AN EFFICIENT PARTICLE FILTERING TECHNIQUE ON THE GRASSMANN MANIFOLD

Quentin Rentmeesters, P.-A. Absil, Paul Van Dooren *

Université catholique de Louvain
B-1348 Louvain-la-Neuve, Belgium

Kyle Gallivan, Anuj Srivastava

Florida State University
Tallahassee, FL, USA 32306

ABSTRACT

Subspace tracking methods are widespread in signal processing and image processing. To reduce the influence of perturbations or outliers on the measurements, some authors have used a stochastic piecewise constant velocity model on the Grassmann manifold. This paper presents an efficient way to simulate such a model using a particular parameterization of the Grassmann manifold. By doing so, we can reduce the spatial and time complexity of filtering techniques based on this model. We also propose an approximation of this system which can be computed in a finite number of operations and show similar results if the subspace variation is slow.

Index Terms— time-varying subspace learning, Grassmann manifold, particle filtering

1. INTRODUCTION

Subspace tracking problems appear in many areas such as signal processing and image processing. For instance, in direction of arrival tracking (DOA), a subspace tracker is required to recursively update the dominant subspace of the estimated covariance matrix of the signal. In image processing, subspace tracking problems appear in moving object tracking on video. In this problem the tracked object is often represented by a linear subspace. This subspace must be updated to cope with change in illumination or viewing angles. Furthermore, to reduce the influence of perturbations or outliers in the measurements, a filter is required. To implement this filter, some authors [1, 2] work on the Grassmann manifold denoted $\mathrm{G}(n, p)$ which is the set of p-dimensional subspaces of R^n and they have introduced a stochastic piecewise geodesic (constant velocity) model as a motion model for the subspaces to reduce the influence of the noise.

In [1], the authors use this model and a particle filtering technique on a DOA tracking problem. This technique approximates the posterior distribution of the subspace at each time step by drawing a large number of samples according to their stochastic piecewise geodesic model. Then, they estimate the subspace by computing the mean of these samples. In [2], the authors use the same model and a Kalman Filter approach on the tangent space of the Grassmann manifold to track an object on a video. Both papers show that better results are obtained using this geometric approach. But one of the main problem of these techniques is the computational cost of the exponential mappings, parallel transports and log-mappings. In fact, the method presented in [1], requires computing the exponential of an n-by-n matrix which is not efficient if n is large.

In this paper, we propose a different implementation of the particle filter introduced in [1] using a computationally more efficient parameterization of the tangent space to the Grassmann manifold yielding a computational complexity of $\mathcal{O}(np^2)$. To reduce again the complexity, we also propose to replace the exponential mapping by a retraction and the parallel transport by a vector transport as introduced in [3]. This is interesting from a computational point of view and yet comparable performances are shown if the subspaces vary slowly.

The paper is organized as follows. Section 2 introduced some notations and explains the main principle of our approach. Section 3 describes our implementation of the particle filtering technique. The approximation of the geodesic process using retractions and vector transports is presented in section 4. Numerical simulations are shown in section 5 and section 6 concludes.

2. PRINCIPLE OF OUR APPROACH

In this section, we describe the main idea behind our complexity reduction. We first recall the definition of a stochastic geodesic process and introduce some notation.

A point $\mathcal{X}$ on the Grassmann manifold is represented by the column space of an orthogonal matrix $X \in R^{n\times p}$; we write $\mathcal{X} = \mathrm{col}(X)$. We restrict ourselves to orthogonal matrices to obtain simpler formulas and to avoid conditioning problems.

In the particle filtering algorithm of [1], the authors use a piecewise-geodesic model where the velocities on all pieces are related by a Markov process. Reformulated in our nota-

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors.

tion, this model (see algorithm 1 in [1]) is:

$$[X_k|X_{\perp,k}] = [X_{k-1}|X_{\perp,k-1}] \exp\left(\begin{bmatrix} 0 & -A_{k-1}^\top \\ A_{k-1} & 0 \end{bmatrix}\right) \quad (1)$$

$$A_k = A_{k-1} + N_k \quad (2)$$

where $X_k \in R^{n\times p}$, $X_{\perp,k} \in R^{n\times(n-p)}$ such that $[X_k|X_{\perp,k}]$ is orthogonal, $A \in R^{(n-p)\times p}$ and N_k is an (n-p)-by-p matrix whose elements are i.i.d. real normals of mean 0 and variance σ^2_{model}.

This process is completely defined by the initial state $(X_0, X_{\perp,0}, A_0)$ and all the realizations of the noise N. In [1], this model is used to simulate the movement of particles and in [2], the same model is used to implement a Kalman filter (the velocity model represented by the matrix A is different). Therefore, it is very important to perform simulations with this model at a very low cost. If this model is implemented as in (1), it is not very efficient.

In fact, the dimension of the Grassmann manifold $\mathrm{G}(n,p)$ is $p(n-p)$. So is the dimension of its tangent space at a given point. Hence, a position-velocity pair on $\mathrm{G}(n,p)$ involves $2p(n-p)$ degrees of freedom. Yet, the data available at time k for one particle in (1) is $(X_k, X_{\perp,k}, A_k)$ which involves $np+n(n-p)+(n-p)p = n^2+(n-p)p$ real numbers. When $n >> p$ this is much larger than the theoretical optimum of $2p(n-p)$ parameters.

It would be possible to achieve the theoretical optimal number of parameters by working in local parameterizations of the Grassmann manifold. This option has drawbacks: the equations are less streamlined, and the transition between local parameterizations has to be managed.

In [4], a method is proposed to compute the exponential mapping in $\mathcal{O}(np^2)$ flops. But this approach does not work for parallel transport. In fact, at time k, only X_k is computed and they build an orthogonal completion of X_k based on Householder vectors. But this orthogonal completion is not equivalent to $X_{\perp,k}$ which is the rotated version of $X_{\perp,k-1}$. So at the next time step we cannot compute X_{k+1} because $X_{\perp,k}$ is required. Therefore, this approach does not solve the problem in this case.

In this paper, we propose to represent the state $(\mathcal{X}_k, \dot{\mathcal{X}}_k)$ of a particle at time k by (X_k, V_k), where $X_k \in R^{n\times p}$ such that $\mathrm{col}(X_k) = \mathcal{X}_k$ and $V_k \in R^{n\times p}$ is such that $X_k^\top V_k = 0$ and $\frac{d}{dt}\mathrm{col}(X_k + tV_k)|_{t=0} = \dot{\mathcal{X}}_k$ is the velocity of the particle. It can be shown that, given X_k, this V_k is unique and moreover, if X_k is replaced by X_kM, then V_k is replaced by V_kM for any orthogonal $M \in R^{p\times p}$. Notice also that V_k is related to A_k in (2) by: $V_k = X_{k,\perp}A_k$.

Observe that our representation uses $2np$ real variables, which is smaller than $n^2+(n-p)p$ for $1 \le p < \frac{n}{2}$ and $n > 1$ and is particularly interesting when $n >> p$.

3. REDUCED-MEMORY IMPLEMENTATION OF PARTICLE FILTERING

In this section, we explain how the subspace tracking method introduced in [1] can be implemented in $\mathcal{O}(np^2)$ flops using the parameterization of the Grassmann manifold introduced in the previews section.

In [5], there are formulas equivalent to (1) to update X_k and V_k:

$$V_{k-1} = U_{k-1}\Sigma_{k-1}W_{k-1}^\top \text{ (compact svd)} \quad (3)$$

$$D_{k-1} = X_{k-1}W_{k-1} \quad (4)$$

$$X_k = (D_{k-1}\cos\Sigma_{k-1} + U_{k-1}\sin\Sigma_{k-1})\,W_{k-1}{}^\top \quad (5)$$

$$V_k = (-D_{k-1}\sin\Sigma_{k-1} + U_{k-1}\cos\Sigma_{k-1})\,\Sigma_{k-1}W_{k-1}{}^\top + (\Omega_k - X_kX_k{}^\top\Omega_k) \quad (6)$$

Equation (5) can be interpreted as the computation of the geodesic and equation (6) as a parallel transport of the velocity vector V_{k-1} at X_{k-1} along the geodesic curve between $\mathcal{X}_{k-1}$ and $\mathcal{X}_k$ where $\mathcal{X}_k$ is represented by X_k. From a statistical point of view, equation (6) is equivalent to (1). Using the fact that $V_k = X_{\perp,k}A_k$, the perturbation in (6) becomes: $(I_n - X_kX_k^\top)\Omega_k = X_{\perp,k}X_{\perp,k}^\top\Omega_k = X_{\perp,k}\tilde{\Omega}_k$ where $\tilde{\Omega}_k = X_{\perp,k}^\top\Omega_k$ is also a Gaussian noise with the same variance than Ω_k because $X_{\perp,k}^\top$ is orthogonal. Thus, this $\tilde{\Omega}_k$ corresponds to the N_k in (2). Notice also that this way of computing the geodesic model requires $\mathcal{O}(np^2)$ flops per step, the most computationally demanding step being the SVD.

In [1], an intrinsic version of the mean called the Chordal mean is used. This mean is the dominant p-dimensional eigenspace of $\tilde{G}_k = \frac{1}{M}\sum_{i=1}^M X_k^iX_k^{i\top}$ and it requires the computation of an SVD of the n-by-n matrix $\tilde{G}_k$ which costs $\mathcal{O}(n^3)$ flops. To reach our computational complexity of $\mathcal{O}(np^2)$, the chordal mean can be replaced by the Karcher mean:

$$\mu(X^1,...,X^M) = \arg\min_{Y\in G}\frac{1}{M}\sum_{i=1}^M \mathrm{geodis}(X^i,Y)$$

where $\mathrm{geodis}(X^i,Y)$ stands for the geodesic distance between X^i and Y. There is no closed form formula to compute this mean. One way to solve this minimization problem is to implement a gradient descent method as follows:

1. compute $\tilde{V} = \frac{1}{M}\sum_{i=1}^M \log_{\mu^k}(X^i)$ where $\log_X(Y)$ stands for the log-mapping, i.e., it returns the velocity vector such that we get Y if we follow the geodesic curve starting at X in the direction of this velocity vector. The method proposed in [4] to compute this log -mapping using a CS decomposition can be slightly modified to work with the parametrization of the Grassmann manifold introduced in section 2:

$$\begin{bmatrix} X^\top Y \\ (I_n - XX^\top)Y \end{bmatrix} = \begin{bmatrix} W_1 \cos(\Sigma) Z^\top \\ W_2 \sin(\Sigma) Z^\top \end{bmatrix}$$
$$\log_X(Y) = W_2 \Sigma W_1^\top$$

2. move along a geodesic curve in the direction $\tilde{V} = U\Sigma W^\top$: set $\mu^{k+1} = \left(\mu^k W \sin(\Sigma) + U\cos(\Sigma)\right) W^\top$

This method requires O(Mnp^2) flops per iteration and is thus more efficient than the chordal mean when n is large.

In the particle filtering technique of [1], the following posterior model is used:

$$p(Y_k|X_k) = K_k \exp \frac{tr(X_k X_k^\top Y_k Y_k^\top)}{\sigma^2_{\text{data}}} \tag{7}$$

where K_k is a normalizer, $\mathcal{Y}_k = \text{col}(Y_k)$ are the measurements and σ^2_{data} is their variance. Notice that this distribution is independent of the particular orthogonal basis we have chosen to represent X_k and Y_k and requires $\mathcal{O}(np^2)$ flops to compute. And consequently, the particle filtering technique introduced in [1] has a total complexity of $\mathcal{O}(np^2)$.

4. FURTHER COMPLEXITY REDUCTION BY APPROXIMATION

The main drawback of this method is its complexity due to the computation of the geodesics and parallel transports. One way to reduce their computational cost is to replace the geodesics by retractions and the parallel transports by vector transports. In fact, this could replace the SVD decomposition of an n-by-p matrix by a QR decomposition which can be performed in a finite number of steps. This is particularly interesting if p is close to $\frac{n}{2}$.

More precisely, we propose to replace the constant velocity model (5-6) by the following retraction process:

$$X_k = q(X_{k-1} + V_{k-1}) \tag{8}$$
$$V_k = \frac{(I_n - X_k {X_k}^\top)V_{k-1}\|V_{k-1}\|_F}{\|(I_n - X_k {X_k}^\top)V_{k-1}\|_F} + (I_n - X_k {X_k}^\top)\Omega_k \tag{9}$$

where $\|.\|_F$ denotes the Frobenius norm and q is a projection of an n-by-p matrix of full rank onto the Stiefel manifold $\text{ST}(n,p)$ (the set of orthogonal n-by-p matrices) such that $\text{col}(X) = \text{col}(q(X))$ and $q(W) = W \quad \forall W \in \text{ST}(n,p)$. This system can be thought as a first order approximation of the geodesic process projected on the Stiefel manifold. Therefore, if $\|V_k\|_F$ is small, we obtain a good approximation of the geodesic. We can attempt to find a tangent vector V such that we get exactly the same subspace as we had followed the geodesic, i.e., if it is easy to compute a $\tilde{V}$ such that:

$$\text{col}(\underbrace{XW\cos(\Sigma) + U\sin(\Sigma))W^\top}_{Y}) = \text{col}(X + \tilde{V}).$$

Using the fact that we can multiply Y by an invertible matrix $W\cos(\Sigma)^{-1}W^\top$ (for $\sigma_i < \pi/2$) without changing its column space, we get:

$$\text{col}(Y) = \text{col}(X + \underbrace{U\tan(\Sigma)W^\top}_{\tilde{V}}). \tag{10}$$

Thus, the exponential mapping using $V = U\Sigma W^\top$ and the retraction (8) using $\tilde{V}$ give the same subspace. Observe also that this $\tilde{V}$ is given by the following log-mapping:

$$\tilde{V} = Y(X^\top Y)^{-1} - X.$$

Consequently, we can compute $\tilde{V}_0$ using this log-mapping and then use (8-9) with this $\tilde{V}_0$ to simulate the stochastic geodesic process. Notice however that the geodesic process (5-6) is not equivalent to the retraction process (8-9) from a statistical point of view but close to it if $\|V\|_F$ is sufficiently small ($\tan(\Sigma) \approx \Sigma$ for small Σ).

The parallel transport associated to a linear connection compatible with the metric preserves the length of vector. Consequently, in our vector transport (9), we normalize the projection of V_k to preserve the length of transported vectors.

It remains to choose the projection q in (8). One possibility is to use the polar decomposition which is the nearest element of the Stiefel manifold in terms of the Frobenius norm. This is the best choice. In fact, as we have $(X + \tilde{V}) = Y(W\cos(\Sigma)^{-1}W^\top)$ the orthogonal factor of the polar decomposition of $(X + \tilde{V})$ will be Y since $(W\cos(\Sigma)^{-1}W^\top)$ is a positive definite matrix ($0 < \sigma_i < \pi/2$). This decomposition also satisfies the following property: $\forall\ W \in \text{ST}(p,p), q(XW) = q(X)W$ which makes the retraction process (8-9) invariant to the specific choice of orthogonal basis used to represent the subspaces. More precisely, if X_1 is the first basis and V_1 is the corresponding tangent vector, this means that a change of basis: $X_1 \leftarrow X_1 W$ and $V_1 \leftarrow V_1 W$ for any $W \in \text{ST}(p,p)$ will not change the trajectory (on the Grassmann manifold) of the retraction process.

Unfortunately, this decomposition is as expensive as an SVD. We therefore propose to use the following iterative method to compute the polar decomposition of an n-by-p matrix A:

Compute the QR-decomposition of $A = QR$
set $r_1 = R$
for $i = 1, ..., N-1$
$r_{i+1} = \frac{1}{2}(r_i + r_i^{-\top})$
end for
set: $q(A) = Qr_N$

The r_i converge quadratically to the orthogonal factor of the polar decomposition of R, see [6] for more details. If the subspaces vary slowly, i.e., if $\|V\|_F$ is small, then $X + V$ will be close to an orthogonal matrix and in this case, it can be more efficient to compute its polar decomposition recursively

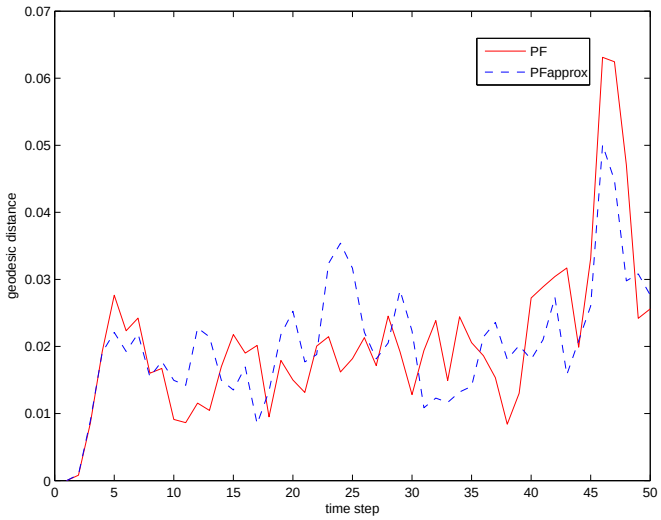

Fig. 1. Geodesic distance between the data and the filtered data returned by the PF and PFapprox methods with $\sigma_{\text{data}} = 0.05$, 1000 particles for $\|A_0\|_F = 0.2236$.

than to compute an SVD. We will show in our numerical experiments that only the QR-decomposition (with positive elements on the diagonal of R to satisfy $q(W) = W \quad \forall W \in \mathrm{ST}(n,p)$) is needed if the subspaces vary sufficiently slowly.

5. EXPERIMENTS AND RESULTS

A stochastic piecewise geodesic trajectory on $\mathrm{G}(4,2)$ with $\sigma_{\text{model}} = 0.05\|A_0\|_F$ was simulated according to (1-2). The particle filtering technique was applied on this data set and the geodesic distance between the data and the filtered data was computed as a measure of error. This was carried out for the particle filtering technique using the geodesic process (5-6) denoted by PF and using our approximation (8-9) with only a QR decomposition denoted by PFapprox. If the norm of A_0 is small, the subspaces vary slowly and the level of the error using the PF or PFapprox methods are equivalent, see Fig.1. But when the norm of A_0 is increased, the subspaces move more rapidly and the time-averaged error of the retraction method PFapprox becomes bigger than for the PF method as shown in Table.1.

6. CONCLUSIONS

We have presented an efficient way to implement the particle filtering technique of [1] that reduces the spatial complexity per particles from $n^2+(n-p)p$ to $2np$ and the time complexity from $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$. An approximation of the geodesic process has been also presented to avoid the computation of the SVD. This approximation is particularly interesting when

$\|A_0\|_F$	mean error PF	mean error PFapprox
0.1118	0.0186	0.0190
0.2236	0.0231	0.0250
0.3354	0.0269	0.0358
0.4472	0.0342	0.1147

Table 1. Time-average errors over 50 time steps averaged over 50 different realizations of a stochastic piecewise geodesic trajectory using the same initial state. All the experiments were carried out with $\sigma_{\text{data}} = 0.05$ and 1000 particles.

p gets close to $\frac{n}{2}$ and when the subspaces vary slowly. In further work, we will consider the use of these approximations in other filtering techniques such as the Luenberger filter introduced in [7]. The use of this filter can be particularly interesting in applications where n or p is large enough such that the particle filtering technique requires too many particles to be computationally tractable.

7. REFERENCES

[1] Anuj Srivastava and Eric Klassen, "Bayesian and geometric subspace tracking," *Adv. in Appl. Probab.*, vol. 36, no. 1, pp. 43–56, 2004.

[2] Tiesheng Wang, Andrew Backhouse, and Irene Y.H. Gu, "Online subspace learning in Grassmann manifold for moving object tracking in video," in *Proceedings of IEEE international conf. Acoustics, Speech, and Signal Processing (ICASSP'08)*, 2008, pp. 969–972.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, Princeton, NJ, 2008.

[4] Kyle Gallivan, Anuj Srivastava, Xiuwen Liu, and Paul Van Dooren, "Efficient algorithms for inferences on Grassmann manifolds," in *Proceedings of 12 th IEEE Workshop on Statistical Signal Processing*, 2003, pp. 315–318.

[5] Alan Edelman, Tomás A. Arias, and Steven T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353 (electronic), 1999.

[6] Nicholas J. Higham, *Functions of matrices*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008, Theory and computation.

[7] Nasradine Aghannan and Pierre Rouchon, "An intrinsic observer for a class of Lagrangian systems," *IEEE Trans. Automat. Control*, vol. 48, no. 6, pp. 936–945, 2003.