Proceedings of the 19th International Symposium on Mathematical Theory of Networks and Systems – MTNS 2010 • 5–9 July, 2010 • Budapest, Hungary

A Filtering Technique on the Grassmann manifold

Quentin Rentmeesters, P.-A. Absil and Paul Van Dooren
Département d'ingénierie mathématique,
Université catholique de Louvain,
B-1348 Louvain-la-Neuve, Belgium
{Quentin.Rentmeesters,PA.Absil,Paul.Vandooren}@uclouvain.be

***Abstract*— In this paper, a filtering technique that deals with subspaces, i.e., points on the Grassmann manifold, is proposed. This technique is based on an observer design where the data points are seen as the outputs of a constant velocity dynamical model. An explicit algorithm is given to efficiently compute this observer on the Grassmann manifold. This approach is compared to a particle filtering technique and similar results are obtained for a lower computational cost. Some extensions of the filter are also proposed.**

I. INTRODUCTION

In many applications related to signal processing and image processing, it is necessary to filter measurements that belong to the Grassmann manifold, i.e., the set of p-dimensional subspaces of R^n or C^n denoted by $\mathrm{G}(n,p)$. This appears in the direction of arrival tracking problem in antenna array processing when a time division multiple access technique is applied to increase the capacity of the antenna array, see [1]. In this case, the measurements are estimations of the signal subspace, which can be used to recover the directions of arrival. In object tracking problems on a video sequence, the tracked object can be represented by the dominant subspace of a covariance matrix, see [2]. To deal with deformations and illumination variations, this subspace must be updated. A filtering technique is then required to update this subspace using the subspaces representing the object at the previous time steps. Subspace tracking problems also arise when a reduced order model must be updated over time. In fact, a reduced order model is often computed as a projection of the full model on a low dimensional subspace. Since this projection is expensive to compute, an updating strategy is then required to predict the reduced order model over time.

A particle filtering technique based on a stochastic piecewise constant velocity model on the Grassmann manifold was introduced in [3]. Efficient implementations were discussed in [4], but even then, this technique is computationally demanding on high dimensional problems due to the number of particles required. In [2], the authors used a Kalman-like filter on the velocity, i.e., on the tangent space to $\mathrm{G}(n,p)$, to update the subspace representing a tracked object on a video.

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 author(s).

In this paper, we present a Luenberger-like observer based on a constant velocity dynamical model on the Grassmann manifold. This model is used to reduce the influence of high frequency noise or outliers in the measurements, i.e., subspaces in this case. This method does not work with particles and is then computationally cheaper than the particle filtering technique introduced in [3]. It has also a lower numerical cost than [2] due to an efficient representation of points and velocities on $\mathrm{G}(n,p)$.

The paper is organized as follows. Section II introduces definitions and gives useful formulas to compute on the Grassmann manifold. Section III describes the problem and introduces our filtering approach. Some numerical examples are shown in section IV. Section V describes some possible generalizations of our observer to other dynamical models and section VI concludes.

II. BACKGROUND

In order to define a dynamical system on the Grassmann manifold, we need to represent the position and the velocity of a subspace. This section introduces our representation of subspaces and some mappings that will be useful to define a discrete-time dynamical model on $\mathrm{G}(n,p)$.

A point $\mathcal{X}$ on the Grassmann manifold is represented by the column space of an orthogonal matrix $X \in R^{n\times p}$:

$$\mathcal{X} = \mathrm{col}(X).$$

So, from now on, a subspace will be denoted by $\mathcal{X}$ and its representation by X, an orthogonal $n\times p$ matrix. We restrict ourselves to orthogonal matrices to obtain simpler formulas and to avoid conditioning problems.

The velocity $\dot{\mathcal{X}}(t) = \frac{d}{dt}\mathcal{X}(t)$ of a smooth curve $t \mapsto \mathcal{X}(t)$ is termed a *tangent vector* to $\mathrm{G}(n,p)$ at $\mathcal{X}(t)$. The term "smooth" and the derivative are well defined because $\mathrm{G}(n,p)$ has a natural manifold structure; see, e.g., [5]. Given a tangent vector $\dot{\mathcal{X}}$ at $\mathcal{X}$, $\exists!\ V \in R^{n\times p}$, $X^\top V = 0$, such that $\dot{\mathcal{X}} = \dfrac{d}{dt}\mathrm{col}(X+tV)|_{t=0}$. This allows us, with a slight abuse of notation, to say that the tangent space at $\mathcal{X}$ has a matrix representation given by:

$$T_X\mathrm{G}(n,p) = \{V \in R^{n\times p} | X^\top V = 0\}.$$

Thus, given X, a tangent vector $\mathcal{V}_{\mathcal{X}}$ at $\mathcal{X}$ is also represented by an $n\times p$ matrix V. Moreover, if X is replaced by XQ, then V is replaced by VQ for any orthogonal $Q \in R^{p\times p}$.

ISBN 978-963-311-370-7

The Grassmann manifold can be turned into a Riemannian manifold by introducing the following inner product on $T_X \mathrm{G}(n,p)$:

$$< V_1, V_2 >_X = \mathrm{trace}((X^\top X)^{-1} V_1^\top V_2).$$

With this inner product or Riemannian metric, we can define the distance between two subspaces $\mathcal{X}$ and $\mathcal{Y}$:

$$d(\mathcal{X},\mathcal{Y}) = \inf_{\{\gamma \in \mathrm{G}(n,p) | \gamma(0)=\mathcal{X}, \gamma(1)=\mathcal{Y}\}} \int_0^1 \sqrt{\langle \dot\gamma(t), \dot\gamma(t) \rangle_{\gamma(t)}} dt.$$

The distance is thus the length of the shortest curve $\gamma(t)$ between $\mathcal{X}$ and $\mathcal{Y}$. It can be shown that this shortest curve is locally a geodesic, i.e., a curve of zero acceleration, where the acceleration is given by the Riemannian covariant derivative of the velocity: $\nabla_{\dot\gamma}\dot\gamma$, see [6]. Notice that this distance corresponds to $\sum_{i=1}^p \sigma_i^2$, where the σ_i's are the principal angles between $\mathcal{X}$ and $\mathcal{Y}$.

Finally, we need to introduce the exponential map, the log-mapping, and the parallel transport, for which efficient computational methods (with a computational complexity of $O(np^2)$) are discussed in [7] and [8].

Exponential map

It returns the point on the manifold obtained by following the geodesic curve $\gamma(t)$ such that $\gamma(0) = \mathcal{X}$ and $\dot\gamma(0) = \mathcal{V}_\mathcal{X}$.

$$\exp_\mathcal{X} : T_X \mathrm{G}(n,p) \mapsto \mathrm{G}(n,p),\ \exp_\mathcal{X}(\mathcal{V}_\mathcal{X}) = \gamma(1)$$

If the subspaces are represented by $n \times p$ matrices, the geodesic can be computed efficiently using

$$\exp_X(V_X) = (XW \cos\Sigma + U \sin\Sigma)\, W^\top,$$

where $V_X = U\Sigma W^\top$ is the compact SVD of V_X.

Log-mapping

The log-mapping is the inverse of the exponential map.

$$\exp_\mathcal{X}^{-1} : \mathrm{G}(n,p) \mapsto T_X \mathrm{G}(n,p),\ \exp_\mathcal{X}^{-1}(\mathcal{Y}) = \mathcal{V}_\mathcal{X}.$$

Notice that the log-mapping is the opposite of the gradient of the function: $\mathcal{X} \mapsto \frac{1}{2} d(\mathcal{X},\mathcal{Y})^2$. The method proposed in [8] to compute this log-mapping using a CS decomposition can be slightly modified to work with the parameterization of the Grassmann manifold introduced in this section

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

$$\exp_X^{-1}(Y) = W_2 \Sigma W_1^\top.$$

Parallel transport

The parallel transport of a tangent vector $T \in T_X \mathrm{G}(n,p)$ along the geodesic curve $\gamma(t)$ joining $\gamma(0) = \mathcal{X}$ to $\gamma(1) = \mathcal{Y}$ is denoted by

$$\Gamma_{\mathcal{X}\to\mathcal{Y}} : T_X \mathrm{G}(n,p) \mapsto T_Y \mathrm{G}(n,p),\ \Gamma_{\mathcal{X}\to\mathcal{Y}}(T) = Z(1)$$

where $Z(t)$ is the solution of $\nabla_{\dot\gamma(t)} Z(t) = 0$ and $Z(0) = T$. An efficient formula to compute this parallel transport on Grassmann is given by:

$$\Gamma_{X\to Y}(T) = (-XW \sin\Sigma + U\cos\Sigma)\, U^\top T + (I - UU^\top) T$$

where $U\Sigma W^\top = V_X$ is the compact SVD of the tangent vector V_X such that $\exp_X(V_X) = Y$. Notice that this vector can be computed using a log-mapping.

III. PROBLEM DESCRIPTION

Let us assume that we measure corrupted data $\mathcal{Y}_k \in \mathrm{G}(n,p)$ for $1 \le k \le T$. The goal is to filter recursively these data to reduce the influence of the noise and the outliers. To achieve this goal, we assume that the data $\mathcal{Y}_k$ are the outputs of the following discrete-time dynamical system whose state is composed of the position (a subspace $\mathcal{X}_k$) and the velocity (a tangent vector $\mathcal{V}_k$ at $\mathcal{X}_k$):

$$\begin{aligned} \mathcal{X}_{k+1} &= \exp_{\mathcal{X}_k}(\mathcal{V}_k), \\ \mathcal{V}_{k+1} &= \Gamma_{\mathcal{X}_k \to \mathcal{X}_{k+1}}(\mathcal{V}_k), \\ \mathcal{Y}_k &= \exp_{\mathcal{X}_k}(\mathcal{U}_k), \end{aligned}$$

where $\mathcal{U}_k$ is an i.i.d. Gaussian vector of mean 0 and variance $\sigma^2 I$ that belongs to the tangent space at $\mathcal{X}_k$. This constant velocity model, i.e., a geodesic model, for the dynamics of the subspaces is chosen to act as a smoother on the subspace trajectory to reduce the influence of the noise. Our filtering method is based on the design of a Luenberger observer for this dynamical system. A Luenberger observer on Riemannian manifolds has been introduced in [9] to observe the state (position and velocity) of a class of nonlinear mechanical systems. We will briefly describe this approach and then will derive our observer.

The continuous dynamical system for a particle moving on a geodesic curve is

$$\begin{aligned} \dot{\mathcal{X}} &= \mathcal{V}, \\ \nabla_{\dot{\mathcal{X}}} \mathcal{V} &= 0. \end{aligned}$$

If the position of the particle is measured, we can build the following Luenberger observer as in [9]:

$$\dot{\hat{\mathcal{X}}} = \hat{\mathcal{V}} + \alpha \exp_{\hat{\mathcal{X}}}^{-1}(\mathcal{Y}), \tag{1}$$

$$\nabla_{\dot{\hat{\mathcal{X}}}} \hat{\mathcal{V}} = \beta \exp_{\hat{\mathcal{X}}}^{-1}(\mathcal{Y}) + R(\hat{\mathcal{V}}, \exp_{\hat{\mathcal{X}}}^{-1}(\mathcal{Y}))\hat{\mathcal{V}}, \tag{2}$$

where $\hat{\mathcal{X}}$ and $\hat{\mathcal{V}}$ are the estimated position and velocity, $\mathcal{Y}$ is the measured trajectory, R is the Riemannian curvature tensor and α, β are two tunable real parameters. Remark that $\exp_{\hat{\mathcal{X}}}^{-1}(\mathcal{Y}) = \nabla_{\hat{\mathcal{X}}} \frac{1}{2} d(\hat{\mathcal{X}},\mathcal{Y})^2$. Notice that the curvature is involved to ensure the convergence of the observer in the negative curvature case. But the curvature of the Grassmann manifold is non negative. Hence, this curvature term is not required. So, we discard this term since the observer is less expensive to compute without this curvature term. The convergence proof in [9] will not hold anymore but our numerical experiments have shown that the observer converges even without this term.

Using an explicit Euler method with $\Delta t = 1$, a discrete version of (1-2) can be derived:

$$\hat{\mathcal{X}}_{k+1} = \exp_{\hat{\mathcal{X}}_k}(\hat{\mathcal{V}}_k + \alpha \exp_{\hat{\mathcal{X}}_k}^{-1}(\mathcal{Y}_k)), \tag{3}$$

$$\hat{\mathcal{V}}_{k+1} = \Gamma_{\hat{\mathcal{X}}_k \to \hat{\mathcal{X}}_{k+1}}(\hat{\mathcal{V}}_k + \beta \exp_{\hat{\mathcal{X}}_k}^{-1}(\mathcal{Y}_k)), \tag{4}$$

where $\hat{\mathcal{X}}_k$ and $\hat{\mathcal{V}}_k$ denote the estimated state and the estimated velocity at time step k. Notice also that this discrete-time

TABLE I

LUENBERGER OBSERVER ALGORITHM ON $G(n, p)$

inputs: the current state representation: (X_k, V_k), the measurement: Y_k, and the parameters α and β
outputs: the next estimation of the state (X_{k+1}, V_{k+1})

$\begin{bmatrix} X_k^\top Y_k \\ (Y_k - X_k X_k^\top Y_k) \end{bmatrix} = \begin{bmatrix} W_1 \cos(D) Z^\top \\ W_2 \sin(D) Z^\top \end{bmatrix}$ (CS Decomposition)
set: $L = W_2 D W_1^\top$
compute: $(V_k + \alpha L) = U \Sigma W^T$ (compact svd)
set: $D_k = X_k W$
set: $X_{k+1} = (D_k \cos(\Sigma) + U \sin(\Sigma)) W^\top$
set: $B = (U^\top V_k + \beta U^\top L)$
set: $V_{k+1} = (-D_k \sin(\Sigma) + U \cos(\Sigma)) B + (V_k + \beta L - UB)$

observer can be seen as the generalization of the following discrete-time observer in R^n:

$$x_{k+1} = x_k + v_k + \alpha(y_k - x_k),$$
$$v_{k+1} = v_k + \beta(y_k - x_k),$$

where x_k, v_k and y_k belong to R^n. In fact, the addition of the position x_k and the velocity v_k is replaced by an exponential map in (3), the difference between two points is computed using a log-mapping, and the parallel transport is used to relate tangent vectors that do not belong to the same tangent space in (4).

This observer can be implemented efficiently using two $n \times p$ matrices: one whose column space represents the subspace and another one that represents the velocity of the subspace. In fact, as we have seen in section II, there exist formulas to compute the parallel transport, the exponential map and its inverse in $O(np^2)$ operations on $G(n, p)$ using this representation. The implementation of the observer is summarized in TABLE I.

IV. SIMULATIONS AND COMPARISONS

We simulated a piecewise geodesic trajectory on $G(10, 5)$. So, the distance between the subspaces is constant on each piece as shown in Fig. 1. The Luenberger observer has an important overshoot when the velocity vector changes but is able to filter the noise, see the red curve in Fig. 1. We have also compared this observer with the particle filtering technique of [3] on $G(4, 2)$. Similar results are obtained and shown in Fig. 2. The main advantage of the Luenberger observer over the particle filtering technique is its low computational cost. In fact, 2000 particles are required to get good results with the particle filtering method and this number increases when n or p increases.

V. EXTENSIONS

This section briefly presents some extensions of the presented observer. We have considered a very simple dynamics but the generalization to other dynamics is straightforward. For instance, it is possible to extend our approach to a constant acceleration model. The acceleration $\hat{\mathcal{A}}_k$ will be also modeled by a tangent vector at $\hat{\mathcal{X}}_k$ and our observer

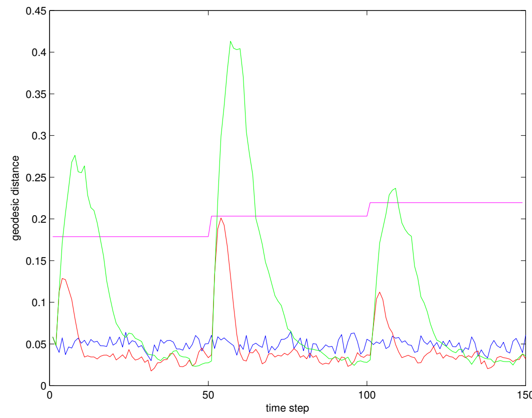

Fig. 1. The magenta curve represents the distance between the subspaces $d(\mathcal{X}_k, \mathcal{X}_{k+1})$ of the true trajectory (unperturbed). The blue curve represents the difference between the perturbed data $\mathcal{Y}$ and the unperturbed data $\mathcal{X}$, i.e., the level of the noise. The difference between the filtered data and the unperturbed data is represented in red for the Luenberger observer (with $\alpha = 0.5$ and $\beta = 0.1$) and in green for the modified observer that deals with measurements in R^n (with $\alpha = 0.1$ and $\beta = 0.01$).

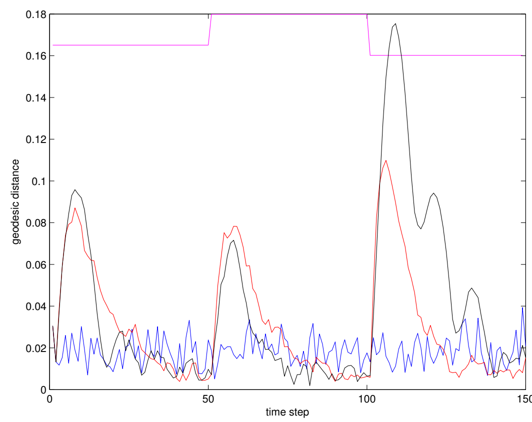

Fig. 2. Difference between the true trajectory and the filtered one for the Luenberger observer with $\alpha = 0.3$ and $\beta = 0.02$ (red curve) and for the particle filtering technique with 2000 particles (black curve). As in Fig. 1, the magenta curve represents the distance between the subspaces $d(\mathcal{X}_k, \mathcal{X}_{k+1})$ of the true trajectory (unperturbed). The blue curve represents the difference between the perturbed data $\mathcal{Y}$ and the unperturbed data $\mathcal{X}$, i.e., the level of the noise.

will become:

$$\hat{\mathcal{X}}_{k+1} = \exp_{\hat{\mathcal{X}}_k}(\hat{\mathcal{V}}_k + \alpha \exp^{-1}_{\hat{\mathcal{X}}_k}(\mathcal{Y}_k)),$$
$$\hat{\mathcal{V}}_{k+1} = \Gamma_{\hat{\mathcal{X}}_k \to \hat{\mathcal{X}}_{k+1}}(\hat{\mathcal{V}}_k + \hat{\mathcal{A}}_k + \beta \exp^{-1}_{\hat{\mathcal{X}}_k}(\mathcal{Y}_k)),$$
$$\hat{\mathcal{A}}_{k+1} = \Gamma_{\hat{\mathcal{X}}_k \to \hat{\mathcal{X}}_{k+1}}(\hat{\mathcal{A}}_k + \gamma \exp^{-1}_{\hat{\mathcal{X}}_k}(\mathcal{Y}_k)),$$

where α, β and γ are real parameters. Notice that this observer will be just a bit more expensive to compute since it requires the parallel transport of the acceleration vector.

If the measurements do not belong to $G(n, p)$, we can replace the function $\frac{1}{2} d(\mathcal{X}, \mathcal{Y})^2$ by another function that depends on the subspace $\mathcal{X}$. Its gradient with respect to $\mathcal{X}$ will be a tangent vector at $\mathcal{X}$. This approach is called gradient based observer and is studied in [10] for observers on Lie groups. In our case, if we measure only points

$y \in R^n$ and not subspaces, one possible choice could be: $\frac{1}{2}||(I - XX^\top)y||_F^2$. The gradient of this function is given by: $-(I - XX^\top)yy^T X$. This approach can be extended to an arbitrary number of points. In this case, the gradient will be $-(I - XX^\top)CX$, where C is the covariance matrix of the data points. We tested this approach by generating, for each measured subspace, p data points y of unit norm. The result is shown in Fig. 1 in green.

VI. CONCLUSIONS AND FURTHER WORK

We have shown how to implement a Luenberger-like observer to filter data on the Grassmann manifold. This approach seems to be more attractive from a computational point of view than the particle filtering approach on problems where the dimension is high as in [2]. Furthemore, this approach can be extended to other dynamical models on Grassmann. The convergence region of our observer seems to be quite large and a particularization of the convergence result obtained in [9] will be considered in further work.

REFERENCES

[1] R. Weber, "Subspace Tracking for Mobile Communications," *Techn. Univ. München, Munich, Germany, Tech. Rep. TUM-LNS-TR-97-7*, 1997.

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

[4] Quentin Rentmeesters, P.-A. Absil, Paul Van Dooren, Kyle Gallivan, and Anuj Srivastava, "An efficient particle filtering technique on the Grassmann manifold," in *Proceedings of IEEE international conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. 3838–3841.

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

[6] M.P. Do Carmo, *Riemannian geometry*, Birkhauser, 1992.

[7] 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.

[8] 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.

[9] 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.

[10] C Lageman, J Trumpf, and R Mahony, "Gradient-like observers for invariant dynamics on a lie group," Tech. Rep. arXiv:0805.0828, May 2008, Comments: 10 pages.