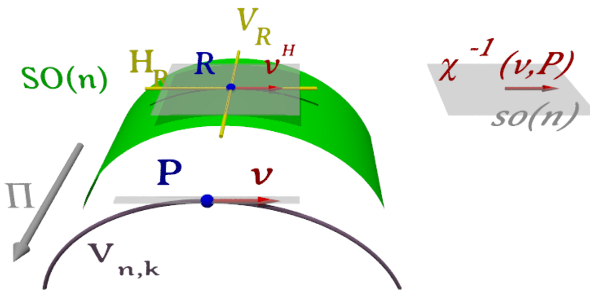

Fig. 1. Illustration of the different notions introduced to describe the Stiefel manifold $V_{n,k}$ as the image from the projection Π of $SO(n)$. The horizontal $\mathcal{H}_R$ and vertical $\mathcal{V}_R$ spaces are dependent on the chosen pre-image but the translation into $\mathfrak{so}(n)$ via χ^{-1} is invariant with respect to the choice of the preimage.

In the absence of noise, by completely observing the rotation process, the angular velocity can be determined [9]. In this article, we consider the case of a partial observation. This observation model can be used when observation sensors are faulty for example. The rotation process is denoted $R_t \in SO(n)$, where t represents time and we consider that only the first k columns are observable. The observation process is denoted $P_t \in V_{n,k}$. By definition, $P_t = \Pi(R_t)$. For a constant angular velocity $x \in \mathfrak{so}(n)$ and in the absence of noise, P_t is solution of the differential equation

$$dP_t = (x dt)\, P_t.$$

Given an initial condition P_0, we have $P_t = \exp(xt)P_0$, with $\exp$ the exponential map $\exp : \mathfrak{so}(n) \to SO(n)$. It is noticeable that knowing the value of x, the value of P_t can then be directly computed. Now, we consider that the observation is corrupted by noise. Despite that the observation is still an element of the Stiefel manifold, we consider a noise model acting like a noisy command. For a non-constant angular velocity x_t, the term xdt in the previous equation is replaced by $x_t dt + \circ dw_t$, where $w_t \in \mathfrak{so}(n)$ is a white noise with a variance σ^2. P_t is then solution of the stochastic differential equation

$$dP_t = (x_t dt + \circ dw_t)\, P_t, \tag{2}$$

where $\circ$ denotes the Stratonovich integral. For the rest of the article, x_t is considered as a Markov process, with a kernel function $q_{\delta t}(\ .\ , x_t)$ for any time interval δt at a time t. It is also assumed that x_t is a random variable with finite variance. In the next section, the filtering problem is considered and treated based on the application χ^{-1} previously introduced.

IV. FILTERING

The problem considered in this article is the estimation of x from the observation of P_t defined in Equation (2).

Because of the noise, the actual value cannot be retrieved and we are looking at π_t, the distribution of x given the observation until time t, *i.e* given $\mathcal{P}_t = \{P_s, s \leq t\}$. From π_t, an estimation based on the first moment of π_t can easily be extracted.

Due to the model used in Equation (2), usual solution like a Kalman filter cannot be used directly. Indeed, these solutions rely on the independence of the increment (conditioned by x) of the observed process. With an observation in $V_{n,k}$, this is not the case anymore as the noise is multiplicative. The solution presented in this article is based on the construction of a process p_t similar to P_t in terms of information but satisfying an additive noise model. Then, classic solutions can be applied, using p_t instead of P_t as an observation process.

Let $p_t \in \mathfrak{so}(n)$ be the solution of the differential equation

$$dp_t = \chi^{-1}(\circ dP_t, P_t) \tag{3}$$

The process p_t is called the antidevelopment of P_t. First of all, we show that this process is indeed similar to P_t in terms of information, or in other words, that this process is one-to-one defined with respect to P_t. As p_t is constructed from P_t, it is enough to show that P_t can be constructed back from its antidevelopment p_t. This is the case as P_t is solution, by definition of χ and χ^{-1} of

$$dP_t = \chi\left(\chi^{-1}(\circ dP_t, P_t), P_t\right)$$

Replacing $\chi^{-1}(\circ dP_t, P_t)$ by $\circ dp_t$ gives

$$dP_t = \chi(\circ dp_t, P_t).$$

Therefore, P_t can be constructed back from p_t.

Second, we need to show that p_t is a process associated to an additive noise model. Replacing the term $\circ dP_t$ in (3) by its expression from (2) gives

$$dp_t = \chi^{-1}(\chi(x_t, P_t), P_t)dt + \chi^{-1}(\chi(\circ dw_t, P_t), P_t).$$

Now, using notations $H_t = \chi(x_t, P_t)$ and $d\beta_t = \chi^{-1}(\chi(\circ dw_t, P_t), P_t)$, the last equation reads

$$dp_t = \chi^{-1}(H_t, P_t) + \circ d\beta_t.$$

It can be shown [9] that β_t is a Brownian motion in $\mathfrak{so}(n)$.

In this case, we can show [4] that the expression of π_t can be obtained in the following form, for an arbitrary test function ϕ

$$\pi_t(\phi) = \frac{\rho_t(\phi)}{\rho_t(1)} \tag{4}$$

where $\pi_t(\phi) = \mathbb{E}[\phi(x_t)|\mathcal{P}_t]$ and $\rho_t(\phi) = \mathbb{E}[\phi(x'_t)L_t(x', P)|\mathcal{P}_t]$. The function ρ_t is a non-normalized version of π_t. In practice, ϕ is usually chosen as $\phi(x) = x_t$ in order to get the first moment of π_t, which is the optimal estimator for x_t with respect to the mean square error. The process x'_t is a copy, in term of distribution, of the process x_t but independent from P_t. The likelihood L_t is defined as

$$L_t(x', P) = \exp\left(\frac{1}{\sigma^2}\int_0^t < H'_s, dP_s > -\frac{1}{2\sigma^2}\int_0^t ||H'_s||^2 ds\right) \tag{5}$$

with $H'_t = \chi(x'_s, P_s)$.

Despite that the solution presented in (4) gives an expression of π_t, the conditional distribution of x_t, we are looking for an adaptive solution to complete real-time filtering.

V. NUMERICAL IMPLEMENTATION

In this section, we present a Monte-Carlo method to generate an adaptive solution. From Equation (4), it is enough to find an expression of $\rho_t(\phi)$, the conditional distribution can then be easily determined.

The expectation in the definition of ρ_t is approximated by an empirical average ρ_t^N

$$\rho_t^N(\phi) = \frac{1}{N}\sum_{i \leq N} \phi(X_t^i) L_t(X^i, P),$$

where the processes X^i are independent copies of x (same distribution), independent from the observation P_t. These processes are called particles [10] and will be used as candidates to the actual value of x_t. To determine $L_t(X^i, P)$, the whole process P_t is needed.

However, in practice, only discrete samples are available. By calling δt the sampling time and $n = \lceil t\delta t \rceil$, the unnormalized distribution ρ_t is then approximated by

$$\rho_n^N(\phi) = \frac{1}{N}\sum_{i \leq N} \phi(X_n^i) L_n(X^i, P)$$

where L_n is the discrete version of L_t, the integrals being replaced by Riemannian sums

$$L_n(X^i, P) = \exp\left(\frac{1}{\sigma^2}\sum_{k \leq n} < H_k^i, \delta P_k > -\frac{1}{2\sigma^2}\sum_{k \leq n} ||H_k^i||^2 \delta t\right).$$

It is noticeable that the likelihood can be written into a recursive form

$$L_n(X^i, P) = L_{n-1}(X^i, P) l_n(X_n^i, P_n)$$

with $l_n(X_n^i, P_n) = \exp\left(\frac{1}{\sigma^2} < H_n^i, \delta P_n > -\frac{1}{2\sigma^2}||H_n^i||^2 \delta t\right)$.

After normalizing ρ_n^N, we get an approximate distribution of π_t. This leads to Algorithm 1 to estimate π_t as a weighted sum of particles.

The normalization step (Step 3) is not only here to compute π_n^N instead of ρ_n^N but also to numerically stabilize the computation of the weights.

Step 4 is called resampling. It is here to prevent a degeneracy due to the finite number of particles.

The resampling step consists in killing the particles far away from x_t (in fact, killing the particles with low weights) and cloning the remaining ones. To measure if the particles

Algorithm 1 Particle filter algorithm

- For the initialization, generate particles from a priori p_0: $X_0^i \sim p_0$ and set $w_0^i = 1/N$.
- At a time $n > 0$:
  1) Propagate the particles $X_n^i \sim q_{\delta t}(\, . \, , X_{n-1}^i)$
  2) Update the weight w_n^i of each particle as:
  $$w_n^i = w_{n-1}^i l_n(X_n^i, P_n).$$
  3) Normalize the weights: $w_n^i = w_n^i / \sum_j w_n^j$
  4) If $\left(\sum_i (w_n^i)^2\right)^{-1} < N/2$, generates
  $$[m^1 \; ... \; m^N] \sim \text{multinomial}(w^1 \; ... \; w^N)$$
  such that $\sum_i m^i = N$. Then, clones X_n^i m^i-times and set $w_n^i = \frac{1}{N}$.
  5) Estimate $\pi_t(\phi)$ as $\pi_n^N(\phi) = \sum_i \phi(X_n^i) w_n^i$

are scattered away from x_t, one commonly used criterion is a threshold based on the Effective Sample Size (ESS_w) defined as
$$ESS_w = \left(\sum_i (w_n^i)^2 \right)^{-1}.$$

Results from a simulation of this algorithm are presented in Figure 2. For this simulation, the chosen Stiefel manifold was the sphere $V_{3,1} = S^2$. The process $x_t \in \mathfrak{so}(3)$ is a stair function. The variance of the noise is fixed to $\sigma^2 = 1$.

To approximate x_t, $N = 500$ particles are generated from a normal prior distribution p_0 centered around the origin with a variance 2. Using more particles does not significantly improve the results. Considering a bad prior for generating the particles is not a big issue as the resampling step quickly eliminates the wrong candidates for the estimation. The time step for the observation is $\delta t = 10^{-2}$.

Figure 2 illustrates the results obtained for the estimation when the state x_t is a stair function. As x_t is constant, the algorithm is able to estimate properly x_t. Even if this has not been implemented for the simulation in Figure 2, one could use a classical algorithm to detect abrupt changes in x_t in order to estimate the time where the particles should be sampled [11]. When a change is detected, the particles are sampled from the initial priori again to converge towards the new value. In our simulation, the particles were simply sampled again from p_0 at the time of variation of x_t.

In the case of a constant angular velocity, the particles will not drift away because they are at a constant position (as they propagate with the same model as x_t). They only merge together when they are resampled (Step 4 in Algorithm 1), letting less and less different candidates for x_t. As the vertical component of x_t has no effect on P_t (by definition of the vertical space), the vertical part can not be estimated at a given time, only the horizontal components can be estimated. As P_t will evolve on $V_{n,k}$, the vertical space will

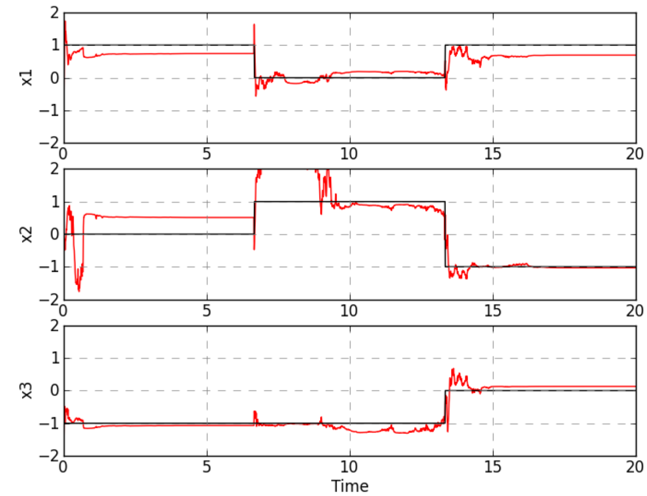

Fig. 2. Evolution of the estimation (red) of each component of x_t (black), namely x_1, x_2 and x_3. The process x_t is here chosen as a stair function. As the present algorithm can estimate a constant state, it is also able to estimate a stair function, by resampling properly the particles when x_t varies abruptly (detected by general decreasing likelihood for example).

change too and the component on the initial vertical space can be estimated. In the end, it will be possible to completely estimate the angular velocity.

However, in the case when x_t is a diffusion for example, if the variance of x_t is too high or the sampling rate too low, the particles will not have the time to converge and make a proper estimation. Indeed, contrary to the previous case, the vertical component is not constant (as x_t is not constant) and during the time that P_t evolves on $V_{n,k}$, the vertical component will be different, making its estimation impossible.

VI. CONCLUSION

In this article, we have presented a model of process on the Stiefel manifold $V_{n,k}$ to represent partial observations of a rotation process on $SO(n)$. The problem of estimating the angular velocity from partial noisy observations was considered. Contrary to the case usually presented in literature, our noise model is multiplicative. This specificity, together with partial observations, prevents us from using traditional resolution methods. A time-continuous solution has been presented and a Monte-Carlo implementation proposed, based on a particle filter. We showed that the angular velocity can be estimated in the case of a stair function model and that only a partial estimation can be obtained for more complicated cases. Further direction for this work consists in validation of the proposed method on experimental datasets.

VII. REFERENCES

[1] A. Astolfi and M. Lovera, "Global spacecraft attitude control using magnetic actuators," in *American Control Conference*, 2002, vol. 2, pp. 1331–1335 Vol.2.

[2] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[3] P. Turaga, A. Veeraraghavan, and R. Chellappa, "Statistical analysis on stiefel and grassmann manifolds with applications in computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[4] A.H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[5] R.E. Kalman and R.S. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME. Series D, Journal of Basic Engineering*, vol. 83, pp. 95–107, 1961.

[6] F. Tompkins and P.J. Wolfe, "Bayesian filtering on the stiefel manifold," in *Computational Advances in Multi-Sensor Adaptive Processing*, dec. 2007, pp. 261 –264.

[7] A. Edelman, T.A. Arias, and S.T. Smith, "The geometry of algorithms with orthogonality constraints," *Siam J. Matrix Anal. Appl*, vol. 20, no. 2, pp. 303–353, 1998.

[8] J.H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Transactions on Signal Processing*, 2002.

[9] S. Said and J.H. Manton, "Filtering with observation in a manifold: reduction to a classical filtering problem," *SIAM Journal on Control and Optimization*, vol. 51, no. 1, pp. 767–783, 2012.

[10] A. Doucet and A.M. Johansen, "A tutorial on particle filtering and smoothing: fifteen years later," 2011.

[11] M. Basseville and I. Nikiforov, *Detection of abrupt changes: theory and application*, System Sciences Series. Prentice Hall Information, 1993.