Submitted to IEEE Transactions on Signal Prcoessing in June, 2000

Geometric Filtering for Subspace Tracking *

Anuj Srivastava † Eric Klassen ‡

Abstract

We address the problem of tracking principal subspaces using ideas from nonlinear filtering. The subspaces are represented by their complex projection-matrices, and time-varying subspaces correspond to trajectories on the Grassmann manifold. Under a Bayesian approach, we impose a smooth prior on the velocities associated with the subspace motion. This prior combined with any standard likelihood function forms a posterior density on the Grassmannian, for filtering and estimation. Using a sequential Monte Carlo method, a recursive nonlinear tracking algorithm is derived and some implementation results are presented.

Keywords: Subspace tracking, Grassmannian, sequential Monte Carlo, particle filtering, homogeneous spaces, principal components
EDICS: 2-ESTM

1 Introduction

Principal-component analysis (PCA), or estimation of principal subspaces, of a large-dimensional observation space, is of interest in many signal and image processing applications. A simple example is the problem of estimating the m-dimensional principal subspace of the sample covariance matrix $K = \frac{1}{p}\sum_{i=1}^{p} y_i y_i^\dagger$, where $y_i \in \mathbb{C}^n$ are the observation vectors, and $\dagger$ denotes the conjugate transpose ($0 \leq m \leq n < \infty$ are integers). Consider an extension in which the observations reflect a time-varying system and therefore the principal subspace is also changing in time. Let $Y_t = [y_{t,1}\ y_{t,2}\ \ldots y_{t,p}]$, $y_{t,i} \in \mathbb{C}^n$ be the set of p observations (p being much smaller than n), collected in a small interval around time t. Then, the time-varying sample covariance matrix is given by

$$K_t = \frac{1}{p}\sum_{i=1}^{p} y_{t,i}\ y_{t,i}^\dagger\ , \quad y_{t,i} \in \mathbb{C}^n\ . \tag{1}$$

The observations $y_{t,i}$ span a complex vector space of n complex dimensions and the problem is now modified to tracking the sequence of m-dimensional principal subspaces of K_t. This paper studies the problem of *tracking m principal components using limited observations from a time-varying n-dimensional system.*

A number of approaches have been presented in the literature for estimating subspaces and also tracking their variations in time, with applications in signal processing (see [10, 25, 7] and the references therein). One focus of the subspace-related research has been the improvement in

*This work is supported in part by ARO DAAG55-98-1-0102, NSF-9871196, and DAAD19-99-1-0267.This paper was presented in part at SAM 2000 workshop held in Boston, MA, on March 2000.

†Department of Statistics, Florida State University. anuj@stat.fsu.edu

‡Department of Mathematics, Florida State University. klassen@math.fsu.edu

computational speed for a given accuracy in estimation, especially for adaptive systems [6, 20, 1, 19, 14]. Another focus is on investigating fundamental representations and resulting formulations which can lead to improved estimators but perhaps at an added computational cost. As an example, a natural way to represent subspaces is as elements of the Grassmann manifold, the set of all (fixed-dimensional) subspaces of a larger space ([5, 8, 11, 9]). A recent paper [23] presents a Bayesian framework for estimating subspaces as elements of complex Grassmann manifold. For a specific choice of distance function on Grassmannian, an MMSE estimator is defined and evaluated for a given posterior.

An important issue in subspace tracking is that, at any given time, there may not be a sufficient number of observations for estimating each of the individual subspaces to a required precision. A related possibility is that the observations are too noisy to provide a reliable estimate at each observation time. A natural choice, in such problems relating to time-series estimation, is to utilize a *temporal structure*, in order to compensate for the lack of excess observations or to account for a large noise level. More formally, the idea is to utilize a Bayesian framework and to derive a prior probability model, on the underlying time series, which encourages smoother trajectories. As an example, in subspace tracking, one possible prior is that which penalizes coarser subspace trajectories and emphasizes smoother trajectories. An even stronger prior can be imposed if the velocities, rather than the subspaces (displacements) themselves, are assumed to vary smoothly in time. In the group-theoretic representations, this has often been accomplished by equating the time derivatives of elements of the Lie algebra (they relate to the velocities) to white noise. For example in [18, 24], in the context of tracking airplanes using remote sensing, the rates of change in rotational and translational velocities (or equivalently the driving torques and forces) are treated as white noise. This procedure imposes a physics-based, smooth prior on airplane trajectories and leads to Bayesian filtering for airplane tracking. Similar to that analysis, an important result in this paper is the derivation of a dynamics-based smoothing prior on the subspace trajectories.

Our motivation for subspace tracking comes from array signal processing where the time-varying principal subspace, associated with the sensor measurements, is tracked. A class of signal processing algorithms, for transmitter tracking and beamforming, rely on estimating the principal subspace spanned by the observation vectors. Consider the situation in which an array of sensors is receiving signals trasmitted by a number of moving transmitters. Due to transmitter motion, there may not be enough sensor measurements at each time to accurately estimate the individual subspaces using likelihood-based techniques. Therefore, a prior supporting smooth changes in the principal subspace may prove useful. This prior is physically motivated since smooth changes in transmitters' locations lead to smooth variation in the resulting subspaces, with a few exceptions. A classical exception is the case when a transmitter disappears from the receiver's view perhaps due to an obstruction, and the signal subspace changes in rank. In most cases, however, sensor physics supports smooth relative variations. As an example, consider a 4-element uniform, linear sensor array (at half wavelength spacing) receiving narrowband signals from two moving transmitters, as shown in Figure 1 left panel. The dominant subspace (for $m = 2$ and $n = 4$) has the projection matrix given by $P_t = D(\theta_t)(D(\theta_t)^\dagger D(\theta_t))^{-1} D(\theta_t)^\dagger$, where $D(\theta_t)$ is the 4×2 complex matrix composed of the direction vectors for transmitters' locations at time t. Shown in the middle panel of Figure 1 are the trajectories taken by the two transmitters, with respect to the linear array, and plotted in the right panel is the distance $\|P_t - P_0\|$, where $\|\cdot\|$ stands for the Frobenious norm. As the picture suggests, for smooth transmitter trajectories the resulting subspace rotations is observed to be smooth.

In this paper, we pursue a geometric approach to Bayesian subspace tracking, using observations from a time-varying system. A posterior density has components from (i) a prior density, and (ii) a data likelihood function. Allowing for any general likelihood function, from popular maximum-likelihood type approaches, we focus on deriving a prior density which establishes a tem-

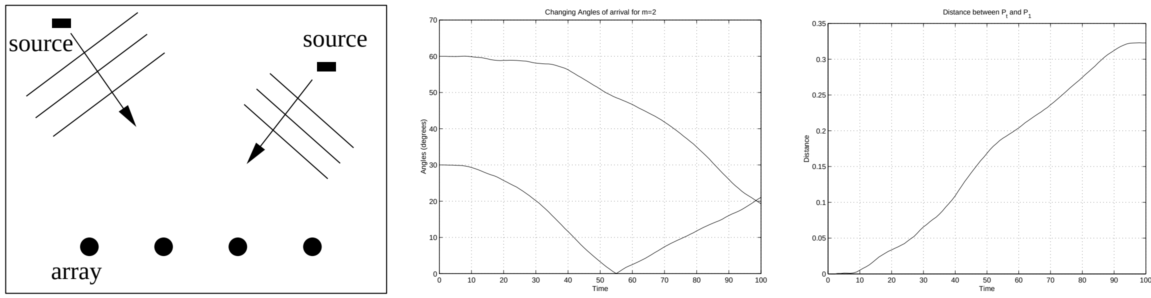

Figure 1: Left panel: array configuration for tracking signal sources; middle panel: the angles of arrival as functions of time; right panel: changes in the signal subspace due to source motion.

poral structure on the rotating subspaces. For this purpose, we study the intrinsic geometry of the Grassmanian, to learn various representations of the subspaces. Let $\mathcal{G}$ be the Grassmann manifold of all m-dimensional complex subspaces of $\mathbb{C}^n$. Any element of this manifold (i.e. any subspace of $\mathbb{C}^n$) can be represented by its projection operator (uniquely) or by a set of orthonormal basis elements (non-uniquely). Using representation via projection matrices, we analyze the velocity vector associated with the trajectories on Grassmanian. To define velocity, we first characterize the geodesic curve connecting any two points on this manifold. Since the Grassmannian is a quotient space of a larger unitary group, modulo two smaller rotations, this geodesic is made explicit by lifting it to a particular geodesic in the unitary group. Furthermore, the tangents to the *lifted* geodesic curve in the unitary group relate to the velocities associated with moving subspaces. This definition is then utilized to impose a prior on the subspace rotation, by equating the rate of change of velocities with white noise. A combination of this prior with any standard likelihood function sets up the Bayesian filtering problem.

Once the representations are chosen, velocities are defined and the probabilities are imposed, the focus shifts to solving for estimates, given the observations. In view of the inherent nonlinearities present in the model, and the problem formulation on curved manifolds, the classical Kalman-filtering framework does not apply. A number of solutions including the extended Kalman filters, interacting multiple models [2], multiple hypothesis tests, and their combinations [27], have been suggested. An MCMC-based sampling procedure for filtering and smoothing is presented in [18]. Here, we take a random sampling approach for computation of (approximate) MMSE estimates. This procedure, based on *particle filtering* or the *sequential Monte Carlo method*, involves sampling from the prior and resampling them according to their likelihoods, in order to generate samples from the posterior, at each observation time. These samples are then averaged to compute the estimates. This emerging family of techniques have been discussed in [12, 3, 16], and a related idea is presented in [18].

Section 2 studies the geometry of Grassmannian in order to define the motion parameters (displacements, velocities) associated with the subspace rotation, and imposes a smooth Markovian prior on the subspace trajectories. Section 3 states a Bayesian nonlinear filtering formulation of the tracking problem and presents a sequential Monte Carlo method for generating MMSE solutions. Some simulation results illustrating the algorithm, for a generic problem in subspace tracking and a particular problem in array signal subspace tracking, are presented in Section 4.

2 Representation of Subspace Motion

Subspace estimation and tracking are well studied problems in the signal processing literature. Even though these problems are naturally posed in terms of manifold-valued parameters, the use of geometric techniques has only been recent [5, 8, 11, 9, 23] . In a recent paper [23], we have presented a geometric approach to estimating a *single* subspace, as an element of a complex Grassmann manifold; we have chosen to represent a point on this manifold by the projection matrix associated with that subspace. In this paper, our goal is to analyze and estimate trajectories, on Grassmann manifold, associated with dynamic systems and changing subspaces. We study the geometric representations of subspaces leading up to a Bayesian formulation of the subspace tracking problem. One reason to study the intrinsic geometry of Grassmanians is *to define the motion parameters such as displacements and velocities* for piecewise-geodesic trajectories on these manifolds. These definitions are analogous to the Euclidean case in that the velocities are determined by the tangent vectors along the geodesics connecting the observed points. At first we introduce the geometry of the Grassmannian and then explicitly characterize the geodesics.

2.1 Geometry of Grassmannian

Let V be an n-dimensional complex vector space equipped with a Hermitian inner product. Assuming $0 \leq m \leq n < \infty$, denote by $\mathcal{G}$ the Grassmanian of all m-dimensional subspaces of V (please refer to [15] p. 133 Ex. 2.4 for a detailed introduction). By fixing m, n throughout the paper, we avoid adding suffixes to index the set $\mathcal{G}$. Using an orthonormal basis $\{v_1, v_2, \ldots, v_n\}$ for V, identify V with $\mathbb{C}^n$, the set of $n \times 1$ column vectors over $\mathbb{C}$. Each point of $\mathcal{G}$ can be identified with a unique $n \times n$ matrix of orthogonal projection onto that m-dimensional subspace of V. Let $\mathbb{P}$ be the set of Hermitian symmetric, idempotent $n \times n$ complex matrices of rank m; $\mathbb{P}$ is the set of all projection matrices and is diffeomorphic to $\mathcal{G}$. $\mathcal{G}$ (or $\mathbb{P}$) is a compact manfold of complex dimension $m(n-m)$. The subspace spanned by the vectors $\{v_1, v_2, \ldots v_m\}$ is identified with the projection matrix Q where

$$Q = \begin{bmatrix} I_m & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and } I_m \text{ is an } m \times m \text{ identity matrix.} \tag{2}$$

Q is fixed throughout the paper. Let $U(n)$ be the Lie group of all $n \times n$ complex unitary matrices. Its subgroup $H = U(m) \times U(n-m)$, of $U(n)$, is the set of matrices of the form $\begin{bmatrix} U_a & 0 \\ 0 & U_b \end{bmatrix}$, where $U_a \in U(m)$ and $U_b \in U(n-m)$. There is a one-to-one correspondence between the quotient space $U(n)/H$ and the Grassmanian $\mathcal{G}$ (or $\mathbb{P}$) (see [15] p.134). The left coset, containing a point $U \in U(n)$, can be explicitly stated as $UH = \{U\tilde{U} : \tilde{U} \in H\} \subset U(n)$. The correspondence between the left cosets (elements of $U(n)/H$) and the projection matrices (elements of $\mathbb{P}$) is given by $UH \mapsto UQU^\dagger$, for any $U \in U(n)$. Denote the projection map from $U(n)$ to $\mathbb{P}$ by $\Phi : U(n) \mapsto \mathbb{P}, \quad \Phi(U) = UQU^\dagger$. Under Φ, each left coset projects to a point in $\mathbb{P}$. Denote this set by $\Phi^{-1}(P)$ (notice that $\Phi^{-1}(P) \equiv UH$ whenever $UQU^\dagger = P$).

Remark 1: An element of $\Phi^{-1}(P)$ is a unitary matrix, whose first m columns form an orthonormal basis of the subspace whose projection is P. Note that the $n \times n$ identity matrix I is an element of $\Phi^{-1}(Q)$, and therefore, $\Phi^{-1}(Q)$ is nothing but H.

The group $U(n)$ acts on the vector space V (from left) by the usual matrix-vector multiplication (please refer to [4] p. 90 Def 7.1 for a definition and some examples of group action). $U(n)$ acts transitively on $\mathbb{P}$ from the left, according to the mapping: $P \mapsto U \cdot P \equiv UPU^\dagger, \;\text{for}\; U \in U(n), \;\; P \in$

$\mathbb{P}$. The transitive group action implies that $\mathbb{P} = \{U \cdot Q : U \in U(n)\}$. Note that Φ is invariant to the group action: $\Phi(U \cdot U_1) = U \cdot \Phi(U_1)$, for all $U,\ U_1 \in U(n)$.

It can be shown that the Lie algebra of $U(n)$ is $\mathcal{U}$, the space of $n \times n$, Hermitian skew-symmetric matrices (see for example [26] p. 107). Similarly, let $\mathcal{H}$ denote the Lie sub-algebra of the subgroup H:

$$\mathcal{H} = \left\{ \begin{bmatrix} Y_a & 0 \\ 0 & Y_b \end{bmatrix} \in \mathbb{C}^{n\times n} : \ \ Y_a \in \mathbb{C}^{m\times m},\ Y_b \in \mathbb{C}^{(n-m)\times(n-m)} \text{ are Hermitian skew-symmetric} \right\} .$$

Let $\mathcal{M}$ be the orthogonal complement of $\mathcal{H}$ in $\mathcal{U}$. $\mathcal{M}$ is given by

$$\mathcal{M} = \left\{ \begin{bmatrix} 0 & A \\ -A^{\dagger} & 0 \end{bmatrix} \in \mathbb{C}^{n\times n} \ : \ A \in \mathbb{C}^{m(n-m)} \right\} . \tag{3}$$

As a compact Lie group, $U(n)$ comes equipped with a unique bi-invariant Riemannian metric, which is inherited by $\mathbb{P}$. On $\mathcal{U}$, this metric is just the inner product $\langle Y_1\,,\, Y_2\rangle = trace(Y_1 Y_2^{\dagger})$. Since $\langle UY_1\,,\, UY_2\rangle = \langle Y_1\,,\, Y_2\rangle$, for any $U \in U(n)$, this metric is invariant to the left translation generated by the group action.

2.2 Geodesics on Grassmannians

Eventually we are interested in estimating trajectories on $\mathbb{P}$ using sensor measurements taken at discrete times. To discretize an underlying trajectory along the observation points, we approximate it by a piecewise-geodesic curve. This corresponds to connecting the values attained at observation times by geodesics. Therefore, we are interested in an explicit description of the geodesic connecting any two given points on $\mathbb{P}$. One characterization of geodesics on Grassmanian comes from the following result. Define the exponential of a square matrix using the infinite series, if it converges, as $\exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots$.

Proposition 1 *The geodesics in $\mathbb{P}$ passing through the point Q (at time $t = 0$) are of the type*

$$\alpha : (-\epsilon, \epsilon) \mapsto \mathbb{P}, \qquad \alpha(t) = \exp(tX) \cdot Q = \exp(tX) Q \exp(-tX) \ ,$$

for some $X \in \mathcal{M}$, where the set $\mathcal{M}$ is specified in Eqn. 3.

Proof: Please refer to [13] (p. 221. Ex 2(i)) for technical details. We will provide a sketch of the proof here. Let α be the geodesic in $\mathbb{P}$ connecting Q (at $t = 0$) with P (at $t = 1$). Geodesics in $\mathbb{P}$ are made explicit via the corresponding geodesics in $U(n)$, since $\mathbb{P}$ can be identified with the quotient space $U(n)/H$. The geodesics in $U(n)$, passing through a point $U \in U(n)$, are the one-parameter subgroups of the type $\beta(t) = \exp(tX) \cdot U$ for any $X \in \mathcal{U}$. The geodesic β (in $U(n)$) projects down to a geodesic α (in $\mathbb{P}$) if and only if β is orthogonal to each coset that it intersects. On the other hand, invariance of the metric implies that if β is orthogonal to one coset, then it is orthogonal to each and every coset it intersects. In particular, if β passes through I (at $t = 0$), it should be orthogonal to H (or $\dot{\beta}(0) \perp \mathcal{H}$). For $\beta(t) = \exp(tX) \cdot I$, this condition implies that X belongs to the orthogonal complement of $\mathcal{H}$ in $\mathcal{U}$, namely $\mathcal{M}$. Finally, the projection of β to $\mathbb{P}$ gives α, using the invariance of Φ,

$$\begin{aligned} \alpha(t) &= \Phi(\beta(t)) = \Phi(\exp(tX) \cdot I) = \exp(tX) \cdot \Phi(I) \\ &= \exp(tX) \cdot Q = \exp(tX) Q \exp(-tX) \ . \end{aligned}$$

In view of this result, we first restrict to finding the geodesics connecting Q to some point $P \in \mathbb{P}$, and later extend it to arbitrary two points in $\mathbb{P}$. Let $\alpha : (-\epsilon, \epsilon) \mapsto \mathbb{P}$ be a geodesic in $\mathbb{P}$ such that

$\alpha(0) = Q$ and $\alpha(1) = P$. According to Proposition 1, α is completely specified by an $X \in \mathcal{M}$ such that $\exp(X) \cdot Q = \exp(X) Q \exp(-X) = P$. Therefore, the problem of finding α becomes:

Problem 1: Given a point $P \in I\!P$, find an $X \in \mathcal{M}$ such that $\exp(X) Q \exp(-X) = P$ and then set $\alpha(t) = \exp(tX) \cdot Q$.

Note that for this X, $\exp(X) \in \Phi^{-1}(P)$. The following theorem provides an algorithm for finding this X, but first we motivate it by a simple example.

Example 1: Consider a two-dimensional vector space V. Let $\langle v_1, w_1 \rangle$ be an ordered basis for V and $Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is the projection matrix associated with the subspace spanned by v_1. Let P denote (the projection matrix of) the one-dimensional subspace spanned by $\cos(\alpha) v_1 + \sin(\alpha) w_1$. Then,

$$P = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} = \begin{bmatrix} \cos^2(\alpha) & \cos(\alpha)\sin(\alpha) \\ \cos(\alpha)\sin(\alpha) & \sin^2(\alpha) \end{bmatrix} \in I\!P \, .$$

A simple calculation shows that the eigenvalues of the difference $Q - P$ are $\sin(\alpha)$ and $-\sin(\alpha)$, with the relation

$$Q - P = W \Sigma W^\dagger \ , \quad \text{where} \ \ \Sigma = \begin{bmatrix} \sin(\alpha) & 0 \\ 0 & -\sin(\alpha) \end{bmatrix} , \tag{4}$$

and where the columns of W are eigenvectors of $Q - P$ corresponding to $\sin(\alpha)$ and $-\sin(\alpha)$, respectively. Since $Q - P$ is Hermitian symmetric, W can be taken to be a unitary matrix. We require that if w_1, w_2 are the columns of W, then Qw_1 and Qw_2 should be *positive real* multiples of each other. This can be achieved simply by multiplying w_1 by an appropriate unit complex number. Returning to the task of finding a geodesic connecting Q to P, as per Problem 1, we have to find an X such that $P = \exp(X) \cdot Q$. It follows that this X is given by

$$X = W \Omega W^\dagger, \ \text{ and } \ \exp(X) = W \exp(\Omega) W^\dagger, \ \text{ where } \ \Omega = \begin{bmatrix} 0 & -\alpha \\ \alpha & 0 \end{bmatrix} . \tag{5}$$

In the 2×2 case $W \Omega W^\dagger = \Omega$.

This example suggests a role for the eigen decomposition of $Q - P$ in finding the X.

Theorem 1 *For a point $P \in I\!P$, let $B = W \Sigma W^\dagger$ be the eigen decomposition of $B = Q - P$. Then,*

1. *the eigen-values of B (or the diagonal entries of Σ) are either 0's or occur in pairs of the form $(\lambda_j, -\lambda_j)$, where $0 < \lambda_j \leq 1$. If needed, modify W such that Qw_j and $Q_{w_{j'}}$ are positive real multiple of each other, where w_j, $w_{j'}$ are the columns of W corresponding to the eigenvalues λ_j and $-\lambda_j$, respectively.*

2. *Let Ω be a $n \times n$ matrix derived from Σ in the following way: replace the 2×2 blocks*

$$\begin{bmatrix} \lambda & 0 \\ 0 & -\lambda \end{bmatrix} \quad by \quad \begin{bmatrix} 0 & -\sin^{-1}(\lambda) \\ \sin^{-1}(\lambda) & 0 \end{bmatrix} ,$$

with the remaining entries staying zeros. Then, X is given by $X = W \Omega W^\dagger \ \in \mathcal{M}$.

3. *The matrix* $\exp(X)$*, can be computed using* $W\tilde{\Omega}W^\dagger \in U(n)$*, where* $\tilde{\Omega}$ *is formed from* Σ *by replacing: (i) the zeros in the diagonal by ones, and (ii) the* 2×2 *blocks*

$$\begin{bmatrix} \lambda & 0 \\ 0 & -\lambda \end{bmatrix} \quad by \quad \begin{bmatrix} \sqrt{1-\lambda^2} & -\lambda \\ \lambda & \sqrt{1-\lambda^2} \end{bmatrix}.$$

Proof: Please refer to the Appendix.
We note that for certain points in $I\!\!P$, the matrix X, and therefore, the resulting geodesic may not be unique. In the context of subspace tracking, these pairs occur with negligible probability, and hence are ignored. Also, note that the computational cost of calculating X is essentially that of finding the eigen-decomposition of $Q - P$.

Remark 2: $\exp(X)$ is the "best rotation" from Q to P. In addition to $\exp(X) \in \Phi^{-1}(P)$, among all the elements of $\Phi^{-1}(P)$, $\exp(X)$ is the nearest to I in terms of their geodesic distances.

The next step is to generalize to the problem of finding a geodesic between arbitrary two points P_1, P_2 in $I\!\!P$. The basic idea is to rotate these points to Q and P (for some $P \in I\!\!P$), respectively and apply earlier steps. Let $U_1 \in \Phi^{-1}(P_1)$ (that is, $P_1 = U_1 Q U_1^\dagger$) and define $P = U_1^\dagger P_2 U_1$. Then, using Theorem 1, we can find an X such that $\alpha(t) = \exp(tX) \cdot Q$ is a geodesic from Q to P. Define a shifted geodesic $\tilde{\alpha}$ according to $\tilde{\alpha}(t) = U_1 \cdot \alpha(t)$; $\tilde{\alpha}(t)$ is the desired geodesic in $I\!\!P$ such that $\tilde{\alpha}(0) = P_1$ and $\tilde{\alpha}(1) = P_2$. Furthermore, the "best possible" rotation from P_1 to P_2, denoted by U, can be found as follows: $P_1 = U_1 Q U_1^\dagger$, $P_2 = U_1 P U_1^\dagger$, and $P = \exp(X) Q \exp(-X)$ imply that $P_2 = U P_1 U^\dagger$, where $U = U_1 \exp(X) U_1^\dagger$. The element X is dependent on the choice of U_1 but the matrix U is independent of U_1. In our tracking procedure, we will make an arbitrary choice for rotation U_1 at the initial time and the rotations for all the following times will be specified accordingly.

Now we are ready to define the motion parameters such as the displacement and the velocity in going from P_1 to P_2 in unit time. As stated above, we can calculate a matrix X which defines a geodesic $\tilde{\alpha}$ from P_1 to P_2. Remember that X is an element of $\mathcal{M}$ ($\mathcal{M}$ is defined in Eqn. 3) and therefore has only $2m(n-m)$ degrees of freedom in the form of the submatrix A in the upper-right corner of X. We define the entries of $A = X(1:m, n+1:m)$ as the *velocities* in rotating from P_1 to P_2 in unit time. A similar characterization of the geodescis and velocities on Grassmannian is suggested in [9]. The matrix $\exp(X)$ denotes the displacement between P_1 and P_2. In summary, given arbitrary P_1 and P_2 we can find the velocity matrix as per above discussion. Conversely, for a point $P_1 \in I\!\!P$, and a given $m \times (n-m)$ complex matrix A, we can find the point $P_2 \in I\!\!P$, that is reached in unit time by starting at P_1 and having the velocity A. This can be accomplished as follows. Let U_1 be any element of $\Phi^{-1}(P_1)$. First, form an $n \times n$ Hermitian skew-symmetric matrix according to $X = \begin{bmatrix} 0 & A \\ -A^\dagger & 0 \end{bmatrix}$, compute $U = U_1 \exp(X) U_1^\dagger$, and then set $P_2 = U P_1 U^\dagger$. Also, note that $U_2 = U_1 \exp(X)$ is an element of $\Phi^{-1}(P_2)$.

2.3 Prior on Subspace Trajectories

Having defined the displacement and the velocity matrix between two arbitrary points in $I\!\!P$, we consider a trajectory on $I\!\!P$ and impose probabilities on the velocities in such a way that the smoother trajectories are more probable than the coarser trajectories. As stated earlier, we discretize smooth trajectories by piecewise-geodesic curves; these curves are completely specified by an initial point $P_1 \in I\!\!P$ (with a choice of $U_1 \in \Phi^{-1}(P_1)$) and the successive velocities $A_1, A_2, \ldots, A_t \in \mathbb{C}^{m(n-m)}$. We impose a conditional prior density on A_t, given A_{t-1}, which favors an A_t with values similar to those of A_{t-1}. Let $\{P_t : t = 1, 2, \ldots\}$ be a discrete-time process in $I\!P$. For each pair (P_{t-1}, P_t), $t = 2, 3, \ldots$, let A_{t-1} be the submatrix of velocities, as defined in the last section. To impose a prior density which results in smoother trajectories, we utilize the dynamic model:

$$A_t = A_{t-1} + \mu_{t-1}, \quad t = 2, 3, \ldots \ , \tag{6}$$

where μ_{t-1} is a $m \times (n-m)$ matrix of i.i.d complex normals (real, imaginary parts are i.i.d normal with mean zero and variance σ_p^2). σ_p is the measure of deviation in values of A_t, away from a given value of A_{t-1}. The conditional density on A_t, conditioned on the previous velocity A_{t-1}, is given by

$$f(A_t|A_{t-1}) = (\frac{1}{2\pi\sigma^2})^{m(n-m)} \exp(-\frac{1}{\sigma^2}\|A_t - A_{t-1}\|^2) \ , \quad t = 2, 3, \ldots \ . \tag{7}$$

As described later, the tracking algorithm will not require the explicit functional form of the prior density; it will depend only on the samples generated from this density. In a Markovian time-series analysis, often there is a standard characterization of a time-varying posterior density, in a convenient recursive form. This characterization relates an underlying Markov process to its observations at each observation time via a pair of filtering equations. To retain the Markov property, we study the filtering problem on the joint space of subspaces and velocities. Define the subspace-velocity pair $J_t = (P_t, A_{t-1}) \in (I\!P \times \mathbb{C}^{m(n-m)})$, for each time t. J_t is a discrete-time Markov process. For the purpose of defining the velocities A_t's, we will keep track of the corresponding U_t's in $\Phi^{-1}(P_t)$'s. Notice that for a given value of the pair $J_{t-1} = (P_{t-1}, A_{t-2})$, only one of the components in J_t, either P_t or A_{t-1}, is random; given one the other is completely specified. This setup leads to the prior density on the joint (Markov) process:

$$\begin{aligned} f(J_t|J_{t-1}) &= f(A_{t-1}|A_{t-2}) f(P_t|A_{t-1}, P_{t-1}, A_{t-2}) \\ &= f(A_{t-1}|A_{t-2}) \delta_{P_t'}(P_t) \quad \text{where} \quad P_t' = W_{t-1} P_{t-1} W_{t-1}^\dagger \\ &\quad \text{and} \quad W_t = U_{t-1} \exp(X_{t-1}) U_{t-1}^\dagger, \quad X_{t-1} = \begin{bmatrix} 0 & A_{t-1} \\ -A_{t-1}^\dagger & 0 \end{bmatrix} \ , \end{aligned}$$

for any $U_{t-1} \in \Phi^{-1}(P_{t-1})$. For the next time step, $U_t = U_{t-1} \exp(X_{t-1}) \ \in \Phi^{-1}(P_t)$. The following algorithm specifies a procedure to sample from the conditional prior $f(J_t|J_{t-1})$:

Algorithm 1 *For some $t = 2, 3, \ldots$, we are given the values for $J_{t-1}^{(i)}$ and points $U_{t-1}^{(i)} \in \Phi^{-1}(P_{t-1}^{(i)})$. For $i = 1, 2, \ldots, M$:*

1. *Generate a sample of $A_{t-1}^{(i)}$, given $A_{t-2}^{(i)}$, according to Eqn. 6.*
2. *For each sample of $A_{t-1}^{(i)}$, set $X_{t-1}^{(i)} = \begin{bmatrix} 0 & A_{t-1}^{(i)} \\ -(A_{t-1}^{(i)})^\dagger & 0 \end{bmatrix}$, and calculate $P_t^{(i)}$ according to $P_t^{(i)} = W_{t-1}^{(i)} P_{t-1}^{(i)} (W_{t-1}^{(i)})^\dagger$, for $W_{t-1}^{(i)} = U_{t-1}^{(i)} \exp(X_{t-1}^{(i)}) (U_{t-1}^{(i)})^\dagger$.*
3. *Define the sampled subspace-velocity pair $J_t^{(i)} = (P_t^{(i)}, A_{t-1}^{(i)})$. Set $U_t^{(i)} = U_{t-1}^{(i)} \exp(X_{t-1}^{(i)})$.*

3 Bayesian Nonlinear Filtering

Now we have a prior density on piecewise-geodesic trajectories in $I\!P$ that favors smoother trajectories. Combining this prior with a likelihood function, we formulate a posterior density on $I\!P$ and pose the filtering problem.

3.1 Filtering Problem

We start by formulating the problem of subspace tracking as a problem in Bayesian nonlinear filtering. For the discrete observation times $t = 1, 2, \ldots$, the subspace trajectory is given by the sequence $P_1, P_2, \ldots \in \mathbb{P}$, and let the observation sequence be given by $Y_1, Y_2, \ldots \in \mathbb{C}^{np}$. A precise problem statement for the filter is:

Problem 2: Given the observation sequence $Y_{1:t} = \{Y_1, \ldots, Y_t\}$, estimate the sequence $P_{1:t} = \{P_1, \ldots, P_t\} \in \mathbb{P}^t$ using a minimum mean-squared error (MMSE) criterion.

As t increases, the underlying parameter space ($\mathbb{P}^t$) grows and the joint posterior, on $P_1, \ldots, P_t$, changes at each time as the new observation is recorded. Solving for the mean of the joint posterior (or the MMSE estimate) at each time, while the parameter space is growing, is a difficult problem. Recent papers [12, 3, 16, 21], describe an efficient procedure, called particle filtering or sequential Monte Carlo method, to solve such Bayesian problems. This procedure is a greedy method in that it restricts estimation to only the parameters at the last time t and utilizes a Monte Carlo technique to sample from the posterior probability associated only with P_t. It does not utilize the current measurement, say Y_t, to improve the estimation of previous states P_1 to P_{t-1}.

Remark 3: Previous estimates $(\hat{P}_1, \ldots, \hat{P}_{t-1})$ remained unchanged in the estimation steps performed at time t. In return, it provides faster speed with a possibility of real-time implementations. One consequence is that a tracking algorithm based on this procedure is useful (beyond a likelihood-based method) mostly when the noise is intermittent. If the noise level is consistently high, a procedure which involves both filtering and smoothing will be needed, at the added computational cost incurred in smoothing.

The Monte Carlo idea is to approximate the posterior density of P_t by a large number of samples drawn from it. Having obtained the samples, any estimate of P_t (MMSE, MAP, MAE etc.) can be approximated using sample averages. An example, described in [23], is a MMSE estimate of P_t given by

$$\hat{P}_t = \hat{U}_t Q (\hat{U}_t)^\dagger \ , \quad \text{where} \ \ G_t = \hat{U}_t \Sigma (\hat{U}_t)^\dagger \ \text{is the SVD of} \ \ G_t = \int_{\mathbb{P}} P_t f(P_t | Y_{1:t}) \gamma(dP_t) \ , \tag{8}$$

and where $f(P_t|Y_{1:t})$ is the posterior density of P_t given all the observations up to time t. Therefore, using the Monte Carlo idea, the samples generated from $f(P_t|Y_{1:t})$ can be used to approximate this integral, and then the MMSE estimate can be computed using SVD. Computational efficiency of these sequential methods comes from the recursion that takes samples from the posterior density of P_{t-1} and generates samples from the posterior density of P_t. We develop such a formulation for the subspace tracking problem.

The filtering equations, on the joint space of subspace-velocity pair, are, for $t = 2, 3, \ldots$

$$f(J_t|Y_{1:t-1}) = \int_{\mathbb{P} \times \mathbb{C}^{m(n-m)}} f(J_t|J_{t-1}) f(J_{t-1}|Y_{1:t-1}) \gamma(dJ_{t-1}) \ , \tag{9}$$

$$f(J_t|Y_{1:t}) = \frac{f(Y_t|J_t) f(J_t|Y_{1:t-1})}{f(Y_t|Y_{1:t-1})} \ . \tag{10}$$

Eqn. 9 is called the prediction equation and Eqn. 10 is called the update equation. The denominator in Eqn. 10 is difficult to compute and, for a given observation set, is a constant; we will denote it by $Z_t \equiv f(Y_t|Y_{1:t-1})$. One distinct advantage of the Monte Carlo approaches is that this normalizing constant need not be explictly evaluated. This relationship between Eqns. 9 and 10 suggests a recursive form for the solutions derived from the posteriors $f(J_{t-1}|Y_{1:t-1})$ and $f(J_t|Y_{1:t})$. That is, given samples from $f(J_{t-1}|Y_{1:t-1})$ it is possible to efficiently generate samples from $f(J_t|Y_{1:t})$, instead of directly sampling from $f(J_t|Y_{1:t})$, which may be complicated and computationally expensive.

Before we present an algorithm for the recursive sampling, we specify a posterior density. As described later, the algorithm requires two components from the posterior: (i) an ability to sample from the prior $f(J_t|J_{t-1})$ for a given value of J_{t-1}, and (ii) the functional form of the likelihood function $f(Y_t|J_t)$.

3.2 Likelihood Function

In principle, this framework allows for any standard likelihood function relating the unknown parameter and the observed data. The two examples that we consider are:

1. In a generic case, let $D \in \mathbb{C}^{n\times m}$ be a unitary matrix ($D^\dagger D = I$) and let the observation be modeled as:

$$Y_t = P_t D + \nu_t \quad \in \mathbb{C}^{n\times m}, \quad t = 1, 2, \dots \ , \tag{11}$$

   where $\nu_t \in \mathbb{C}^{n\times m}$ is additive noise and $\{P_t\}$ is the unknown trajectory on $\mathbb{P}$. If ν_t has i.i.d normal elements (with independent real and imaginary parts) with mean zero and variance σ^2, the likelihood function is $f(Y_t|J_t) = \frac{1}{L_t}\exp(\frac{1}{\sigma^2}(trace(P_t Y_t Y_t^\dagger)))$, L_t is the normalizer.

2. In the case of array signal processing, the sensor observations are modeled as superpositions of signals received from multiple transmitters and the ambient noise. Let there be n sensors and m signal transmitters ($m \le n$), and the angular location vector is denoted by $\theta_t = [\theta_{1,t}, \ \dots, \theta_{m,t}] \in [0, \pi]^m$. The observation vector is given by:

$$y_{t,i} = D(\theta_t)s_t + \nu_{t,i}, \ i = 1, 2, \dots, p \quad \in \mathbb{C}^n \ , \tag{12}$$

   where $D(\theta_t) = [d(\theta_{1,t}), \dots, d(\theta_{m,t})] \in \mathbb{C}^{n\times m}$, for $d(\theta) = [1\exp(-j\phi) \ \exp(-j2\phi) \ \dots \ \exp(-j(n-1)\phi)]^T$, $\phi = \pi\cos(\theta)$. s_t is the m-vector of signal amplitudes and, as earlier, $\nu_{t,i}$ is $i.i.d.$ complex normal noise. If K_t is the sample covariance, as defined in Eqn. 1, the likelihood function is given by $f(Y_t|J_t) = \frac{1}{L_t}\exp(\frac{1}{\sigma^2}(trace(K_t P_t)))$, where L_t is the normalizer. This normalizer need not be specified in Monte Carlo based inference.

3.3 Sequential Monte Carlo Approach

Our concern is to find the MMSE estimate of P_t given the observations $Y_{1:t}$, for each t. Taking a Monte Carlo simulation approach, we first generate samples from the posterior $f(J_t|Y_{1:t})$, extract the subspace components from the joint samples, and then compute sample averages to approximate the integral in Eqn. 8. In view of the complicated relationship between J_t and the observation set, $Y_{1:t}$, it is difficult and often inefficient to sample directly from the posterior $f(J_t|Y_{1:t})$.

A recursive formulation, which takes samples from $f(J_{t-1}|Y_{1:t-1})$ and generates the samples from $f(J_t|Y_{1:t})$ in an efficient fashion, is desirable. We accomplish this task using ideas from sequential methods and importance sampling. Assume that, at the observation time $t-1$, we have a set of M samples from the posterior, $S_{t-1} = \{J_{t-1}^{(i)} : i = 1, 2, \dots, M\}$, $J_{t-1}^{(i)} \sim f(J_{t-1}|Y_{1:t-1})$. Following are the steps which utilize elements of S_{t-1} to generate the set S_t.

1. **Prediction**: The first step is to sample from $f(J_t|Y_{1:t-1})$ given the samples from $f(J_{t-1}|Y_{1:t-1})$. We take a *compositional approach* by treating $f(J_t|Y_{1:t-1})$ as a mixture density. According to Eqn. 9, $f(J_t|Y_{1:t-1})$ is the integral of the product of a marginal and a conditional density. This implies that, for each element $J_{t-1}^{(i)} \in S_{t-1}$, by generating a sample from the conditional, $f(J_t|J_{t-1}^{(i)})$, we can generate a sample from $f(J_t|Y_{1:t-1})$. Of course, this method is practical only when there is an efficient technique to sample from the prior density $f(J_t|J_{t-1})$. In our

case, this is accomplished using Algorithm 1. Now we have samples $\{\tilde{J}_t^{(i)}\}$ from $f(J_t|Y_{1:t-1})$; similar to the Kalman-filtering notation these samples are called *predictions*.

2. **Resampling**: Given these predictions, the next step is to generate samples from the posterior $f(J_t|Y_{1:t})$. For this, we utilize the notion of *importance sampling* in the following way. The samples from the prior ($f(J_t|Y_{1:t-1})$) are resampled (see reference [16]) according to the probabilities that are proportional to the likelihoods $f(Y_t|\tilde{J}_t^{(i)})$. Form a discrete probability mass function on the set $\{\tilde{J}_t^{(i)} : i = 1, 2, \ldots, M\}$ according to

$$\beta_{t,i} = \frac{f(Y_t|\tilde{J}_t^{(i)})}{\sum_{j=1}^{M} f(Y_t|\tilde{J}_t^{(j)})} \ , \ \text{ and set } \ \beta_t = [\beta_{t,1} \ \beta_{t,2} \ \ldots \ \beta_{t,M}] \ . \tag{13}$$

   Then, resample M values from the set $\{\tilde{J}_t^{(1)}, \tilde{J}_t^{(2)}, \ldots, \tilde{J}_t^{(M)}\}$ according to the mass function β_t. These values are the desired samples from the posterior $f(J_t|Y_{1:t})$. The resampled set is denoted by $S_t = \{J_t^{(i)} : i = 1, 2, \ldots, M\}$, $J_t^{(i)} \sim f(J_t|Y_{1:t})$. It must be remarked that after resampling, the indices (i) are renamed so that the sequence $J_{t-1}^{(i)}, J_t^{(i)}, J_{t+1}^{(i)}, \ldots$, for the same i, may not be consistent anymore. In other words, it is possible that the velocity $A_{t-1}^{(i)}$ does not take $P_{t-1}^{(i)}$ to $P_t^{(i)}$ in a unit time. This inconsistency has no bearing on the estimation procedure since the past samples are not used in estimating future parameters, only the current samples are used.

3. **Averaging**: Now that we have M samples from the posterior $f(J_t|Y_{1:t})$, we can average them appropriately to approximate the MMSE estimate of P_t. As described in the paper [23], the MMSE estimate, of P_t, is given by Eqn. 8. Using Monte Carlo sampling, we approximate E by

$$\hat{G}_{t,M} = \frac{1}{M} \sum_{i=1}^{M} P_t^{(i)} \ \in \mathbb{C}^{n \times n} \ , \tag{14}$$

   and compute SVD of $\hat{G}_{t,M}$ to obtain the MMSE estimate $\hat{P}_{t,M}$.

In numerous papers, the ergodic properties of sequential Monte Carlo samples has been studied. It has been shown that the elements of the set S_t are exact samples from the posterior and the ergodic result (that is, sample averages converge to the expected values as the sample size gets larger) holds. It should be noted that due to the resampling step, the resulting samples are not independent of each other.

Error Analysis: There are two sources of error in this tracking procedure. First, there is the sampling error in estimating G_t by a finite sample mean $\hat{G}_{t,M}$ and it is quantified by the variance of the estimator. One can use the delta method to asymptotically estimate this variance (see for example [17]) and probabilistically bound the resulting sampling error. Conversely, the same result can also be used for sample size determination, for a desired estimation performance. The second source of error is due to the difference between the underlying true value P_t and its exact MMSE estimate obtained from G_t. This error can been quantified using Hilbert-Schmidt lower bounds (HSB) (see [23] for reference) on errors for estimation on matrix Lie groups. HSB is a lower bound on the expected squared error, and is achieved by the MMSE estimate defined in Eqns. 8 and 8. Another way to lower bound this error is using Cramer-Rao derivation as described in [22].

3.4 Algorithm

In this section, we write a step-by-step procedure for subspace tracking. Assume that for any time $t-1$ we have the samples $\{J_{t-1}^{(i)} : i = 1, 2, \ldots, M\} \sim f(J_{t-1}|Y_{1:t-1})$. The following algorithm outlines the steps to generate the samples at time t, and then, to estimate $\hat{P}_{t,M}$.

Algorithm 2 1. **Sample Conditional**: *Draw* $\{\tilde{J}_t^{(i)},\ i = 1, 2, \ldots, M\}$ *from the conditional prior according to Algorithm 1.*

2. **Importance Weights**: *Compute the probability mass function* $\beta_t^{(i)}$, $i = 1, 2 \ldots, M$, *according to Eqn. 13.*

3. **Resampling**: *Generate* M *samples from the set* $\{\tilde{J}_t^{(i)},\ i = 1, 2, \ldots, M\}$ *with the associated probabilities* $\{\beta_t^{(i)},\ i = 1, 2, \ldots, M\}$. *Denote these samples by* $\{J_t^{(i)},\ i = 1, 2, \ldots, M\}$.

4. **MMSE Averaging**: *Calculate the sample average* $\hat{G}_{t,M}$ *according to Eqn. 14 and compute the subspace estimate* $\hat{P}_{t,M}$ *using the SVD of* $\hat{G}_{t,M}$. *Set* $t \leftarrow t + 1$ *and go to step 1.*

A schematic diagram of the algorithm is shown in Figure 2.

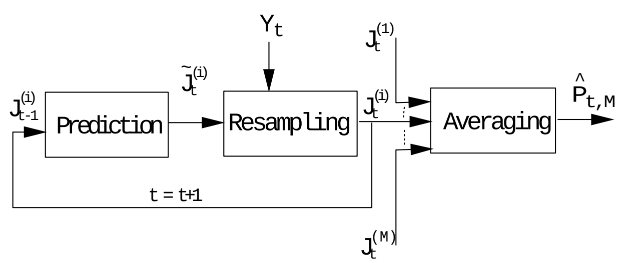

Figure 2: Schematic diagram of a Bayesian filter for subspace tracking.

4 Simulation Results

In this section, we present some experimental results on subspace tracking. These experiments are of two types, corresponding to the two data models given in Eqns. 11 and 12. According to Remark 3, in these experiments we restrict to the situations with intermittent noise, i.e. the additive noise has two levels: corresponding to σ and 1000σ. For any t, the data contains either low or high noise randomly, with probability 0.5 each.

In the diplays, each plot shows the estimation error $\|P_t - \hat{P}\|$ for three different estimation procedures. First, the error associated with the maximum-likelihood estimate (MLE), obtained by SVD of the covariance matrix K_t, is shown in the broken line. The error resulting from an adaptive procedure, relying on the SVD of the matrix $R_t = \gamma K_t + (1-\gamma)K_{t-1}$, is shown in the dotted line (for $\gamma = 0.3$). Finally, the estimation error for tracking resulting from Algorithm 2 is plotted in bold. Since the prior is based on smooth velocities, MLEs at $t = 1, 2$ are used to initialize the algorithm and Bayes' filter starts at $t = 3$.

1. First, consider a generic problem in subspace estimation, with the observation model given by Eqn. 11 for $n = 4$ and $m = 2$. D is chosen to be a fixed $n \times m$ unitary matrix. Let $P_1 = Q$ ($U_1 = I$) be the initial point, and $A_1 \in \mathbb{C}^{m(n-m)}$ be some initial velocity. Successive velocities, $A_2, A_3, \ldots$, are generated using the dynamic model given in Eqn. 6. The velocities give rise to

a subspace trajectory according to Step 2 of Algorithm 1. For each time t, the observations are generated according to the model given in Eqn. 11, and are used in Algorithm 2 to calculate the estimates.

In the following results we generated random trajectories on $\mathbb{P}$ using the prior energy at $\sigma_p = 0.001$ and the noise energy at $\sigma = 0.0002$. Shown in Figure 3 are the tracking results for four sample trajectories. To quantify the change in the underlying subspaces, $\|P_t - P_0\|$ is plotted using the cross marks. These curves show that the smoothness constraint imposed in form of the prior helps guard against intermittent noise.

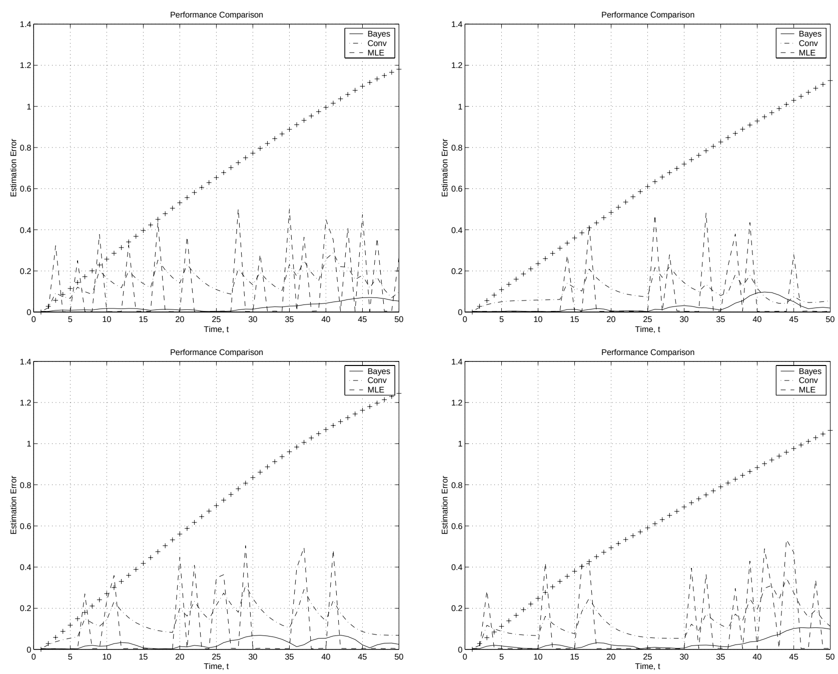

Figure 3: Estimation error versus time of three methods: MLE (broken line), adaptive filtering (dotted line) and the Bayesian filtering (solid line). Actual motion in the underling subspace is quantified by plotting $\|P_t - P_0\|$ using crosses.

2. Consider the problem of subspace estimation using a narrowband, uniform linear-array (ULA) consisting of five elements at half-wavelength spacing each ($n = 5$) (similar to Figure 1). Furthermore, assume that there are three signal transmitters ($m = 3$) moving with respect to the array, and transmitting signals that are received at the sensor according to the data model in Eqn. 12. For these experiments, the transmitter motion is generated according to the equation: for $i = 1, 2, \ldots, m$

$$\theta_{i,t} = [\theta_{i,t-1} + \vartheta_{i,t-1}]_{\text{mod } \pi} \ , \quad \text{where} \quad \vartheta_{i,t-1} = \vartheta_{i,t-2} + u_i \ ,$$

and where $u_i \sim N(0, \sigma_p^2)$. The initial conditions, $\theta_{1,0}, \ldots \theta_{m,0}$, are chosen uniformly between $[0, \pi)$. The lower panels of Figure 4 show some example trajectories of the transmitter motion according to this model for $t = 1, 2, \ldots, 50$. For each $\theta_t = [\theta_{1,t}, \ldots, \theta_{m,t}]$, the observation vector $y_{i,t}$ calculated according to Eqn. 12, with $p = m$. As earlier, the additive noise standard deviation is either σ or 1000σ, selected randomly with equal probability. For each t, m data vectors are generated and utilized to compute the sample covariance matrix K_t according to Eqn. 1. The tracking algorithm then estimates $\hat{P}_{t,M}$ for $M = 200$ at each time t.

Shown in Figure 4 are the estimation results in form of the estimation error $\|\hat{P}_t - P_t\|$ as a function of time, for three sample trajectories of the transmitter motion ($\sigma_p = 0.001$ and $\sigma = 0.0003$). The estimation error associated with the MLE is plotted in the broken line, the error associated with the adaptive tracking is plotted in the dotted line, and the error for Bayesian tracking is depicted by the solid line. Similar to the earlier experiment, the prior on

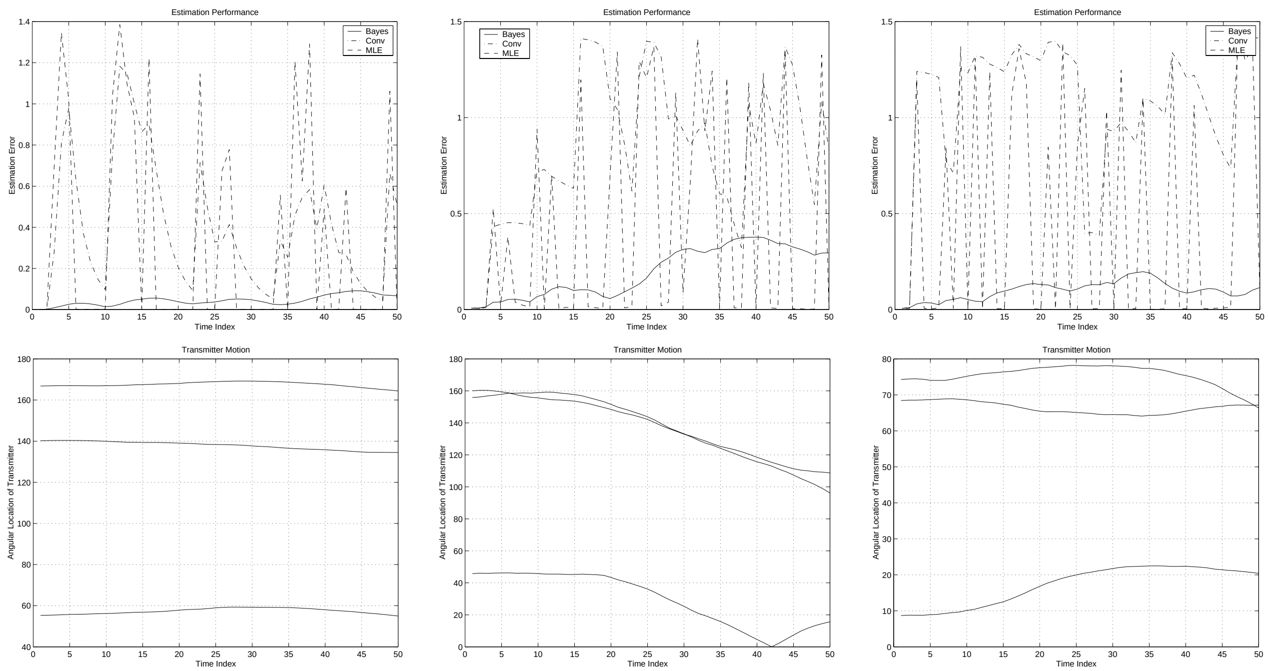

Figure 4: The upper panels plots the error in subspace tracking ($\|\hat{P}_t - P_t\|$) as a function of t for: (i) MLE (broken line), (ii) adaptive tracking (dotted line), and (iii) Bayesian tracking (solid line). The lower panels show the corresponding transmitter trajectories.

subspace motion improves tracking performance in the presence of intermittent noise.

Algorithm 2 is based on a greedy implementation of the Bayesian filter: it estimates only the current state based on the accumulated data. It gains in speed by fixing and not improving upon the past estimates once the current data are observed. A slower algorithm for joint estimation of all the states, using all the observations, is given in [18], and improvements are demonstrated even for situations with consistently high observation noise.

In view of the added smoothness constraint, in form of the prior density, this Bayesian method takes more computational effort than a standard likelihood based procedure. This cost is mostly incurred in Step 1 of Algorithm 2 where the exponential of a $n \times n$ matrix is calculated for each of the M samples.. The second reason for computational expense is the price paid for the ability to handle arbitrary nonlinear densities. Thsi approach is applicable to a large class of posterior densities on Grassmannian manifolds. In this nonlinear filter, one generates M samples at each observation time, for a large value of M. The total cost is linear in M.

5 Conclusion

In this paper, we have proposed a geometric approach to tracking principal subspaces of the observations taken from time-varying systems. It relies on imposing a smooth variation in the velocities associated with the subspace rotation. A recursive, computational technique to sample from the posterior and to generate MMSE estimates is described. The computational complexity of this technique remains to be investigated.

6 Acknowledgement

The authors wish to thank Prof. Daniel R. Fuhrmann of Washington University for many useful discussions on this research.

References

[1] K. Abed-Meraim, A. Chkeif, and Y. Hua. Fast orthonormal past algorithm. *IEEE Signal Processing Letters*, 7(3):60–62, March 2000.

[2] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association.* Academic Press, 1988.

[3] A. Blake and M. Isard. *Active Contours.* Springer, 1998.

[4] William M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry.* Academic Press, Inc., 1986.

[5] R. S. Bucy. Geometry and multiple direction estimation. *Information Sciences*, 57-58:145–58, 1991.

[6] J. Dehaene, M. Moonen, and J. Vandewalle. An improved stochastic gradient algorithm for principal component analysis and subspace tracking. *IEEE Transactions on Signal Processing*, 45(10):2582–2586, October 1997.

[7] J. P. Delmas and J. F. Cardoso. Performance analysis of an adaptive algorithm for tracking dominant subspaces. *IEEE transactions on signal processing*, 46(11):3045–3057, November 1998.

[8] A. Edelman, T. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM Journal of Matrix Analysis and Applications*, 20(2):303–353.

[9] D. R. Fuhrmann. The subspace tracking loop (poster presentation). *Proceedings of the ASAP*, pages 29–44, 11-12 March, 1999.

[10] D. R. Fuhrmann. An algorithm for subspace computation with applications in signal processing. *SIAM Jour. Matrix Anal. Appl.*, 9(2), April 1988.

[11] D. R. Fuhrmann, A. Srivastava, and H. Moon. Subspace tracking via rigid body dynamics. *Proceedings of Sixth Statistical Signal and Array Processing Workshop*, June, 1996.

[12] N. J. Gordon, D. J. Salmon, and A. F. M. Smith. A novel approach to nonlinear/non-gaussian bayesian state estimation. *IEEE Proceedings on Radar Signal Processing*, 140:107–113, 1993.

[13] S. Helgason. *Differential Geometry, Lie Groups and Symmetric Spaces.* Academic Press, 1978.

[14] Y. Hua, X. Yong, T. Chen, K Abed-Meraim, and Y. Miao. A new look at the power method for fast subspace tracking. *Digital Signal Processing*, 9(4):297–314, October 1999.

[15] S. Kobayashi and K. Nomizu. *Foundations of Differential Geometry, vol 2.* Interscience Publishers, 1969.

[16] J. S. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–44, September 1998.

[17] Jun S. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *website*, 1999.

[18] M. I. Miller, A. Srivastava, and U. Grenander. Conditional-expectation estimation via jump-diffusion processes in multiple target tracking/recognition. *IEEE Transactions on Signal Processing*, 43(11):2678–2690, November 1995.

[19] P. Pango and B. Champagne. On the efficient use of givens rotations in svd-based subspace tracking algorithms. *Signal Processing*, 74(3):253–77, May 1999.

[20] E. C. Real, D. W. Tufts, and J. W. Cooley. Two algorithms for fast approximate subspace tracking. *IEEE Transactions on Signal Processing*, 47(7):1936–45, July 1999.

[21] C. P. Robert and G. Castella. *Monte Carlo Statistical Methods.* Springer Text in Statistics, 1999.

[22] S. Smith. Intrinsic cramer-rao bounds and subspace estimation accuracy. *Proceedings of IEEE Sensor Array and Multichannel signal processing workshop*, March 16-17, 2000.

[23] A. Srivastava. A bayesian approach to geometric subspace estimation. *IEEE Transactions on Signal Processing*, 48(5):1390–1400, 2000.

[24] A. Srivastava, U. Grenander, G. R. Jensen, and M. I. Miller. Jump-diffusion markov processes on orthogonal groups for object recognition. *accepted for publication by Journal of Statistical Planning and Inference*, December, 1999.

[25] L. Tong and S. Perreau. Multichannel blind estimation: From subspace to maximum likelihood methods. *Proceedings of the IEEE*, 86(10):1951–1968, October 1998.

[26] Frank W. Warner. *Foundations of Differential Manifolds and Lie Groups.* Springer-Verlag, New York, 1994.

[27] Editor: Y. Bar-Shalom. *Multitarget-Multisensor Tracking.* Artech House, 1990.

A Proof of Theorem 1

Let $\tilde{P}$ and $\tilde{Q}$ be the m-dimensional subspaces of V, represented by projection matrices P and Q, respectively. We prove the theorem in three steps: At first, we prove it for the case $n = 2$ and $m = 1$. Next, we show that if there exists a basis of V such that $\tilde{P} = \exp(\Omega)\tilde{Q}$ for a specific form of $\Omega \in \mathcal{M}$, then the theorem is just an extension of the $n = 2$, $m = 1$ case. And finally, for any given $\tilde{P}$, we show that there exists a basis of V such that the requirements of the second step are met.

1. Start with the case where $n = 2$ and $m = 1$. Rule out the cases $\tilde{Q} \perp \tilde{P}$ and $\tilde{Q} = \tilde{P}$ since they are easy to handle. Let v_1 be the unit vector in $\tilde{Q}$, then Pv_1 is an element of $\tilde{P}$. Let w_1 be the unit vector in the real span of $\{v_1, Pv_1\}$ such that $w_1 \cdot Pv_1 > 0$. Let α_1 be the (positive) angle between v_1 and Pv_1. As shown in Example 1 (Eqn. 4), $Q - P$ can be decomposed as $W \begin{bmatrix} \lambda_1 & 0 \\ 0 & -\lambda_1 \end{bmatrix} W^\dagger$ for $\lambda_1 = \sin(\alpha_1)$.

   The resulting $X \in \mathcal{M}$ and $\exp(X) \in U(n)$ are given in Eqn. 5.

2. In a general case of an arbitrary $P \in I\!\!P$, we will essentially decompose V as orthonormal direct sum of two-dimensional subspaces to obtain the best rotation from $\tilde{Q}$ to $\tilde{P}$. The above result will apply independently to each two-dimensional component. Let there be an orthonormal basis of V of the form

$$\{u_1, \ldots, u_k, v_1, \ldots, v_r, w_1, \ldots, w_r, x_1, \ldots, x_p\} \tag{15}$$

   where k, r, p are three nonnegative integers such that $k + 2r + p = n$ and $k + r = m$. Also, let $\{u_1, \ldots, u_k, v_1, \ldots, v_r\}$ be an orthonormal basis of $\tilde{Q}$. For $\alpha_1, \ldots, \alpha_r$ positive real numbers, define an element of $\mathcal{U}$ by

$$\Omega = \left[\begin{array}{c|cc|c} 0_k & 0 & \cdots & 0 \\ \hline 0 & 0_r & diag(\alpha_1, \alpha_2, \ldots, \alpha_r) & \\ 0 & -diag(\alpha_1, \alpha_2, \ldots, \alpha_r) & 0_r & 0 \\ \hline 0 & 0 & \cdots & 0_p \end{array}\right] \in \mathcal{M},$$

   and define a subspace $\tilde{P} = \exp(\Omega)\tilde{Q}$. We can also write $\tilde{P} = \text{span}\{u_1, \ldots, u_k, \cos(\alpha_1)v_1 + \sin(\alpha_1)w_1, \ldots, \cos(\alpha_m)v_m + \sin(\alpha_m)w_m\}$. $u_1, \ldots u_k$ is a basis of $(\tilde{Q} \cap \tilde{P})$, and $x_1, \ldots, x_p$ is a basis of the space $(\tilde{Q}^\perp \cap \tilde{P}^\perp)$. With respect to the basis given in Eqn. 15, we can factor the rotation from $\tilde{Q}$ to $\tilde{P}$ into a sequence of 2×2 rotations in 2-planes orthogonal to each other. The planes are spanned by v_j, w_j and the rotation angles are α_j's. The results from Part 1 apply to each 2×2 rotation independently. Therefore, the eigen decomposition of $Q - P$ takes the form

$$W \left[\begin{array}{c|cc|c} 0_k & 0 & \cdots & 0 \\ \hline 0 & diag(\lambda_1, \ldots, \lambda_r) & 0_r & 0 \\ 0 & 0_r & -diag(\lambda_1, \ldots, \lambda_r) & 0 \\ \hline 0 & 0 & \cdots & 0_p \end{array}\right] W^\dagger, \quad \lambda_j = \sin(\alpha_j).$$

   We require that the eigen vectors corresponding to the eigen values λ_j and $-\lambda_j$ should project to a positive real multiple of each other, when multiplied by Q. If λ_j's are distinct, and w_j and $w_{j'}$ are the eigen vectors corresponding to λ_j and $-\lambda_j$, this can be achieved by multiplying $w_{j'}$ by an appropriate unit complex number. If several λ_j's are the same; for example, suppose that $\lambda_1 = \ldots = \lambda_s =: \lambda$. Then the columns $\{w_1, \ldots, w_s\}$ of W form an orthonormal basis for the eigenspace of the eigenvalue λ, while the columns $\{w_{1'}, \ldots, w_{s'}\}$ form an orthonormal basis for the eigenspace of the eigenvalue $-\lambda$. We will alter the columns $\{w_{1'}, \ldots, w_{s'}\}$ to form a new basis for this eigenspace. For each $i = 1 \ldots, s$, there is a unique unit vector $y_i \in \text{span}\{w_{1'}, \ldots, w_{s'}\}$ with the property that Qy_i and Qw_i differ from each other only by a positive real multiple. For each such i, replace $w_{i'}$ by y_i. Continue to call the resulting matrix W. Perform this procedure for each repeated λ. This completes the necessary modification of W.

Using the 2×2 example, $X = W\Omega W^\dagger$ is the desired $X \in \mathcal{M}$, and $\exp(X) = W\exp(\Omega)W^\dagger$. Therefore, if there exists a basis of V of the type given in Eqn. 15 such that $\tilde{Q}$ and $\tilde{P}$ can be written in these specific forms, the result follows from two-dimensional analysis.

3. Next, we show that for any arbitrary projection matrix $P \in I\!P$, there exists an orthonormal basis of V of the form given in Eqn. 15. Let $k = \dim(\tilde{Q} \cap \tilde{P})$ and $p = \dim(\tilde{Q}^\perp \cap \tilde{P}^\perp)$. Choose any orthonormal bases $\{u_1, \ldots, u_k\}$ for $\tilde{Q} \cap \tilde{P}$ and $\{x_1, \ldots, x_p\}$ for $\tilde{Q}^\perp \cap \tilde{P}^\perp$.

   Lemma 1 $\ker(Q - P) = (\tilde{Q} \cap \tilde{P}) \oplus (\tilde{Q}^\perp \cap \tilde{P}^\perp)$.

   Proof: If $v \in \tilde{Q} \cap \tilde{P}$, then $(Q - P)v = v - v = 0$. If $v \in (\tilde{Q} \cap \tilde{P})^\perp$, then $(Q - P)v = 0 - 0 = 0$. On the other hand, suppose $(Q - P)v = 0$. Then $Qv = Pv \in \tilde{Q} \cap \tilde{P}$. But, likewise, $v - Qv = v - Pv \in \tilde{Q}^\perp \cap \tilde{P}^\perp = (\tilde{Q} + \tilde{P})^\perp$. Since $v = Qv + (v - Qv)$, it follows that $v \in (\tilde{Q} \cap \tilde{P}) \oplus (\tilde{Q} \cap \tilde{P})^\perp$, proving the lemma. Furthermore, if we diagonalize $Q - P$, the multiplicity of 0 as an eigenvalue will be precisely $k + p$.

   For the remainder, replace V by the orthogonal complement in V of the subspace $(\tilde{Q} \cap \tilde{P}) \oplus (\tilde{Q} + \tilde{P})^\perp$, $\tilde{P}$ by the orthogonal complement in $\tilde{P}$ of $\tilde{P} \cap \tilde{Q}$, and $\tilde{Q}$ by the orthogonal complement in $\tilde{Q}$ of $\tilde{P} \cap \tilde{Q}$. Now, $\dim(\tilde{P}) = \dim(\tilde{Q}) = r$ and $\dim(V) = 2r$; $\tilde{P}$ and $\tilde{Q}$ are now transverse and span V. Let S_P and S_Q denote the unit spheres in $\tilde{P}$ and $\tilde{Q}$, respectively and let $\epsilon = \inf\{|v - w| : v \in S_Q \text{ and } w \in S_P\}$. Since S_P and S_Q are disjoint compact sets, $\epsilon > 0$ and there exist vectors $v \in S_Q$ and $w \in S_P$ satisfying $|v - w| = \epsilon$. It follows that Pv is a positive real multiple of w and Qw is a positive real multiple of v. Furthermore, the projection operators P and Q, and hence their difference $Q - P$, all map the 2-dimensional subspace of V spanned by $\{v, w\}$ to itself. Call this subspace Z. We conclude that P and Q each restrict to rank 1 projection operators on Z. From Example 1, the restriction of $Q - P$ to Z has two eigenvalues $\sin(\alpha_1)$ and $-\sin(\alpha_1)$, where α_1 is the angle between the image of P and the image of Q (restricted to Z). One of these images may be moved to the other by a rotation of angle α_1 in the space Z. Hence, we choose basis elements of Z as: Let $v_1 = v$, and let w_1 be the unique unit vector in the *real* span of $\{v, w\}$ such that (1) $w_1 \perp v$ and (2) $w_1 \cdot Pv > 0$. To construct the rest of the basis, inductively replace $\tilde{P}$, $\tilde{Q}$, and V by the orthogonal complement of Z in each of them, and repeat this step procedure to find $(v_2, w_2), \ldots, (v_r, w_r)$.