Diffusion Maps and Coarse-Graining:
A unified framework for dimensionality reduction, graph partitioning and data set parameterization

Authors: Stephane Lafon and Ann B. Lee (PAMI, Sept. 2006)

Presented by Shihao Ji
Duke University Machine Learning Group
Sept. 28, 2006
Outline

• Diffusion distances and Maps
• Graph partitioning and subsampling
• Numerical examples
Diffusion distances

- Let $G = (\Omega, W)$ be a finite graph with $n$ nodes, and weight matrix $W$ satisfies the following conditions:
  - symmetry: $W = W^T$
  - positivity: $w(x, y) \geq 0 \quad \forall x, y \in \Omega$

  i.e. Gaussian kernel

- Markov random walk $P$
  
  \[ p_1(y|x) = \frac{w(x, y)}{\sum_{z \in \Omega} w(x, z)} \]

- $t$ step random walk $P^t = \overbrace{P \cdot P \cdots P}^{t}$
  
  \[ p_t(y|x) \]
Diffusion distances (cont’d)

• Definition

\[ D^2_t(x, z) = \|p_t(\cdot \mid x) - p_t(\cdot \mid z)\|_{1/\phi_0}^2 = \sum_{y \in \Omega} \frac{(p_t(y \mid x) - p_t(y \mid z))^2}{\phi_0(y)} \]

where \( \phi_0 \) is the unique stationary distribution of \( P \).

\[ \phi_0 P = \phi_0 \quad \rightarrow \quad \phi_0(x) = \frac{d(x)}{\sum_{z \in \Omega} d(z)} \]
Diffusion Maps

• The transition matrix $P$ is adjoint to a symmetric matrix
  \[ P_s = D^{1/2} P D^{-1/2} \]
  thus, $P$ and $P_s$ share the same eigenvalues.

• Since $P_s$ is a symmetric matrix
  
  eigenvalues: \( \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{n-1} \geq 0 \)
  
  eigenvectors: \( \{v_j\}_{j=0}^{n-1} \) form the orthonormal basis.
Diffusion Maps (cont’d)

- The left and right eigenvectors of $P$:

  \[ \phi_j = D^{1/2}v_j, \quad \psi_j = D^{-1/2}v_j \]

  \[ \langle \phi_j \cdot \psi_k \rangle = \delta_{jk} \]

- Biorthogonal spectral decomposition

  \[ p_t(y|x) = \sum_{j \geq 0} \lambda_j^t \psi_j(x) \phi_j(y) \]

- Diffusion distances

  \[ D_t^2(x, z) = \sum_{j=1}^{n-1} \lambda_j^{2t} (\psi_j(x) - \psi_j(z))^2 \]

  \[ \approx \sum_{j=1}^{m} \lambda_j^{2t} (\psi_j(x) - \psi_j(z))^2 \quad \text{if} \quad \lambda_m^t > \delta \]
Diffusion Maps (cont’d)

- Diffusion Maps

\[ \psi_t(x) = \begin{pmatrix} 
\lambda_1^t \psi_1(x) \\
\lambda_2^t \psi_2(x) \\
\vdots \\
\lambda_m^t \psi_m(x) 
\end{pmatrix} \]

- Diffusion distances

\[ D_t^2(x, z) \approx \sum_{j=1}^{m} \lambda_j^{2t} (\psi_j(x) - \psi_j(z))^2 = \| \psi_t(x) - \psi_t(z) \|^2 \]
Graph partitioning

- Consider an arbitrary partition $\{S_i\}_{1 \leq i \leq k}$

\[ P, \phi, \psi \quad \longrightarrow \quad \tilde{P}, \tilde{\phi}, \tilde{\psi} \]

i.e., $\tilde{\phi}_l(S_i) = \sum_{x \in S_i} \phi_l(x)$
Graph partitioning (cont’d)

• **Definition** (*geometric centroid*):

\[
c(S_i) = \sum_{x \in S_i} \frac{\phi_0(x)}{\tilde{\phi}_0(S_i)} \psi_t(x)
\]

• **Theorem**: for \(0 \leq l \leq n - 1\), we have

\[
\tilde{\phi}_l^T \tilde{P} = \lambda_l^t \tilde{\phi}_l^T + e_l \quad \text{and} \quad \tilde{P} \tilde{\psi}_l = \lambda_l^t \tilde{\psi}_l + f_l
\]

where

\[
\|e_l\|_{1/\tilde{\phi}_0}^2 \leq 2\mathcal{D} \quad \text{and} \quad \|f_l\|_{1/\tilde{\phi}_0}^2 \leq 2\mathcal{D}
\]

and

\[
\mathcal{D} = \sum_i \sum_{x \in S_i} \phi_0(x) \|\psi_t(x) - c(S_i)\|^2
\]
Graph partitioning (Cont’d)

• This theorem tells us

1. If \( \lambda_i^t \gg \sqrt{D} \) then \( \phi_l \) and \( \psi_l \) are approximate left and right eigenvectors of \( \tilde{P} \) with approximate eigenvalue \( \lambda_i^t \).

2. In order to maximize the quality of approximation, we need to minimize the following distortion in diffusion space:

\[
D = \sum_i \sum_{x \in S_i} \phi_0(x) \| \Psi_t(x) - c(S_i) \|^2
\]

• This provides a rigorous justification for k-means clustering in diffusion space.
Numerical examples

- Diffusion distance vs. Euclidean distance

The Swiss roll, and its quantization by k-means ($k=4$)
Numerical examples (cont’d)

- Robustness of the diffusion distance
Numerical examples (cont’d)

Averaged on 1000 instances
Messages

- Diffusion maps provide a unified framework for dimensionality reduction, graph partitioning and data set parameterization.

- Coarse-graining gives a rigorous justification of k-means clustering in diffusion space.

- Diffusion distance is robust to noise and small perturbations of the data.