Lecture 10
Geometric Data Analysis: Laplacian, Diffusion, and Hessian LLE

姚 远

2011.5.3

Manifold Learning

Learning when data $\sim \mathcal{M} \subset \mathbb{R}^N$

- Clustering: $\mathcal{M} \to \{1, \ldots, k\}$

  connected components, min cut

- Classification/Regression: $\mathcal{M} \to \{-1, +1\}$ or $\mathcal{M} \to \mathbb{R}$

  P on $\mathcal{M} \times \{-1, +1\}$ or P on $\mathcal{M} \times \mathbb{R}$

- Dimensionality Reduction: $f : \mathcal{M} \to \mathbb{R}^n \quad n << N$

- $\mathcal{M}$ unknown: what can you learn about $\mathcal{M}$ from data?

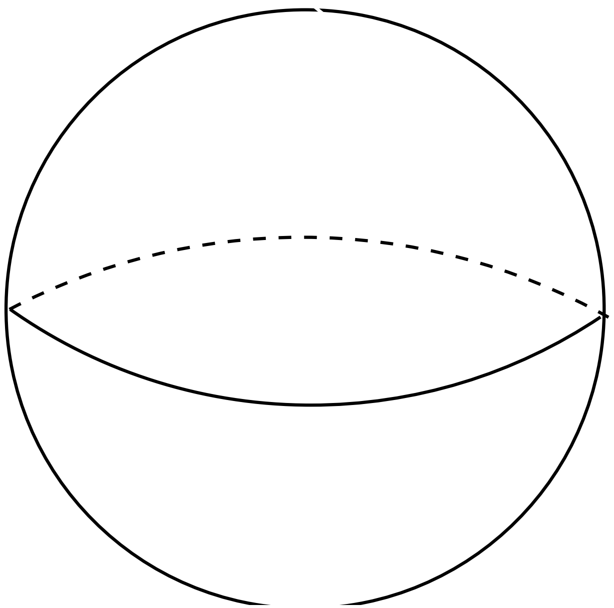
  e.g. dimensionality, connected components
  holes, handles, homology
  curvature, geodesics

All you wanna know about differential geometry but were afraid to ask, in 9 easy slides

Embeded Manifolds

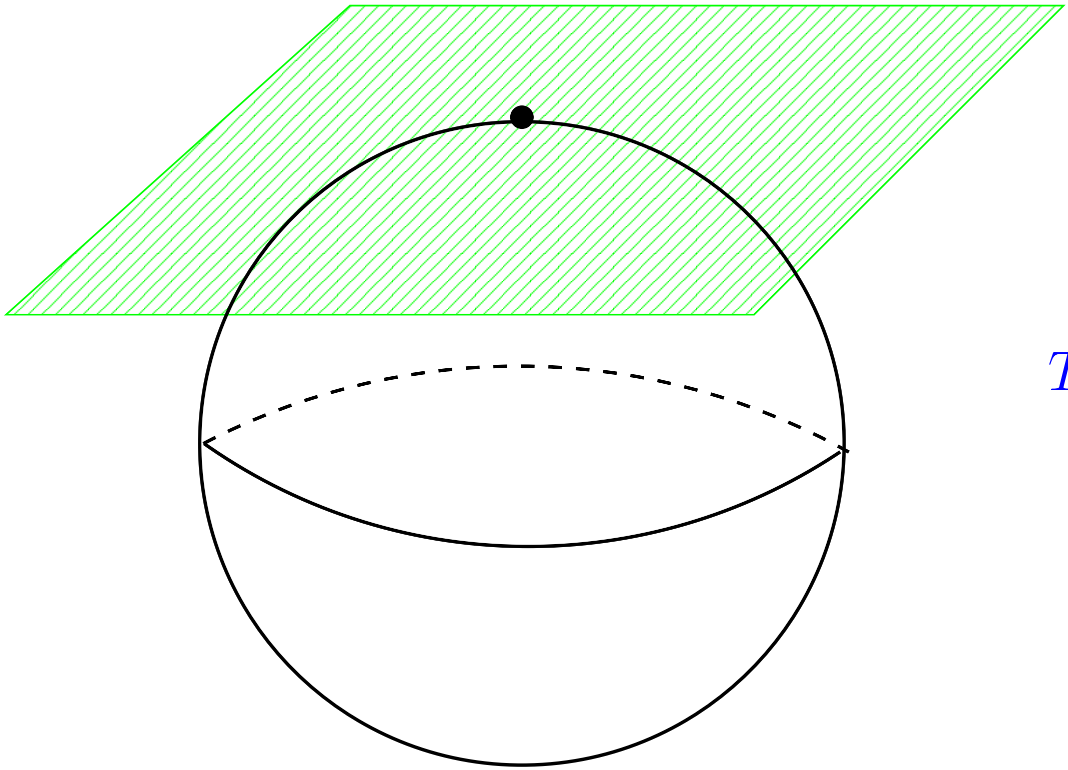
$$\mathcal{M}^k \subset \mathbb{R}^N$$

Locally (not globally) looks like Euclidean space.



$$S^2 \subset \mathbb{R}^3$$

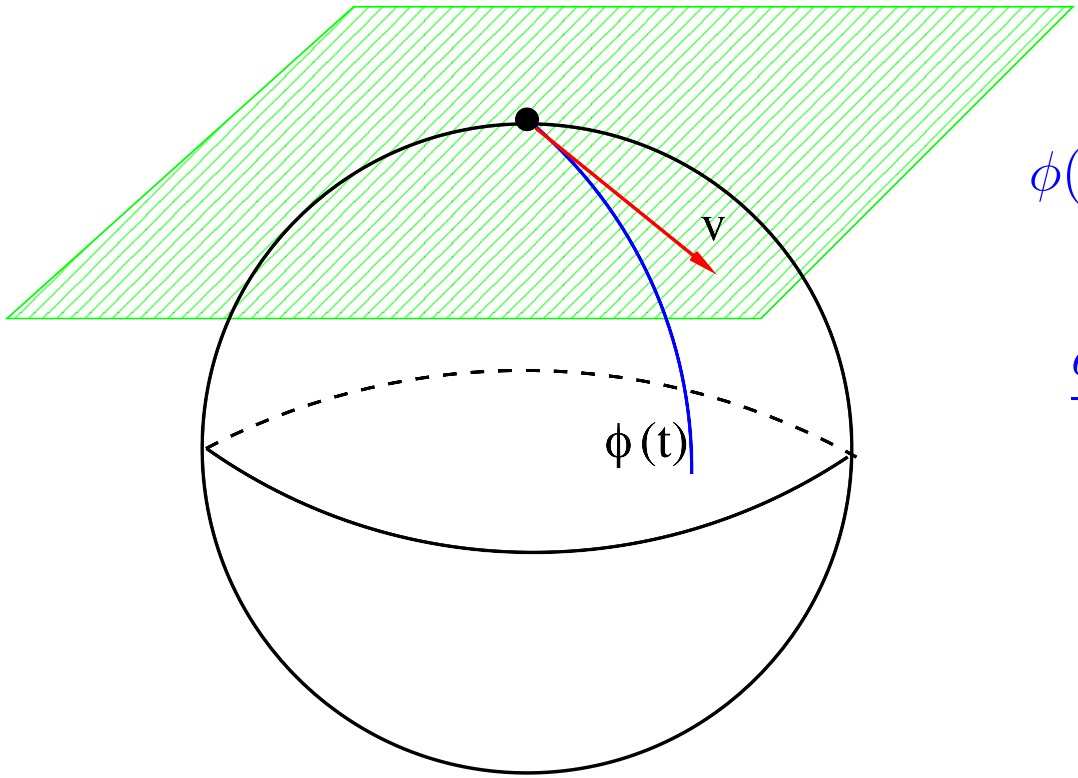
Tangent Space



$T_p\mathcal{M}^k \subset \mathbb{R}^N$

k-dimensional affine subspace of $\mathbb{R}^N$.

Tangent Vectors and Curves



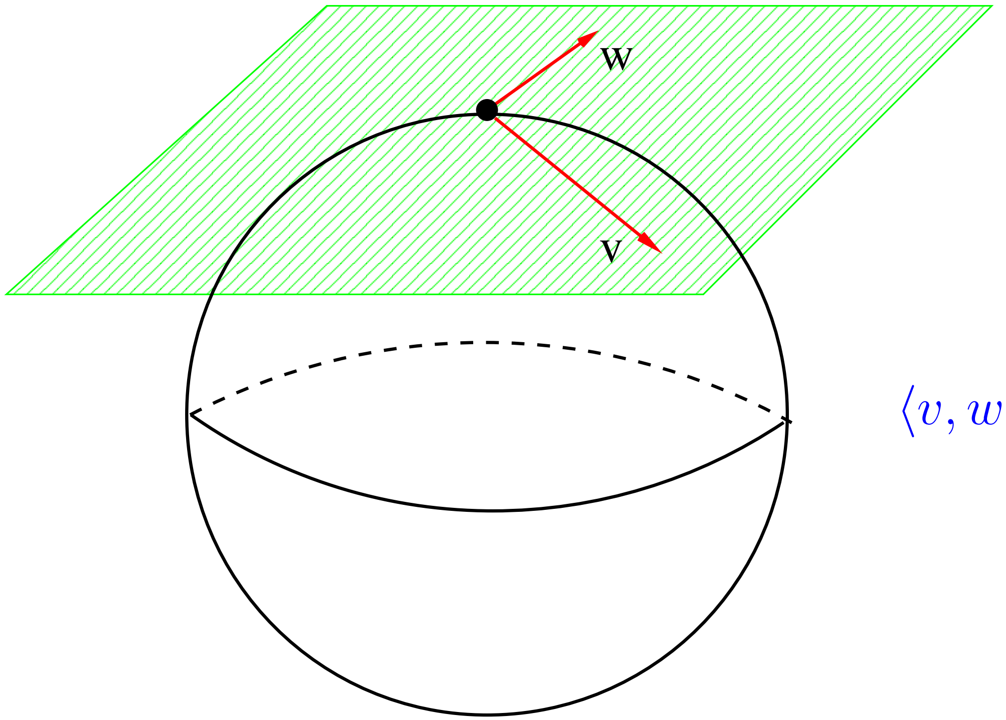
$$\phi(t) : \mathbb{R} \to \mathcal{M}^k$$

$$\left.\frac{d\phi(t)}{dt}\right|_0 = V$$

Tangent vectors <−−−> curves.

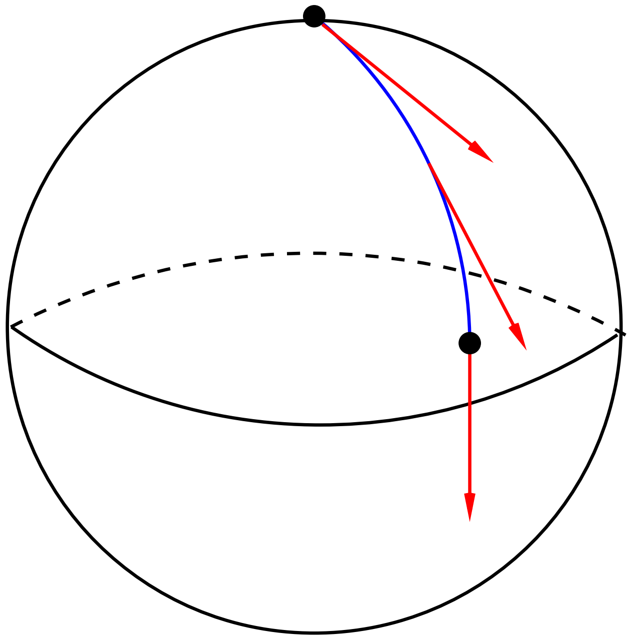
Riemannian Geometry

Norms and angles in tangent space.



$\langle v, w \rangle \qquad \|v\|, \ \|w\|$

Geodesics



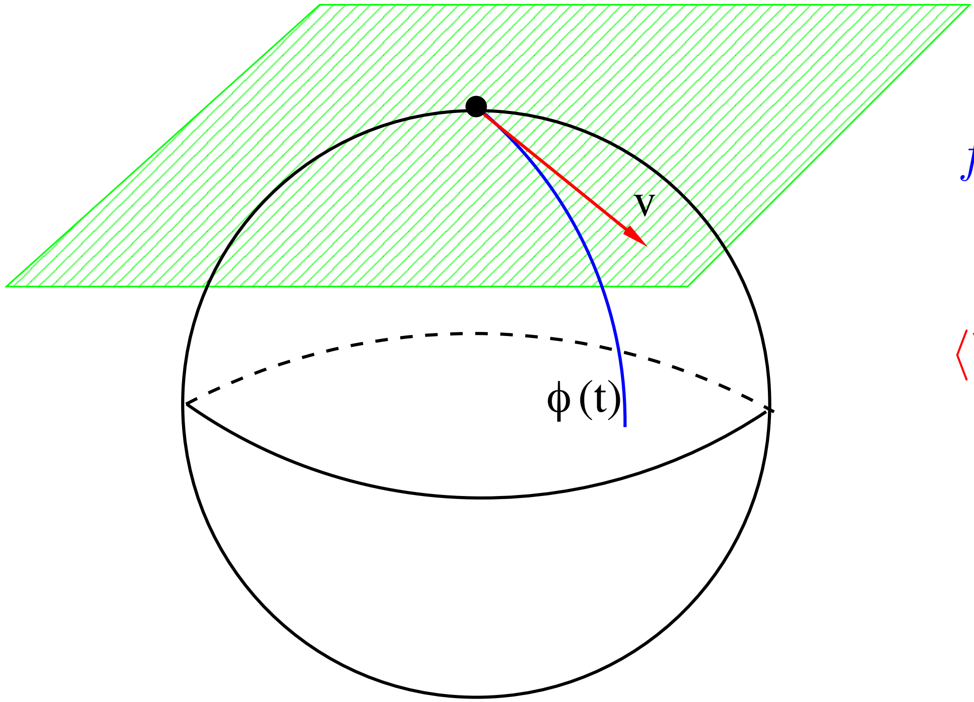
$$\phi(t) : [0, 1] \rightarrow \mathcal{M}^k$$

$$l(\phi) = \int_0^1 \left\| \frac{d\phi}{dt} \right\| dt$$

Can measure length using norm in tangent space.

Geodesic — shortest curve between two points.

Gradients



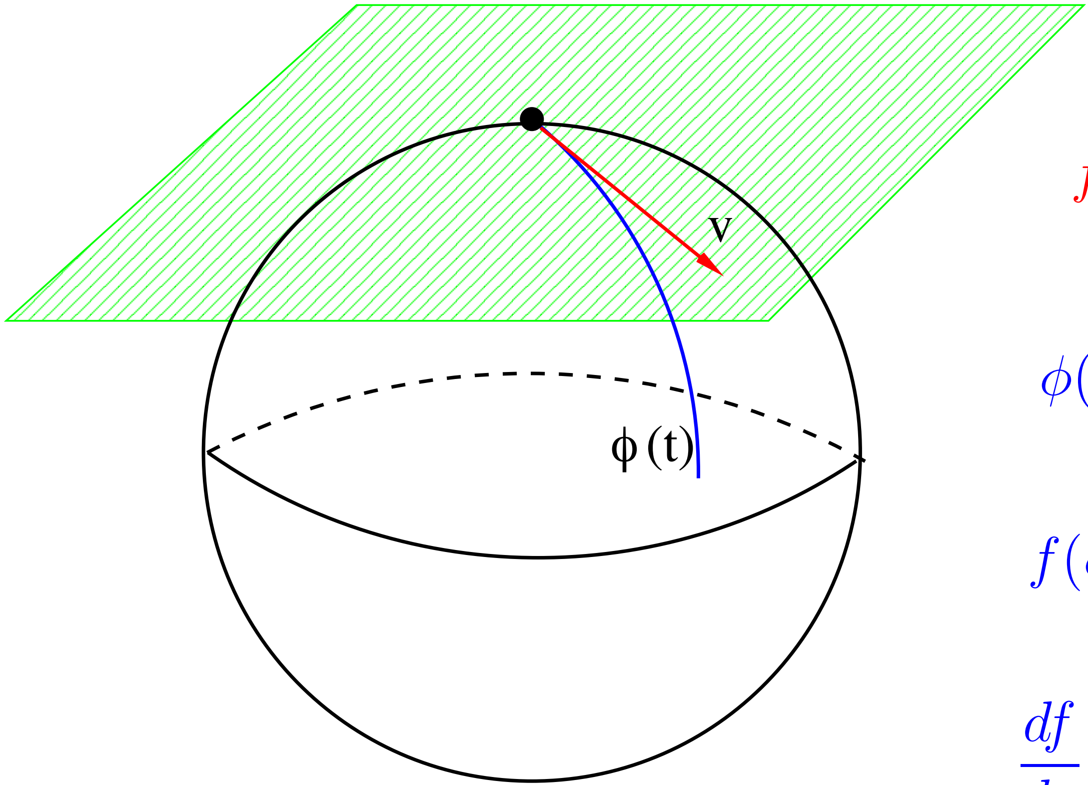
$$f : \mathcal{M}^k \rightarrow \mathbb{R}$$

$$\langle \nabla f, v \rangle \equiv \frac{df}{dv}$$

Tangent vectors <— — —> Directional derivatives.

Gradient points in the direction of maximum change.

Tangent Vectors vs. Derivatives



$$f : \mathcal{M}^k \rightarrow \mathbb{R}$$

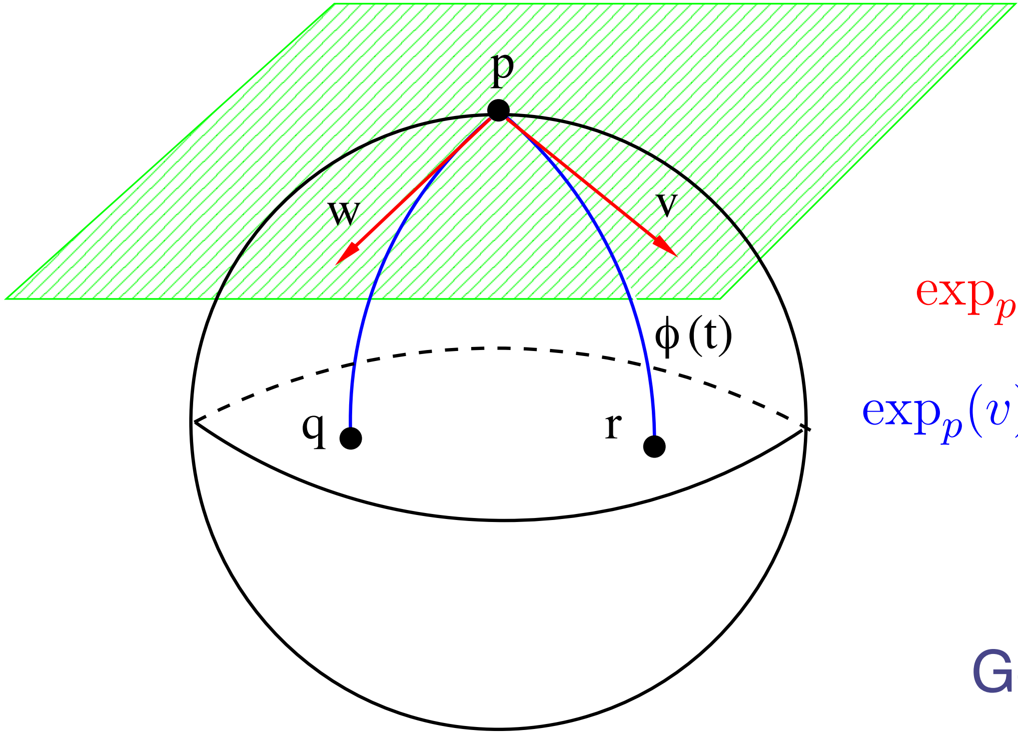
$$\phi(t) : \mathbb{R} \rightarrow \mathcal{M}^k$$

$$f(\phi(t)) : \mathbb{R} \rightarrow \mathbb{R}$$

$$\frac{df}{dv} = \left.\frac{d\,f(\phi(t))}{d\,t}\right|_0$$

Tangent vectors <———> Directional derivatives.

Exponential Maps



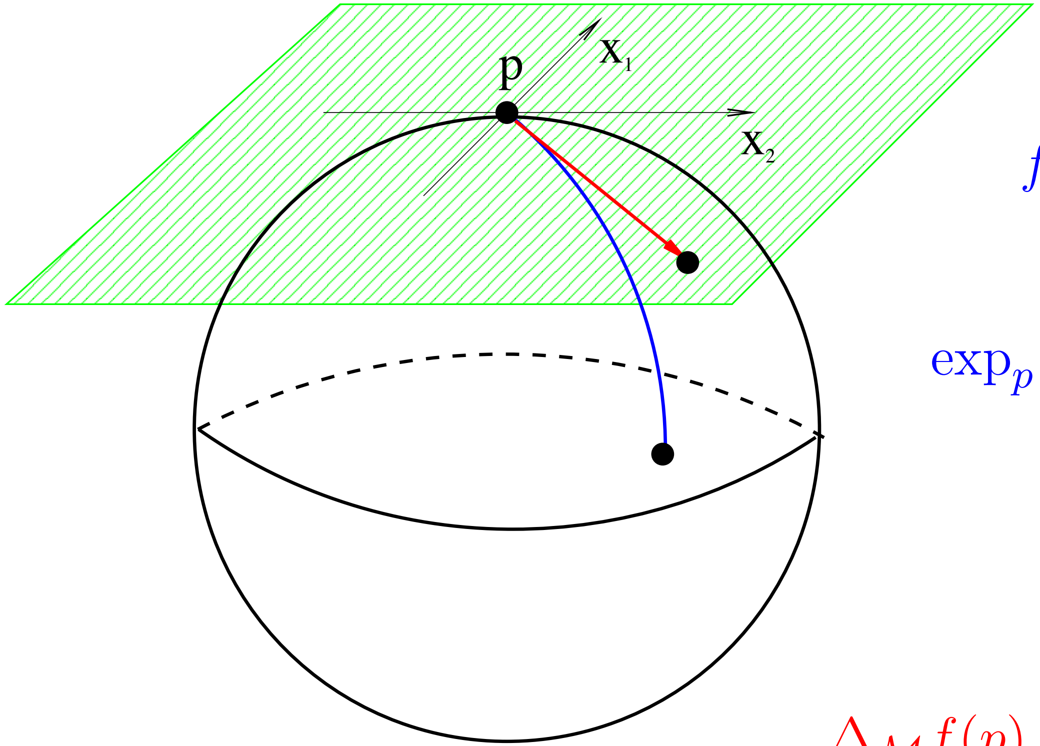
$\exp_p : T_p\mathcal{M}^k \to \mathcal{M}^k$

$\exp_p(v) = r \quad \exp_p(w) = q$

Geodesic $\phi(t)$

$$\phi(0) = p, \quad \phi(\|v\|) = q \qquad \left.\frac{d\phi(t)}{dt}\right|_0 = v$$

Laplacian-Beltrami Operator



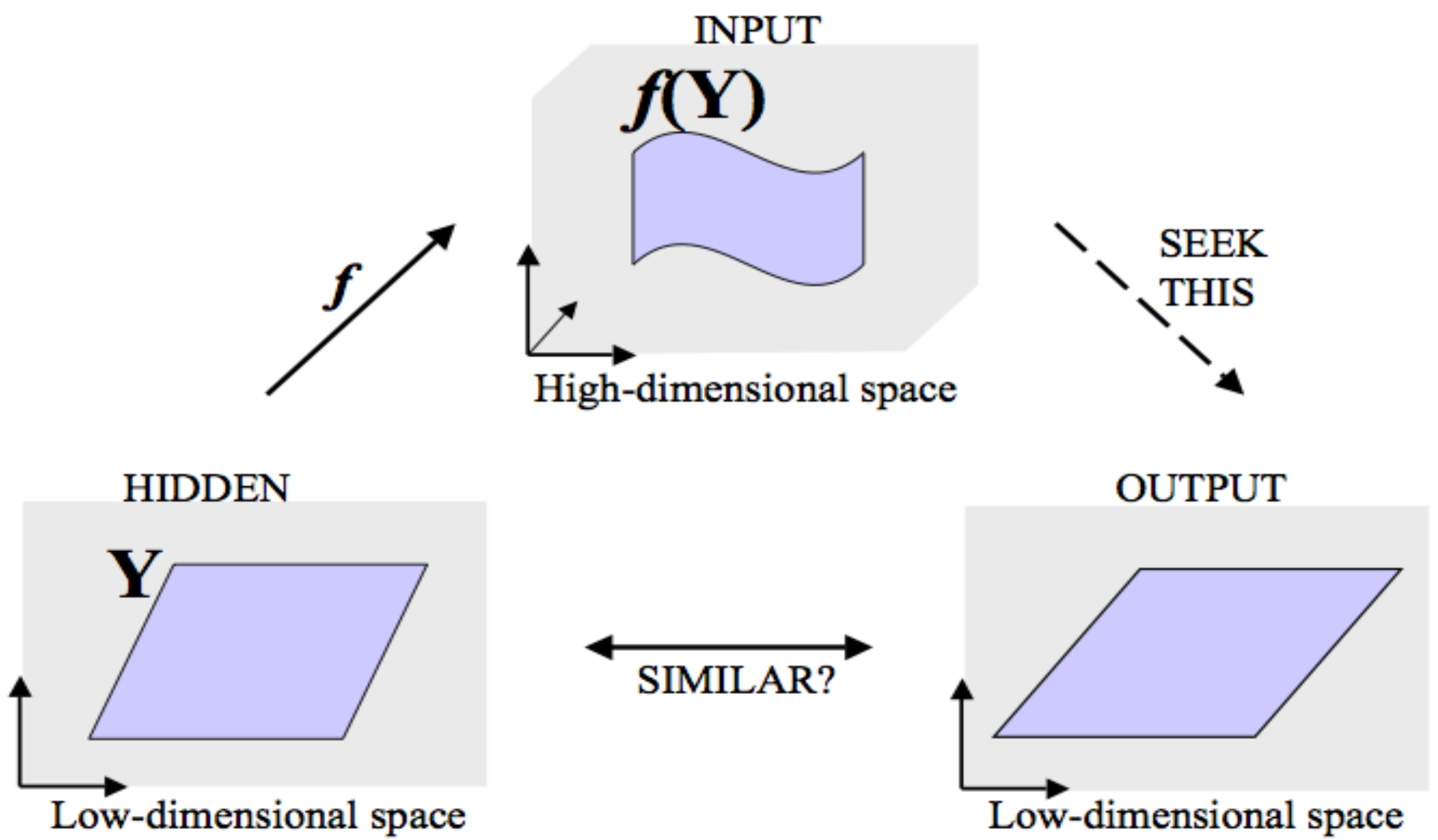
$$f : \mathcal{M}^k \to \mathbb{R}$$

$$\exp_p : T_p\mathcal{M}^k \to \mathcal{M}^k$$

$$\Delta_{\mathcal{M}} f(p) \equiv \sum_i \frac{\partial^2 f(\exp_p(x))}{\partial x_i^2}$$

Orthonormal coordinate system.

Generative Models in Manifold Learning



Spectral Geometric Embedding

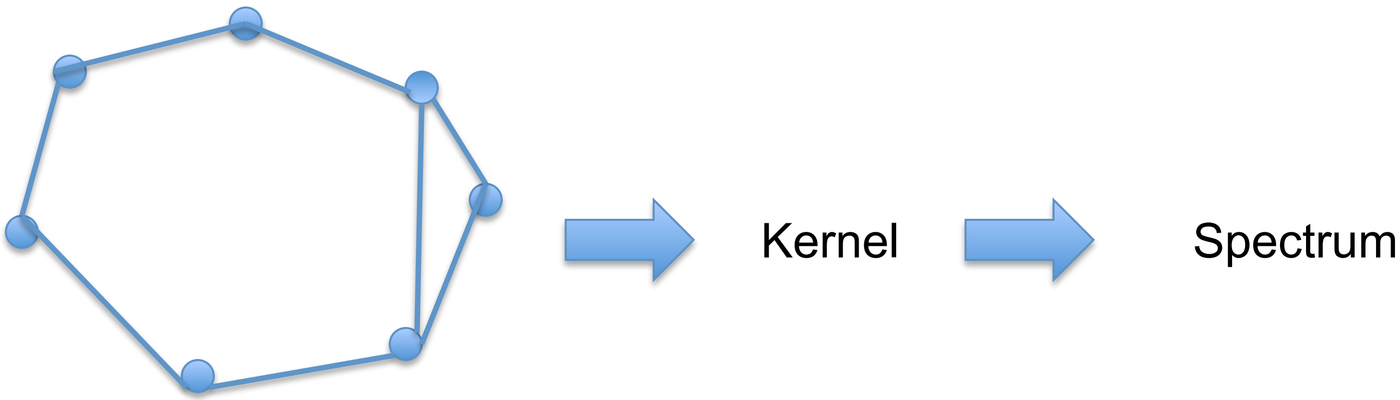
Given $x_1, \ldots, x_n \in \mathcal{M} \subset \mathbb{R}^N$,
Find $y_1, \ldots, y_n \in \mathbb{R}^d$ where $d << N$

- 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)

Meta-Algorithm

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



Recall: ISOMAP

1. Construct Neighborhood Graph.
2. Find shortest path (geodesic) distances.

   D_{ij} is $n \times n$

3. Embed using Multidimensional Scaling.

Recall: MDS

- Idea: Distances -> Inner Products -> Embedding
- Inner Product:

$$\|x-y\|^2 = \langle x,x\rangle + \langle y,y\rangle - 2\langle x,y\rangle$$

$$D_{ij} = Y_{ii} + Y_{jj} - 2Y_{ij}$$

$$\Rightarrow \quad A = -\frac{1}{2}HDH^T, \qquad H = I - \frac{1}{n}11^T$$

- *A* is positive semi-definite with

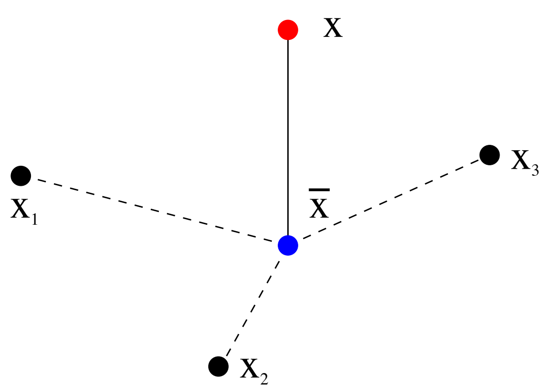
$$A = U\Lambda U^T = YY^T, \qquad Y = U\Lambda^{1/2}$$

Recall: LLE (I)

1. Construct Neighborhood Graph.

2. Let $x_1, \ldots, x_n$ be neighbors of x. Project x to the span of $x_1, \ldots, x_n$.

3. Find barycentric coordinates of $\bar{x}$.



$$\bar{x} = w_1 x_1 + w_2 x_2 + w_3 x_3$$

$$w_1 + w_2 + w_3 = 1$$

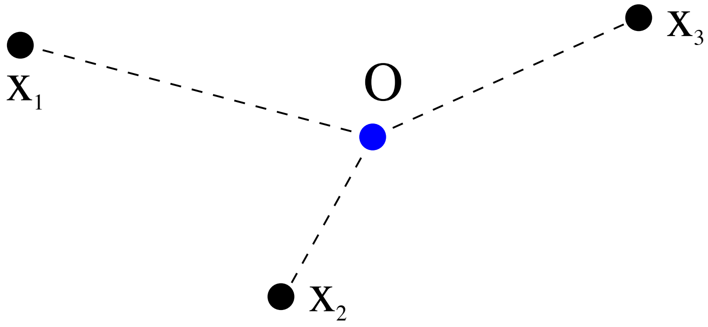
Weights w_1, w_2, w_3 chosen,
so that $\bar{x}$ is the center of mass.

Recall: LLE (II)

4. Construct sparse matrix W. i th row is barycentric coordinates of $\bar{x}_i$ in the basis of its nearest neighbors.

5. Use lowest eigenvectors of $(I - W)^t(I - W)$ to embed.

Laplacian and LLE



$$\sum w_i x_i = 0$$

$$\sum w_i = 1$$

Hessian H. Taylor expansion :

$$f(x_i) = f(0) + x_i^t \nabla f + \frac{1}{2} x_i^t H x_i + o(\|x_i\|^2)$$

$$(I - W)f(0) = f(0) - \sum w_i f(x_i) \approx f(0) - \sum w_i f(0) - \sum_i w_i x_i^t \nabla f - \frac{1}{2} \sum_i x_i^t H x_i =$$

$$= -\frac{1}{2} \sum_i x_i^t H x_i \approx -trH = \Delta f$$

Laplacian Eigenmaps (I) [Belkin-Niyogi]

Step 1 [*Constructing the Graph*]

$$e_{ij} = 1 \Leftrightarrow \mathbf{x}_i \text{ “close to” } \mathbf{x}_j$$

1. ϵ-neighborhoods. [*parameter* $\epsilon \in \mathbb{R}$] Nodes i and j are connected by an edge if

$$||\mathbf{x}_i - \mathbf{x}_j||^2 < \epsilon$$

2. n nearest neighbors. [*parameter* $n \in \mathbb{N}$] Nodes i and j are connected by an edge if i is among n nearest neighbors of j or j is among n nearest neighbors of i.

Laplacian Eigenmaps (II)

Step 2. [*Choosing the weights*].

1. Heat kernel. [*parameter* $t \in \mathbb{R}$]. If nodes i and j are connected, put

$$W_{ij} = e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{t}}$$

2. Simple-minded. [*No parameters*]. $W_{ij} = 1$ if and only if vertices i and j are connected by an edge.

Laplacian Eigenmaps (III)

Step 3. [*Eigenmaps*] Compute eigenvalues and eigenvectors for the generalized eigenvector problem:

$$Lf = \lambda D f$$

D is diagonal matrix where

$$D_{ii} = \sum_j W_{ij}$$

$$L = D - W$$

Let $\mathbf{f}_0, \ldots, \mathbf{f}_{k-1}$ be eigenvectors.

Leave out the eigenvector $\mathbf{f}_0$ and use the next m lowest eigenvectors for embedding in an m-dimensional Euclidean space.

Connection to Markov Chain

- $L = D\text{-}W$
- $P = I - D^{-1}L = D^{-1}W$ is a markov matrix
- v is generalized eigenvector of L: $L\,v = \lambda\,D\,v$
- v is also a right eigenvector of P with eigenvalue $1\text{-}\lambda$
- P is lumpable iff v is piece-wise constant
- So Laplacian eigenmaps have Markov Chain interpretations (Diffusion Map)

Another choice of eigenmaps

- Normalized positive semi-definite Laplacian

$$L_n = D^{-1/2}(D - W)D^{-1/2} = I - D^{-1/2}WD^{-1/2}$$

- ϕ_i is an eigenvector of L_n with eigenvalue λ_i
- Laplacian eigenmap/Diffusion map:

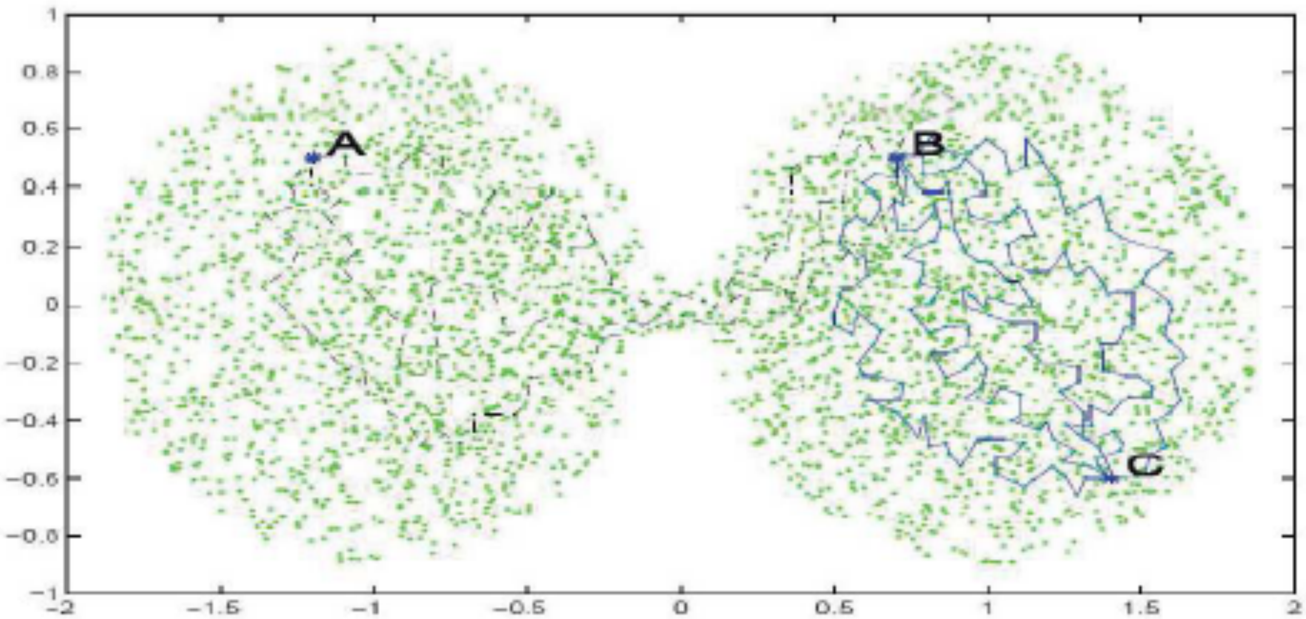
$$Y = \begin{pmatrix} \lambda_1^{1/2}\phi_1 & \lambda_2^{1/2}\phi_2 & \dots & \lambda_d^{1/2}\phi_d \end{pmatrix}$$

Laplacian on Graphs

Given a weighted graph (G, W, E), the combinatorial Laplacian is defined by $L = D - W$, where $(D)_{ii} = \sum_j W_{ij}$, and the normalized Laplacian is defined by

$$\mathcal{L} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}.$$

These are self-adjoint positive-semi-definite operators, let λ_i and ϕ_i be the eigenvalues and eigenvectors. Fourier analysis on graphs. The heat kernel is of course defined by $H_t = e^{-t\mathcal{L}}$; the natural random walk is $D^{-1}W$.



$d_{geod.}(A, B) \sim d_{geod.}(C, B)$, however $d^{(t)}(A, B) >> d^{(t)}(C, B)$.

A simple empirical diffusion matrix A can be constructed as follows

Let X_i represent normalized data ,we “soft truncate” the covariance matrix

as

$$A_0 = [X_i * X_j]_\varepsilon = \exp\{-(1 - X_i * X_j)/\varepsilon\}$$

$$\|X_i\| = 1$$

$A = D^{-1/2} A_0 D^{-1/2}$ ($D = \mathrm{diag}(\mathrm{sum}(A_0,1))$) is a renormalized version of this matrix

The eigenvectors of this matrix provide a local non linear principal component analysis of the data . Whose entries are the diffusion coordinates These are also the eigenfunctions of the discrete Graph Laplace Operator.

$$A = \sum \lambda_l^2 \phi_l(X_i)\phi_l(X_j)$$

$$X_i^{(t)} \rightarrow (\lambda_1{}^t \phi_1(X_i), \lambda_2{}^t \phi_2(X_i), \lambda_3{}^t \phi_3(X_i), ..)$$

This map is a diffusion (at time t) embedding into Euclidean space

Kernel PCA and Diffusion Map

Let k be a positive definite kernel whose restriction to the data set is expanded in eigenfunctions

$$k(x,y) = \sum \lambda_i^{\,2} \varphi_i(x)\varphi_i(y)$$

Let

$$D^2(x,y) = \sum \lambda_i^{\,2} (\varphi_i(x) - \varphi_i(y))^2$$

Then

$$k(x,x) + k(y,y) - 2k(x,y) = D^2(x,y)$$

Clearly D is a distance on the data induced by the Geometric short time Diffusion map

$$\mathrm{x} \in \Gamma \rightarrow \widehat{\mathrm{X}}^{\mathrm{t}}(\mathrm{x}) = \{\lambda_i^{\,t} \varphi_i(x)\} \in \mathrm{l}^2 \ .$$

.

Heat Diffusion Map

- Find Gaussian kernel

$$K_\varepsilon(x,y) = \exp\left(-\frac{\|x-y\|^2}{\varepsilon^2}\right)$$

- Normalize kernel

$$K^{(\alpha)}(x,y) = \frac{K_\varepsilon(x,y)}{p^\alpha(x)p^\alpha(y)} \quad \textit{where} \quad p(x) = \int K_\varepsilon(x,y)d\mu(y)$$

- Renormalized kernel

$$A_\varepsilon(x,y) = \frac{K^{(\alpha)}(x,y)}{\sqrt{d^{(\alpha)}(x)}\sqrt{d^{(\alpha)}(y)}} \quad \textit{where} \quad d^{(\alpha)}(x) = \int K^{(\alpha)}(x,y)d\mu(y)$$

  - α=1, Laplacian-Beltrami operator, separate geometry from density
  - α=0, classical normalized graph Laplacian
  - α=1/2, backward Fokkar-Planck operator

Heat Diffusion Distance

$$H^t = \exp(-tL_n) \quad where \quad L_n = I - D^{-1/2}WD^{-1/2}$$

Heat diffusion operator H^t.

δ_x and δ_y initial heat distributions.

Diffusion distance between x and y:

$$\|H^t\delta_x - H^t\delta_y\|_{L^2}$$

Difference between heat distributions after time t.

Heat Diffusion Maps

Embed using weighted eigenfunctions of the Laplacian:

$$x \to (e^{-\lambda_1 t}\mathbf{f}_1(x), e^{-\lambda_2 t}\mathbf{f}_2(x), \ldots)$$

Diffusion distance is (approximated by) the distance between the embedded points.

Closely related to random walks on graphs.

Justification

Find $y_1, \ldots, y_n \in R$

$$\min \sum_{i,j} (y_i - y_j)^2 W_{ij}$$

Tries to preserve locality

A Fundamental Identity

But

$$\frac{1}{2}\sum_{i,j}(y_i - y_j)^2 W_{ij} = \mathbf{y}^T L \mathbf{y}$$

$$\sum_{i,j}(y_i - y_j)^2 W_{ij} = \sum_{i,j}(y_i^2 + y_j^2 - 2y_i y_j) W_{ij}$$

$$= \sum_i y_i^2 D_{ii} + \sum_j y_j^2 D_{jj} - 2\sum_{i,j} y_i y_j W_{ij}$$

$$= 2\mathbf{y}^T L \mathbf{y}$$

Embedding

$$\lambda = 0 \rightarrow \mathbf{y} = \mathbf{1}$$

$$\min_{\mathbf{y}^T\mathbf{1}=0} \mathbf{y}^T L \mathbf{y}$$

Let $Y = [\mathbf{y}_1 \mathbf{y}_2 \ldots \mathbf{y}_m]$

$$\sum_{i,j} ||Y_i - Y_j||^2 W_{ij} = \text{trace}(Y^T L Y)$$

subject to $Y^T Y = I$.

Use eigenvectors of L to embed.

On the Manifold

smooth map $f : \mathcal{M} \rightarrow R$

$$\int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 \approx \sum_{i \sim j} W_{ij}(f_i - f_j)^2$$

Recall standard gradient in $\mathbb{R}^k$ of $f(z_1, \ldots, z_k)$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial z_1} \\ \frac{\partial f}{\partial z_2} \\ \cdot \\ \cdot \\ \frac{\partial f}{\partial z_k} \end{bmatrix}$$

Stokes Theorem

A Basic Fact

$$\int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 = \int f \cdot \Delta_{\mathcal{M}} f$$

This is like

$$\sum_{i,j} W_{ij}(f_i - f_j)^2 = \mathbf{f}^T \mathbf{L} \mathbf{f}$$

where

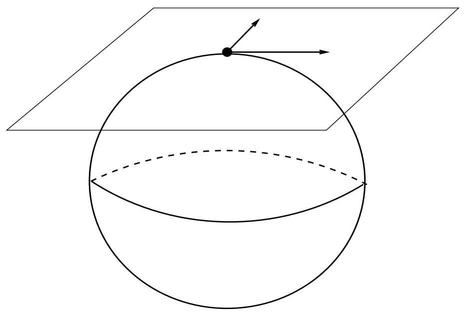
$\Delta_{\mathcal{M}} f$ is the manifold Laplacian

Manifold Laplacian

Recall ordinary Laplacian in $\mathbb{R}^k$
This maps

$$f(x_1, \ldots, x_k) \to \left(- \sum_{i=1}^{k} \frac{\partial^2 f}{\partial x_i^2} \right)$$

Manifold Laplacian is the same on the tangent space.



Manifold Laplacian Eigenvectors

Eigensystem

$$\Delta_{\mathcal{M}} f = \lambda_i \phi_i$$

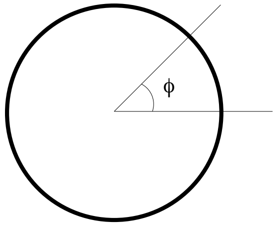
$\lambda_i \geq 0$ and $\lambda_i \to \infty$

$\{\phi_i\}$ form an orthonormal basis for $L^2(\mathcal{M})$

$$\int \|\nabla_{\mathcal{M}} \phi_i\|^2 = \lambda_i$$

Manifold Laplacian is non-compact!

Example: Circle



$$-\frac{d^2u}{dt^2} = \lambda u \text{ where } u(0) = u(2\pi)$$

Eigenvalues are

$$\lambda_n = n^2$$

Eigenfunctions are

$$\sin(nt), \cos(nt)$$

Spherical Harmonics in high-D sphere!

Spectral Growth

$$\lambda_1 \leq \lambda_2 \ldots \leq \lambda_j \leq \ldots$$

Then

$$A + \frac{2}{d}\log(j) \leq \log(\lambda_j) \leq B + \frac{2}{d}\log(j+1)$$

Example: on S^1

$$\lambda_j = j^2 \implies \log(\lambda_j) = \frac{2}{1}\log(j)$$

(Li and Yau; Weyl's asymptotics)

From Graph to Manifolds

$$f : \mathcal{M} \to \mathbb{R} \quad x \in \mathcal{M} \quad x_1, \ldots, x_n \in \mathcal{M}$$

Graph Laplacian:

$$L_n^t(f)(x) = f(x) \sum_j e^{-\frac{\|x - x_j\|^2}{t}} - \sum_j f(x_j) e^{-\frac{\|x - x_j\|^2}{t}}$$

Theorem [pointwise convergence] $t_n = n^{-\frac{1}{k+2+\alpha}}$

$$\lim_{n \to \infty} \frac{(4\pi t_n)^{-\frac{k+2}{2}}}{n} L_n^{t_n} f(x) = \Delta_{\mathcal{M}} f(x)$$

Belkin 03, Lafon Coifman 04, Belkin Niyogi 05, Hein et al 05

From Graph to Manifolds

Theorem [convergence of eigenfunctions]

$$\lim_{t \to 0, n \to \infty} Eig[L_n^{t_n}] \to Eig[\Delta_{\mathcal{M}}]$$

Belkin Niyogi 06

Recall

Heat equation in $\mathbb{R}^n$:

$u(x,t)$ – heat distribution at time t.
$u(x,0) = f(x)$ – initial distribution. $x \in \mathbb{R}^n, t \in \mathbb{R}$.

$$\Delta_{\mathbb{R}^n} u(x,t) = \frac{du}{dt}(x,t)$$

Solution – convolution with the heat kernel:

$$u(x,t) = (4\pi t)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(y) e^{-\frac{\|x-y\|^2}{4t}} dy$$

Proof Idea
(pointwise convergence)

Functional approximation:

Taking limit as $t \to 0$ and writing the derivative:

$$\Delta_{\mathbb{R}^n} f(x) = \frac{d}{dt}\left[(4\pi t)^{-\frac{n}{2}} \int_{\mathbb{R}^n} f(y) e^{-\frac{\|x-y\|^2}{4t}} dy\right]_0$$

$$\Delta_{\mathbb{R}^n} f(x) \approx -\frac{1}{t}(4\pi t)^{-\frac{n}{2}} \left(f(x) - \int_{\mathbb{R}^n} f(y) e^{-\frac{\|x-y\|^2}{4t}} dy\right)$$

Empirical approximation:

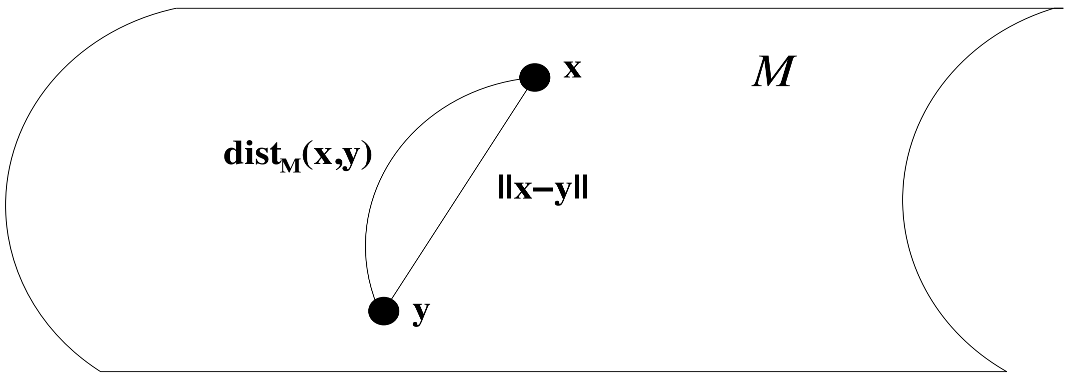
Integral can be estimated from empirical data.

$$\Delta_{\mathbb{R}^n} f(x) \approx -\frac{1}{t}(4\pi t)^{-\frac{n}{2}} \left(f(x) - \sum_{x_i} f(x_i) e^{-\frac{\|x-x_i\|^2}{4t}}\right)$$

Some Difficulties

Some difficulties arise for manifolds:

- Do not know distances.
- Do not know the heat kernel.



Careful analysis needed.

The Heat Kernel Approximation

- $H_t(x,y) = \sum_i e^{-\lambda_i t}\phi_i(x)\phi_i(y)$
- in $\mathbb{R}^d$, closed form expression

$$H_t(x,y) = \frac{1}{(4\pi t)^{d/2}} e^{-\frac{\|x-y\|^2}{4t}}$$

- Goodness of approximation depends on the gap

$$\left| H_t(x,y) - \frac{1}{(4\pi t)^{d/2}} e^{-\frac{\|x-y\|^2}{4t}} \right|$$

- H_t is a Mercer kernel intrinsically defined on manifold. Leads to SVMs on manifolds.

Three remarks on noises

1. Arbitrary probability distribution on the manifold: convergence to weighted Laplacian.
2. Noise off the manifold:
   $\mu = \mu_{\mathcal{M}^d} + \mu_{\mathbb{R}^N}$
   Then
   $$\lim_{t \to 0} L^t f(x) = \Delta f(x)$$
3. Noise off the manifold:
   $$z = x + \eta \; (\sim N(0, \sigma^2 I))$$
   We have
   $$\lim_{t \to 0} \lim_{\sigma \to 0} L^{t,\sigma} f(x) = \Delta f(x)$$

General Diffusion Map

- P.S.D. Radial basis kernel $K_\varepsilon(x,y) = h\left(\frac{\|x-y\|^2}{\varepsilon^2}\right)$
- Normalize kernel

$$K^{(\alpha)}(x,y) = \frac{K_\varepsilon(x,y)}{p^\alpha(x)p^\alpha(y)} \quad \textit{where} \quad p(x) = \int K_\varepsilon(x,y)d\mu(y)$$

- Markov kernel

$$a_\varepsilon^{(\alpha)}(x,y) = \frac{K^{(\alpha)}(x,y)}{d^{(\alpha)}(x)} \quad \textit{where} \quad d^{(\alpha)}(x) = \int K^{(\alpha)}(x,y)d\mu(y)$$

- Diffusion Operator:

$$A_\varepsilon^{(\alpha)} f(x) = \int a_\varepsilon^{(\alpha)}(x,y) f(y) p(y) dy, \qquad p(x) = \frac{\exp(-U(x))}{Z}$$

$$\Delta_\varepsilon^{(\alpha)} = \frac{I - A_\varepsilon^{(\alpha)}}{\varepsilon}$$

Convergence of Diffusion Map [Coifman et al. 2005]

- Uniform sampling: Laplacian eigenmap converges to Laplacian-Beltrami operators [Belkin-Niyogi]
- Nonuniform sampling with $p(x)$
  - α=1: $\Delta_\varepsilon^{(1)} = \frac{I - A_\varepsilon^{(1)}}{\varepsilon} = \Delta_0 + O(\varepsilon^{1/2})$ where Δ_0 is Laplacian-Beltrami operator on Riemannian manifolds
  - α=1/2: backward Fokkar-Planck operator
  - α=0: classical normalized graph laplacian

Two Assumptions on ISOMAP

(ISO1) *Isometry.* The mapping ψ preserves geodesic distances. That is, define a distance between two points m and m' on the manifold according to the distance travelled by a bug walking along the manifold M according to the shortest path between m and m'. Then the isometry assumption says that

$$G(m, m') = |\theta - \theta'|, \qquad \forall m \leftrightarrow \theta, m' \leftrightarrow \theta',$$

where $|\cdot|$ denotes Euclidean distance in $\mathbb{R}^d$.

(ISO2) *Convexity.* The parameter space Θ is a convex subset of $\mathbb{R}^d$. That is, if θ, θ' is a pair of points in Θ, then the entire line segment $\{(1-t)\theta + t\theta' : t \in (0,1)\}$ lies in Θ.

Convexity is hard to meet: consider two balls in an image which never intersect, whose center coordinate space (x_1,y_1,x_2,y_2) must have a hole.

Relaxations (Donoho-Grimes'2003)

(LocISO1) *Local Isometry.* In a small enough neighborhood of each point m, geodesic distances to nearby points m' in M are identical to Euclidean distances between the corresponding parameter points θ and θ'.

(LocISO2) *Connectedness.* The parameter space Θ is a open connected subset of $\mathbb{R}^d$.

Summary of Laplacian LLE

- Summary
  - Build graph from K Nearest Neighbors.
  - Construct weighted adjacency matrix with Gaussian kernel.
  - Compute embedding from normalized Laplacian.
    - $\text{minimize } \int \|\nabla f\|^2 dx \text{ subject to } \|f\| = 1$
- Predictions
  - Assumes each point lies in the convex hull of its neighbors. So it might have trouble at the boundary.
  - Will have difficulty with non-uniform sampling.

Hessian LLE

- Summary
  - Build graph from K Nearest Neighbors.
  - Estimate tangent Hessians.
  - Compute embedding based on Hessians.

$$f : X \to \Re \qquad Basis\left(null\left(\int \left\| H_f(x) \right\| dx\right)\right) = Basis(X)$$

- Predictions
  - Specifically set up to handle non-convexity.
  - Slower than LLE & Laplacian.
  - Will perform poorly in sparse regions.
  - Only method with convergence guarantees.

Note that: $\Delta(f) = trace\left(H(f)\right)$

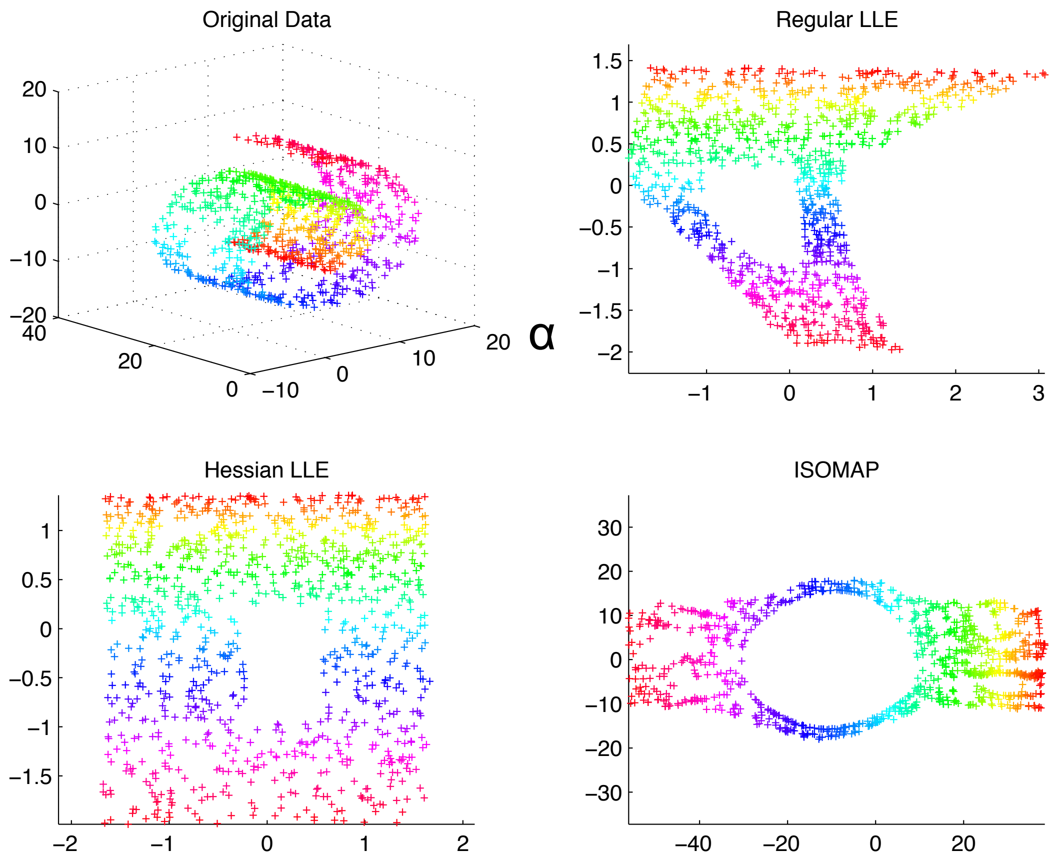
Convergence of Hessian LLE (Donoho-Grimes)

Theorem 1 *Suppose $M = \psi(\Theta)$ where Θ is an open connected subset of $\mathbb{R}^d$, and ψ is a locally isometric embedding of Θ into $\mathbb{R}^n$. Then $\mathcal{H}(f)$ has a $d+1$ dimensional nullspace, consisting of the constant function and a d-dimensional space of functions spanned by the original isometric coordinates.*

We give the proof in Appendix A.

Corollary 2 *Under the same assumptions as Theorem 1, the original isometric coordinates θ can be recovered, up to a rigid motion, by identifying a suitable basis for the null space of $\mathcal{H}(f)$.*

Comparisons on Swiss Roll with holes



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.
- M. Hein, J.-Y. Audibert, and U. von Luxburg. From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians. In P. Auer and R. Meir, editors, Proc. of the 18th Conf. on Learning Theory (COLT), pages 486–500, Berlin, 2005. Springer.
- Ulrike von Luxburg, Mikhail Belkin, and Olivier Bousquet. Consistency of spectral clustering. Ann. Statist. Volume 36, Number 2 (2008), 555-586.

Acknowledgement

- Slides stolen from M. Belkin, R. Coifman, G. Lerman, et al.