Diffusion map

problem setup: a finite dataset $\{x^{(i)}\}_{i=1}^{N} \in R^D$

Two viewpoints:

1: $\vec{x}^{(i)}$ are sampled from arbitrary distribution. $\pi(\vec{x})$

$\pi(\vec{x})$ in a compact domain $\Omega \subset R^D$

we can define the potential

$U(\vec{x}) = -\log \pi(\vec{x})$ so $\pi(\vec{x}) = \exp(-U(\vec{x}))$

2: $\vec{x}^{(i)}$ are sampled from a stochastic dynamical system from equilibrium (i.e. order matters)

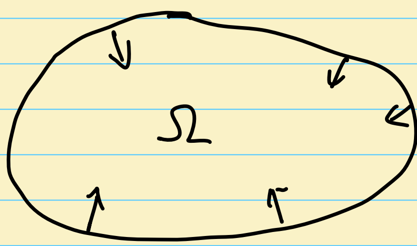
$\vec{x}^{(1)}, \vec{x}^{(2)} \cdots, \vec{x}^{(N)}$

Assume $\vec{X}_t \in \Omega$ at time t

is the sample path of the following SDE.

Langevin equation: $d\vec{X}_t = -\nabla U(\vec{X}_t)\,dt + \sqrt{2}\,dW_t$ with reflecting BC at $\partial\Omega$

W_t: D-dim Brownian Motion



Another way to describe SDE to use transition probability density $P(\vec{x}, t \mid \vec{y}, s)$ with $t > s$ of finding the system at location $\vec{x}$ at time t given initial location $\vec{y}$ at s.

Transition Probability density satisfies the following PDE: (Fokker-Planck eq)

$\nabla \cdot (\nabla P)$
$= \Delta P$
is the Laplacian operator

$$\frac{\partial P}{\partial t} = \underbrace{\nabla \cdot P \nabla U(\vec{x}) + \nabla \cdot \nabla P}_{\triangleq L_F P}$$

$L_F P$ → infinitesimal generator of LF operator

with IC $\lim_{t \to s^+} P(\vec{x}, t | \vec{y}, s) = \delta(\vec{x} - \vec{y})$

Treat $P(\vec{x}, t | \vec{y}, s)$ $s \le t$
→ backward variable

it satisfies the backward eq.

$$-\frac{\partial P}{\partial s} = \underbrace{-\nabla P \cdot \nabla U(\vec{y}) + \nabla \cdot \nabla P}_{\triangleq L_B P}$$

$L_B P$ → infinitesimal generator of backward operator

The steady state solution (equilibrium)

$$\lim_{t \to +\infty} P(\vec{x}, t | \vec{y}, s) = \frac{\exp(-U(\vec{x}))}{Z} \triangleq \vec{\pi}(x)$$

$$Z = \int_{\Omega} \exp(-U(\vec{x}))\, d\vec{x}$$

If $Z \neq 1$, we can shift $U(\vec{x})$ by a constant to make $Z = 1$

If we don't consider the time dependence of the data in both scenarios, The steady state density is the same for ① and ②

Only the potential $U(\vec{x})$ and the geometry Ω matters.

Note: L_F is $\boxed{\text{self-adjoint operator}}$ with the inner product

$$\langle f, g\rangle_{1/\pi} = \int_\Omega \frac{f(\vec{x})g(\vec{x})}{\pi(\vec{x})}d\vec{x}, \quad f(\vec{x}), g(\vec{x}) \in L^1$$

L_B is $\boxed{\text{self-adjoint operator}}$ with the inner product.

$$\langle f, g\rangle_\pi = \int_\Omega f(\vec{x})g(\vec{x})\pi(\vec{x})d\vec{x},$$

for. $f(\vec{x}), g(\vec{x}) \in L^\infty$

The solution of the Fokker-Planck eq is

$$P(\vec{x},t) = \sum_{j=0}^{\infty} a_j \exp(-\lambda_j t)\varphi_j(\vec{x})$$

$\varphi_0(\vec{x}) = \vec{\pi}(\vec{x})$

where $-\lambda_j$ are spectrum of FP operator and $\varphi_j(\vec{x})$ are corresponding eigenfunctions

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots$$

The solution of the backward eq is

$$p(\vec{y}, s) = \sum_{j=0}^{\infty} b_j \exp(-\lambda_j t)\phi_j(\vec{y})$$

$\phi_0(\vec{y}) = \vec{1}$ // eigenvalues of L_F and L_B are the same

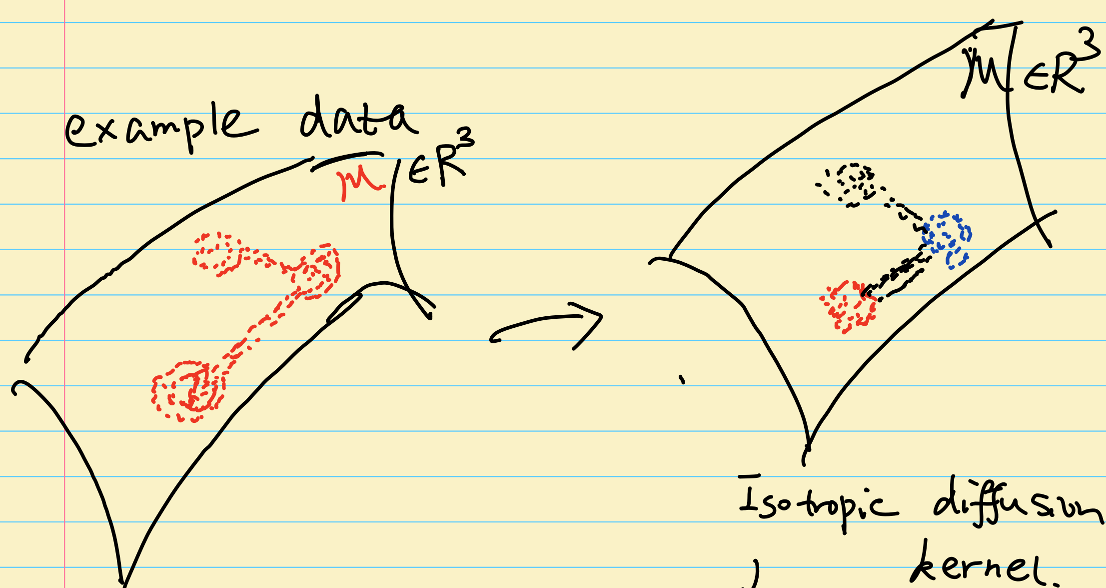
and eigenfunctions are $\boxed{\phi_j = \varphi_j / \pi.}$

Long time behavior of the system is governed by first few eigenfunctions, $\varphi_0, \varphi_1, \cdots \varphi_k$

and $\lambda_k >> \lambda_{k+1}$ spectrum gap.

or $\phi_0, \phi_1, \cdots \phi_k$.

example data



Isotropic diffusion kernel.

Diffusion Map

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

Assume the transition probability between $x^{(i)}$ and $x^{(j)}$ is proportional to $K_\varepsilon(\vec{x}^{(i)}, \vec{x}^{(j)})$

and construct $N \times N$ Markov Matrix

$$P_{ij} = \frac{K_\varepsilon(\vec{x}^{(i)}, \vec{x}^{(j)})}{\sum_j K_\varepsilon(\vec{x}^{(i)}, \vec{x}^{(j)})} \qquad \sum_j K_\varepsilon(\vec{x}^{(i)}, \vec{x}^{(j)}) = P_\varepsilon(x^{(i)})$$

normalization constant

interpretation

$$\Downarrow \quad \triangleq Pr(S_\varepsilon = \vec{x}^{(j)} \mid S_0 = \vec{x}^{(i)})$$

Note $\sum_j P_{ij} = 1$

$$P_{ij}^{t} \triangleq Pr(S_{\varepsilon t} = \vec{x}^{(j)} \mid S_0 = \vec{x}^{(i)})$$

ε : has dual interpretation.

unit (m^2) ① ε square radius of neighbourhood to infer local geometry and density information.

unit (s) ② discrete time step ε at random walk jumps from point to point.

$$\varepsilon = \frac{\varepsilon \rightarrow \text{squared distance}}{1 \rightarrow \text{diffusion coeffient } (m^2/s)}$$

time (s)

If ε is large enough, M is fully connected in numerical sense.

$\Rightarrow \lambda_0 = 1$ is simple.

Analog to continuous case.

a row probability vector $\vec{v}_0 \in L^1$,

$\vec{v}_1 = \vec{v}_0 P$ describe the evolution of the probability at one step.

row vector

$(P, \vec{v}) : \xrightarrow{\text{representation}}$ FP operator of MC

a column vector $\vec{u}_0 \in L^\infty$.

$\vec{u}_1 = P \vec{u}_0$ backward operator of MC.

$(P, \vec{u}) \xrightarrow{\text{representation}}$

Denote $U = [\vec{u}_1^T, \cdots \vec{u}_N^T]$ and

$V = [\vec{V}_1, \cdots \vec{V}_N]$ are left and right eigenvectors of P.

Note $\vec{u}_1 = \vec{\pi}/\sqrt{N}$ $\vec{V}_1 = \vec{1}/\sqrt{N}$ Moreover $\vec{u}_j^T / \vec{\pi} = \vec{V}_j$

$$P = V \Lambda U^T \quad \text{with } V U^T = I.$$

$\langle \vec{V}_i, \vec{u}_j \rangle = \delta_{ij}$ $\langle \vec{u}_i, \vec{u}_j \rangle_{D^{-1}}$

At time t: $P^t = V \Lambda^t U^T$

Ex: $\vec{u}_i^T \begin{bmatrix} 1/d_1 & & \\ & \ddots & \\ & & 1/d_N \end{bmatrix} \vec{u}_j = \delta_{ij}$

$$\boxed{P_{ij}^t = \sum_{k=1}^{N} \lambda_k^t \vec{v}_k(i) \vec{u}_k(j)}$$

$\vec{V}_i^T \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_N \end{bmatrix} \vec{V}_j = \delta_{ij}$ $\langle \vec{V}_i, \vec{V}_j \rangle_D$.

In diffusion map

all equal.

$$\vec{X} \to (\vec{Z})^{(t)} \doteq V \Lambda^t = [\boxed{\lambda_1^t \vec{V}_1}, \lambda_2^t \vec{V}_2, \cdots \lambda_N^t \vec{V}_N]$$

Lemma: The diffusion distance is equal to ℓ^2 distance between two probability vector P_{i*}^t, P_{j*}^t with the weight D^{-1}.

$$D_t^2(i,j) \triangleq \| Z_t^{(i)} - Z_t^{(j)} \|_2^2$$

$$= \| \underset{\Downarrow}{P_{i*}^t} - \underset{\Downarrow}{P_{j*}^t} \|_{D^{-1}}^2.$$

$$p(t, x^{(k)} | x^{(i)}) \quad p(t, x^{(k)} | x^{(j)}).$$

Proof: $$\| P_{i*}^t - P_{j*}^t \|_{1/\pi}^2 = \sum_{\ell=1}^{N} (P_{i\ell}^t - P_{j\ell}^t)^2 \cdot \frac{1}{d_L}$$

$$= \sum_{\ell=1}^{N} \left[\sum_{k=1}^{N} \lambda_k^t \vec{v}_k(i) \vec{u}_k(\ell) - \lambda_k^t \vec{v}_k(j) \vec{u}_k(\ell) \right]^2 \cdot \frac{1}{d_L}.$$

$$= \sum_{k=1}^{N} \left(\lambda_k^t\right)^2 \left[\vec{v}_k(i) - \vec{v}_k(j)\right]^2 \boxed{\sum_{\ell=1}^{N} \frac{\vec{u}_k^{\,2}(\ell)}{d_\ell}} = 1$$

$$= \sum_{k=1}^{N} \left(\lambda_k^t\right)^2 \left[\vec{v}_k(i) - \vec{v}_k(j)\right]^2.$$

$$= \| Z_t^{(i)} - Z_t^{(j)} \|_2^2.$$

Define truncated distance $D_t^{\delta}(i,j)$

$$= \left[\sum_{k:\ |\lambda_k|^t > \delta} \lambda_k^{2t} \left(v_k(i) - v_k(j) \right)^2 \right]^{1/2}.$$

Lemma:

$$D_t^2(i,j) - \frac{2\delta^2}{d_{min}}(1-\delta_{ij}) \leq \left[D_t^{\delta}(i,j) \right]^2 \leq D_t^2(i,j)$$

where $d_{min} = \min d_i$

proof: $\langle \vec{V}_k, \vec{V}_j \rangle_D = \delta_{kj}.$

$$\Rightarrow V^T D V = I. \Rightarrow (D^{1/2} V)^T (D^{1/2} V) = I.$$

$$\Rightarrow (D^{1/2} V)(D^{1/2} V)^T = I$$

$$\Rightarrow D^{1/2} V V^T D^{1/2} = I.$$

$$\Rightarrow V V^T = D^{-1}.$$

$$\sum_{k=1}^{N} \vec{v}_k(i) \vec{v}_k(j) = \frac{\delta_{ij}}{d_i}$$

$$\sum_{k=1}^{N}\left(\vec{V}_k(i)-\vec{V}_k(j)\right)^2$$

$$= \sum_{k=1}^{N}\left(\vec{V}_k(i)\right)^2 + \sum_{k=1}^{N}\left(\vec{V}_k(j)\right)^2 - 2\sum_{k=1}\vec{V}_k(i)\,\vec{V}_k(j)$$

$$= \frac{1}{d_i} + \frac{1}{d_j} - \frac{2\delta_{ij}}{d_i} \le \frac{2}{d_{min}} - \frac{2\delta_{ij}}{d_{min}}$$

$$= \frac{2}{d_{min}}(1-\delta_{ij})$$

$$[D_t^{\delta}(i,j)]^2 = D_t(i,j) - \sum_{k:|\lambda_k|^t<\delta} \lambda_k^{2t}\left(\vec{V}_k(i)-\vec{V}_k(j)\right)^2$$

$$\geq D_t(i,j) - \delta^2 \sum_{k:|\lambda_k|^t<\delta}\left(\vec{V}_k(i)-\vec{V}_k(j)\right)^2$$

$$\geq D_t(i,j) - \delta^2 \sum_{k=1}^{N}\left(\vec{V}_k(i)-\vec{V}_k(j)\right)^2.$$

$$\geq D_t(i,j) - \frac{2\delta^2}{d_{min}}(1-\delta_{ij}).$$

proposition: $D_t(i,j)=0$ with $X_i \neq X_j$

occurs $\Longleftrightarrow$ node i and j

have the exact same neighbors

and proportional weights,

i.e. $W_{ik} = \alpha W_{jk}$ $\alpha > 0$ forall $k \in V$.

proof: $\Rightarrow$ if $D_t(i,j) = 0$,

then $\sum_{k=1}^{N} (\lambda_k^t)^2 (\vec{v}_k(i) - \vec{v}_k(j))^2 = 0$

forall k either ① or ②

① $\lambda_k^t \neq 0$, and $\vec{v}_k(i) = \vec{v}_k(j)$

② $\lambda_k^t = 0 \Rightarrow \lambda_k = 0$

This implies $t=1$, $D_1(i,j) = 0$

Note $D_1(i,j) = \| P_{i*} - P_{j*} \|_{D^{-1}} = 0$

$\|\cdot\|_{D^{-1}}$ is a norm so $P_{i*} = P_{j*}$

Then $\frac{W_{ik}}{d_i} = \frac{W_{jk}}{d_j}$ for all $k \in V$.

define $\alpha = d_i / d_j$, then $W_{ik} = \alpha W_{jk}$.

$\Leftarrow$ If $P_{i*} = P_{j*}$, then.

$\sum_{k=1}^{N} (P_{ik} - P_{jk})^2 / d_k = 0$

then $D_1(i,j) = 0$

then $\sum_{k=1}^{N} \lambda_k^2 (\vec{v}_k(i) - \vec{v}_k(j))^2 = 0$

So for $\lambda_k \neq 0$, we have $\vec{v}_k(i) = \vec{v}_k(j)$.

Then we can conclude.

$$\sum_{k=1}^{N} (\lambda_k^2)^t (\vec{v}_k(i) - \vec{v}_k(j))^2 = 0$$

case	operator	stochastic process.
$\varepsilon > 0$ N finite $\downarrow$	finite $N \times N$ Matrix with left and right vectors	Random walk discrete in space and discrete in time.
$\varepsilon > 0$ $N \to \infty$ $\downarrow$	operators T_f, T_b.	Random walk. continuous space discrete in time.
$\varepsilon \to 0$ $N \to \infty$	infinitestimal gerator. L_f, L_b.	diffusion process continuous. in time and space.

$\hookrightarrow$ fixed ε and $N \to \infty$

$$\mu(x) = \exp(-U(x))$$

local probability density estimate around x

$$d_i = \sum_{j=1}^{N} W_{ij}$$

Normalization factor is

$$P_\varepsilon(y) = \int_\Omega k_\varepsilon(x,y)\,\mu(x)\,dx$$

The forward transition probability

$$M_f(x|y) = P_r(X(t+\varepsilon)=x \mid X(t)=y)$$

$$= \frac{k_\varepsilon(x,y)}{P_\varepsilon(y)}. \Rightarrow M_{ij} = \frac{W_{ij}}{d_i}$$

A symmetric kernel $M_s(x,y)$.

$$M_s(x,y) = \frac{k_\varepsilon(x,y)}{\sqrt{P_\varepsilon(x)\,P_\varepsilon(y)}}.$$

Define forward, backward, symmetric (FP) operator on the function space.

① $T_f[\varphi](x) = \int_\Omega M_f(x|y)\,\varphi(y)\,\mu(y)\,dy$

② $T_b[\psi](x) = \int_\Omega M_f(y|x)\,\psi(y)\,\mu(y)\,dy$

③ $T_s[\varphi](x) = \int_\Omega M_s(x,y)\,\varphi(y)\,\mu(y)\,dy$

① $\varphi(x) \in L^1$: probability distribution of finding system at location x.

$T_f^m[\varphi]$ at $t = m\varepsilon$.

$T_f[\varphi](x)$: the evolution of probability distribution at $t = \varepsilon$.

$T_b^m[\psi]$

② $\psi \in L^\infty$

$T_b[\psi](x)$: the average value of that function ψ at time ε for the random walk that started x.

By definition. T_f and T_b are adjoint under the inner product with weight μ.

$$\langle T_f \varphi, \psi \rangle_\mu = \langle \varphi, T_b \psi \rangle_\mu$$

T_s is self-adjonit.

$$\langle T_s \varphi, \psi \rangle_\mu = \langle \varphi, T_s \psi \rangle_\mu$$

① T_s, T_f, T_b have the same eigenvalues

② If $T_s \varphi_s = \lambda \varphi_s$

the corresponding eigenfunctions for T_f and T_b are $\varphi_f = \varphi_s \sqrt{P_\varepsilon}$ and

$$\varphi_b = \varphi_s / \sqrt{P_\varepsilon}.$$

③ $\sqrt{P_\varepsilon}$ is the first eigenfunction for T_s. with $\lambda_0 = 1$.

Then for T_f, it is P_ε.

for T_b it is 1.