Diffusion Maps

– The Finite Set Case

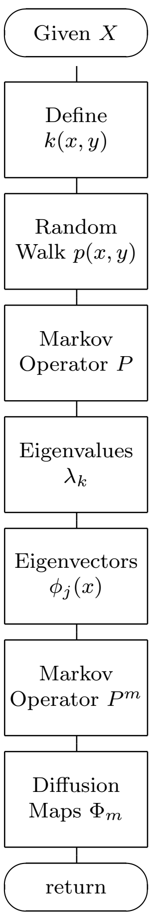
(Dr. Xin Li, UCF Math, 10/20/2008)

In this lecture, we take a closer look at the method of diffusion maps for the case when X is a finite set. This is an important special case since many applications (in particular in image analysis) come with a finite X.

It is also hoped that after this lecture, you will be able to implement diffusion maps as a tool for meaningful applications.

Here is quick recap of the method of diffusion maps:

We are given a set X of points from a space of "high" dimension.



General Idea of Diffusion Maps (**Homework**: Put your notes and remarks beside each box)

1 Finite Set X

Let X be a given set of N elements.

$$X = \{x_1, x_2, ..., x_N\}.$$

For each pair of points from X, x_i and x_j, we assign a value $k(x_i, x_j)$ to reflect our knowledge about the relationships between the two points. For example, their distance from each other, the "closeness" of certain feature parameters of the two points, etc. For convenience, let us label the points in X by integers. So, x_i will be referred to as the ith point of X. Construct the matrix

$$K = \begin{pmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_N) \\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_2,x_N) \\ \cdots & & & \\ k(x_N,x_1) & k(x_N,x_2) & \cdots & k(x_N,x_N) \end{pmatrix}.$$

So, $K(i,j) = k(x_i, x_j)$. There are several requirements on K:

1. Symmetry: $K^t = K$
2. Positivity: $K(i,j) \geq 0$ for all $i, j = 1, 2, ..., N$
3. Positive Definiteness: $\mathbf{v}^t K \mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^N$ and $\mathbf{v} \neq 0$

We now construction *normalized graph Laplacian*: Set

$$d_i = \sum_{j=1}^{N} K(i,j)$$

so d_i is just the column (= row) sum of K and define

$$p_{ij} = \frac{K(i,j)}{d_i}.$$

Note that p is positive and

$$\sum_{j=1}^{N} p_{ij} = 1.$$

So, we interpret p_{ij} as a probability for a random walk to move from x_i to y_j. Define

$$P = (p_{ij}).$$

Then P is a stochastic matrix (meaning: its rows summed to the same constant, 1, in our case).

Note that P is not symmetric. Consider

$$\tilde{p}_{ij} = \sqrt{\frac{d_i}{d_j}} p_{ij} = \frac{p_{ij}}{\sqrt{d_i d_j}}.$$

Then $\tilde{P} = (\tilde{p}_{ij})$.

Now, $\tilde{P}$ is a symmetric matrix.

2 Spectral Decomposition

Reminder: *The finite-dimensional* **spectral theorem** *(which you proved in your first project) says that any symmetric matrix whose entries are real can be diagonalized by an orthogonal matrix.*

As a real (with actually positive entries) symmetric matrix, $\tilde{P}$ can be diagonalized and there is an orthogonal matrix Q whose columns are the eigenvectors of $\tilde{P}$ such that

$$\tilde{P} = Q \Lambda Q^t, \quad QQ^t = I, \tag{1}$$

where Λ is the diagonal matrix forms by the eigenvalues of $\tilde{P}$:

$$1 = \lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_N.$$

Write $\mathbf{v}_i$ for the ith column of Q and an eigenvector associated with λ_i. Then (1) can be written as

$$\tilde{P} = \sum_{k=1}^{N} \lambda_k \mathbf{v}_k \mathbf{v}_k^t.$$

Homework: Verify the identity.

Note that

$$\tilde{P} = DPD^{-1},$$

where

$$D = \begin{pmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & 0 & \\ & 0 & \ddots & \\ & & & \sqrt{d_N} \end{pmatrix}.$$

Homework: Verify the equality.

So,

$$P = D^{-1}\tilde{P}D = \sum_{k=1}^{N} \lambda_k D^{-1}\mathbf{v}_k\mathbf{v}_k^t D = \sum_{k=1}^{N} \lambda_k \psi_k \varphi_k^t,$$

with

$$\psi_k = D^{-1}\mathbf{v}_k$$

and

$$\varphi_k = D\mathbf{v}_k.$$

This is the decomposition we are trying to arrive at.

Remarks: There are alternative ways to derive the decomposition. If you remember the Singular Value Decomposition theorem, then you may come up with a shorter solution.

The two families of vectors $\{\psi_k\}$ and $\{\varphi_k\}$ have the following nice property:

$$\psi_p^t \varphi_q = 0 \quad \text{if} \quad p \neq q$$

Homework: Verify the equality above.

Let $\tilde{P}^m$ be the mth power of $\tilde{P}$, then

$$\tilde{P}^m = \sum_{i=1}^{N} \lambda_i^m \mathbf{v}_k \mathbf{v}_k^t.$$

Homework: Let $\mathbf{v}_k = (v_{k1}, v_{k2}, ..., v_{kN})^t$. Show that if $\tilde{p}_m(i,j)$ denote the (i,j) entry of the matrix $\tilde{P}^m$, then

$$\tilde{p}_m(i,j) = \sum_{k=1}^{N} \lambda_k^m v_{ki} v_{kj}.$$

2.1 Diffusion Maps and Distances

Let $\mathbf{v}_k = (v_{k1}, v_{k2}, ..., v_{kN})^t$. We introduce the family of *diffusion maps* $\{\Phi_m\}$ by

$$\Phi_m(x_i) = \begin{pmatrix} \lambda_1^m v_{1i} \\ \lambda_2^m v_{2i} \\ \vdots \\ \lambda_N^m v_{Ni} \end{pmatrix},$$

and the family of *diffusion distances* $\{D_m\}$ defined by

$$D_m(x_1, x_j) = \|\Phi_m(x_i) - \Phi_m(x_j)\|.$$

Homework. 1. Show that

$$D_m^2(x_i, x_j) = \tilde{p}_m(i,i) - 2\tilde{p}_m(i,j) + \tilde{p}_m(j,j).$$

2. Show that

$$D_{2m}(x_i, y_j)^2 = \sum_{k=1}^{N} |\tilde{p}_m(i,k) - \tilde{p}_m(j,k)|^2.$$

3. What are the interpretations of the previous two identities?

3 What's Next

We will examine several special cases where the idea of diffusion maps has been used. We will also look at some techniques that were introduced before the diffusion maps method and employed similar (but not the same) approaches. In particular, we will look at a solution to the correspondence problem in images.