Lecture 11
Geometric Data Analysis: Local Tangent Space Alignment (LTSA)

姚 远

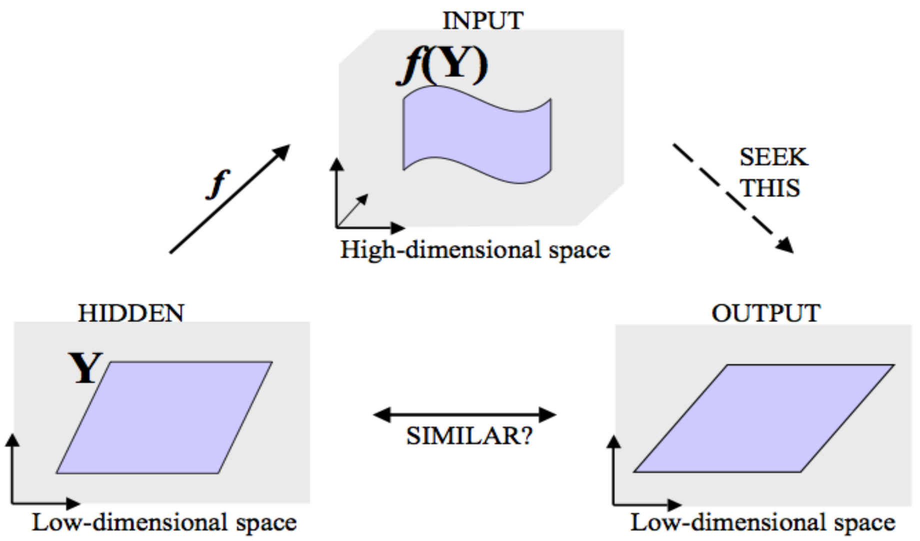
2011.5.10

Manifold Learning

- Given a set of data points $x_1,\ldots,x_N$ in R^m

$$x_i = f(\tau_i) + \varepsilon_i, \qquad i = 1,\ldots,N$$

- Manifold Learning: reconstruct f from x_i



Spectral Geometric Embedding

Given $x_1, \ldots, x_n \in \mathcal{M} \subset \mathbb{R}^N$,
Find $y_1, \ldots, y_n \in \mathbb{R}^d$ where $d << N$

- ISOMAP (Tenenbaum, et al, 00)
- LLE (Roweis, Saul, 00)
- Laplacian Eigenmaps (Belkin, Niyogi, 01)
- Local Tangent Space Alignment (Zhang, Zha, 02)
- Hessian Eigenmaps (Donoho, Grimes, 02)
- Diffusion Maps (Coifman, Lafon, et al, 04)

Related: Kernel PCA (Schoelkopf, et al, 98)

Recall: PCA

- 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 Spaces

- 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 $J_f(\tau)$.

Local vs. 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

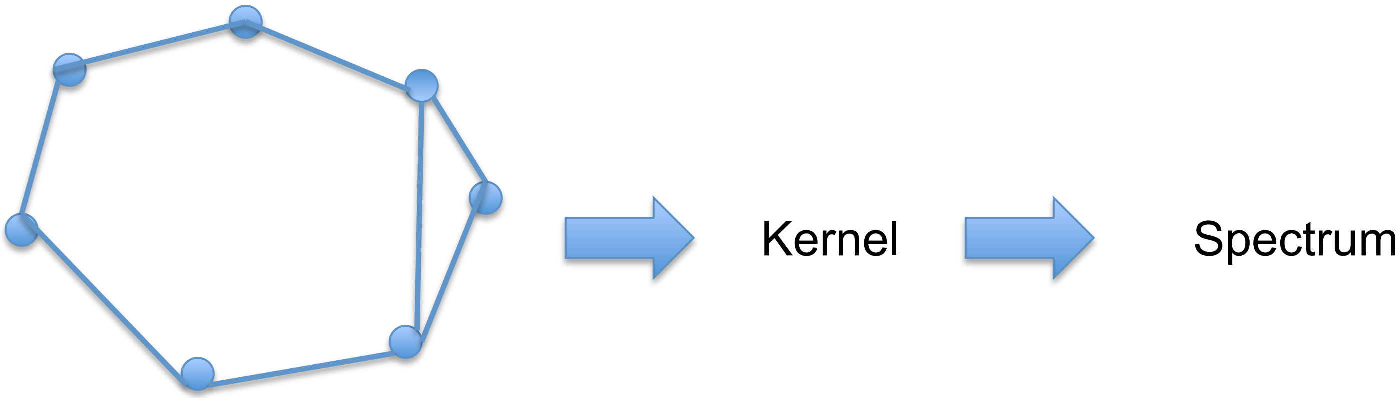
Find global coordinate τ and local affine transformation L_τ to minimize (Symbolically),

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

over all possible nonsingular L_τ.

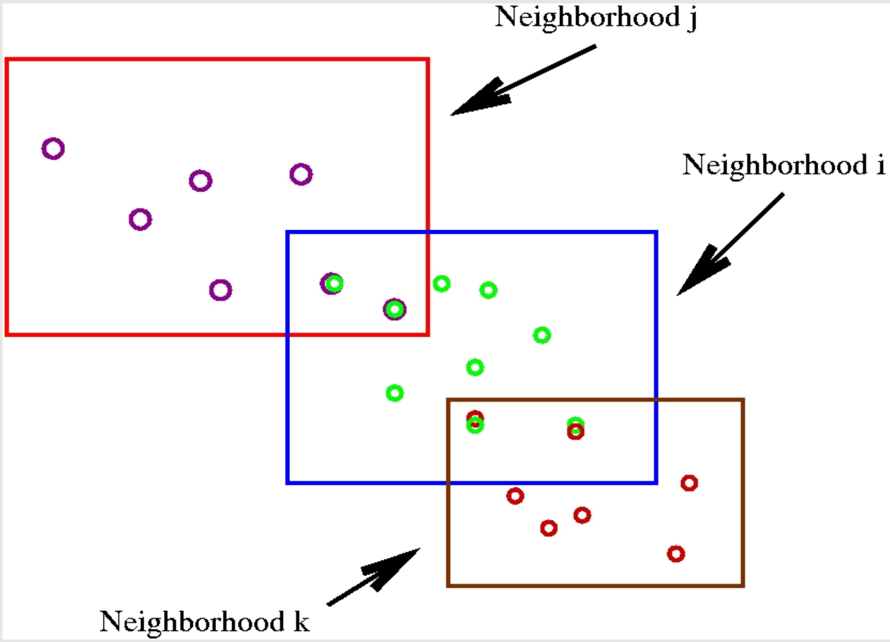
Meta-Algorithm

- Construct a neighborhood graph
- Construct a positive semi-definite kernel
- Find the spectrum decomposition



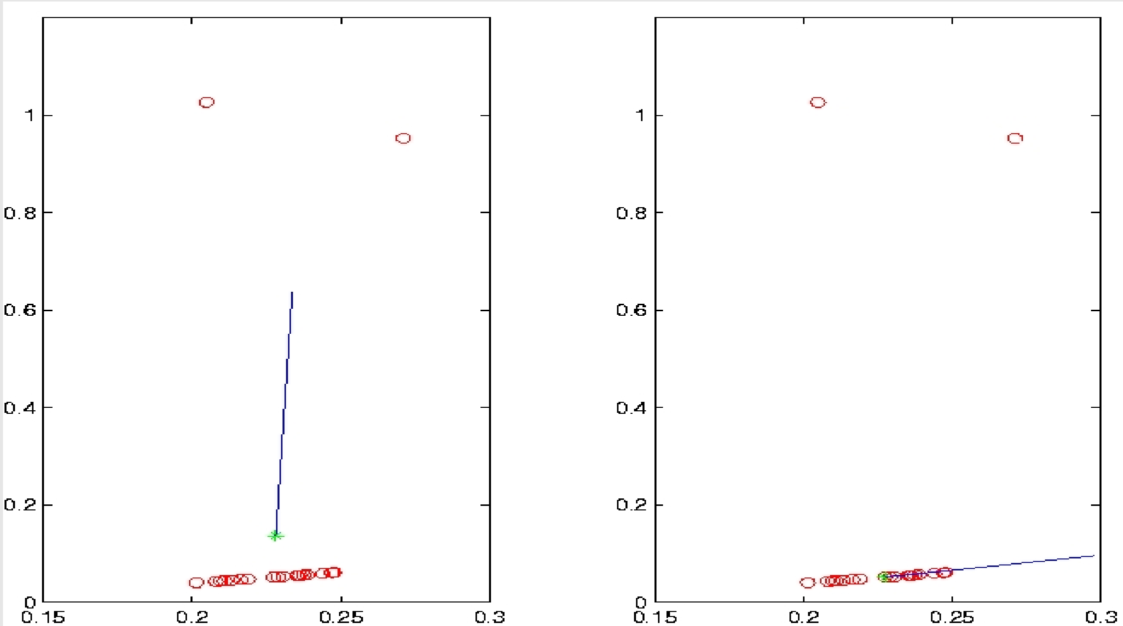
K-NN Neighborhood Graph

- 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.)



PCA = Approximate Tangent Space

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



Weighted PCA as Approximate Tangent Space

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

Weight selection

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

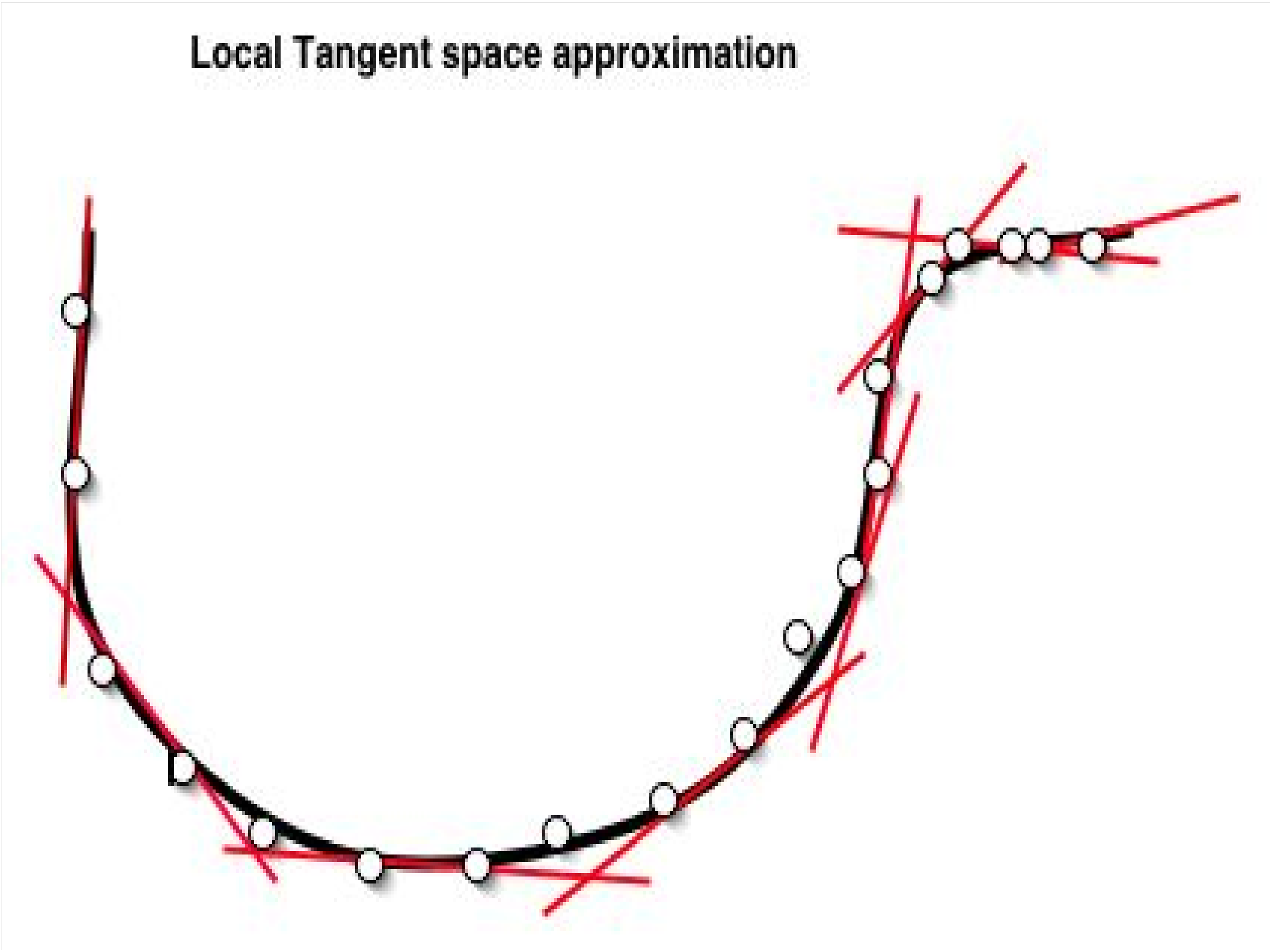
1. Compute the current weights,

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

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.

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

Alignment (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

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

- 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 ($\Phi = \Psi\Psi^T$)

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

- 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

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 $T = [u_2,\ \cdots,\ u_{d+1}]^T$.

Comparisons of Manifold Learning Techniques

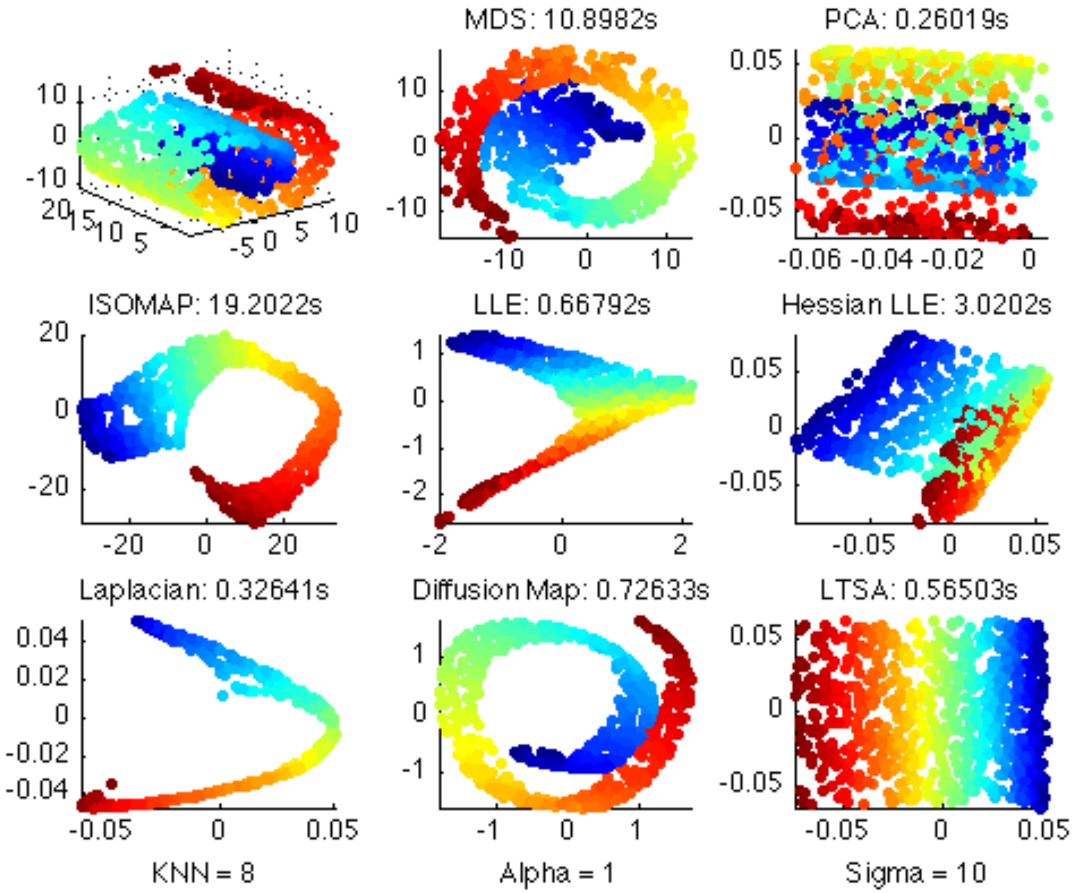
- MDS
- PCA
- ISOMAP
- LLE
- Hessian LLE
- Laplacian LLE
- Diffusion Map
- Local Tangent Space Alignment
- Matlab codes: mani.m

Courtesy of Todd Wittman

How To Compare

- Speed
- Manifold Geometry
- Non-convexity
- Curvature
- Corners
- Noise
- Non-uniform Sampling
- Sparse Data
- Clustering
- High-Dimensional Data: *Can the method process image manifolds?*
- Sensitivity to Parameters
  - K Nearest Neighbors: *Isomap, LLE, Hessian, Laplacian, KNN Diffusion*
  - Sigma: *Diffusion Map, KNN Diffusion*

Speed on Swiss Roll



Now go to Todd Wittman's slides, page 13th…

Reference

- Tenenbaum, de Silva, and Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science* 290:2319-2323, 22 Dec. 2000.
- Roweis and Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290:2323-2326, 22 Dec. 2000.
- M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation, 15:1373–1396, 2003.
- D. L. Donoho and C. Grimes. Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data. PNAS 100 (10): 5591–5596 2003.
- R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. PNAS 102 (21):7426-7431, 2005.
- Zhenyue Zhang and Hongyuan Zha, Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment, SIAM Journal of Scientific Computing, 2002

Acknowledgement

- Slides stolen from H. Zha and G. Lerman, et al.