

Geodesics for Subspace Tracking

ICCOPT 2025

Laura Balzano

with Soo Min Kwon, Cameron Blocker, Haroon Raja,
Jeffrey Fessler

Overview

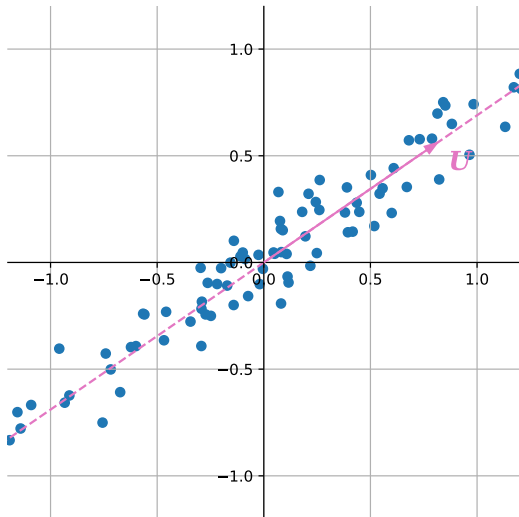
- 1 Static and Dynamic PCA**
- 2 Optimization Formulation and Algorithm**
- 3 Experiments**

Subspaces for data modeling

Subspaces are used for modeling low-dimensional structure in many signals and datasets:

Examples:

- lines, planes containing origin
- B bandlimited signals
- images upsampled with a fixed interpolation kernel
- images with common support

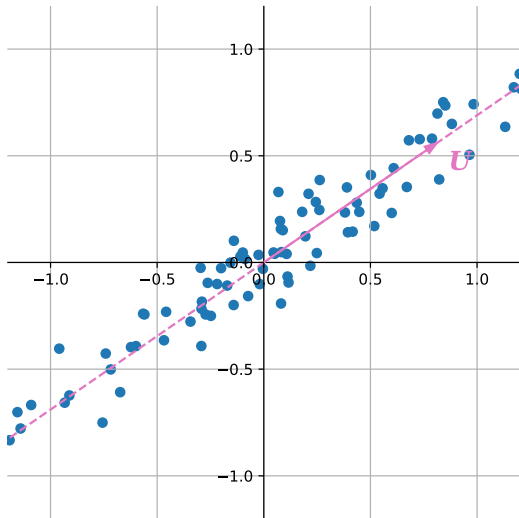


Subspaces for data modeling

Let $\mathbf{X} \in \mathbb{R}^{d \times \ell}$ be measured data.
One can estimate the best-fitting
subspace with optimization:

$$\min_{\mathbf{U}, \mathbf{G}} \|\mathbf{X} - \mathbf{U}\mathbf{G}\|_F^2$$

where $\mathbf{U}$ describes a basis for a
rank- k subspace of $\mathbb{R}^d$ and
 $\mathbf{G} \in \mathbb{R}^{k \times \ell}$ are weights of projection
onto that subspace.



Dynamic subspaces for data modeling

Now let $\mathbf{X}_i \in \mathbb{R}^{d \times \ell}$ be measured data for $i = 1, \dots, T$.

We are interested in estimating closest subspaces to each $\mathbf{X}_i$ while also constraining those subspaces to be smoothly related in some sense.

$$\min_{\mathbf{U}_i, \mathbf{G}_i} \sum_{i=1}^T \|\mathbf{X}_i - \mathbf{U}_i \mathbf{G}_i\|_F^2 \text{ s.t. } \mathbf{U}_i \in \mathcal{V}^{d \times k} \text{ are smoothly related}$$

where $\mathcal{V}^{d \times k}$ denotes the Stiefel manifold of $d \times k$ matrices with orthonormal columns.

Dynamic subspaces for data modeling

Now let $\mathbf{X}_i \in \mathbb{R}^{d \times \ell}$ be measured data for $i = 1, \dots, T$.

We are interested in estimating closest subspaces to each $\mathbf{X}_i$ while also constraining those subspaces to be smoothly related in some sense.

$$\min_{\mathbf{U}_i, \mathbf{G}_i} \sum_{i=1}^T \|\mathbf{X}_i - \mathbf{U}_i \mathbf{G}_i\|_F^2 \text{ s.t. } \mathbf{U}_i \in \mathcal{V}^{d \times k} \text{ are smoothly related}$$

This is useful in applications such as:

- Wireless communications, where signal channels and interference signals (modeled as lying in subspaces) change over time
- Direction-of-arrival estimation in radar and sonar sensing
- Reconstruction of dynamic medical images
- Graph structure that changes over time (e.g. social networks)

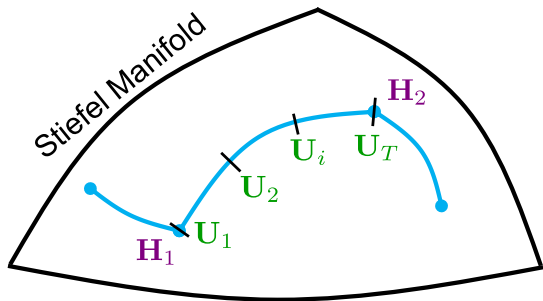
Dynamic subspaces for data modeling

$\mathbf{X}_i \in \mathbb{R}^{d \times \ell}$ are measured data for $i = 1, \dots, T$.

Optimize:

$$\min_{\mathbf{U}_i, \mathbf{G}_i} \sum_{i=1}^T \|\mathbf{X}_i - \mathbf{U}_i \mathbf{G}_i\|_F^2$$

s.t. $\mathbf{U}_i \in \mathcal{V}^{d \times k}$ lie on a (piecewise) geodesic of the Stiefel manifold.



Stiefel Geodesic Model

Optimize:

$$\min_{U_i, G_i} \sum_{i=1}^T \|X_i - U_i G_i\|_F^2$$

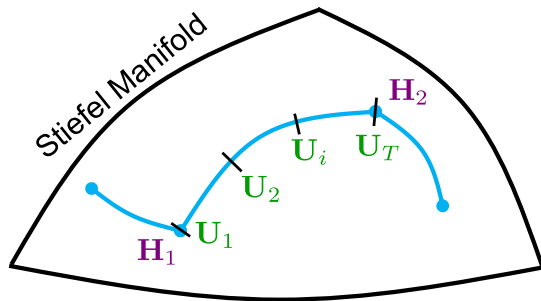
s.t. $U_i \in \mathcal{V}^{d \times k}$ lie on a (piecewise) geodesic of the Stiefel manifold:

$$U_i = U(t_i) = H \cos(\Theta t_i) + Z \sin(\Theta t_i),$$

- $H \in \mathcal{V}^{d \times k}$ is a basis for the “starting” endpoint of the geodesic,
- $Z \in \mathcal{V}^{d \times k}$ is a basis for the “ending” endpoint projected onto the orthogonal complement of H ,
- Θ is a diagonal matrix with principal angles $\theta_j \in [0, \pi/2]$,
- $t_i \in [0, 1]$ represents a 1d parameterization of the geodesic.

Stiefel Geodesic Model

$$U_i = U(t_i) = H \cos(\Theta t_i) + Z \sin(\Theta t_i) ,$$



In this picture, $H = H_1$, and $Z = \text{orth} \{ (I - HH^\top) H_2 \}$.

Overview

- 1 Static and Dynamic PCA**
- 2 Optimization Formulation and Algorithm**
- 3 Experiments**

Optimization Formulation

$\mathbf{X}_i \in \mathbb{R}^{d \times \ell}$ are measured; Assume we are given $t_i, i = 1, \dots, T$.

$$\min_{\mathbf{U}_i, \mathbf{G}_i} \sum_{i=1}^T \|\mathbf{X}_i - \mathbf{U}_i \mathbf{G}_i\|_F^2$$

$$\Leftrightarrow \min_{\mathbf{U}_i} \sum_{i=1}^T \|\mathbf{X}_i - \mathbf{U}_i \mathbf{U}_i^\top \mathbf{X}_i\|_F^2 \quad \text{since } \mathbf{U}_i^\top \mathbf{U}_i = \mathbf{I}$$

$$\Leftrightarrow \min_{\mathbf{U}_i} - \sum_{i=1}^T \|\mathbf{U}_i^\top \mathbf{X}_i\|_F^2 \quad \text{dropping constant terms}$$

$$\Leftrightarrow \min_{\mathbf{H}, \mathbf{Z}, \Theta} - \sum_{i=1}^T \|(\mathbf{H} \cos(\Theta t_i) + \mathbf{Z} \sin(\Theta t_i))^\top \mathbf{X}_i\|_F^2 \quad \text{geodesic constraint}$$

Optimization formulation

Letting $Q = \begin{bmatrix} H & Z \end{bmatrix}$:

$$\begin{aligned} & \min_{H, Z, \Theta} - \sum_{i=1}^T \| (H \cos(\Theta t_i) + Z \sin(\Theta t_i))^\top \mathbf{X}_i \|_F^2 \\ \Leftrightarrow & \min_{Q \in \mathcal{V}^{d \times 2k}, \Theta} - \sum_{i=1}^T \left\| \left(Q \begin{bmatrix} \cos(\Theta t_i) \\ \sin(\Theta t_i) \end{bmatrix} \right)^\top \mathbf{X}_i \right\|_F^2 \end{aligned}$$

Algorithm

$$\min_{\mathbf{Q}, \Theta} - \sum_{i=1}^T \left\| \left(\mathbf{Q} \begin{bmatrix} \cos(\Theta t_i) \\ \sin(\Theta t_i) \end{bmatrix} \right)^\top \mathbf{X}_i \right\|_F^2$$

Block MM [Kwon et al., 2023]:

- Fix Θ , form a linear majorizer and minimize over $\mathbf{Q}$ using a Procrustes step.
- Fix $\mathbf{Q}$, then the objective is separable in each entry of Θ , denoted θ_j , $j = 1, \dots, k$. Majorize for each θ_j and iteratively minimize the majorizer.

Algorithm: Q update

Following [Breloy et al., 2021], we form a linear majorizer for our loss at $Q^{(n)}$ and minimize it with a Stiefel constraint. The update is then given by

$$Q^{(n+1)} = \arg \min_{Q \in \mathcal{V}^{d \times 2k}} \|Q - \sum_{i=1}^T X_i X_i^\top Q^{(n)} \Gamma_i \Gamma_i^\top\|_F^2 = W V^\top, \quad (1)$$

where

- $\Gamma_i \triangleq \begin{bmatrix} \cos(\Theta t_i) \\ \sin(\Theta t_i) \end{bmatrix}$
- $W \Sigma V^\top$ is the SVD of $\sum_{i=1}^T X_i X_i^\top Q^{(n)} \Gamma_i \Gamma_i^\top$.

Algorithm: Θ update

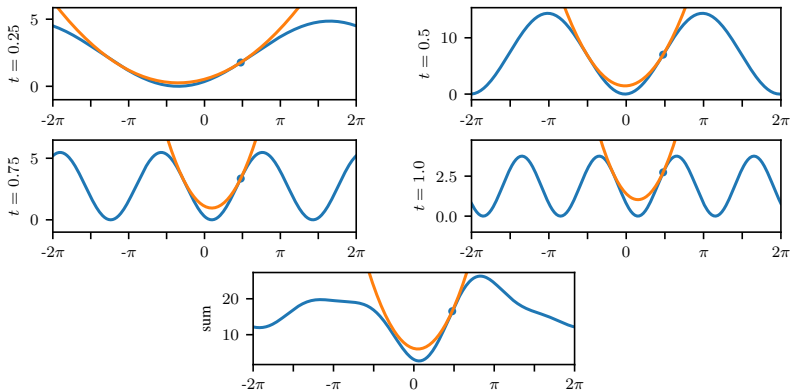


FIG – An example of four cosines (top two rows, blue) that sum to form the (nonconvex) loss for a single θ_j (bottom row, blue). For each cosine function we construct a quadratic majorizer; they sum to a quadratic majorizer for the loss (orange) at a point $\theta_j^{(n)}$ (blue dot). The loss is often well-behaved on $\theta_j \in [-\pi/2, \pi/2]$ (here, quasi-convex).

Algorithm convergence

- As an MM algorithm, we have monotonic descent.
- In [Li et al., 2023], we proved convergence to a stationary point for a slight variant of our algorithm, where proximal terms are included in majorizers.
- Moreover, if both proximal parameters $\lambda_Q, \lambda_\Theta > 0$, then the iteration complexity is $\tilde{O}(\varepsilon^{-2})$, where ε is the distance to the stationary point and $\tilde{O}(\cdot)$ is big- O notation ignoring logarithmic factors.

Overview

- 1 Static and Dynamic PCA**
- 2 Optimization Formulation and Algorithm**
- 3 Experiments**

Synthetic Experiments

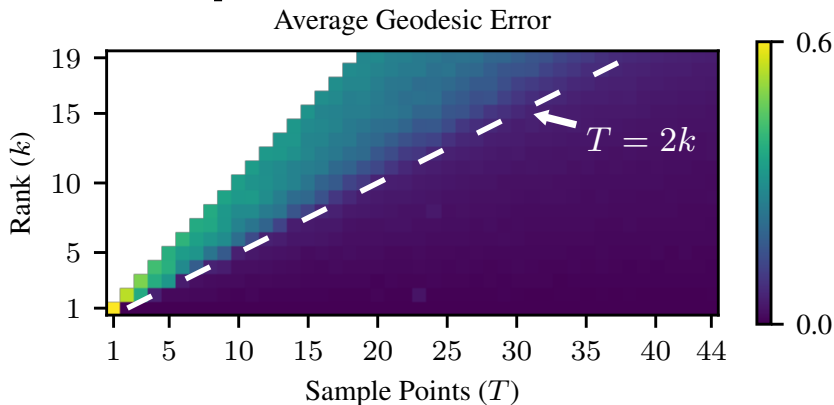


FIG – The average geodesic error over 15 trials for varying rank k and number of sample points T . One vector was sampled at each of T points on a rank- k geodesic; $d = 40, \ell = 1$, with AWGN $\sigma = 10^{-5}$. We see a phase transition at $T = 2k$; with at least this many samples, we recover the true subspace with low error.

Synthetic Experiments

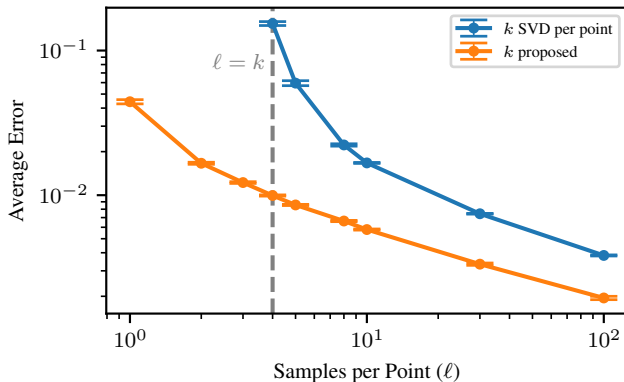


FIG – Average geodesic error over 100 trials, with standard error bars, for varying number of samples (ℓ) collected at each time point for a fixed number of time points ($T = 11$) on a planted rank-4 geodesic with AWGN $\sigma = 10^{-2}$.

Oscillating Steady-State Imaging

We have an “inverse problem” version of our problem formulation and algorithm:

We observe $\mathbf{y}_i = \mathcal{A}(\mathbf{X}_i)$ where the linear operator $\mathcal{A} : \mathbb{R}^{d \times \ell} \rightarrow \mathbb{R}^m$.

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{Y}, \boldsymbol{\Theta}, \mathbf{X}_i} \quad & \sum_{i=1}^T \|\mathbf{X}_i - \mathbf{U}_i \mathbf{U}_i^\top \mathbf{X}_i\|_F^2 + \frac{\lambda}{2} \|\mathbf{y}_i - \mathcal{A}(\mathbf{X}_i)\|_2^2 \\ \text{s.t.} \quad & \mathbf{U}_i = \mathbf{H} \cos(\boldsymbol{\Theta} t_i) + \mathbf{Z} \sin(\boldsymbol{\Theta} t_i) \end{aligned}$$

We may now apply our approach to functional MRI, where scanner drift and patient breathing affects the images in a smooth time-varying way.

Oscillating Steady-State Imaging

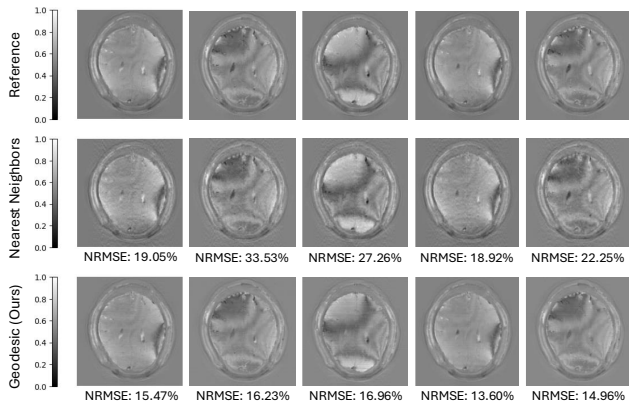


FIG – fMRI reconstruction results, where each frame was masked by a 50% sampled k -space mask. Our method outperforms nearest-neighbors reconstruction, which is commonly used in dynamic image reconstruction.

Dynamic Communities in Graphs

Spectral methods are commonly used for identifying communities in graphs, but many graphs are dynamic with snapshots over time.

This data captures an interaction network in an elementary school segmented into 10-minute intervals.

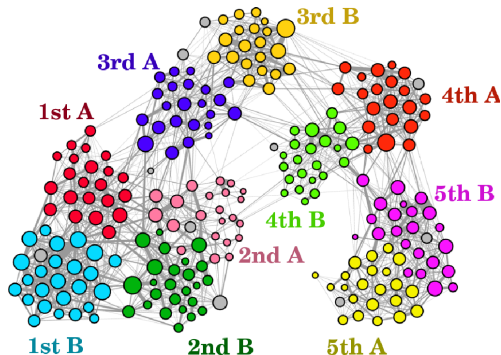


FIG — Network of contacts aggregated over the first day showing those who interacted at least 2 minutes. Link width corresponds to duration of contact, colors correspond to classes, and teachers are shown in grey.

Dynamic Communities in Graphs

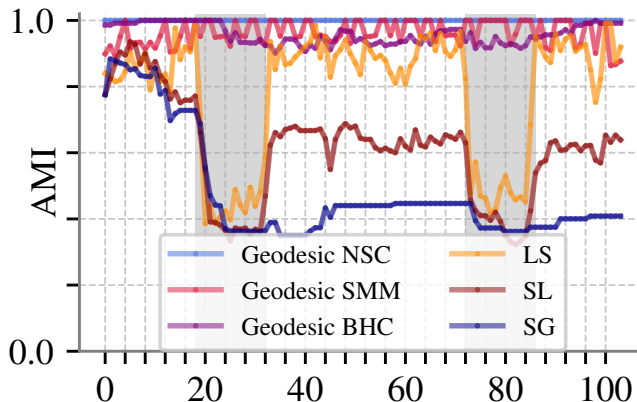


FIG — Evaluation of Adjusted Mutual Information (AMI) clustering performance on a two-day elementary school face-to-face interaction network. Benchmarks are Label Smoothing (LS), Smoothed Louvain (SL), and Graph Smoothing (SG). The three geodesic extensions of static clustering methods (Normalized Spectral Clustering, Spectral Modularity Maximization, Bethe Hessian Clustering) uniformly outperform the benchmarks. See [Hume and Balzano, 2024] for references and more results.

Conclusion

- We modeled the dynamic subspace estimation problem as geodesic learning, and formulated a nonconvex Riemannian optimization problem
- We developed a Riemannian Block-MM algorithm to solve it and proved convergence to a stationary point
- We demonstrated the algorithm on several applications

Future:

- Learn a piecewise geodesic from data
- Simultaneously learn the geodesic parameter t_i
- Understand the initialization (PCA on first/last half of the data works)
- Understand when we will find a global minimizer

Thank you!



Breloy, A., Kumar, S., Sun, Y., and Palomar, D. P. (2021).

Majorization-minimization on the Stiefel manifold with application to robust sparse PCA.
IEEE Trans. Sig. Proc., 69:1507–20.



Hume, J. and Balzano, L. (2024).

A spectral framework for tracking communities in evolving networks.
In *The Third Learning on Graphs Conference*.



Kwon, S. M., Blocker, C. J., Raja, H., Fessler, J. A., and Balzano, L. (2023).

Dynamic subspace estimation with stiefel geodesics.
In preparation; preliminary version at arXiv preprint arXiv:2303.14851.



Li, Y., Balzano, L., Needell, D., and Lyu, H. (2023).

Convergence and complexity of block majorization-minimization for constrained block-riemannian optimization.
arXiv preprint arXiv:2312.10330.