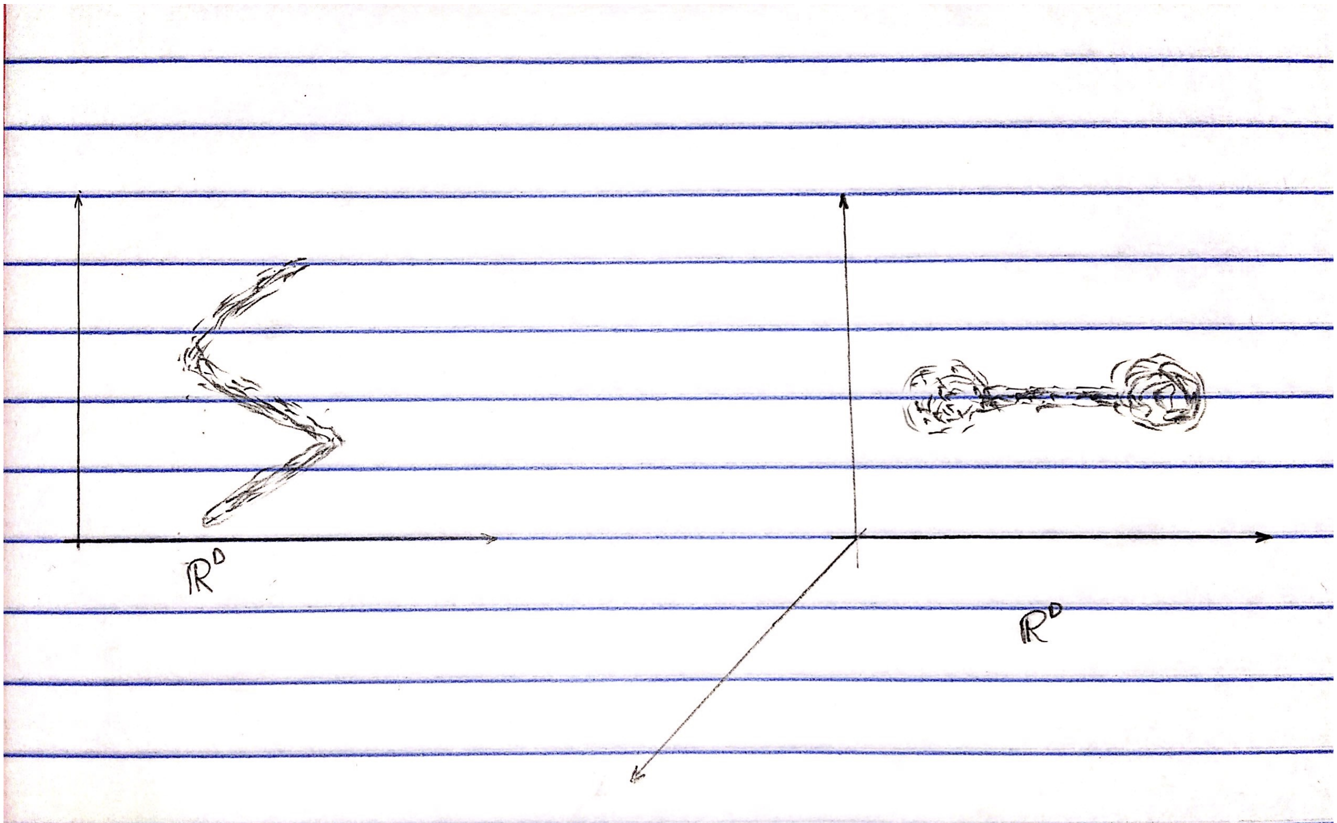
Diffusion Geometry

Consider a set of points $\{x_i\}_j^n$ = 1 in $\mathbb{R}^n$



Goals:-

* Find nonlinear low dimensional structures.
  - Parameterization.
* Clustering/Classification.

We seek a nonlinear map $\phi = \mathbb{R}^n \to \mathbb{R}^k$

now, consider a random walk on the points. Suppose the probability of jumping from X_j to X_i is $\approx k(X_i, X_j)$ for some kernel function k.

Common choices:

* $k(x, y) = k(y, x)$.
* $e^{-\frac{1}{\varepsilon^2}|x-y|^2}$ (usually the best one)

We Need:

* $k(x, y) = k(y, x)$.
* $k(x, y) \geq 0 \, \forall x, y$.
* Often, we require $\sum_{i,j=1}^n q : k(x_i, x_j) q_j \geq 0$ for every vector q

Set:

* $k(x, y) = k(y, x)$.
* $k(x, y) \geq 0 \, \forall x, y$.
* Often, we require $\sum_{i,j=1}^n q : k(x_i, x_j) q_j \geq 0$ for every vector q

Set $L(i,j) = k(x_i, x_j)$. L is an n x n matrix but not of transition probabilities since the columns do not sum to 1.

Let us fix this:

Define diagonal matrix D via

$$D(i,j) = \sum_{j=1}^{n} L(i,j)$$

Then,

$$M = LD^{-1}$$

where M is of transition probabilities.

For $t = 1, 2, 3, ...$, the matrix M^t holds the transition probabilities for t steps of the random walk. We seek to compute the eigenvectors and eigenvalues of M^t.

$$\text{Set } \widetilde{M} = D^{-\frac{1}{2}} M D^{\frac{1}{2}} = D^{-\frac{1}{2}} L D^{-1} D^{\frac{1}{2}} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}.$$

Note: Since L is symmetric, $\widetilde{M}$ is also symmetric.

Recall:

- $*$ Let B be an invertable matrix.
- $*$ Let A be a matrix of the same size as B.

Then λ is an eval of A if and only if λ is an eval of $B^{-1}AB$. Proof:

$$Av = \lambda v \Leftrightarrow B^{-1}Av = B^{-1}\lambda v \Leftrightarrow B^{-1}ABB^{-1}v = \lambda B^{-1}v$$

Thus, $\widetilde{M}$ is symmetric and has the same evals as M.

$$M^t = (D^{\frac{1}{2}} \widetilde{M} D^{-\frac{1}{2}})^t = D^{\frac{t}{2}} \widetilde{M}^t D^{-\frac{t}{2}}$$

Let us compute the eigenvalue decomposition of $\widetilde{M}$ so that $\widetilde{M} = V \Lambda V^*$

$$\text{where } \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

$$\text{So } M^t = D^{\frac{1}{2}} V \Lambda^t V^* D^{-\frac{1}{2}}$$

The evals typically decay. The faster they decay, the better the points organize on a low rank structure.

Let us define the diffusion distance:

$$d(i,j) := (\textstyle\sum_{p=1}^{n} \lambda_p^{2t} \, |V_p(i) - V_p(j)|^2)^{\frac{1}{2}}$$

$\left\{\lambda_j^{2t}\right\}_{j=1}^{n}$ decays quickly, so we truncate:

$$d(i,j) \approx (\textstyle\sum_{p=1}^{k} \lambda_p^{2t} \, |V_p(i) - V_p(j)|^2)^{\frac{1}{2}}$$

$$= \left\|\phi_{(i)} - \phi_{(j)}\right\|$$

$$\text{where } \phi = \begin{bmatrix} \lambda_1^t V_1(i) \\ \lambda_2^t V_2(i) \\ \lambda_3^t V_3(i) \\ \vdots \\ \lambda_k^t V_k(i) \end{bmatrix}.$$

phi is a nonlinear map from $\|X_i\|_{i=1}^{n}$ to $\mathbb{R}^k$