PARTICLE FILTERING ON THE COMPLEX STIEFEL MANIFOLD WITH APPLICATION TO SUBSPACE TRACKING

Claudio J. Bordin Jr.[1] Marcelo G. S. Bruno[2]

[1]Universidade Federal do ABC
claudio.bordin@ufabc.edu.br

[2]Instituto Tecnológico de Aeronáutica
bruno@ita.br

ICASSP 2020 - SPTM-L7: Bayesian Signal Processing II (FR2.L4.4)

- In many signal processing tasks (e.g. target tracking, processing of hyperspectral imagery), estimates must satisfy geometrical constraints.
- Challenge: derive principled estimators taking into account nonlinear restrictions.
- In this paper, we extend previous particle filtering methods [Tompinks 2007], [Bordin 2019] to deal with densities on the complex Stiefel manifold.
- The new method is applied to a Bayesian version of the subspace tracking problem, formulated as a matrix tracking problem on the complex Stiefel manifold.

- Several Bayesian subspace tracking algorithms have been already proposed:
  - [Srivastava 2004] and [Rentmeesters 2010] modeled the subspaces to be estimated as time-variant according to geodesics on the Grassmann manifold;
  - In [Besson 2011], real subspaces are represented as the span of matrices on the Stiefel manifold and maximum *a posteriori* estimates are obtained analytically and via MCMC methods.
- The signal model considered in this paper is a superset of the model of [Besson 2011]: all involved quantities are complex and time-variant according to random walks.
- Main contributions in this paper:
  1. describing a new MCMC method to simulate from the Von Mises-Fisher distribution on the complex Stiefel manifold;
  2. proposing an extension of the subspace averaging method of [Fiori 2015] to the complex case;
  3. formulating subspace tracking as a Bayesian estimation problem suitable to be solved via a Rao-Blackwellized particle filter.

- We consider a complex, dynamic version of the signal model deployed in [Besson 2011], namely,

$$\mathbf{Y}_l = \mathbf{U}_l \mathbf{S}_l + \mathbf{N}_l,$$

  where $l \in \mathbb{N}$ denotes the time index, $\mathbf{Y}_l \in \mathbb{C}^{N \times K}$ is the observed matrix, $\mathbf{U}_l \in \mathbb{C}^{N \times p}$, $N > p$, is a matrix whose columns span the subspace of interest, $\mathbf{S}_l \in \mathbb{C}^{p \times K}$ is the waveform matrix, assumed to be a matrix Gaussian random process, and $\mathbf{N}_l \in \mathbb{C}^{N \times K}$ denotes the additive noise.
- To guarantee that $\mathbf{U}_l$ is full rank, we impose that it belongs to the (compact) complex Stiefel manifold $\mathcal{V}_{N,p}$, i.e., $\mathbf{U}_l^H \mathbf{U}_l = \mathbf{I}_p$, where $\mathbf{I}_p$ denotes the $p \times p$ identity matrix and $()^H$ the conjugate transpose of a matrix.

- The subspace of interest spanning matrix $\mathbf{U}_l$ evolves in time according to the random walk

$$\mathbf{U}_l|\mathbf{U}_{l-1} \sim \mathrm{VMF}_c(\mathbf{U}_l|\kappa\mathbf{U}_{l-1}),$$

where $\kappa \in \mathbb{R}^+$ is a hyperparameter and VMF_c stands for a complex matrix-variate Von Mises-Fisher distribution, defined as

$$\mathrm{VMF}_c(\mathbf{X}|\mathbf{A}) \triangleq \frac{\mathrm{etr}\left(\Re\left(\mathbf{X}^H\mathbf{A}\right)\right)}{{}_0\tilde{F}_1\left(r_{\mathbf{A}}, \frac{1}{4}\mathbf{A}^H\mathbf{A}\right)},$$

where etr denotes the exponential of the trace of a square matrix, $\Re$ the real part of the argument, ${}_0\tilde{F}_1$ is the hypergeometric function with complex matrix argument, and $r_{\mathbf{A}}$ denotes the number of rows of $\mathbf{A}$.

- The additive noise is assumed to form an i.i.d. random process that follows a complex matrix-variate Gaussian distribution

$$\mathbf{N}_l \sim \mathrm{N}_c(\mathbf{0}_{N,K}, \mathbf{I}_N, \mathbf{I}_K\sigma^2),$$

where $\sigma^2 > 0$ is a known parameter, $\mathbf{0}_{N,K}$ denotes an $N \times K$ matrix with null entries, and

$$\mathrm{N}_c(\mathbf{X}|\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Psi}) \triangleq \frac{\mathrm{etr}\left(\mathbf{\Sigma}^{-1}(\mathbf{X}-\mathbf{M})\mathbf{\Psi}^{-1}(\mathbf{X}-\mathbf{M})^H\right)}{\pi^{-NK}|\mathbf{\Sigma}|^{-K}|\mathbf{\Psi}|^{-N}},$$

where $\mathbf{\Sigma} \in \mathbb{C}^{N\times N}$ and $\mathbf{\Psi} \in \mathbb{C}^{K\times K}$ are Hermitian, positive-definite matrices.

- We wish to design a particle filter to approximate the probability

$$\mathbf{Pr}(\{\mathbf{U}_{0:l} \in \Delta\}|\mathbf{Y}_{0:l}) \approx \sum_{q=1}^{Q} w_l^{(q)} \delta_{\mathbf{U}_{0:l}^{(q)}}(\Delta),$$

where Δ is a subset of the $(l+1)$-ary Cartesian power of $\mathcal{V}_{N,K}$, $\delta_x(X)$ is a Dirac measure, $\mathbf{U}_k^{(q)}$, $0 \leq k \leq l$, are the particles, $Q \gg 1$ is the number of particles, and $w_l^{(q)}$ are the particle weights, recursively evaluated as

$$w_l^{(q)} \propto w_{l-1}^{(q)} \frac{p(\mathbf{Y}_l|\mathbf{U}_{0:l}^{(q)}, \mathbf{Y}_{1:l-1}) p(\mathbf{U}_l^{(q)}|\mathbf{U}_{0:l-1}^{(q)}, \mathbf{Y}_{1:l-1})}{\pi\left(\mathbf{U}_l^{(q)}|\mathbf{U}_{0:l-1}^{(q)}, \mathbf{Y}_{1:l}\right)},$$

where $\pi(\cdot)$ denotes the importance funcion, $p(\cdot)$ the p.d.f. of the indicated random variables and $\propto$ proportionality, such that $\sum_{q=1}^{Q} w_l^{(q)} = 1$.

- We do not aim to estimate $\mathbf{U}_I$ in itself, but its range $\mathbf{R}(\mathbf{U}_I)$ instead.
- This constitutes an estimation problem in the Grassmann manifold $\mathcal{G}_{N,p}$, the set of $p-$dimensional linear subspaces of $\mathbb{C}^N$.
- The elements of $\mathcal{G}_{N,p}$, which are subspaces spanned by Stiefel matrices $\mathbf{X} \in \mathcal{V}_{N,p}$, can be represented by such matrices, but this representation is not unique.

- Using the prior importance function,

$$\pi\left(\mathbf{U}_l|\mathbf{U}_{0:l-1},\mathbf{Y}_{1:l}\right)=p(\mathbf{U}_l|\mathbf{U}_{0:l-1},\mathbf{Y}_{1:l-1})=p(\mathbf{U}_l|\mathbf{U}_{l-1})=\mathrm{VMF}_c(\mathbf{U}_l|\kappa\mathbf{U}_{l-1}),$$

and the weights can be updated as

$$w_l^{(q)}\propto w_{l-1}^{(q)}\,p(\mathbf{Y}_l|\mathbf{U}_{0:l}^{(q)},\mathbf{Y}_{1:l-1}), \tag{1}$$

- The p.d.f. on the right-hand side (r.h.s.) of (1) is evaluated analytically by Rao-Blackwellizing the unknown amplitudes matrix $\mathbf{S}_l$:

$$\begin{aligned}&p(\mathbf{Y}_l|\mathbf{U}_{0:l},\mathbf{Y}_{1:l-1})=\\&=\int_{\mathbb{R}^{pK}}\int_{\mathbb{R}^{pK}}p(\mathbf{y}_l,\mathbf{s}_l|\mathbf{u}_{0:l},\mathbf{y}_{1:l-1})\;d[\Re\mathbf{s}_l]\;d[\Im\mathbf{s}_l]\\&=\int_{\mathbb{R}^{pK}}\int_{\mathbb{R}^{pK}}p(\mathbf{y}_l|\mathbf{s}_l,\mathbf{u}_l)p(\mathbf{s}_l|\mathbf{u}_{0:l-1},\mathbf{y}_{1:l-1})\;d[\Re\mathbf{s}_l]\;d[\Im\mathbf{s}_l],\end{aligned}$$

where $\mathbf{y}_l$, $\mathbf{s}_l$ and $\mathbf{u}_l$ denote the vectors obtained by vectorizing (i.e., stacking the columns) of the random matrices $\mathbf{Y}_l$, $\mathbf{S}_l$ and $\mathbf{U}_l$, respectively.

- After manipulations, the p.d.f. required to update the weights boils down to

$$p(\mathbf{Y}_l|\mathbf{U}_{0:l},\mathbf{Y}_{1:l-1}) = \mathcal{N}_c(\mathbf{y}_l|(\mathbf{I}_K\otimes\mathbf{U}_l)\bar{\mathbf{s}}_{l|l-1},(\mathbf{I}_K\otimes\mathbf{U}_l)\boldsymbol{\Sigma}_{l|l-1}(\mathbf{I}_K\otimes\mathbf{U}_l)^H + \mathbf{I}_{NK}\sigma^2),$$

where $\otimes$ the Kronecker matrix product, $\mathcal{N}_c$ a NK-variate (vector) complex Gaussian p.d.f., and $\bar{\mathbf{s}}_{l|l-1}$ and $\boldsymbol{\Sigma}_{l|l-1}$ are determined the via Kalman Filter-like recursions:

$$\bar{\mathbf{s}}_{l|l-1} = \mathbf{F}_l\,\bar{\mathbf{s}}_l,$$

$$\boldsymbol{\Sigma}_{l|l-1} = \mathbf{F}_l\boldsymbol{\Sigma}_l\mathbf{F}_l^H + \mathbf{Q}_l,$$

$$\mathbf{K}_l = \boldsymbol{\Sigma}_{l|l-1}(\mathbf{I}_K\otimes\mathbf{U}_l)^H\left((\mathbf{I}_K\otimes\mathbf{U}_l)\boldsymbol{\Sigma}_{l|l-1}(\mathbf{I}_K\otimes\mathbf{U}_l)^H + \mathbf{I}_{pK}\sigma^2\right)^{-1},$$

$$\bar{\mathbf{s}}_l = \bar{\mathbf{s}}_{l|l-1} + \mathbf{K}_l\left(\mathbf{y}_l - (\mathbf{I}_K\otimes\mathbf{U}_l)\bar{\mathbf{s}}_{l|l-1}\right),$$

$$\boldsymbol{\Sigma}_l = (\mathbf{I}_{pK} - \mathbf{K}_l(\mathbf{I}_K\otimes\mathbf{U}_l))\,\boldsymbol{\Sigma}_{l|l-1}.$$

- To run the proposed particle filter, one needs to draw samples from $\mathrm{VMF}_c(\mathbf{U}_l|\kappa\mathbf{U}_{l-1})$.
- To this aim, we adapted the Gibbs sampling algorithm of [Hoff 2009], designed to simulate from the real matrix Von Mises-Fisher distribution.
- The proposed method runs as follows: let $\mathbf{X}$ and $\mathbf{A} \in \mathbb{C}^{N\times p}$. From definitions, we get that

$$\mathrm{VMF}_c(\mathbf{X}|\mathbf{A}) \propto \prod_{m=1}^{p} \exp\left[\Re\left(\mathbf{A}[,m]^H\mathbf{X}[,m]\right)\right],$$

where $[,m]$ stands for the $m-$th column of a matrix.
- As the columns of $\mathbf{X}$ are orthogonal, we can write

$$\mathbf{X} = [\mathbf{X}[,1:m-1]\ \mathbf{Oz}\ \mathbf{X}[,m+1:p]],$$

where $\mathbf{X}[,a:b]$ collects the columns of $\mathbf{X}$ with index a to b, $\mathbf{O} \in \mathbb{C}^{N\times(N-p+1)}$ is an orthonormal basis for the left null space of $\mathbf{X}[,-m]$, defined as the matrix formed by removing the m-th column of $\mathbf{X}$, and $\mathbf{z}$ is an $(N-p+1)$ unit-norm column vector.

- The conditional p.d.f. of $\mathbf{z}$ is then given by

$$p(\mathbf{z}|\mathbf{X}[,-m],\mathbf{A}) \propto \exp\left[\Re\left(\mathbf{A}[,-m]^H\mathbf{O}\mathbf{z}\right)\right] \triangleq \mathrm{vmf}_c(\mathbf{z}|\mathbf{O}^H\mathbf{A}[,-m]),$$

where vmf_c stands for a Von Mises-Fisher density on the complex unit sphere.

- Using the conditional p.d.f. above, a Markov chain with stationary p.d.f. $\mathrm{VMF}_c(\mathbf{X}|\mathbf{A})$ can be obtained via the Gibbs sampler:
  - For $m \in 1,\ldots,p$, in a random order, do
    - Compute O, an orthonormal basis for the left null space of $\mathbf{X}[,-m]$.
    - Sample $\mathbf{z} \sim \mathrm{vmf}_c(\mathbf{z}|\mathbf{O}^H\mathbf{A}[,-m])$.
    - Set $\mathbf{X}[,m] = \mathbf{O}\mathbf{z}$.

- To draw samples $\mathbf{z}' \sim \mathrm{vmf}_c(\mathbf{z}|\tilde{\mathbf{a}})$ to run the Gibbs sampler, observe that

$$\exp\left[\Re\left(\tilde{\mathbf{a}}^H\mathbf{z}\right)\right] = \exp\left(\left[\tilde{\mathbf{a}}_R^T\ \tilde{\mathbf{a}}_I^T\right]\left[\begin{array}{c}\mathbf{z}_R\\ \mathbf{z}_I\end{array}\right]\right) \propto \mathrm{vmf}\left(\left[\begin{array}{c}\mathbf{z}_R\\ \mathbf{z}_I\end{array}\right]\middle|\left[\begin{array}{c}\tilde{\mathbf{a}}_R\\ \tilde{\mathbf{a}}_I\end{array}\right]\right),$$

where subscripts R and I denote the real and imaginary parts of the vectors, respectively, and vmf stands for a Von Mises-Fisher density on the real unit sphere.

- The samples $\mathbf{z}' = \mathbf{z}'_R + i\ \mathbf{z}'_I$ can then be obtained as

$$\left[\begin{array}{c}\mathbf{z}'_R\\ \mathbf{z}'_I\end{array}\right] \sim \mathrm{vmf}\left(\left[\begin{array}{c}\mathbf{z}_R\\ \mathbf{z}_I\end{array}\right]\middle|\left[\begin{array}{c}\tilde{\mathbf{a}}_R\\ \tilde{\mathbf{a}}_I\end{array}\right]\right),$$

where i denotes the imaginary unit.

- Samples from a Von Mises-Fisher density on the real unit sphere can be obtained via the algorithm introduced in [Wood 1994].

- To compute an estimate of $\mathcal{R}(\mathbf{U}_l)$ given the particle approximation $\left\{w_l^{(q)}, \mathbf{U}_l^{(q)}\right\}_{q=1}^{Q}$, one should ideally determine its Karcher mean on the complex Grassmann manifold.
- To reduce the computational burden, we adapted the empirical averaging procedure of [Fiori] 2015 using Thin-QR-Decomposition-Based Maps.
- The Thin-QR-Decomposition Map and its inverse are defined as

$$P_{\mathbf{X}}(\mathbf{V}) \triangleq \mathrm{qf}(\mathbf{X}+\mathbf{V}),$$
$$P_{\mathbf{X}}^{-1}(\mathbf{Y}) \triangleq \mathbf{Y}(\mathbf{X}^H\mathbf{Y})^{-1} - \mathbf{X}, \quad \text{(There is a typo in the paper's Eq. 31!)}$$

where qf denotes the Q factor of a QR decomposition.
- Thin-QR-Decomposition Map satisfy $P_{\mathbf{X}}(P_{\mathbf{X}}^{-1}(\mathbf{Y})) \sim \mathbf{Y}$ (if $\mathbf{X}^H\mathbf{Y}$ is invertible).

- The Thin-QR-Decomposition Map $P_{\mathbf{X}}(\mathbf{V})$ maps a point on the tangent space onto $\mathcal{G}_{N,p}$.
    - Replaces the Exponential Map, without the same distance preserving properties.
- Similarly, the Inverse Thin-QR-Decomposition Map is related to the Logarithmic Map.
- The resulting averaging algorithm is given by iterating

$$\hat{\mathbf{U}}_l^{<j+1>} = P_{\hat{\mathbf{U}}_l^{<j>}} \left(\sum_{q=1}^{Q} w_l^{(q)} \, P^{-1}_{\hat{\mathbf{U}}_l^{<j>}} \left(\mathbf{U}_l^{(q)} \right) \right), \; j \geq 0$$

where $\mathbf{U}_l^{<j>}$ is the $j-$th estimate of the weighted average, with $\mathbf{U}_l^{<0>}$ (arbitrarily) chosen as the particle with maximum weight.

- For performance evaluation, the proposed method was run for 200 independent trials.
- In each run, 30 successive samples were generated according to the signal model and processed.
- The particle filter employed $Q = 300$ particles, and the remaining parameters were set to $N = 6$, $p = 2$, $\kappa = 200$, $\mathbf{F}_l = 0.999\,\mathbf{I}_{pK}$, and $\mathbf{Q}_l = 0.001\,\mathbf{I}_{pK}$.
- To compute the particle filter estimates, the averaging algorithm of was run until $\|\mathbf{U}_n^{<i+1>} - \mathbf{U}_n^{<i>}\|_F < 10^{-9}$.
- To draw each sample from the importance function, the Gibbs sampler was run for $j = 50$ iterations.
- For comparison, we evaluated the performance, under the same hypotheses, of a competing SVD-based estimator [Adali 2010].

- To measure the similarity between the subspaces spanned by $\mathbf{U}_l$ and that spanned by the estimates, we employed the so-called fractional energy [Besson 2011], defined as

$$\mathrm{FE}(\mathbf{U}_l, \hat{\mathbf{U}}_l) \triangleq \frac{1}{p}\mathrm{tr}\left(\mathbf{U}_l^H \hat{\mathbf{U}}_l \hat{\mathbf{U}}_l^H \mathbf{U}_l\right)$$

  where tr denotes the trace of a matrix. Note that $\mathrm{FE}(\mathbf{U}_l, \hat{\mathbf{U}}_l)$ is inversely related to the distance between projective matrices

$$d^2(\mathbf{U}_l, \hat{\mathbf{U}}_l) \triangleq \|\mathbf{U}_l\mathbf{U}_l^H - \hat{\mathbf{U}}_l\hat{\mathbf{U}}_l^H\|_F^2 = 2p[1 - \mathrm{FE}(\mathbf{U}_l, \hat{\mathbf{U}}_l)],$$

  which is more adequate to measure distances between subspaces than the Euclidean distance, where $\|\cdot\|_F$ denotes the Frobenius norm.
- Figure 1 displays the fractional energy results at instant $l = 30$ obtained as functions of K, the number of simultaneous measurements (snapshots) and $1/\sigma^2$, the inverse of variance of the additive noise.

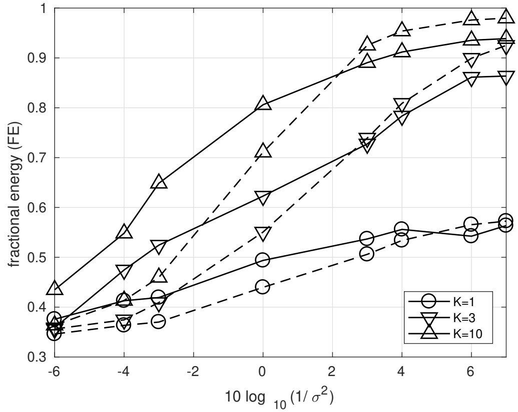

Figure: Mean fractional energy of the estimates provided by the proposed method (solid line) and an alternative SVD-based method (dashed line) as a function of $1/\sigma^2$ and of the number of *snapshots* K.

- We described in this paper a new particle filtering algorithm designed to estimate the complex subspace of a sequence of observations contaminated by additive noise.
- The proposed algorithm draws Stiefel matrices whose ranges span the subspace to be estimated. The sought subspace is then estimated by averaging such Stiefel matrices on the represented complex Grassmannian.
- As we verified via a numerical experiment with synthetic data, the proposed method outperforms a traditional SVD-based subspace tracking algorithm for scenarios with low signal-to-noise ratio.
- We are currently investigating the causes of the performance plateau observed for high signal-to-noise ratios.

Thank You!

claudio.bordin@ufabc.edu.br

bruno@ita.br