BAYESIAN ESTIMATION AND TRACKING OF DYNAMIC SIGNAL SUBSPACES

Anuj Srivastava

Department of Statistics

Florida State University

CAVIS
Center for Applied Vision and Imaging Sciences
@ Florida State University

SPAWC, Tuesday, July 13^{th}, Lisbon.

TOOLS FROM DIFFERENTIAL GEOMETRY

Quick Comments:

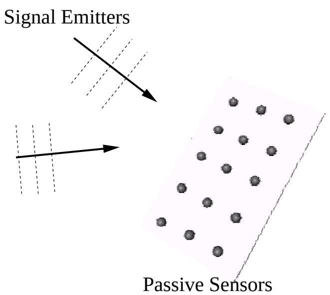
1. Timely for signal processing. Statistics was a new tool two to three decades ago; differential geometry is a new tool now.
2. Difficult topic for engineering community but there are rewards. Starting cost is high. Perhaps, there should be tutorials at conferences.
3. There is a lot more research to be done, especially in applications, and even in theoretical areas.
4. Such methods are increasingly popular in other areas such as computer vision and pattern recognition.

OUTLINE

1. Subspace Estimation/Tracking in Signal Processing
2. Formulation of Nonlinear Filtering for Subspace Tracking
3. Geometry of Grassmann Manifolds
4. Prior: Subspace Dynamics
5. Monte Carlo methods for Nonlinear Filtering
6. A Simple Experiment
7. Summary and Open Issues

Motivation, Explanation, and Demonstration

SIGNAL SUBSPACE TRACKING



For d signal transmitters

observed using n sensors:

$$y_{t,i} = D(\Theta_t)s_t + \mu_{t,i}\ ,\ i = 1, \ldots, k$$

- $s_t \in \mathbb{C}^d$ is the signal amplitude,
- $\Theta_t \in [0, \pi)^d$ is the vector of transmitter locations at time t,
- $D(\Theta_t) \in \mathbb{C}^{n \times d}$ is the matrix of direction vectors, and
- $\mu_{t,i} \in \mathbb{C}^n$ is additive noise.
- $y_{t,i} \in \mathbb{C}^n$ is the i^{th} observation vector at time t.
  $\mathbf{y}_t = [y_{t,1}, \ldots, y_{t,k}] \in \mathbb{C}^{nk}$.

SIGNAL SUBSPACE ESTIMATION

Classical Approach

- Assuming, additive white Gaussian noise, $\mu_{t,i} \sim CN(0, \sigma^2 I_n)$

  Each time can be treated independently from other times.

- Maximum likelihood estimate (MLE) is derived as follows:

$$(\hat{s}_t, \hat{\Theta}_t) = \underset{s_t, \Theta_t}{\operatorname{argmin}} \|\mathbf{y}_t - D(\Theta_t)s_t\|^2 .$$

  - First, fix Θ_t and maximize over s_t:

$$\hat{s}_t = (D^\dagger D)^{-1} D^\dagger \mathbf{y}_t \ , \quad \text{where} \ \ D \equiv D(\Theta_t).$$

- Substituting it back, we get

$$\hat{\Theta}_t = \underset{\Theta}{\operatorname{argmin}} \|(I - D(D^\dagger D)^{-1} D^\dagger)\mathbf{y}_t\|^2 \ .$$

  Or,

$$\hat{\Theta}_t = \underset{\Theta}{\operatorname{argmax}} \|D(D^\dagger D)^{-1} D^\dagger \mathbf{y}_t\|^2 \ .$$

  $D(D^\dagger D)^{-1} D^\dagger$ is the projection onto the subspace (of $\mathbb{C}^n$) spanned by d columns of D.

- If the columns of D are linearly independent (signal transmitters are not too close), then a d-dimensional subspace of $\mathbb{C}^n$ is of interest.

SIGNAL SUBSPACE ESTIMATION

So, the problem can be solved in two steps:

1. Step 1: **Subspace Estimation**: A subspace can be represented by an orthonormal basis (non-uniquely) or by a projection matrix (uniquely).

   Task:

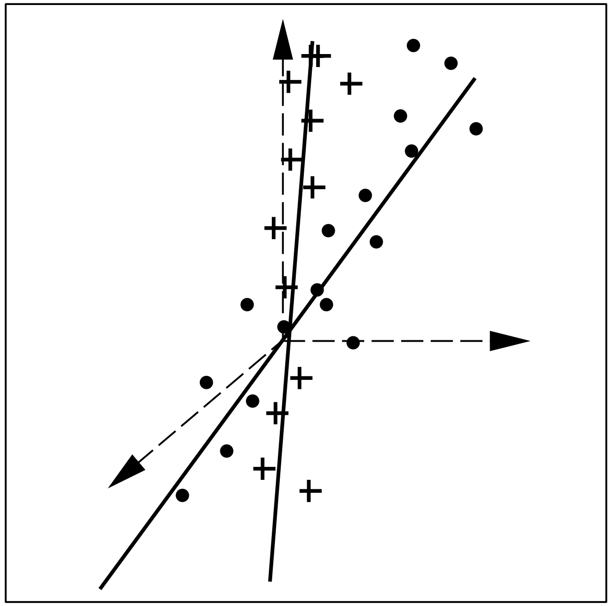
   - Solve for a basis S_t such that span($S_t) = \text{span}(D_t)$.
     S_t is an $n \times d$ unitary matrix.
   - Or, estimate a projection matrix P_t that best fits the data.
     $P_t \in \mathbb{C}^{n\times n}$, $P_t^\dagger = P_t$, $P_t^2 = P_t$, rank(P_t) = d.

   For computational reasons, we choose a basis representation.

2. Step 2: **Angle Estimation**: Solve for the angles Θ_t that best match the estimated subspace.

This talk is about Step 1.

SIGNAL SUBSPACE TRACKING



plusses denote observations at time t_1

dots denote observations at time t_2

An interesting problem is subspace tracking when data is occasionally very bad (transmitter obscuration, noise burst, etc).

SIGNAL SUBSPACE TRACKING

Goal: To utilize dynamics of subspace to improve estimation. To estimate temporal evolution of subspace rather than individual estimation. Path estimation instead of point estimation.

Problem Statement: Given measurements $\mathbf{y}_{[1:T]} \equiv \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T$, estimate the subspace sequence (via unitary bases) $S_1, S_2, \ldots, S_T$.

Filtering: A greedy solution is that given all the measurements up to time t, estimate the basis S_t.

Some Issues:

1. On what space this filtering problem is to be solved?

   The set of all (complex) d-dimensional subspaces of $\mathbb{C}^n$ is called *Grassmann manifold* $\mathcal{G}_{n,d}$.

2. What tools are available to us to solve such a filtering problem?

   Kalman Filter? Nonlinear Filtering, or Non-Euclidean Filtering?

KALMAN FILTERING

Use a state equation and an observation equation:

$$\begin{aligned} \text{State equation :} \quad & x_{t+1} = f(x_t) + \nu_t \\ \text{Observation equation :} \quad & y_{t+1} = g(x_{t+1}) + \mu_t \end{aligned}$$

Then, find the mean and covariance associated with the posterior at time t

$$P(x_t | y_{[1:t]}) \ .$$

If f and g are known linear functions, (and μ_t, ν_t are independent Gaussian), then Kalman filtering provides exact solution.

Two Problems:

1. Kalman Filter does not directly apply to nonlinear manifolds.
2. We need some form of a state equation on $\mathcal{G}_{n,d}$. We do have an observation model.

BAYESIAN NONLINEAR FILTERING

Nonlinear Manifolds: Kalman filtering does not apply, but Monte Carlo methods do!

Use nonlinear filtering equations: Under usual Markov assumptions,

$$\begin{aligned}
\text{Predict}: P(S_{t+1}|\mathbf{y}_{[1:t]}) &= \int_{S_t} P(S_{t+1}|S_t)P(S_t|\mathbf{y}_{[1:t]})dS_t \\
\text{Update}: P(S_{t+1}|\mathbf{y}_{[1:t+1]}) &= \frac{P(\mathbf{y}_{t+1}|S_{t+1})P(S_{t+1}|\mathbf{y}_{[1:t]})}{P(\mathbf{y}_{t+1}|\mathbf{y}_{[1:t]})}
\end{aligned}$$

- $P(S_{t+1}|S_t)$: comes from the state equation, or subspace dynamics
- $P(\mathbf{y}_t|S_t)$ comes from the likelihood function.

Next, we develop a model for subspace dynamics.

GRASSMANN MANIFOLDS: BACKGROUND

- Let the columns of $S \in \mathbb{C}^{n\times d}$ denote an orthonormal basis of a d-dimensional subspace of $\mathbb{R}^n$.
- $\mathbb{U}(d) = \{O \in \mathbb{C}^{d\times d} | O^\dagger O = I_d\}$.
- Define an equivalence class of bases:

$$[S] = \{SO : O \in \mathbb{U}(d)\} .$$

  $[S]$ contains all orthonormal bases that span the same subspace. $[S]$ denotes a subspace and S denotes a particular basis.

- Grassmann manifold:

$$\mathcal{G}_{n,d} = \{[S] : S \in \mathbb{C}^{n\times d}, S^\dagger S = I_d\} .$$

  $\mathcal{G}_{n,d}$ has real dimension = $2d(n-d)$.

GEOMETRY OF GRASSMANN MANIFOLD

- Grassmann manifolds are quotient spaces of unitary groups.
- Consider the embedding: $\phi : (\mathbb{U}(d) \times \mathbb{U}(n-d)) \mapsto \mathbb{U}(n)$, given by:

$$\phi(V_1, V_2) = \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix}.$$

  This way $\phi(\mathbb{U}(d), \mathbb{U}(n-d))$ is a subset of $\mathbb{U}(n)$.

- Consider an equivalence relation in $\mathbb{U}(n)$:

$$Q_1 \sim Q_2 \text{ if and only if } Q_1 = Q_2 \begin{bmatrix} V_1 & 0 \\ 0 & V_2 \end{bmatrix},$$

  for some $V_1 \in \mathbb{U}(d)$, and $V_2 \in \mathbb{U}(n-d)$.

  The quotient space $\mathbb{U}(n)/\sim$ or $\mathbb{U}(n)/(\mathbb{U}(d) \times \mathbb{U}(n-d))$ is a complex Grassmann manifold $\mathcal{G}_{n,d}$.

GEOMETRY OF UNITARY GROUP

The set of unitary matrices is not a vector space. It is a group under matrix multiplication.

1. **Tangent Spaces**:
   - At $I_n \in \mathbb{C}^{n\times n}$, the space of vectors (actually, matrices) is given by:

     $$T_{I_n}(\mathbb{U}(n)) = \{A \in \mathbb{C}^{n\times n} | A^\dagger + A = 0\}\ .$$

     The set of Hermitian skew-symmetric matrices.
   - At any other point $Q \in \mathbb{U}(n)$, the tangent space is:

     $$T_Q(\mathbb{U}(n)) = \{QA \in \mathbb{C}^{n\times n} | A^\dagger + A = 0\}\ .$$

   Assume the usual Euclidean inner product on tangent spaces.

2. **Geodesic Flows**: Geodesics are one-parameter flows

$$t \mapsto Q\exp(tA), \quad \text{where} \quad A \in \mathbb{C}^{n\times n}, \quad A^\dagger + A = 0\,.$$

This is a geodesic flow starting at Q and in the direction of A.

We want tangent spaces and geodesic flows on Grassmann manifolds.

GEOMETRY OF GRASSMANN MANIFOLD

- Let $J = \begin{bmatrix} I_d \\ 0 \end{bmatrix} \in \mathbb{C}^{n \times d}$ denote the first d columns of I_n.
- For any $Q \in \mathbb{U}(n)$, the matrix $[QJ] \in \mathcal{G}_{n,d}$. Multiplication on right projects from $\mathbb{U}(n)$ to $\mathcal{G}_{n,d}$.
- ϕ induces an embedding $d\phi$ from tangent spaces to tangent spaces:

$$d\phi : (T_{V_1}(\mathbb{U}(d)) \times T_{V_2}(\mathbb{U}(n-d))) \mapsto T_{\phi(V_1,V_2)}(\mathbb{U}(n)) \ ,$$

given by:

$$d\phi(A_1, A_2) = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix} ,$$

for some $V_1 \in \mathbb{U}(d)$, and $V_2 \in \mathbb{U}(n-d)$.

- A vector tangent to $\mathbb{U}(n)$ projects to a tangent to $\mathcal{G}_{n,d}$ if and only if it is orthogonal to the range of $d\phi$.

GEOMETRY OF GRASSMANN MANIFOLD

1. **Tangent Space**: For an $[S] \in \mathcal{G}_{n,d}$, let $Q \in \mathbb{U}(n)$ such that $QJ = S$. Q is not unique; it has to be computed efficiently. Then, the tangent space

$$T_{[S]}(\mathcal{G}_{n,d}) = \{QAJ | A = \begin{bmatrix} 0 & B \\ -B^\dagger & 0 \end{bmatrix}, B \in \mathbb{C}^{d\times(n-d)}\} \; .$$

2. **Geodesic Flow**:

   A Geodesic in $\mathbb{U}(n)$ is also a geodesic in $\mathcal{G}_{n,d}$ as long as it is orthogonal to every equivalence class it meets

   Geodesics on $\mathcal{G}_{n,d}$ is given by one parameter flows:

$$t \mapsto Q \exp(tA) J \; .$$

   We will denote the flow by $\Psi_S(t, A) = Q \exp(tA) J$.

TWO IMPORTANT TASKS

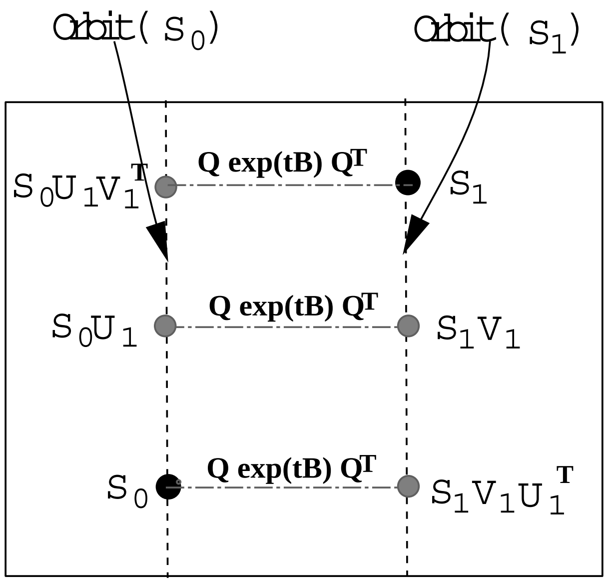
To perform differential calculus and statistics on $\mathcal{G}_{n,d}$, two operations are repeatedly used:

1. **Exponentiation**: Given $[S_0]$ and a tangent direction A, find elements along the geodesic $\Psi(t)$ at $t_1, t_2, \ldots$.

2. **Logarithm**: Given two subspaces $[S_0]$ and $[S_1]$, find the direction A (or B inside A) such that $\Psi(0) = [S_0]$ and $\Psi(1) = [S_1]$.

We want to be able to perform these tasks efficiently.

GENERAL BACKGROUND



- Let the SVD of $S_0^\dagger S_1 \in \mathbb{C}^{d\times d}$ be given by $(S_0^\dagger S_1) = U_1 \Gamma V_1$.
- Define canonical bases $\overline{S}_0 = S_0 U_1$ and $\overline{S}_1 = S_1 V_1$, and canonical flow $\overline{\Psi}(t) = \Psi(t) U_1 = Q \exp(tA) J U_1$.
- $\|\Gamma\|$ is the length of geodesic between two subspaces.

TASK 1: EXPONENTIATION

Goal: Given S_0, Q and $B \in \mathbb{C}^{(n-d)\times d}$, construct $\overline{\Psi}(t)$.

- Let $\tilde{U}_2 \Theta U_1^\dagger$ be the SVD of $B^\dagger$.
- Determine $\Gamma(t) = \cos(t\Theta)$ and $\Sigma(t) = \sin(t\Theta)$.
- Since $\overline{\Psi}(t) = Q \exp(tA) J U_1$, we have

$$\dot{\overline{\Psi}}(0) = Q \begin{pmatrix} 0 \\ -B^\dagger \end{pmatrix} U_1 = -C_0 \tilde{U}_2 \Theta = -D\Theta \tag{1}$$

  for $D \equiv C_0 \tilde{U}_2$. Compute D using Q, B, Θ, and U_1

- Then,

$$\overline{\Psi}(t) = S_0 U_1 \Gamma(t) - (C_0 \tilde{U}_2)\Sigma(t) = S_0 U_1 \Gamma(t) - D\Sigma(t).$$

Exponentiation can be accomplished in $O(nd^2)$ computations

TASK 2: LOGARITHM

Goal: Given two subspaces $[S_0]$ and $[S_1]$, and Q, find the direction matrix $B \in \mathbb{C}^{d\times(n-d)}$.

- Compute $QS_1 \in \mathbb{C}^{n\times d}$. Compute its thin CS decomposition, i.e.,

$$QS_1 = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} \begin{pmatrix} \Gamma(1) \\ -\Sigma(1) \\ 0 \end{pmatrix} V_1^T$$

$$= \begin{pmatrix} U_1 & 0 \\ 0 & \tilde{U}_2 \end{pmatrix} \begin{pmatrix} \Gamma(1) \\ -\Sigma(1) \end{pmatrix} V_1^T$$

  This decomposition costs $O(nd^2)$ (as a generalized SVD).

- Determine Θ via the $\arcsin$ or $\arccos$ that is numerically reliable, and evaluate $B = \tilde{U}_2 \Theta U_1^\dagger$.

SUBSPACE DYNAMICS

We want to model (statistically) evolution of $[S_0], [S_1], \ldots, [S_t]$ in $\mathcal{G}_{n,d}$.

Let $A_t = \begin{bmatrix} 0 & B_t \\ -B_t^\dagger & 0 \end{bmatrix}$ be the direction such that:

$$S_{t+1} = Q_t \exp(A_t) J, \;\; S_t = Q_t J \; .$$

- **Zero velocity or Stationary subspace model**:

$$\frac{dB(t)}{dt} = \nu_t, \quad B_t \in \mathbb{C}^{d \times (n-d)} \; ,$$

  where ν_t is zero-mean, complex Gaussian noise.

- **Constant velocity model**:

$$\frac{d^2 B(t)}{dt^2} = \nu_t, \quad B_t \in \mathbb{C}^{d \times (n-d)} \; .$$

Leads to one-step prior model, $P(S_{t+1}|S_t)$, that is easy to sample from:

SEQUENTIAL MONTE CARLO METHOD

Given samples from the posterior at time t:

$$S_t^{(i)} \sim P(S_t|\mathbf{y}_{[1:t]}), \;\; i = 1, 2, \ldots, N \; .$$

The goal is to generate samples from the posterior at times $t + 1$.

1. **Prediction**: Use the prediction equation

$$P(S_{t+1}|\mathbf{y}_{[1:t]}) = \int_{S_t} P(S_{t+1}|S_t) P(S_t|\mathbf{y}_{[1:t]}) dS_t \; .$$

For an $S_t^{(i)}$, let $\tilde{S}_{t+1}^{(i)} \sim P(S_{t+1}|S_t)$.

For the constant velocity model, set $A_t = A_{t-1} + \mu_{t-1}$, and set

$$\tilde{S}_{t+1}^{(i)} = Q_t^{\dagger} \exp(A_t) J \; .$$

2. **Update**: Importance Sampling

$$P(S_{t+1}|\mathbf{y}_{[1:t+1]}) = \frac{P(\mathbf{y}_{t+1}|S_{t+1})P(S_{t+1}|\mathbf{y}_{[1:t]})}{P(\mathbf{y}_{t+1}|\mathbf{y}_{[1:t]})}.$$

   Set $w_{t+1}^{(i)} = P(\mathbf{y}_{t+1}|\tilde{S}_{t+1}^{(i)})$, and set $\tilde{w}_{t+1}^{(i)} = w_{t+1}^{(i)} / \sum_i w_{t+1}^{(i)}$.

3. **Resampling**: Generate

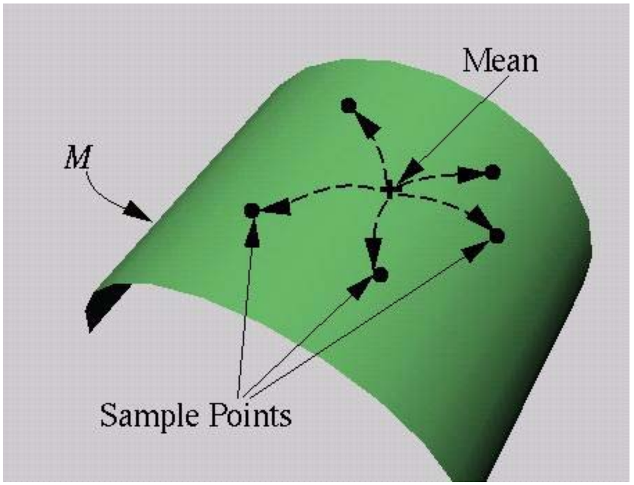
   $S_{t+1}^{(i)} \sim \{\tilde{S}_{t+1}^{(i)},\ i = 1, \ldots, N\}$ with probabilities$\{\tilde{w}_{t+1}^{(i)},\ i = 1, \ldots, N\}$

We have recursively generated samples from the posterior at time $t+1$.

We still have to compute an estimated mean subspace, or covariance, or other statistics.

ESTIMATION OF MEAN SUBSPACE

Definition and estimation of mean on $\mathcal{G}_{n,d}$.



Let $d(S_1, S_2)$ be the geodesic distance between S_1 and S_2 in $\mathcal{G}_{n,d}$. Also, let $\mathcal{G}_{n,d}$ be embedded inside $\mathbb{C}^{n \times d}$, and let $\|S_1 - S_2\|$ denote the Euclidean distance after embedding.

Mean under a probability density function $f(p)$:

- Intrinsic mean:

$$\hat{p} = \operatorname*{argmin}_{p \in M} \int_M d(p,u)^2 f(u)\gamma(du) ,$$

- Extrinsic mean:

$$\hat{p} = \operatorname*{argmin}_{p \in M} \int_M \|p - u\|^2 f(u)\gamma(du) ,$$

  This can also be viewed as computing the mean in $\mathbb{R}^n$ and then projecting it back to M. Extrinsic analysis implies emedding the manifold in a larger Euclidean space, computing the estimate there, and projecting the solution back on the manifold.

Bound on Estimation Error: Hilbert-Schmidt bound

$$HSB = \int_M d(\hat{p},u)^2 f(u)\gamma(du) .$$

EXAMPLE OF MEAN ESTIMATION ON A CIRCLE

Given $\theta_1, \theta_2, \ldots, \theta_n \in S^1$.

1. Intrinsic Mean:

$$d(\theta_1, \theta_2) = min\{|\theta_1 - \theta_2|, |\theta_1 - \theta_2 + 2\pi|, |\theta_1 - \theta_2 - 2\pi|$$

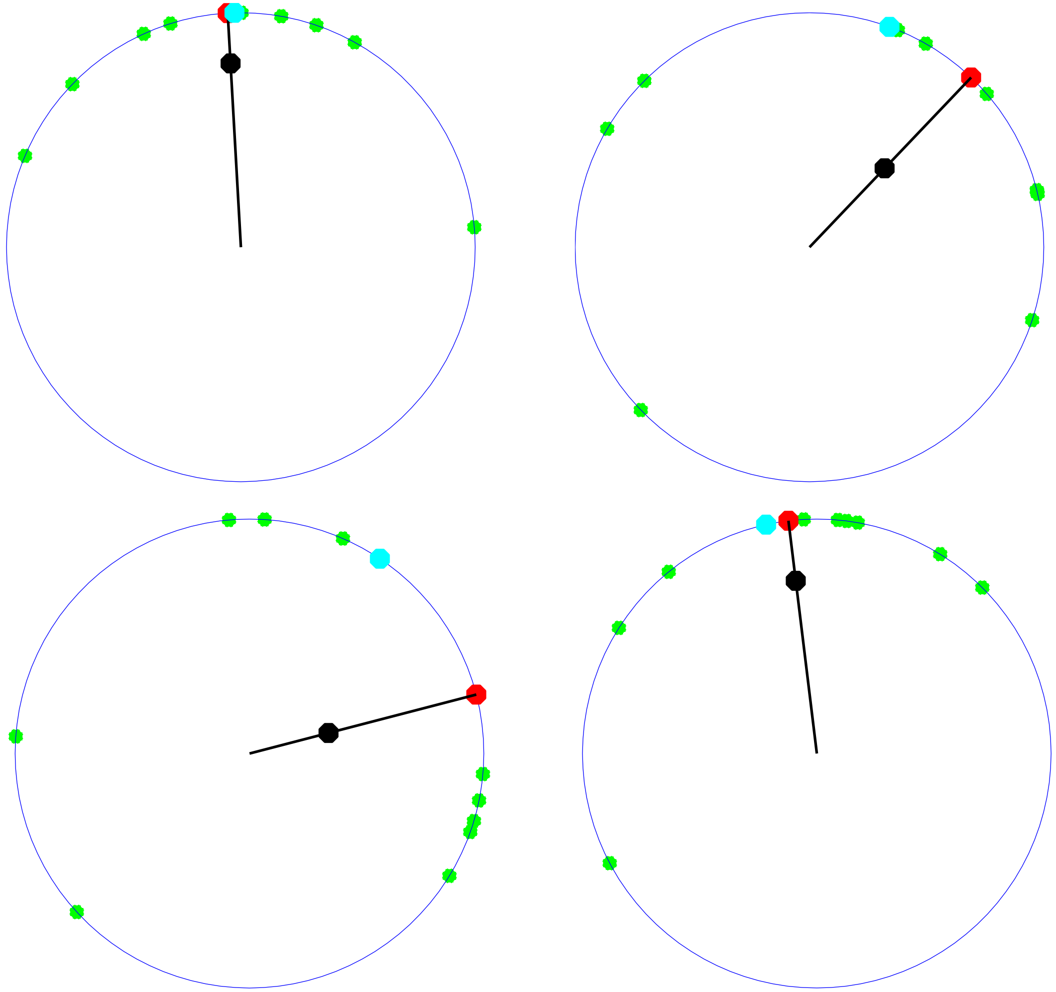
Then, find the mean

$$\hat{\theta} = \mathrm{argmin}\, \theta \sum_{i=1}^{n} d(\theta, \theta_i)^2 \ .$$

Solve this problem numerically.

2. An Extrinsic Mean: Let $z_i = e^{j\theta_i} \in \mathbb{C}$, $i = 1, \ldots, n$. Then, $\hat{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$, and set $\hat{\theta} = arg(\hat{z})$.

SAMPLE MEANS ON CIRCLE



BAYESIAN SUBSPACE TRACKING

Toy Experiment:

- Generated signal motion using a smooth Markov process on angular locations.
- Simulated data using $\mathbf{y} = D(\Theta_t)s_t + \nu_t$.

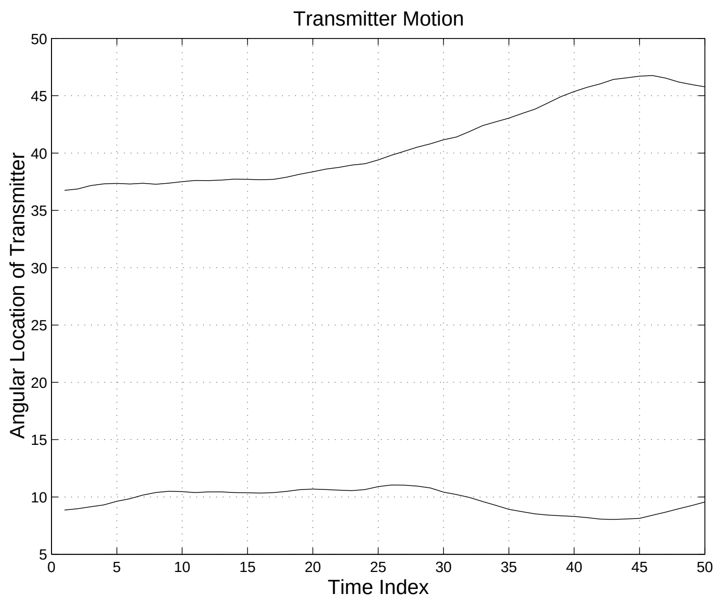
Models Assumed:

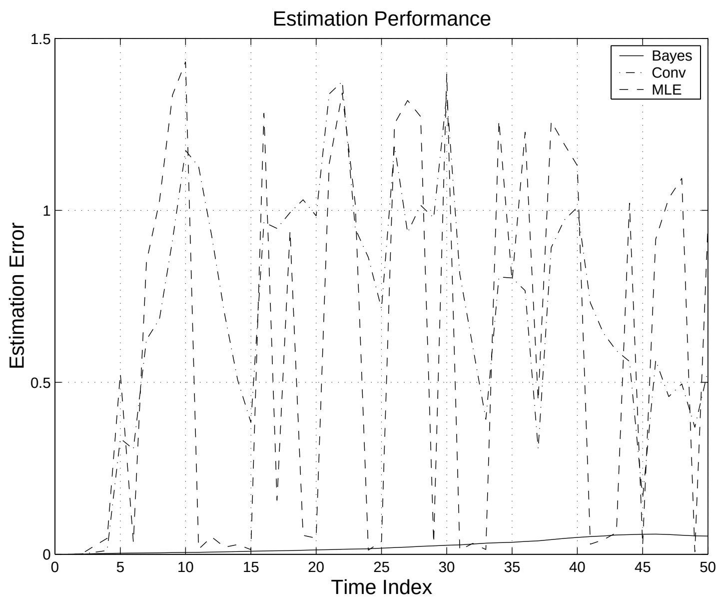
- **Prior Model**: Zero velocity model
- **Likelihood Model**: Additive white Gaussian noise model.

Performance Analysis: Compared three different methods

- Maximum likelihood estimation,
- Adaptive tracking (uses a moving window on data, followed by MLE), and
- Nonlinear tracking on $\mathcal{G}_{n,d}$.

TRACKING EXPERIMENT 1

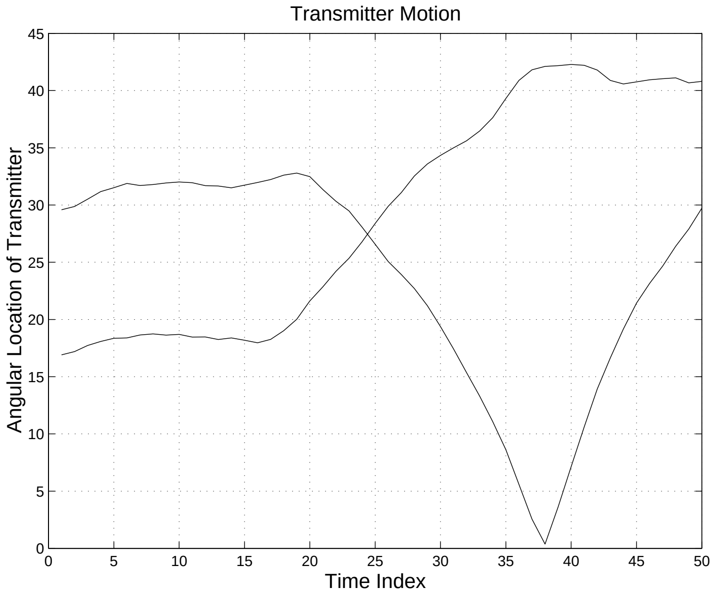


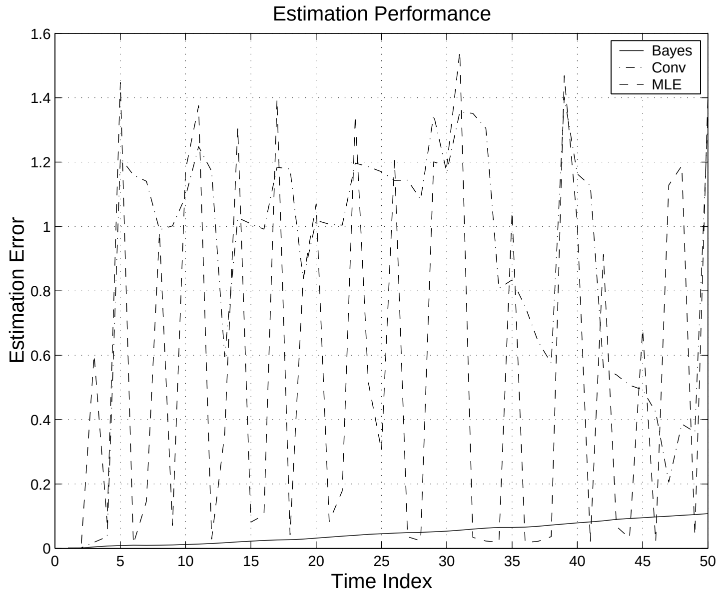


Left panel: Motion of two signal sources

Right panel: Error in subspace estimation using three different methods

TRACKING EXPERIMENT 2





Left panel: Motion of two signal sources

Right panel: Error in subspace estimation using three different methods

SUMMARY

- Signal subspace tracking can be posed as that of estimating a stochastic process on a Grassmann manifold.
- It is important, and interesting, to utilize the intrinsic geometry of the underlying manifold.
- A prior on subspace motion, maybe a smoothing prior, can help compensate for bad data.
- Monte Carlo idea can be used to handle the non-Euclidean nature of filtering problem.

OPEN ISSUES

- Filtering may not be enough. Filtering is a greedy procedure that estimates one subspace at a time. Need procedures that jointly estimate subspaces for some neighboring times.
- Need more interesting dynamic models for use in nonlinear filtering on Grassmann manifolds. Zero velocity, constant velocity models are only to get started. Remember, past observations form prior for future.
- Need more efficient algorithms for performing calculus on Grassmann manifolds.
- Need to focus on that subset of $\mathcal{G}_{n,d}$ that relates to the array manifold.

REFERENCES TO OUR WORK

- Estimation, filtering and MCMC on special Euclidean groups.

  Application: Pose estimation and target tracking using signals and images

  IEEE Trans. on SP, Nov. 1995, J. of Stat. Planning and Inference, March 2002

- Statistical bounds (extrinsic) for Bayesian estimators on orthogonal groups.

  Application: Pose estimation in image analysis

  IEEE Trans. on Pattern Analysis and Machine Intelligence, August 98

- Estimation, MCMC and statistical bounds (extrinsic) on Grassmann manifolds.

  Application: Signal subspace estimation in a Bayesian framework using projection matrices

  IEEE Trans. on SP, May 00

- Bayesian nonlinear filtering and SMC on Grassmann manifolds.

  Applications: Signal subspace tracking

  Advances in Applied Probability, March 04

- Extrinsic statistics of estimators on Lie groups and their quotient spaces.

  Applications: several from signal and image processing

  IEEE Trans. on SP, Feb 02

- Intrinsic stochastic optimization on Grassmann manifolds.

  Application: Optimal component analysis for image based object recognition

  IEEE Trans. on Pattern Analysis and Machine Intelligence, May 04

- Efficient algorithms for computations on Grassmann manifolds.

  Proc. of 12^{th} Statistical Signal Processing Workshop, 2003