NONLINEAR STATE ESTIMATION USING PARTICLE FILTERS ON THE STIEFEL MANIFOLD

Claudio J. Bordin Jr., Universidade Federal do ABC**, and Marcelo G. S. Bruno,** Instituto Tecnológico de Aeronáutica.

1. Introduction

- Many engineering applications, such as attitude estimation, image processing, robotics, lead to models whose states are constrained to the Stiefel manifold $\mathcal{V}_{k,l}$.
- This work extends [1] in two ways:
  - The observations are nonlinear functions of the state.
  - We approximate the optimal importance function.

2. Problem Setup

- Let $\mathbf{S}_n$ denote the state of a system on the Stiefel manifold $\mathcal{V}_{k,l}$, i.e., $\{\mathbf{V} \in \mathbb{R}^{k\times l} : \mathbf{V}^T\mathbf{V} = \mathbf{I}_l\}$, $k > l$, according to

$$\mathbf{S}_n|\mathbf{S}_{n-1} \sim \text{vMF}(\mathbf{S}_n|\kappa\mathbf{S}_{n-1}) = \frac{\text{etr}\left(\kappa\mathbf{S}_{n-1}^T\mathbf{S}_n\right)}{{}_0F_1\left(\frac{k}{2}, \frac{\kappa^2}{4}\mathbf{I}_l\right)},$$

where $\kappa \in \mathbb{R}^+$ is a fixed hyperparameter and ${}_0F_1$ is the hypergeometric function with matrix argument.
- $\{\mathbf{S}_n\}$ gives rise to the observation sequence $\{\mathbf{Y}_n\}$, $\mathbf{Y}_n \in \mathbb{R}^{k\times l}$,

$$\mathbf{Y}_n|\mathbf{S}_n \sim \mathbf{N}_{k,l}(\mathbf{Y}_n|\mathbf{G}(\mathbf{S}_n), \boldsymbol{\Omega}, \boldsymbol{\Gamma}),$$

where $\mathbf{G} : \mathbb{R}^{k\times l} \to \mathbb{R}^{k\times l}$ is a possibly nonlinear function, and $\mathbf{N}_{k,l}$ is a matrix normal distribution on $\mathbb{R}^{k\times l}$.
- The particle filtering algorithm of [1] was restricted to $\mathbf{G}(\mathbf{S}_n) = \mathbf{S}_n$ and used the prior importance function:

$$\mathbf{S}_n^{(q)} \sim p\left(\mathbf{S}_n|\mathbf{S}_{n-1}^{(q)}, \mathbf{Y}_n\right) = \text{vMF}(\mathbf{S}_n|\kappa\mathbf{S}_{n-1}^{(q)}).$$

- The restriction on $\mathbf{G}$ can be trivially lifted, leading to the weight update equation

$$w_n^{(q)} \propto w_{n-1}^{(q)}\mathbf{N}_{k,l}(\mathbf{Y}_n|\mathbf{G}(\mathbf{S}_n^{(q)}), \boldsymbol{\Omega}, \boldsymbol{\Gamma}).$$

3. Proposed Method

- Optimal importance function:

$$\mathbf{S}_n^{(q)} \sim p(\mathbf{S}_n|\mathbf{Y}_n, \mathbf{S}_{n-1}^{(q)}) = \frac{p(\mathbf{Y}_n|\mathbf{S}_n)p(\mathbf{S}_n|\mathbf{S}_{n-1}^{(q)})}{\int_{\mathcal{V}_{k,l}} p(\mathbf{Y}_n|\mathbf{S}_n)p(\mathbf{S}_n|\mathbf{S}_{n-1}^{(q)})d\mathcal{V}_{k,l}(\mathbf{S}_n)}. \quad (1)$$

- The integral in (1) cannot be analytically evaluated if $\mathbf{G}(\mathbf{S}_n)$ is a general nonlinear function. By linearizing $\mathbf{G}(\mathbf{S}_n)$ around $\mathbf{S}_{n-1}$, we get

$$\mathbf{g}(\mathbf{s}_n) \approx \mathbf{g}(\mathbf{s}_{n-1}) + \mathbf{J}(\mathbf{s}_{n-1})\left[\mathbf{s}_n - \mathbf{s}_{n-1}\right], \quad (2)$$

where $\mathbf{s}_n \triangleq \text{vec}(\mathbf{S}_n)$, $\mathbf{g}(\mathbf{s}_n) \triangleq \text{vec}(\mathbf{G}(\mathbf{S}_n))$ and $[\mathbf{J}(\mathbf{s}_{n-1})]_{kl} \triangleq \left.\frac{\partial[\mathbf{g}(\mathbf{s})]_k}{\partial[\mathbf{s}]_l}\right|_{\mathbf{s}=\mathbf{s}_{n-1}}$ is a Jacobian matrix.

- As a result of (2):

$$p(\mathbf{S}_n|\mathbf{Y}_n, \mathbf{S}_{n-1}^{(q)}) \approx \text{FB}(\mathbf{S}_n|\mathbf{A}_n^{(q)}, \mathbf{B}_n^{(q)}),$$

- The weights are then exactly propagated as

$$w_n^{(q)} \propto w_{n-1}^{(q)}\frac{\text{vMF}(\mathbf{S}_n^{(q)}|\kappa\mathbf{S}_{n-1}^{(q)})\ \mathbf{N}_{k,l}(\mathbf{Y}_n|\mathbf{G}(\mathbf{S}_n^{(q)}), \boldsymbol{\Omega}, \boldsymbol{\Gamma})}{\text{FB}(\mathbf{S}_n^{(q)}|\mathbf{A}_n^{(q)}, \mathbf{B}_n^{(q)})}. \quad (3)$$

where FB stands for the matrix Fisher-Bingham p.d.f. on $\mathcal{V}_{n,k}$:

$$\text{FB}(\mathbf{S}_n|\mathbf{A}_n, \mathbf{B}_n) = \frac{\exp\left\{\text{tr}(\mathbf{A}_n^T\mathbf{S}_n) + \text{vec}(\mathbf{S}_n)^T\mathbf{B}_n\text{vec}(\mathbf{S}_n)\right\}}{c_{FB}(\mathbf{A}_n, \mathbf{B}_n)},$$

where $c_{FB}(\mathbf{A}_n, \mathbf{B}_n)$ is the matrix Fisher-Bingham p.d.f. normalization constant, and

$$\mathbf{A}_n \triangleq \text{vec}^{-1}(\tilde{\mathbf{y}}_n^T\boldsymbol{\Sigma}^{-1}\mathbf{J}(\mathbf{s}_{n-1})) + \kappa\mathbf{S}_{n-1},$$
$$\mathbf{B}_n \triangleq -\frac{1}{2}\mathbf{J}(\mathbf{s}_{n-1})^T\boldsymbol{\Sigma}^{-1}\mathbf{J}(\mathbf{s}_{n-1}),$$
$$\tilde{\mathbf{y}}_n \triangleq \mathbf{y}_n - \mathbf{g}(\mathbf{s}_{n-1}) + \mathbf{J}(\mathbf{s}_{n-1})\mathbf{s}_{n-1},$$
$$\boldsymbol{\Sigma} \triangleq \boldsymbol{\Gamma} \otimes \boldsymbol{\Omega}.$$

4. Sampling from a matrix Fisher-Bingham p.d.f.

- We adapted from the algorithm in [2, Sec. 3.3], originally developed for the matrix Bingham-Von Mises-Fisher p.d.f.
- Under the restriction that $\mathbf{B}_n$ is a block-diagonal matrix,

$$\text{FB}(\mathbf{S}_n|\mathbf{A}_n, \mathbf{B}_n) \propto \prod_{m=1}^{l}\exp\left(\mathbf{A}_n[,m]^T\mathbf{S}_n[,m] + \mathbf{S}_n[,m]^T\mathbf{B}_n(m)\mathbf{S}_n[,m]\right),$$

where $[,m]$ stands for the $m-$th column of a matrix and $\mathbf{B}(m) \in \mathbb{R}^{k\times k}$ denotes the $m-$th block of the diagonal of $\mathbf{B}_n$.
- As the columns of $\mathbf{S}_n$ are orthogonal with probability 1, $\mathbf{S}_n = [\mathbf{N}\mathbf{z}\ \ \mathbf{S}_n[,-1]]$, where $\mathbf{S}_n[,-1]$ is the matrix formed by removing the first column of $\mathbf{S}_n$, $\mathbf{N} \in \mathbb{R}^{k\times(k-l+1)}$ is an orthonormal basis for the null space of $\mathbf{S}_n[,-1]$, and $\mathbf{z}$ is a $(k-l+1)$ unit-norm column vector.
- The conditional p.d.f. of $\mathbf{z}$ is then given by [2]

$$p(\mathbf{z}|\mathbf{S}_n[,-1]) \propto \exp\left(\mathbf{A}_n[,1]^T\mathbf{N}\mathbf{z} + \mathbf{z}^T\mathbf{N}^T\mathbf{B}_n(1)\mathbf{N}\mathbf{z}\right)$$

which is a Fisher-Bingham density on the unit sphere.
- A Markov chain with stationary p.d.f. $\text{FB}(\mathbf{S}_n|\mathbf{A}_n, \mathbf{B}_n)$ can be obtained via the Gibbs sampler:
  - Given $\mathbf{S}_n^{<j>} = \mathbf{S}$, the $j-$th element of the chain, compute steps 1 to 4 for each $m \in \{1, \ldots, l\}$ in a random order:
    1. compute $\mathbf{N}$, an orthonormal basis for the null space of $\mathbf{S}[,-m]$;
    2. compute $\tilde{\mathbf{a}} = \mathbf{A}_n[,m]^T\mathbf{N}$ and $\tilde{\mathbf{B}} = \mathbf{N}^T\mathbf{B}_n(m)\mathbf{N}$;
    3. sample $\mathbf{z}$ from a Fisher-Bingham density on the unit sphere with parameters $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{B}}$.
    4. set $\mathbf{S}[,m] = \mathbf{N}\mathbf{z}$.
  - Set $\mathbf{S}_n^{<j+1>} = \mathbf{S}$.

5. Computation of the matrix Fisher-Bingham p.d.f. normalization constant

- To update the weights (Equation 3), it is necessary to compute the normalization constants

$$c_{FB}(\mathbf{A}_n, \mathbf{B}_n) \triangleq \int_{\mathcal{V}_{k,l}}\exp\left\{\text{tr}(\mathbf{A}_n^T\mathbf{S}) + \text{vec}(\mathbf{S})^T\mathbf{B}_n\text{vec}(\mathbf{S})\right\}d\mathcal{V}_{k,l}(\mathbf{S}), \quad (4)$$

$$c_{MF}(k\mathbf{S}_{n-1}) = c_{FB}(k\mathbf{S}_{n-1}, \mathbf{0}).$$

- As existing approaches did not perform adequately, we introduced the Quasi Monte Carlo algorithm:
  1. Generate low-discrepancy samples uniformly distributed on $\mathcal{V}_{k,l}$.
  2. Approximate (4) as

$$\text{Vol}(\mathcal{V}_{k,l})\frac{1}{N_S}\sum_{i=1}^{N_S}\left\{\text{tr}(\mathbf{A}_n^T\mathbf{X}^{<i>}) + \text{vec}(\mathbf{X}^{<i>})^T\mathbf{B}_n\text{vec}(\mathbf{X}^{<i>})\right\}$$

where N_S is the number of samples, $\mathbf{X}^{<i>}$ is the $i-$th generated sample, and $\text{Vol}(\mathcal{V}_{k,l})$ is the volume of $\mathcal{V}_{k,l}$.

6. Computation of the weighted averages on the Stiefel manifold

- Ideally, one would estimate the state as a Karcher mean, i.e., the value of $\hat{\mathbf{S}}_n$ that minimizes the weighted mean square geodesic distance to the particle set.
- To reduce computational complexity, we evaluated the weighted averages over the Stiefel manifold as

$$\mathbf{S}_n^{<i+1>} = \mathcal{M}_{\mathbf{S}_n^{<i>}}\left(\sum_{q=1}^{Q}w_n^{(q)}\mathcal{M}_{\mathbf{S}_n^{<i>}}^{-1}\left(\mathbf{S}_n^{(q)}\right)\right), \quad i \geq 0,$$

where $\mathbf{S}_n^{<i>}$ denotes the $i-$th estimate of the weighted average, with $\mathbf{S}_n^{<0>}$ chosen as a random element of the particle set, and $\mathcal{M}$ and $\mathcal{M}^{-1}$ are the orthographic retraction and lifting maps [3].

7. Numerical Experiment

- We performed numerical simulations with 150 independent trials of 100 synthetic data samples. Particle filters used 300 particles.

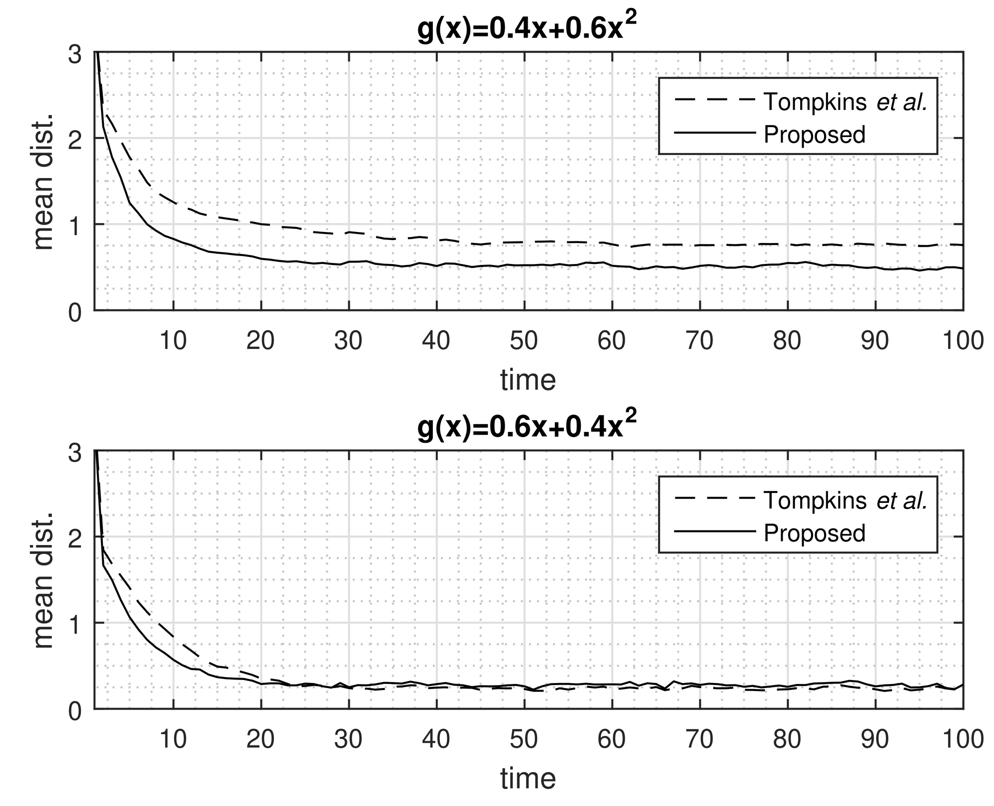

Figure 1: *Mean geodesic distance for the proposed algorithm and that of [1] as a function of time, for distinct nonlinear observation functions $g(x)$.*

- We assumed that $\mathbf{G}$ acts element-wise, so that the Jacobian $\mathbf{J}(\mathbf{s}_{n-1})$ is diagonal. The parameters were set to $\kappa = 150$, $\boldsymbol{\Omega} \triangleq \mathbf{I}_l$, $\boldsymbol{\Gamma} \triangleq \mathbf{I}_k\sigma^2$, and $\boldsymbol{\Sigma} = \mathbf{I}_{kl}\sigma^2$, with $\sigma^2 = 0.05$, $k = 3$, $l = 2$.
- The algorithms' performance was evaluated in terms of the mean geodesic distance from the true state $\mathbf{S}_n$ to the estimated state $\hat{\mathbf{S}}_n$, i.e., $d(\mathbf{S}_n, \hat{\mathbf{S}}_n) = \|\text{Exp}_{\mathbf{S}_n}^{-1}(\hat{\mathbf{S}}_n)\|_F$.
- For stronger nonlinearity (top), the proposed method exhibited an asymptotic error about 30% smaller than the method of [1]

8. Conclusions

- For certain choices of $\mathbf{G}$, the proposed method outperforms that of [1] at the expense of increased computational complexity.
- Most of the computational complexity of the proposed method is related to drawing samples from and computing normalization constants for the matrix Fisher-Bingham density.

References

[1] F. Tompkins and P. J. Wolfe, "Bayesian Filtering on the Stiefel Manifold," in *Computational Advances in Multi-Sensor Adaptive Processing, 2007. CAMPSAP 2007. 2nd IEEE International Workshop on*. IEEE, 2007, pp. 261–264.

[2] P. D. Hoff, "Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data," *Journal of Computational and Graphical Statistics*, vol. 18, no. 2, pp. 438–456, 2009.

[3] T. Kaneko, S. Fiori, and T. Tanaka, "Empirical arithmetic averaging over the compact Stiefel manifold," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 883–894, 2013.

ICASSP 2019, Brighton - United Kingdom.