Tangent Space Alignment for Manifold Learning

Hongyuan Zha & Zhenyue Zhang

Department of Computer Science and Engineering
The Pennsylvania State University

Department of Mathematics
Zhejiang University

IPAM, Oct. 29, 2004

Acknowledgment

Collaborators on spectral analysis of alignment: Qiang Ye and Ren-cang Li, Math/U. of Kentucky.

Collaborators on CAMLET project: Qiang Du, Math/PSU, Richard Li, Statistics/PSU, Jorge Sofo, Physics/PSU.

Outline

- Introduction
- Tangent spaces and their global alignment
- Spectral analysis of alignment
- Applications to molecular dynamics simulations

Linear and nonlinear dimension reduction

- Principal component analysis and probabilistic extensions (PCA)
- Kernel PCA
- Kahonen's self-organizing maps (SOM)
- Topology-preserving networks
- Principal curves, surfaces and manifolds
- Multi-dimensional scaling (MDS)
- Many more ...

Prior/Related Work

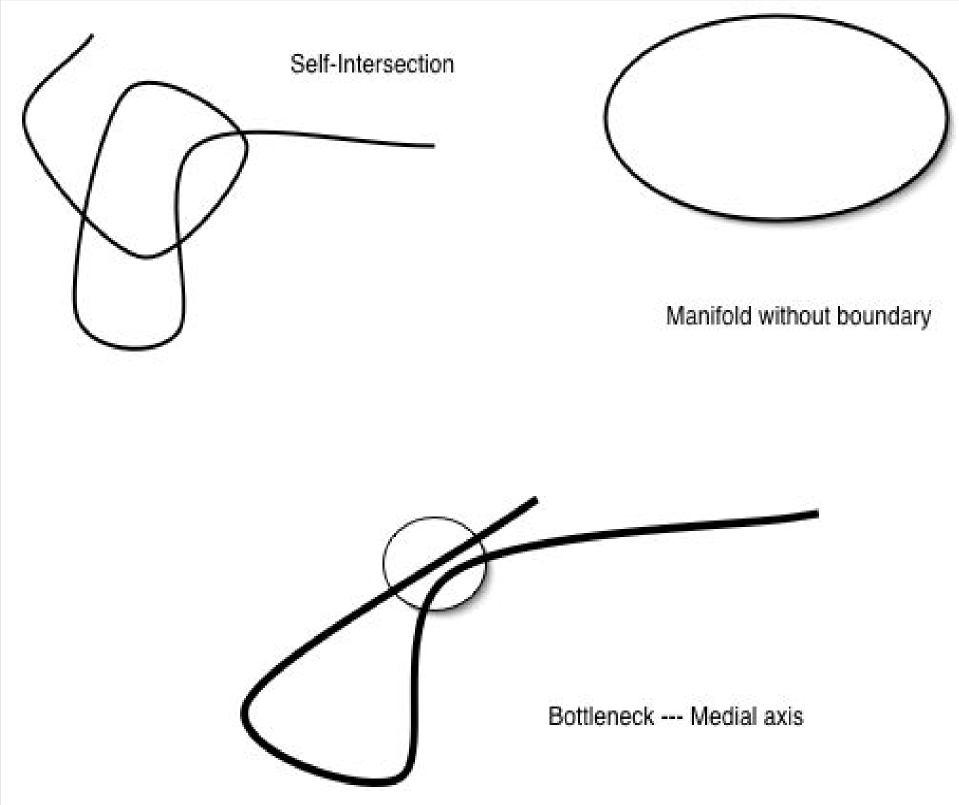
- Isomap, J. Tenenbaum, V. De Silva and J. Langford. *Science*, 2000.
- LLE, S. Roweis and L. Saul *Science*, 2000.
- Automatic alignment of local representations, Y. W. Teh and S. Roweis, *NIPS*, 2002.
- Charting a manifold, Geodesic nullspace method, M. Brand, *NIPS*, 2002/2004.
- Laplacian eigenmap, M. Belkin and P. Niyogi, 2002.
- Hessian LLE, D. Donoho and C. Grimes, PNAS, 2003.

Introduction

- Assume a d-dimensional Parametrized manifold $\mathcal{F}$ embedded in an m-dimensional space $(d < m)$,

$$f : C \subset \mathcal{R}^d \to R^m,$$

where C is a compact and connected subset of $\mathcal{R}^d$ with open interior.
($\mathbf{Note.}$ $\mathcal{F}$ well-behaved, no self-intersection etc.)


Self-Intersection
Manifold without boundary
Bottleneck --- Medial axis

- Given a set of data points $x_1, \cdots, x_N$, where $x_i \in \mathcal{R}^m$,

$$x_i = f(\tau_i) + \epsilon_i, \quad i = 1, \ldots, N,$$

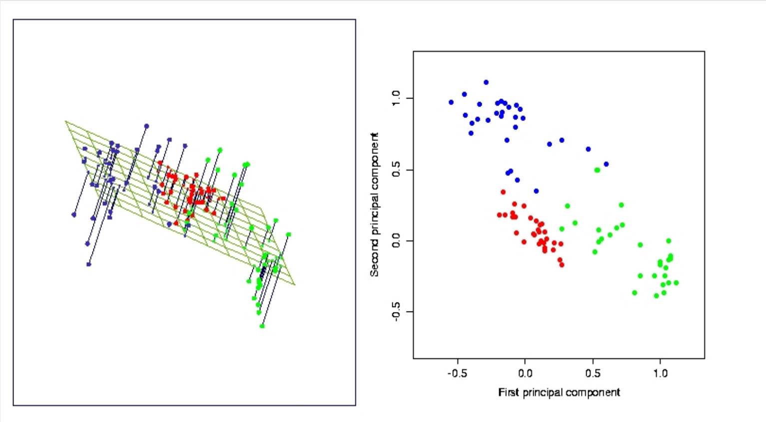
  where ϵ_i represent noise.

- By DIMENSION REDUCTION we mean the estimation of the unknown lower dimensional parameter vectors τ_i's from the x_i's

- By MANIFOLD LEARNING we mean the reconstruction of f from the x_i's.

PCA and Orthogonal Projections

(Figure from Hastie et. al. *Element of Statistical Learning*)



PCA for Linear Case

- Data points sampled from a d-dimensional affine subspace, i.e.,

$$x_i = c + U\tau_i + \epsilon_i, \quad i = 1, \ldots, N,$$

  U orthonormal columns. In matrix format, let

$$X = [x_1, \cdots, x_N], \quad T = [\tau_1, \cdots, \tau_N], \quad E = [\epsilon_1, \cdots, \epsilon_N].$$

- Find c, U and T to minimize the reconstruction error E, i.e.,

$$\min \|E\| = \min_{c,U,T} \|X - (c\,e^T + UT)\|_F.$$

- Solutions are given by

$$c = \bar{x}$$
$$\tau_i = V_d^T (x_i - c)$$
$$V_d = d \text{ largest } \textit{left} \text{ singular vectors of } \textit{centered } X$$

Tangent space

- At a reference point τ, first-order Taylor expansion,

$$f(\tilde{\tau}) = f(\tau) + J_f(\tau) \cdot (\tilde{\tau} - \tau) + O(\|\tilde{\tau} - \tau\|^2)$$

with $J_f(\tau) \in \mathcal{R}^{m \times d}$ the Jacobi matrix,

$$f(\tau) = \begin{bmatrix} f_1(\tau) \\ \vdots \\ f_m(\tau) \end{bmatrix}, \quad \text{then} \quad J_f(\tau) = \begin{bmatrix} \partial f_1/\partial \tau_1 & \cdots & \partial f_1/\partial \tau_d \\ \vdots & \vdots & \vdots \\ \partial f_m/\partial \tau_1 & \cdots & \partial f_m/\partial \tau_d \end{bmatrix}.$$

- Local linear approximation in a neighborhood of τ,

$$f(\tilde{\tau}) \approx f(\tau) + J_f(\tau) \cdot (\tilde{\tau} - \tau)$$

Points in the neighborhood lie close to a d-dimensional affine subspace spanned by columns of $J_f(\tau)$.

Relation between local coordinates and global coordinates

- Q_τ: orthonormal basis of tangent space at τ

$$J_f(\tau) \cdot (\tilde{\tau} - \tau) = Q_\tau \theta_\tau, \quad \tilde{\tau} - \tau = J_f^+(\tau) Q_\tau \theta_{\tilde{\tau}} \equiv L_\tau \theta_{\tilde{\tau}}$$

- Local vs. global

$$\tilde{x} = x + Q_\tau \theta_{\tilde{\tau}}, \quad \tilde{\tau} = \tau + L_\tau \theta_{\tilde{\tau}},$$

  i.e., local coordinates $\theta_{\tilde{\tau}}$ and global coordinates $\tilde{\tau}$ are related by an affine transformation.

- Note. If f is locally isometric, J_f is orthonormal, and L_τ is orthogonal.

Alignment

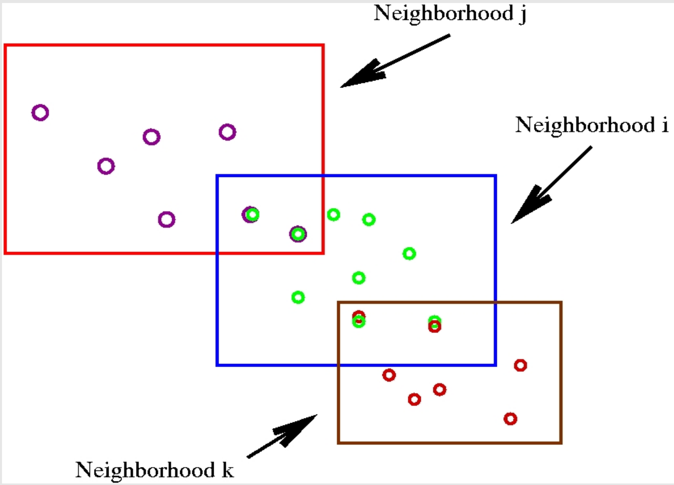
Find global coordinate τ and local affine transformation L_τ to minimize (Symbolically),

$$\int_{\Omega}\Big(\int_{\Omega(\tau)}\|\bar{\tau}-\tau-L_\tau\theta(\bar{\tau})\|d\bar{\tau}\Big/\int_{\Omega(\tau)}d\bar{\tau}\Big)d\tau$$

over all possible nonsingular L_τ.

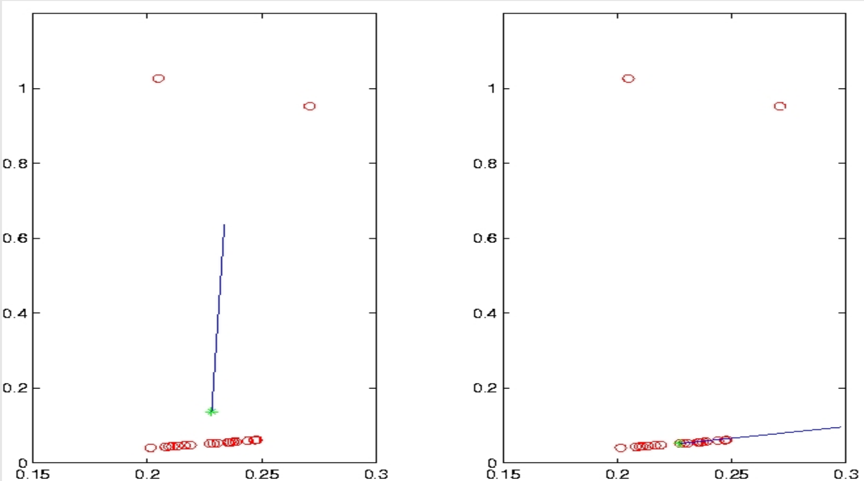
Overlay a K-NN graph on the sample points

- For each x_i, let $X_i = [x_{i_1}, \ldots, x_{i_k}]$ be its k-nearest neighbors including x_i, say in terms of the Euclidean distance. (Other possibilities and acceleration.)



Constructing approximate tangent space

Apply PCA to each neighborhood $X_i = [x_{i_1}, \ldots, x_{i_k}] \Rightarrow$ sensitive to outliers.



Weighted (robust) PCA

$$\sum_j w_{i,j} \|x_{i_j} - (\bar{x}_i^w + U_i \theta_j^{(i)})\|_2^2 = \min_{c, U, \theta_j} \sum_j w_{i,j} \|x_{i_j} - (c + U\theta_j)\|_2^2,$$

Weight selection

Choose the initial vector $\bar{x}_{w^{(0)}}$ as the mean of the k vectors $x_{i_1}, \ldots, x_{i_k}$,

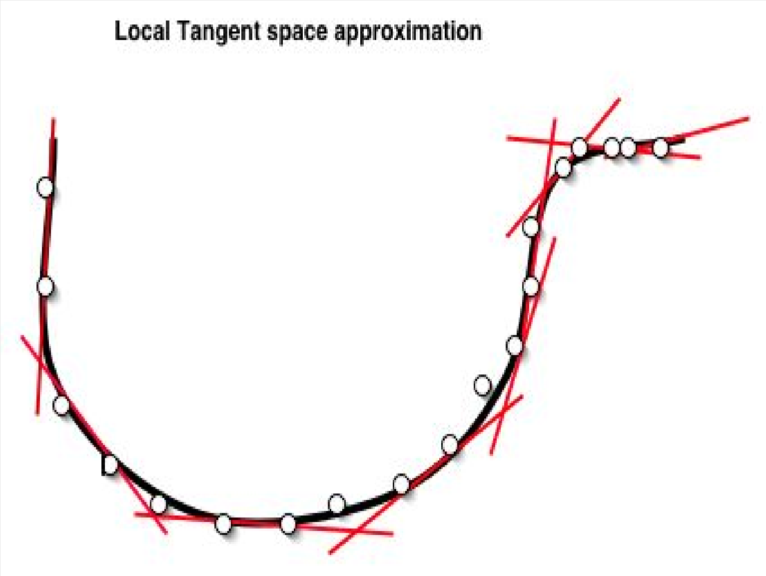
1. Compute the current weights,

$$w_s^{(j)} = \exp(-\gamma \|x_{i_s} - \bar{x}_{w^{(j-1)}}\|_2^2).$$

2. Compute a new weighted center

$$\bar{x}_{w^{(j)}} = \sum_{s=1}^k w_s^{(j)} x_{i_s}.$$

Illustration



Alignment

- In each nbhd, apply (weighted) PCA to $X_i = [x_{i_1}, \ldots, x_{i_k}]$,

$$x_{i_j} = \bar{x}_i + V_i \theta_j^{(i)}, \quad j = 1, \ldots, k$$

V_i orthonormal basis.

- Global vs. local,

$$\tau_{i_j} = \bar{\tau}_i + L_i \theta_j^{(i)}, \quad j = 1, \ldots, k$$

Let $T_i = [\tau_{i_1}, \ldots, \tau_{i_k}]$ and $\Theta_i = [\theta_1^{(i)}, \cdots, \theta_k^{(i)}]$

$$T_i J_k - L_i \Theta_i \approx 0, \quad i = 1, \ldots, N$$

with $J_k = I_k - ee^T/k$, centering matrix.

(Con't)

- A minimization problem (over T, L_i)

$$\sum_i \|T_i J_k - L_i \Theta_i\|^2 = \min$$

- Fix T_i and minimize

$$\|T_i J_k - L_i \Theta_i\|$$

w.r.t. $L_i \Longrightarrow \|T_i J_k (I - \Theta_i^+ \Theta_i)\|$.

- Let $W_i = J_k(I - \Theta_i^+ \Theta_i)$. Note.

$$W_i W_i^T = J_k (I - \Theta_i^+ \Theta_i) J_k,$$

orthogonal projection onto $\mathrm{span}^\perp([e, \Theta_i^T])$.

(Con't)

- Define S_i a selection matrix such that

$$T_i = TS_i, \quad T = [\tau_1, \ldots, \tau_N].$$

- Let

$$[TS_1W_1, \ldots, TS_NW_N] \equiv T\Psi$$

leading to

$$\min_T \|T\Psi\|_F^2 = \min_T \text{trace}\left(T(\Psi\Psi^T)T^T\right).$$

- Normalization $TT^T = I_d$. Solution T given by the eigenvectors of $\Phi \equiv \Psi\Psi^T$ corresponding to the 2nd to d+1st smallest eigenvalues. (more on normalization later).

Computational Issues

- Forming Krylov subspaces ($\Phi = \Psi\Psi^T$)

$$K_p(\Phi, v_0) = \mathrm{span}\{v_0, \Phi v_0, \Phi^2 v_0, \ldots, \Phi^{p-1} v_0\}.$$

- Matrix-vector multiplications Φx

$$\Phi x = S_1 W_1 W_1^T S_1^T x + \cdots + S_N W_N W_N^T S_N^T x,$$

where

$$W_i = (I - \frac{1}{k} e e^T)(I - \Theta_i^+ \Theta_i).$$

Each term involves the x_i's in *one neighborhood*.

- With the SVD of $X_i - \bar{x}_i e^T = Q_i \Sigma_i H_i^T$

$$W_i = I - \frac{1}{k} e e^T - H_i H_i^T = I - [e/\sqrt{k},\ H_i][e/\sqrt{k},\ H_i]^T \equiv I - G_i G_i^T.$$

Local Tangent Space Alignment (LTSA)

Given N m-dimensional points sampled possibly with noise from an underlying d-dimensional manifold, this algorithm produces N d-dimensional coordinates $T \in \mathcal{R}^{d\times N}$ for the manifold constructed from k local nearest neighbors.

Step 1. [Extracting local information.] For each $i = 1, \cdots, N$,

1.1 Determine k nearest neighbors x_{i_j} of x_i, $j = 1, \ldots, k$.

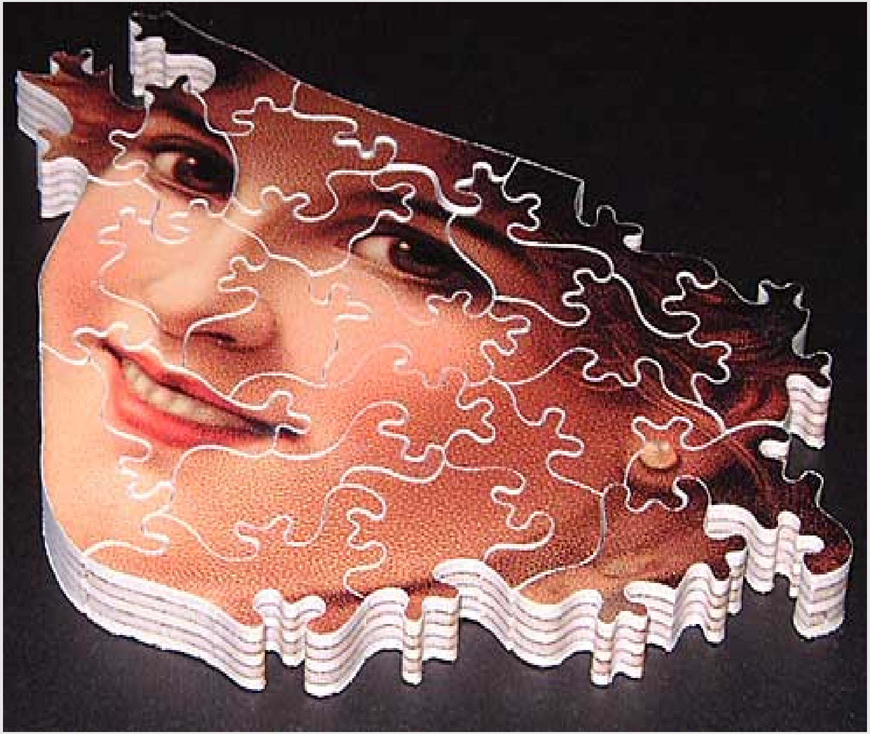
1.2 Compute the d largest eigenvectors $g_1, \cdots, g_d$ of the correlation matrix $(X_i - \bar{x}_i e^T)^T (X_i - \bar{x}_i e^T)$, and set

$$G_i = [e/\sqrt{k},\ g_1,\ \cdots,\ g_d].$$

Step 2. [Constructing the alignment matrix.] Form the the alignment matrix Φ by locally summation if a direct eigen-solver will be used. Otherwise implement a routine that computes matrix-vector multiplication Bu for an arbitrary vector u.

Step 3. [Computing global coordinates.] Compute the $d+1$ smallest eigenvectors of Φ and pick up the eigenvector matrix $[u_2,\ \cdots,\ u_{d+1}]$ corresponding to the 2nd to $d+1$st smallest eigenvalues, and set $T = [u_2,\ \cdots,\ u_{d+1}]^T$.

Solving jigsaw puzzles



(Better, we allow each piece be affinely transformed.)

Examples

Consider $d = 1$, and

$$x_i = c + u\tau_i, \quad i = 1, \ldots, N.$$

PCA on $X = [x_1, \ldots, x_N]$ is equivalent to finding the nullspace of

$$\Phi = I - J_N(I - X^+X)J_N = I - J_N(I - T^+T)J_N, \quad J_N = I - ee^T/N.$$

Here $T = [\tau_1, \ldots, \tau_N]$, all distinct.

Φ is the orthogonal projection onto $\mathrm{span}^\perp([e, T^T])$.

(Con't)

Split T into two parts $T = [T_1, T_2]$, and build the matrix

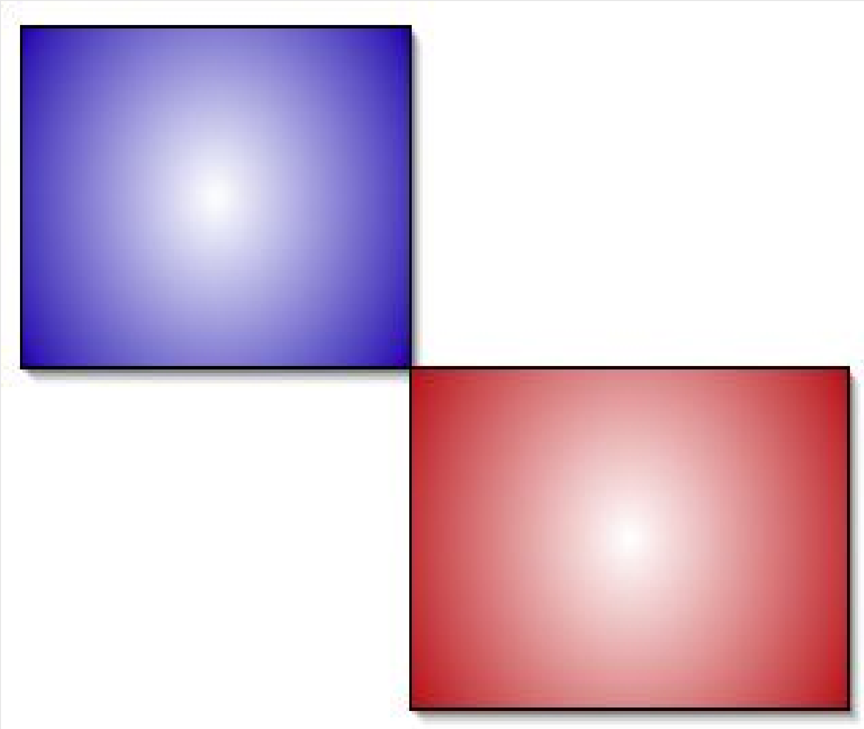
$$\Phi = \mathrm{diag}(\Phi_1, \Phi_2), \quad \Phi_i = I - [e, T_i^T]^+[e, T_i^T], \;\; i = 1, 2.$$

It is easy to check

$$e, T^T \in \mathcal{N}(\Phi) = \mathrm{span}\left\{ \begin{bmatrix} e \\ 0 \end{bmatrix}, \begin{bmatrix} T_1^T \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ e \end{bmatrix}, \begin{bmatrix} 0 \\ T_2^T \end{bmatrix} \right\}$$

but $\dim(\mathcal{N}(\Phi)) = 4$. How to get rid of the unwanted info in $\mathcal{N}(\Phi)$?

Illustration



(Con't)

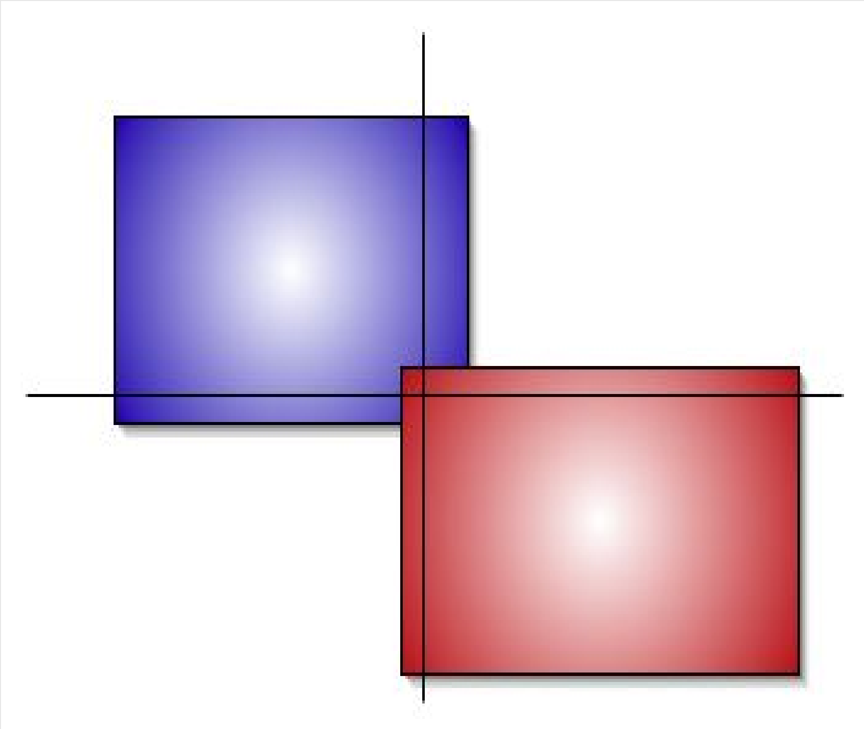
Now let T_1 and T_2 share a point. Build Φ as before,

$$\Phi = \Phi_1 + \Phi_2.$$

Φ_1 and Φ_2 overlap one row and one column. Again it is easy to see that

$$e, T^T \in \mathcal{N}(\Phi) = \text{span}\left\{ e, T^T, \begin{bmatrix} T_1^T \\ 2T_1(n_1) - T_2(2:n_2)^T \end{bmatrix} \right\}$$

Illustration



(Con't)

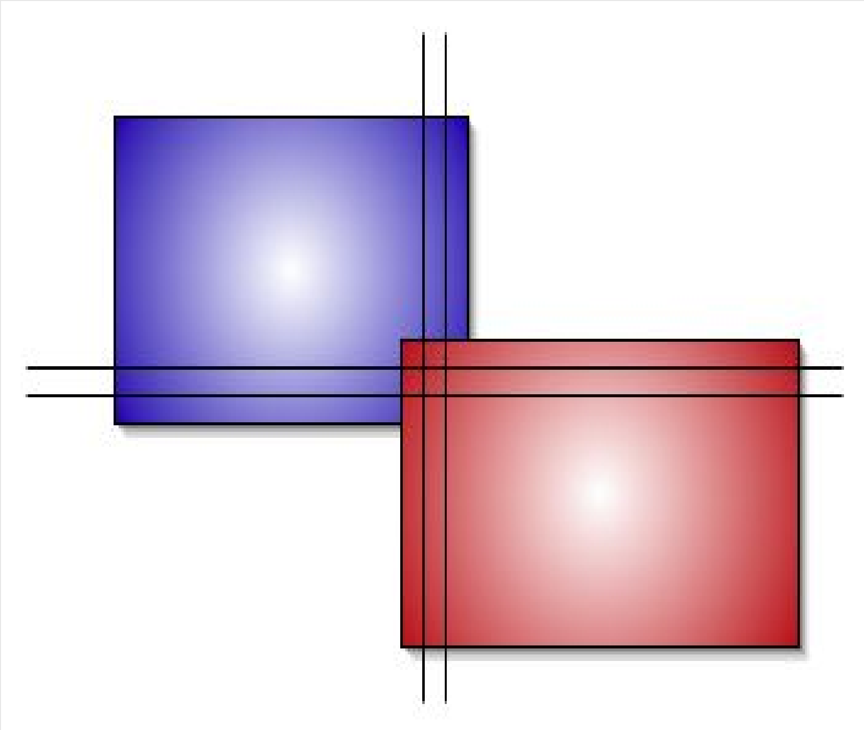
Now let T_1 and T_2 share two distinct points. Build Φ as before,

$$\Phi = \Phi_1 + \Phi_2,$$

Φ_1 and Φ_2 overlap two rows and two columns. Then

$$\mathcal{N}(\Phi) = \mathrm{span}\{e, T^T\}.$$

Illustration



Spectral analysis of the alignment

- Consider $T = [\tau_1, \ldots, \tau_N] \in \mathcal{R}^{d\times N}$, d-dimensional parameter vectors with each neighborhood (patch) corresponding to a submatrix of T, called a section.
- Assume we have computed sections $T_i = [\tau_{i_1}, \ldots, \tau_{i_{k_i}}] \in \mathcal{R}^{d\times k_i}$. (Actually up to an affine transformation, or a rigid motion).
- Given a collection of sections $\{T_1, \ldots, T_s\}$ of T, build an alignment matrix:

$$\Phi = \sum_{i=1}^{s} \Phi_i,$$

here P_i orthogonal projection onto $\mathrm{span}^{\perp}([e, T_i^T])$, *stretch* to obtain Φ_i.

Reminder

Recall,

$$\Phi = S_1(W_1 W_1^T)S_1^T + \cdots + S_N(W_N W_N^T)S_N^T,$$

where

$$W_i = J_k(I - \Theta_i^+ \Theta_i),$$

and Θ_i local coordinates. Now

$$W_i W_i^T = J_k(I - \Theta_i^+ \Theta_i)J_k,$$

orthogonal projection onto $\mathrm{span}^\perp([e, T_i^T])$.

Null space of Φ

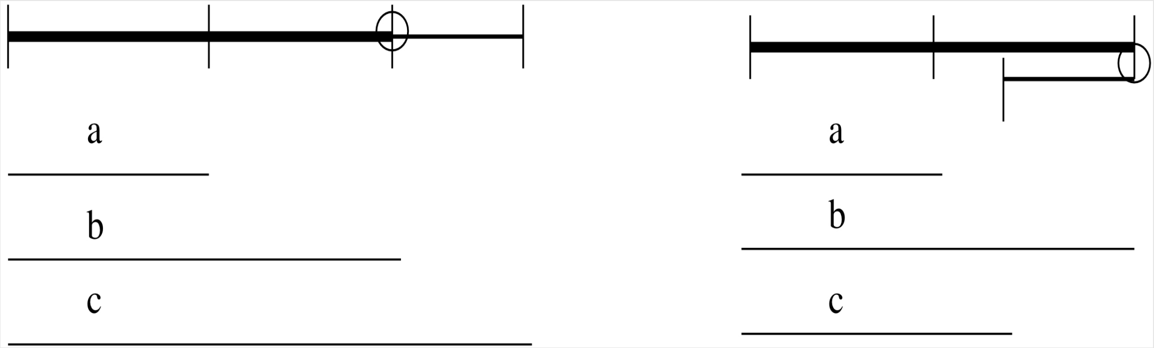
Fully overlapped. Two sections T_1 and T_2 are fully overlapped, if the vectors in the intersection part are in general position, i.e., dimension of the spanned affine subspace is d.

Theorem. Assume two sections, then

1. $\mathrm{span}([e,\, T^T]) \subset \mathcal{N}(\Phi)$.
2. $N(\Phi) = \mathrm{span}([e, T^T])$ iff $\{T_1,\, T_2\}$ is fully overlapped.

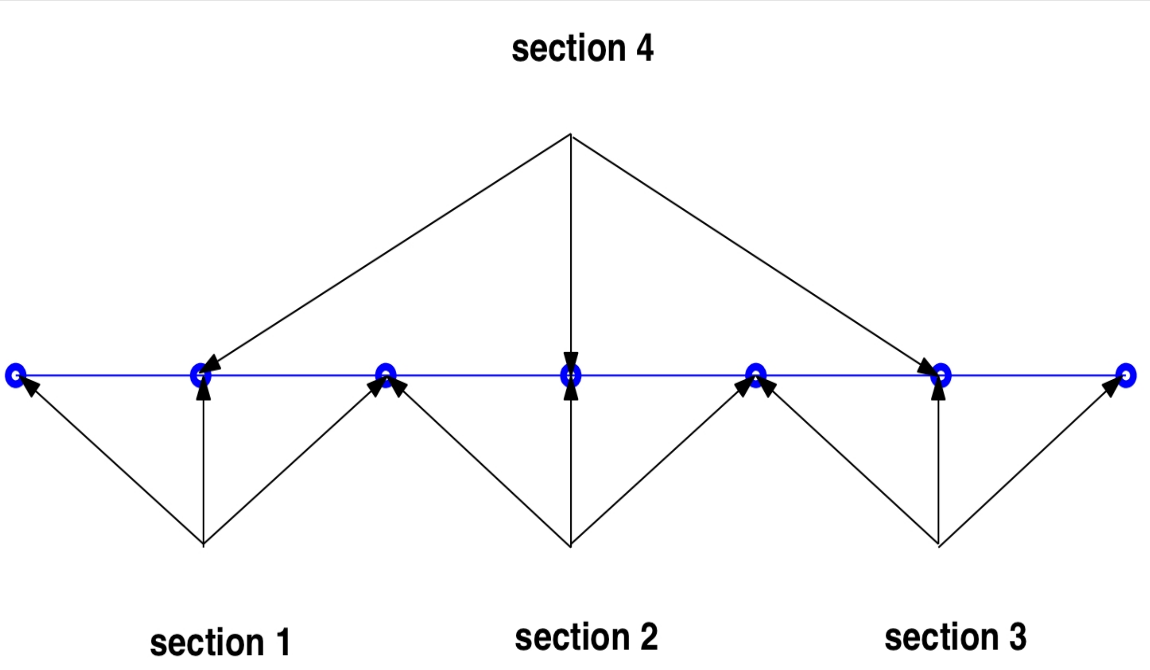
Recall in LTSA, we extract the 2nd to $d + 1$st smallest eigenvectors of Φ.

An example

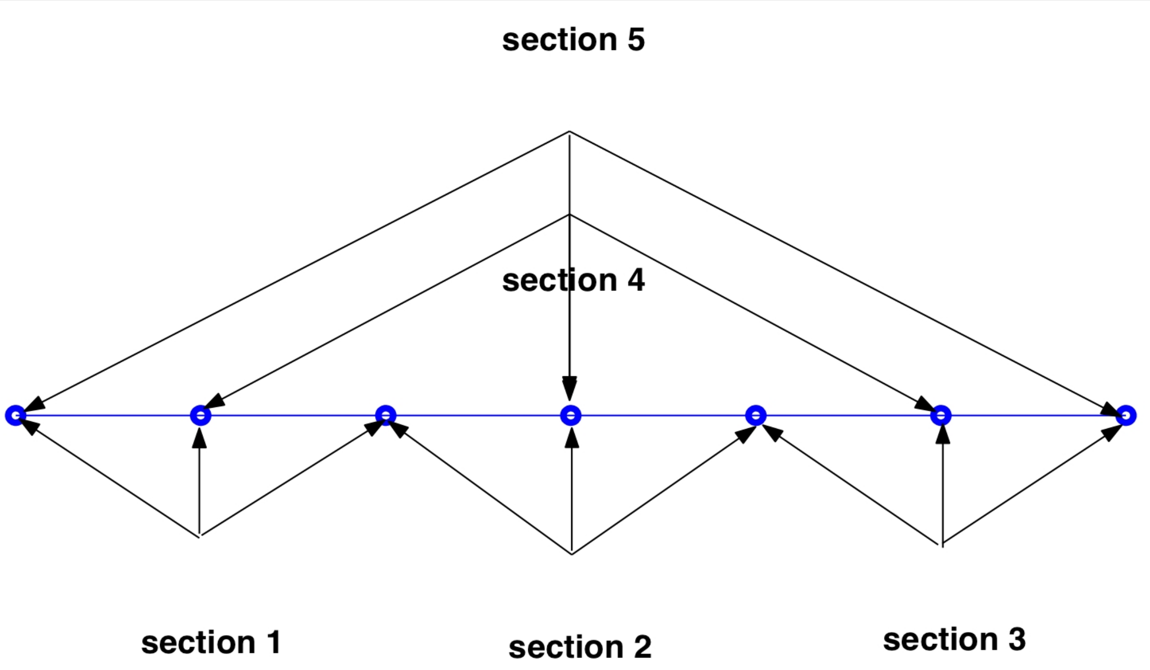


Nullspace contains unwanted information, and embedding has manifold folds upon itself.

Multiple Sections



Unsuccessful recovery.



Successful recovery.

Necessary Conditions

Theorem. Let $[e,\, T_i^T]$ be of full column-rank for $i = 1, \ldots, s$. If

$$\mathcal{N}\{\Phi\} = \operatorname{span}\{[e,\, T^T]\},$$

then

1) $\{T_1,\, \ldots,\, T_s\}$ is connectedly overlapped, or

2) $\{T_1,\, \ldots,\, T_s\}$ is not connectedly overlapped, and for any maximally connectedly overlapped subset $\{T_{i_1},\, \ldots,\, T_{i_k}\}$,

$$\{[T_{i_1},\, \ldots,\, T_{i_k}], [T_{i_{k+1}},\, \ldots,\, T_{i_s}]\}$$

are fully overlapped.

Sufficient Conditions

Theorem.

1) $\{T_1, \ldots, T_s\}$ is connectedly overlapped, or

2) $\{T_1, \ldots, T_s\}$ is not connectedly overlapped, but there is a maximally connectedly overlapped subset $\{T_{i_1}, \ldots, T_{i_k}\}$ such that

$$\{T_{i_{k+1}}, \ldots, T_{i_s}\}$$

is also a connectedly overlapped subset and

$$\{[T_{i_1}, \ldots, T_{i_k}], [T_{i_{k+1}}, \ldots, T_{i_s}]\}$$

are fully overlapped, then

$$\mathcal{N}\{\Phi\} = \text{span}\{[e, T^T]\}.$$

Spectral gap

Recall, two sections fully overlapped iff $\sigma_d(V - \bar{v}e^T) > 0$, where V, vectors in the intersection. Quantitatively,

Theorem. The smallest nonzero eigenvalue of Φ is $O(\sigma_d^2(V - \bar{v}e^T))$. (Only $d = 1$ case is proved, working on the more general case.)

Theorem. For $i = 1, 2$, $P_i = Q_i Q_i^T$ orthogonal projections onto the orthogonal complements of $[e, T_i^T]$, and $H_i = Q_i(I_1 \cap I_2, :)$. Then

$$\lambda(\Phi) = \{0, 1, 1 \pm \sigma_i(H_1^T H_2)\}.$$

Practical situation

Each patch is handled separately, local coordinates $\Longrightarrow$ transformations and approximations w.r.t. global coordinates.

Proposition. Let $T_i, i = 1, 2$ be two sections of T, and $\Theta_i, i = 1, 2$ are the same as T_i up to an affine transformation. P_i be the orthogonal projection onto the orthogonal complement of $[e, \Theta_i^T]$. Then

$$\mathrm{span}\{[e, T^T]\} \subset \mathcal{N}\{\Phi\},$$

where Φ is built from P_i. Furthermore, if T_i fully overlapped, then

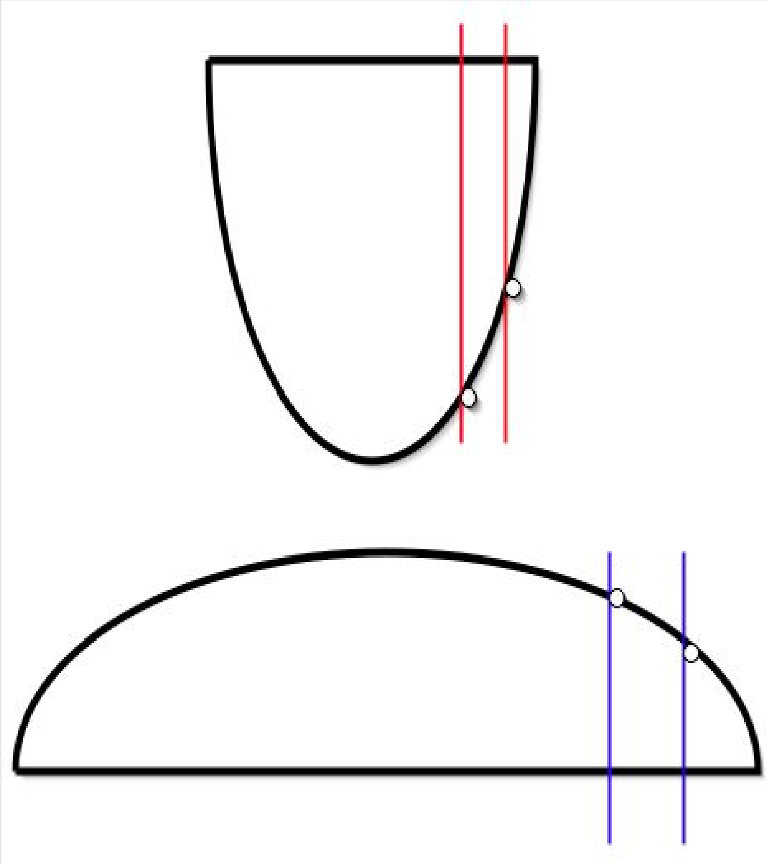
$$\mathrm{span}\{[e, T^T]\} = \mathcal{N}\{\Phi\}.$$

LTSA recovers isometry

- Assume f is an isometry.
- Up to local approximation errors, local coordinates are isometric to global coordinates: Jacobi matrix is orthonormal.
- Nullspace of Φ gives global coordinates.
- Normalization of nullspace basis vectors.

Two competing requirements

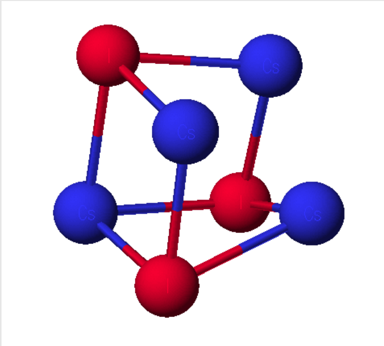
- Large overlap favors large neighborhood
- Large neighborhood results in large approximation errors

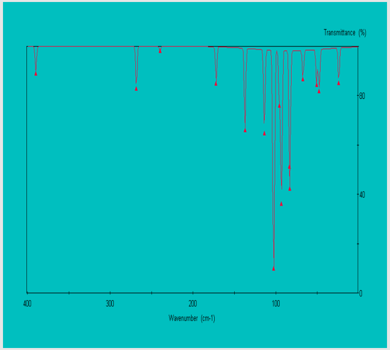


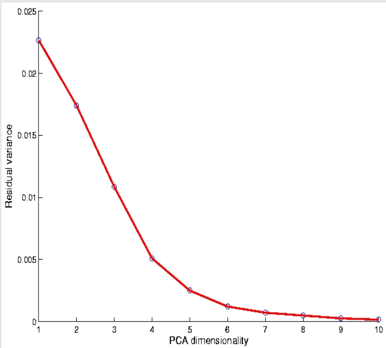
Adaptive manifold learning

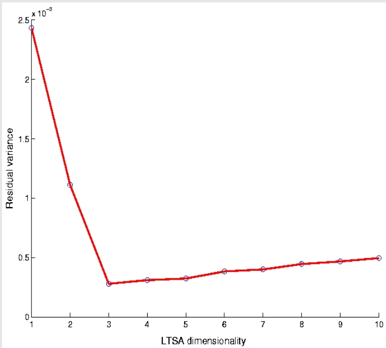
- Adaptive neighborhood size selection.
- Bias reduction in local coordinate estimation.

Molecular Dynamics Simulation/Example









Energy Landscape Interpolation

- N-particles simulation system $\Rightarrow$ configuration space $d = 3N$
- Not all $3N$ degrees of freedom are *activated*
- Trajectories of the particles occupy a low-dimensional manifold
- More efficient and accurate approaches for the potential energy surface and force computation

CAMLET: A Combined Ab-initio Manifold LEarning Toolbox

- Explore the low-dimensional characteristics particle trajectories
- Identify the suitable clusters in the reduced dimension spaces
- Efficient energy and force interpolations

Thanks! Questions?

Papers/preprints can be found at

http://www.cse.psu.edu/∼zha/papers.html