# The Koopman Operator, Diffusion Maps, and Partially Known Dynamics

UCLA
IPAM
Workshop: Operator Theoretic Methods in Dynamical Data Analysis and Control
02/11/2019

Ryan Mohr, Igor Mezic

AIMdyn, Inc., Santa Barbara, CA 93101

# Outline

1. Mathematical background
    1. The Koopman operator
    2. Diffusion Maps
2. Integral operator approximations to Koopman
    1. RKHS
    2. Invariance
3. Converge to the Koopman operator
    1. mu-sumability
    2. Pointwise convergence
4. Partial dynamics
5. The unscrambling problem

# Mathematical Setting

$\mathbb{X} \subset \mathbb{R}^n$ Compact, connected

$\Phi : \mathbb{X} \to \mathbb{X}$ Discrete time map, invertible

$\mu$ Borel probability measure on X

$$\forall \delta > 0, \inf\{\mu(B_\delta(w)) : w \in \mathbb{X}\} > 0$$

# Mathematical Setting

$\mathbb{X} \subset \mathbb{R}^n$ Compact, connected

$\Phi : \mathbb{X} \to \mathbb{X}$ Discrete time map, invertible

$\mu$ Borel probability measure on X

$\forall \delta > 0, \inf\{\mu(B_\delta(w)) : w \in \mathbb{X}\} > 0$

---

Koopman (composition) operator

$\mathbb{F}$ Function space invariant under the dynamics

$C_\Phi : \mathbb{F} \to \mathbb{F}$

$(C_\Phi f)(x) = (f \circ \Phi)(x) = f(\Phi(x))$ Composition operator

# Diffusion Maps

$\mathbb{F}$

Function space invariant under the diffusion below

$$a : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$$
$$a(x,y) = a(y,x)$$
$$a(x,y) \geq 0$$

Similarity (affinity) kernel (e.g.) a gaussian function

$$m(x) = \int_{\mathbb{X}} a(x,y) d\mu(u)$$

“mass” at x

$$d(x,y) = \frac{a(x,y)}{m(x)}$$

Diffusion kernel

$$(Df)(x) = \int_{\mathbb{X}} d(x,y) f(y) d\mu(y)$$

Diffusion operator

# Integral approximations of the Koopman Operator

- Goal: Approximate the Koopman operator with an integral operator
- Take inspiration from Diffusion Maps

$$g_\sigma(x, y) = \exp(-\|x - y\|^2/\sigma^2)$$

Similarity (affinity) kernel

# Integral approximations of the Koopman Operator

- Goal: Approximate the Koopman operator with an integral operator
- Take inspiration from Diffusion Maps

$$g_\sigma(x,y) = \exp(-\|x-y\|^2/\sigma^2)$$

Similarity (affinity) kernel

$$\phi_\sigma(x,y) = \frac{g_\sigma(\Phi(x),y)}{m_\sigma(\Phi(x))}$$

$$m_\sigma(\Phi(x)) = \int_{\mathbb{X}} g_\sigma(\Phi(x),y)\,d\mu(y)$$

Asymmetric gaussian

# Integral approximations of the Koopman Operator

- <u>Goal</u>: Approximate the Koopman operator with an integral operator
- Take inspiration from Diffusion Maps

$$g_\sigma(x,y) = \exp(-\|x-y\|^2/\sigma^2)$$

Similarity (affinity) kernel

$$\phi_\sigma(x,y) = \frac{g_\sigma(\Phi(x),y)}{m_\sigma(\Phi(x))}$$

Koopman kernel

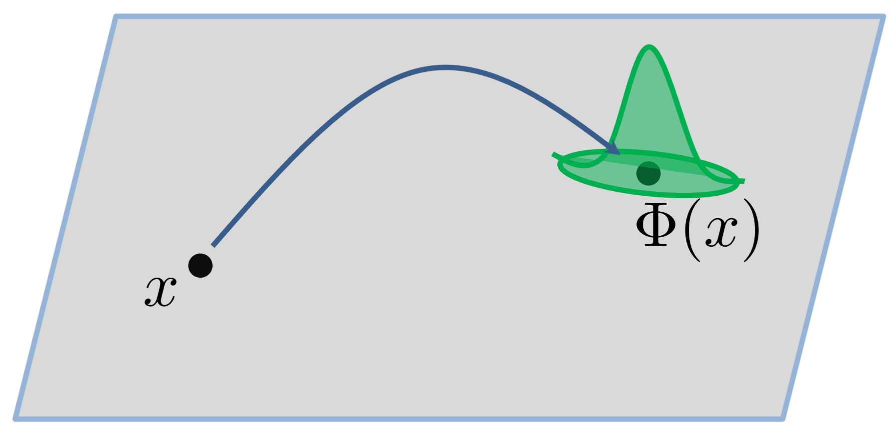
$$m_\sigma(\Phi(x)) = \int_{\mathbb{X}} g_\sigma(\Phi(x),y)\,d\mu(y)$$

$$\boxed{(A_\sigma f)(x) = \int_{\mathbb{X}} \phi_\sigma(x,y) f(y)\,d\mu(y)}$$

Integral Koopman

**Aside**: $A_\sigma : L^2(\mu) \to L^2(\mu)$

Is a Hilbert-Schmidt integral operator



# Reproducing Kernel Hilbert Space

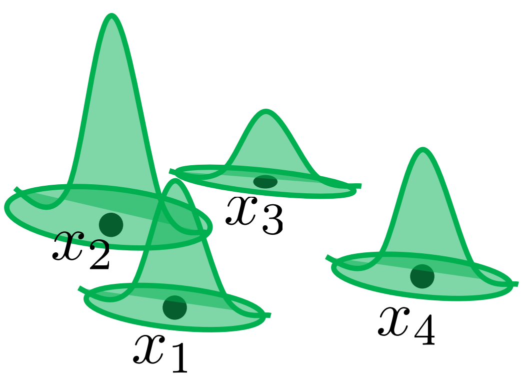
- Function space has been unspecified so far
- We care about pointwise evaluations of functions
  - RKHS with gaussian kernel

$$\langle g_\sigma(x,\cdot), g_\sigma(y,\cdot)\rangle := g_\sigma(x,y)$$

- The Hilbert space is the closure of the vector space spanned by

$$\{k_x(\cdot) = g_\sigma(x,\cdot) : x \in \mathbb{X}\}$$

$$f = \sum_{i=1}^{m} \alpha_i k_{x_i}$$



# How do we tell if a function is in the RKHS?

**Def:** A function K(x,y) is positive semidefinite if for any n and every choice of n distinct points $x_1, \dots, x_n$, the matrix $(K(x_i, x_j))$ is positive semidefinite.

$$K(x,y) \geq 0 \iff \sum_{i=1}^{n}\sum_{j=1}^{n}\overline{\alpha_i}\alpha_j K(x_i,x_j) \geq 0$$

**Theorem 3.1** ([2], Theorem 3.11)**.** *Let $\mathcal{H}$ be an RKHS with a kernel $K$. If $f : \mathbb{X} \to \mathbb{C}$, then the following are equivalent:*

*(i)* $f \in \mathcal{H}$;

*(ii) there exists a constant $c > 0$ such that for every finite subset $F = \{w_1, \dots, w_n\} \subset \mathbb{X}$, there exists a function $h \in \mathcal{H}$ with $\|h\|_{\mathcal{H}} \leq c$ and $f(w_i) = h(w_i)$, $i = 1, \dots, n$;*

*(iii) there exists a constant $c > 0$, such that the function $M(x,y) = c^2K(x,y) - f(x)\overline{f(y)}$ is a kernel function.*

**Theorem 5.7** (Pull-back theorem)**.** *Let $X$ and $S$ be sets, let $\varphi : S \to X$ be a function and let $K : X \times X \to \mathbb{C}$ be a kernel function. Then $\mathcal{H}(K \circ \varphi) = \{f \circ \varphi : f \in \mathcal{H}(K)\}$, and for $u \in \mathcal{H}(K \circ \varphi)$ we have that $\|u\|_{\mathcal{H}(K\circ\varphi)} = \min\{\|f\|_{\mathcal{H}(K)} : u = f \circ \varphi\}$.*

$$(K \circ \varphi)(x,y) := K(\varphi(x), \varphi(y))$$

[2] Vern I Paulsen and Mrinal Raghupathi. *An Introduction to the Theory of Reproducing Kernel Hilbert Spaces.* Cambridge University Press, Cambridge, 2016.

# RKHS invariance under Koopman

$$\boxed{C_\Phi(\mathcal{H}(g_\sigma)) \subset \mathcal{H}(g_\sigma)}$$

The idea is to let the dynamics induce a new kernel function and then show that this kernel function defines the same space as the original kernel function.

$$f \in \mathcal{H}(g_\sigma)) \implies f(x)\overline{f(y)} \leq g_\sigma(x,y)$$ (positive semidefinite)

$$\implies f(\Phi(x))\overline{f(\Phi(y))} \leq g_\sigma(\Phi(x),\Phi(y))$$

$$\implies f \circ \Phi \in \mathcal{H}(g_\sigma \circ \Phi)$$

$$C_\Phi : \mathcal{H}(g_\sigma) \to \mathcal{H}(g_\sigma \circ \Phi)$$

Since $\Phi$ is a bijection, the spaces spanned by the sets of functions

$$\{h_y(\cdot) = g_\sigma(\Phi(y),\Phi(\cdot)) : y \in \mathbb{X}\} \qquad \{k_x(\cdot) = g_\sigma(x,\cdot) : x \in \mathbb{X}\}$$

are the same.*

*Needs a proof

# RKHS invariance under integral Koopman

$$\boxed{A_\sigma(\mathcal{H}(g_\sigma)) \subset \mathcal{H}(g_\sigma)}$$

**Idea**: Show $h = A_\sigma f$ is in $\mathcal{H}(g_\sigma \circ \Phi)$, then use the same equivalence argument.

Find c such that: $\sum_{i,j=1}^{n} \overline{z_i} z_j \left[ c^2 J(x_i, x_j) - h(x_i)\overline{h(x_j)} \right] \geq 0$ where $J(x,y) = g_\sigma(\Phi(x), \Phi(y))$

1: $\sum_{i,j=1}^{n} \overline{z_i} z_j c^2 J(x_i, x_j) = c^2 \left\langle \sum_{i=1}^{n} \overline{z_i} k_{\Phi(x_i)}, \sum_{j=1}^{n} \overline{z_j} k_{\Phi(x_j)} \right\rangle = c^2 \left\| \sum_{i=1}^{n} \overline{z_i} k_{\Phi(x_i)} \right\|^2$

2: $\sum_{i,j=1}^{n} \overline{z_i} z_j h(x_i)\overline{h(x_j)} = \left| \sum_{i=1}^{n} \overline{z_i} h(x_i) \right|^2 = \left| \sum_{i=1}^{n} \overline{z_i} \int \frac{g_\sigma(\Phi(x_i), y)}{m_\sigma(\Phi(x_i))} f(y) d\mu(y) \right|^2$

# RKHS invariance under integral Koopman

$$\boxed{A_\sigma(\mathcal{H}(g_\sigma)) \subset \mathcal{H}(g_\sigma)}$$

**Idea**: Show $h = A_\sigma f$ is in $\mathcal{H}(g_\sigma \circ \Phi)$, then use the same equivalence argument.

Find c such that: $\sum_{i,j=1}^{n} \overline{z_i} z_j \left[c^2 J(x_i,x_j) - h(x_i)\overline{h(x_j)}\right] \geq 0$ where $J(x,y) = g_\sigma(\Phi(x),\Phi(y))$

1: $$\sum_{i,j=1}^{n} \overline{z_i} z_j c^2 J(x_i,x_j) = c^2 \left\langle \sum_{i=1}^{n} \overline{z_i} k_{\Phi(x_i)}, \sum_{j=1}^{n} \overline{z_j} k_{\Phi(x_j)} \right\rangle = c^2 \left\| \sum_{i=1}^{n} \overline{z_i} k_{\Phi(x_i)} \right\|^2$$

2: $$\sum_{i,j=1}^{n} \overline{z_i} z_j h(x_i)\overline{h(x_j)} = \left| \sum_{i=1}^{n} \overline{z_i} h(x_i) \right|^2 = \left| \sum_{i=1}^{n} \overline{z_i} \int \frac{g_\sigma(\Phi(x_i),y)}{m_\sigma(\Phi(x_i))} f(y) d\mu(y) \right|^2 \leq \beta^2 \left| \sum_{i=1}^{n} \overline{z_i} \int \langle k_{\Phi(x_i)}, k_y \rangle f(y) d\mu(y) \right|^2$$

$$\forall \delta > 0, \inf\{\mu(B_\delta(w)) : w \in \mathbb{X}\} > 0$$

# RKHS invariance under integral Koopman

$$\boxed{A_\sigma(\mathcal{H}(g_\sigma)) \subset \mathcal{H}(g_\sigma)}$$

**Idea**: Show $h = A_\sigma f$ is in $\mathcal{H}(g_\sigma \circ \Phi)$, then use the same equivalence argument.

Find c such that: $\sum_{i,j=1}^{n} \overline{z_i} z_j \left[c^2 J(x_i,x_j) - h(x_i)\overline{h(x_j)}\right] \geq 0$ where $J(x,y) = g_\sigma(\Phi(x),\Phi(y))$

1: $$\sum_{i,j=1}^{n} \overline{z_i} z_j c^2 J(x_i,x_j) = c^2 \left\langle \sum_{i=1}^{n} \overline{z_i} k_{\Phi(x_i)}, \sum_{j=1}^{n} \overline{z_j} k_{\Phi(x_j)} \right\rangle = c^2 \left\| \sum_{i=1}^{n} \overline{z_i} k_{\Phi(x_i)} \right\|^2$$

2: $$\sum_{i,j=1}^{n} \overline{z_i} z_j h(x_i)\overline{h(x_j)} = \left| \sum_{i=1}^{n} \overline{z_i} h(x_i) \right|^2 = \left| \sum_{i=1}^{n} \overline{z_i} \int \frac{g_\sigma(\Phi(x_i),y)}{m_\sigma(\Phi(x_i))} f(y) d\mu(y) \right|^2 \leq \beta^2 \left| \sum_{i=1}^{n} \overline{z_i} \int \langle k_{\Phi(x_i)}, k_y \rangle f(y) d\mu(y) \right|^2$$

3: $$\sum_{i,j=1}^{n} \overline{z_i} z_j h(x_i)\overline{h(x_j)} \leq (\beta \|f\|_\infty \mu(\mathbb{X}))^2 \left\| \sum_{i=1}^{n} \overline{z_i} k_{\Phi(x_i)} \right\|^2$$

# Convergence to the Koopman operator

**Definition 3.3.** A family of maps $\{s_\sigma : \mathbb{X} \times \mathbb{X} \to \mathbb{R} \,|\, \sigma > 0\}$ is called a summability kernel if for all $x \in \mathbb{X}$

(1) $\forall \sigma > 0$, $\int_{\mathbb{X}} s_\sigma(x,y)\,d\mu(y) = 1$, and

(2) $\exists K > 0, \forall \sigma > 0$, $\int_{\mathbb{X}} |s_\sigma(x,y)|\,d\mu(y) \leq K$.

(3) $\forall \delta > 0$,

$$\lim_{\sigma \to 0} \int_{\mathbb{X} - B_\delta(\Phi(x))} s_\sigma(x,y)\,d\mu(y) = 0. \tag{3.20}$$

where $B_\delta(x) = \{y \in \mathbb{X} \,|\, \|x - y\|_2 < \delta\}$.

# Convergence to the Koopman operator

**Definition 3.3.** A family of maps $\{s_\sigma : \mathbb{X}\times\mathbb{X}\to\mathbb{R} \,|\, \sigma > 0\}$ is called a summability kernel if for all $x\in\mathbb{X}$

(1) $\forall\sigma > 0$, $\int_{\mathbb{X}} s_\sigma(x,y)d\mu(y) = 1$, and

(2) $\exists K > 0, \forall\sigma > 0$, $\int_{\mathbb{X}} |s_\sigma(x,y)|d\mu(y) \le K$.

(3) $\forall\delta > 0$,

$$\lim_{\sigma\to 0}\int_{\mathbb{X}-B_\delta(\Phi(x))} s_\sigma(x,y)d\mu(y) = 0. \tag{3.20}$$

where $B_\delta(x) = \{y\in\mathbb{X} \,|\, \|x-y\|_2 < \delta\}$.

**Lemma 3.4.** *If $\mu$ satisfies* (2.1)*, then $\{\phi_\sigma \,|\, \sigma > 0\}$, where $\phi_\sigma$ is given by* (3.7)*, is a summability kernel. Furthermore, the convergence*

$$\lim_{\sigma\to 0}\int_{\mathbb{X}-B_\delta(\Phi(x))} \phi_\sigma(x,y)d\mu(y) = 0 \tag{3.21}$$

*is uniform in x.*

Proof requires (2.1): $\forall\delta > 0, \inf\{\mu(B_\delta(w)) : w\in\mathbb{X}\} > 0$

# Convergence to the Koopman operator

**Pointwise convergence on continuous functions**

**Lemma 3.5** (Pointwise convergence)**.** *For any continuous* $f : \mathbb{X} \to \mathbb{C}$

$$\lim_{\sigma \to 0} \Big| (A_\sigma f)(x) - (C_\Phi f)(x) \Big| = 0$$

*uniformly in* $x \in \mathbb{X}$.

# Convergence to the Koopman operator

AIMdynInc.

**Pointwise convergence on continuous functions**

**Lemma 3.5** (Pointwise convergence)**.** *For any continuous $f : \mathbb{X} \to \mathbb{C}$*

$$\lim_{\sigma \to 0} \left| (A_\sigma f)(x) - (C_\Phi f)(x) \right| = 0$$

*uniformly in $x \in \mathbb{X}$.*

$$\begin{aligned}\left| \int \phi_\sigma(x,y) f(y) d\mu(y) - f(\Phi(x)) \right| &= \left| \int \phi_\sigma(x,y) [f(y) - f(\Phi(x))] d\mu(y) \right| \\ &\le 2\|f\|_\infty \int_{\mathbb{X} - B_\delta(\Phi(x))} \phi_\sigma(x,y) d\mu(y) + \int_{B_\delta(\Phi(x))} \phi_\sigma(x,y) |f(y) - f(\Phi(x))| d\mu(y) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \int_{B_\delta(\Phi(x))} \phi_\sigma(x,y) d\mu(y) \\ &< \varepsilon.\end{aligned}$$

Bound using summability kernel property

Bound using continuity of f

# Partial Dynamics

- Assume that the dynamics are only known for

$$\mathbb{X}_m = \{x_1, \ldots, x_m\}$$

- **Goal**: Define a transition kernel $p_{\sigma,\delta}^{(m)} : \mathbb{X} \times \mathbb{X} \to [0, \infty)$ such that

# Partial Dynamics

- Assume that the dynamics are only known for

$$\mathbb{X}_m = \{x_1, \dots, x_m\}$$

- **Goal**: Define a transition kernel $p^{(m)}_{\sigma,\delta} : \mathbb{X} \times \mathbb{X} \to [0, \infty)$ such that

(1) if $w \in \mathbb{X}$ is such that $\Phi(w)$ is known, then $\lim_{u \to w} p^{(m)}_{\sigma,\delta}(u, y) \approx \phi_\sigma(w, y)$, the Koopman kernel

# Partial Dynamics

- Assume that the dynamics are only known for

$$\mathbb{X}_m = \{x_1, \ldots, x_m\}$$

- **Goal**: Define a transition kernel $p^{(m)}_{\sigma,\delta} : \mathbb{X} \times \mathbb{X} \to [0, \infty)$ such that

(1) if $w \in \mathbb{X}$ is such that $\Phi(w)$ is known, then $\lim_{u \to w} p^{(m)}_{\sigma,\delta}(u, y) \approx \phi_\sigma(w, y)$, the Koopman kernel

(2) if there is no $w \in \mathbb{X}$ for which $\Phi(w)$ is known, then $p^{(m)}_{\sigma,\delta} \approx d_\sigma$, the pure diffusion kernel

# Partial Dynamics

- Assume that the dynamics are only known for

$$\mathbb{X}_m = \{x_1, \ldots, x_m\}$$

- **Goal**: Define a transition kernel $p^{(m)}_{\sigma,\delta} : \mathbb{X} \times \mathbb{X} \to [0,\infty)$ such that

(1) if $w \in \mathbb{X}$ is such that $\Phi(w)$ is known, then $\lim_{u\to w} p^{(m)}_{\sigma,\delta}(u,y) \approx \phi_\sigma(w,y)$, the Koopman kernel

(2) if there is no $w \in \mathbb{X}$ for which $\Phi(w)$ is known, then $p^{(m)}_{\sigma,\delta} \approx d_\sigma$, the pure diffusion kernel

(3) as the set of $w$'s for which $\Phi(w)$ is known becomes dense in $\mathbb{X}$, the operator defined via $p^{(m)}_{\sigma,\delta}$ will converge to $A_\sigma$ for any appropriately chosen decreasing sequence of $\delta$'s.

# Continuous extension of known dynamics

$$\mathbb{X}_m = \{x_1, \ldots, x_m\}$$

Choose delta such that the gaussians satisfy

$$g_\delta(x_i, x_j) \approx 0, \quad \forall x_i, x_j \in \mathbb{X}_m$$

Transition kernel

$$\boxed{p^{(m)}_{\sigma,\delta}(u, y|\mathbb{X}_m) = \sum_{x \in \mathbb{X}_m} \alpha_\delta(u, x) p^{(m)}_{\sigma,\delta}(u, y|x)}$$

# Continuous extension of known dynamics

$$\mathbb{X}_m = \{x_1, \ldots, x_m\}$$

Choose delta such that the gaussians satisfy

$$g_\delta(x_i, x_j) \approx 0, \quad \forall x_i, x_j \in \mathbb{X}_m$$

Transition kernel

$$\boxed{p^{(m)}_{\sigma,\delta}(u,y|\mathbb{X}_m) = \sum_{x \in \mathbb{X}_m} \alpha_\delta(u,x) p^{(m)}_{\sigma,\delta}(u,y|x)}$$

$$\alpha_\delta(u,x) = \frac{g_\delta(u,x)}{\sum_{x' \in \mathbb{X}_m} g_\delta(u,x')}$$

$$p^{(m)}_{\sigma,\delta}(u,y|x) = \begin{cases} d_\sigma(u,y) & ,m=0 \\ (1-g_\delta(u,x))d_\sigma(u,y) + g_\delta(u,x)\phi_\sigma(x,y) & ,m>0 \end{cases}$$

Mixed kernel

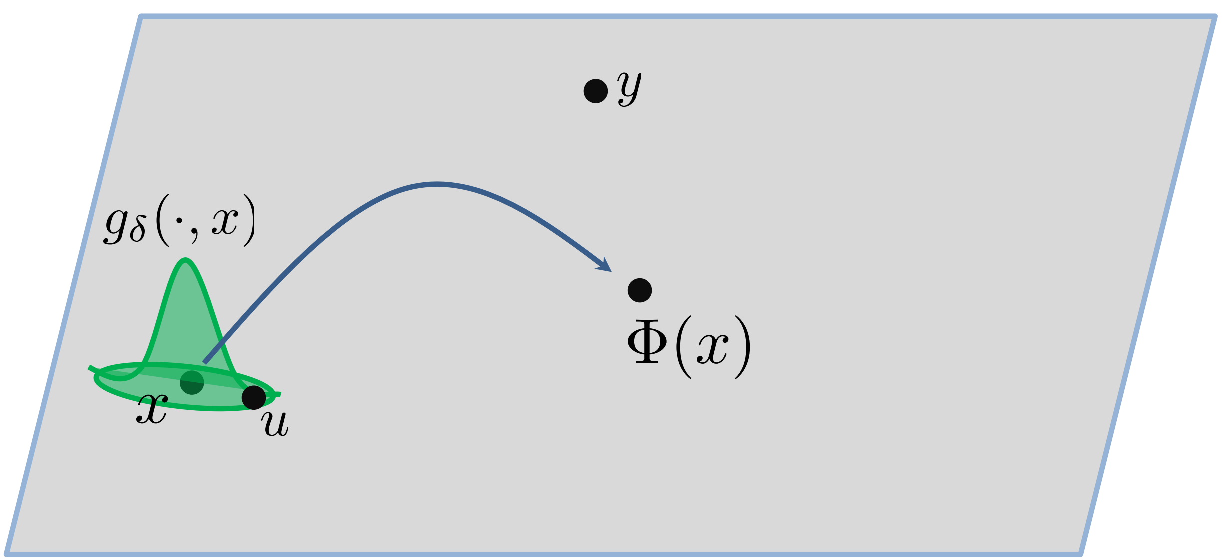
# Continuous extension of known dynamics

$$\mathbb{X}_m = \{x_1, \ldots, x_m\}$$

$$p_{\sigma,\delta}^{(m)}(u,y|x) = \begin{cases} d_\sigma(u,y) & , m = 0 \quad \text{(Zero knowledge)} \\ (1-g_\delta(u,x))d_\sigma(u,y) + g_\delta(u,x)\phi_\sigma(x,y) & , m > 0 \quad \text{(Some knowledge)} \end{cases}$$

Pure diffusion kernel

Koopman kernel

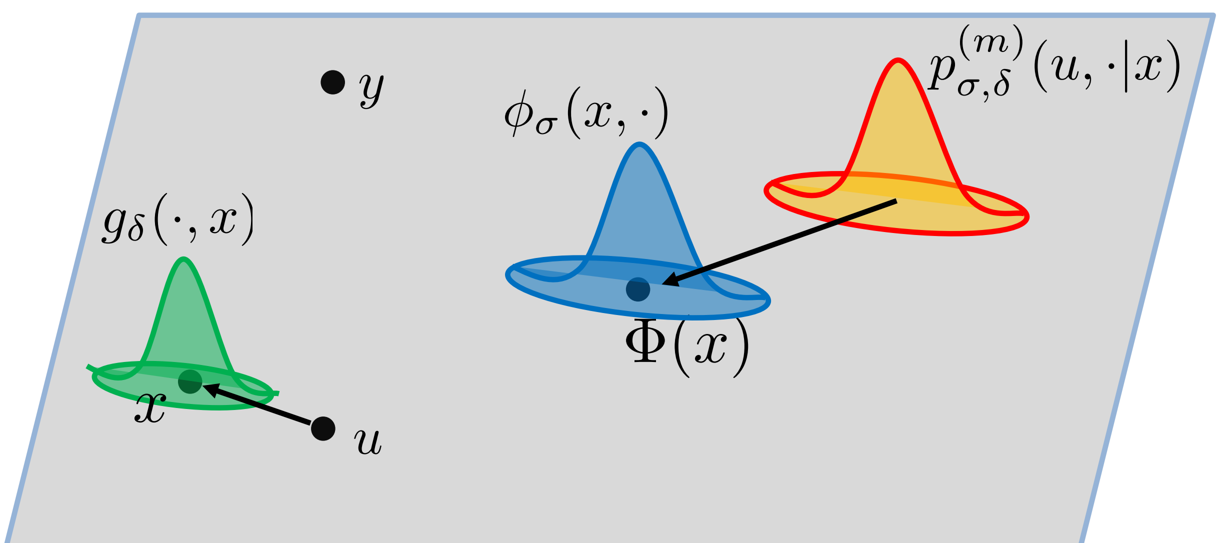


# Arbitrary approximation of known dynamics

**Lemma 4.1.** *For any $m > 0$, $\delta > 0$, and $x \in \mathbb{X}_m$,*

$$\lim_{u \to x} p^{(m)}_{\sigma,\delta}(u, y|x) = \phi_\sigma(x, y).$$

$$\left| p^{(m)}_{\sigma,\delta}(u, y|x) - \phi_\sigma(x, y) \right| = |(1 - g_\delta(u, x))(d_\sigma(u, y) - \phi_\sigma(x, y))|$$



# Arbitrary approximation of known dynamics

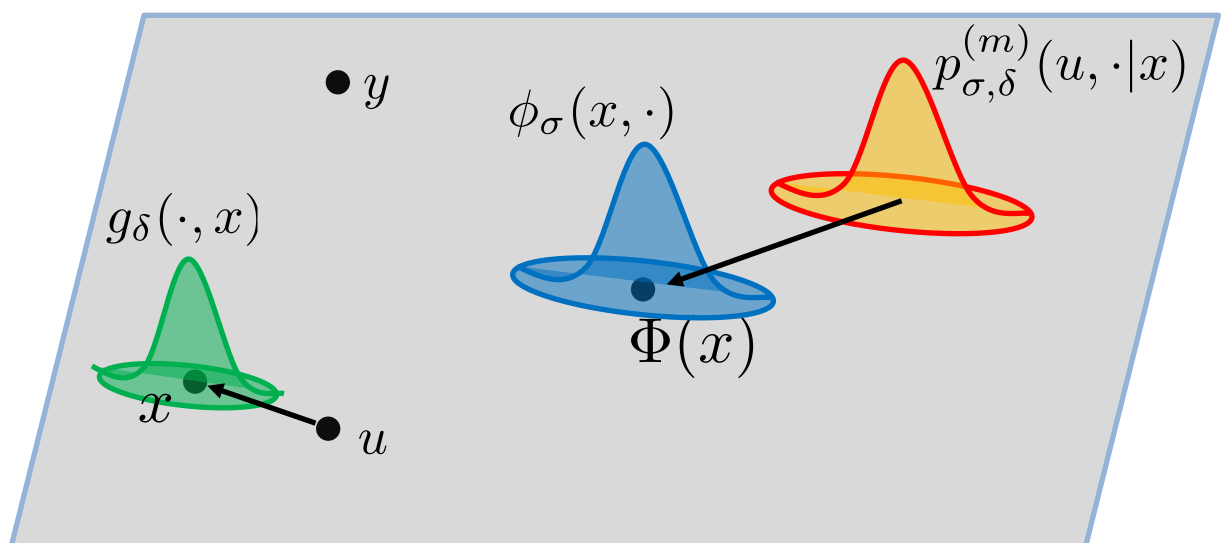
**Lemma 4.2.** *Fix $x \in \mathbb{X}_m$. For any $m > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\lim_{u \to x} \left| p^{(m)}_{\sigma,\delta}(u, y | \mathbb{X}_m) - \phi_\sigma(x, y) \right| = O(\varepsilon).$$

$$\lim_{u \to x} p^{(m)}_{\sigma,\delta}(u, y | \mathbb{X}_m) = p^{(m)}_{\sigma,\delta}(x, y | \mathbb{X}_m) = a_\delta(x, x) p^{(m)}_{\sigma,\delta}(x, y | x) + \sum_{x' \neq x} a_\delta(x, x') p^{(m)}_{\sigma,\delta}(x, y | x')$$

$$= \left( \frac{1}{1 + \sum_{x' \neq x} g_\delta(x, x')} \right) p^{(m)}_{\sigma,\delta}(x, y | x) + \sum_{x' \neq x} \left( \frac{g_\delta(x, x')}{1 + \sum_{x'' \neq x} g_\delta(x, x'')} \right) p^{(m)}_{\sigma,\delta}(x, y | x')$$

Uniformly bounded



# Arbitrary approximation of known dynamics

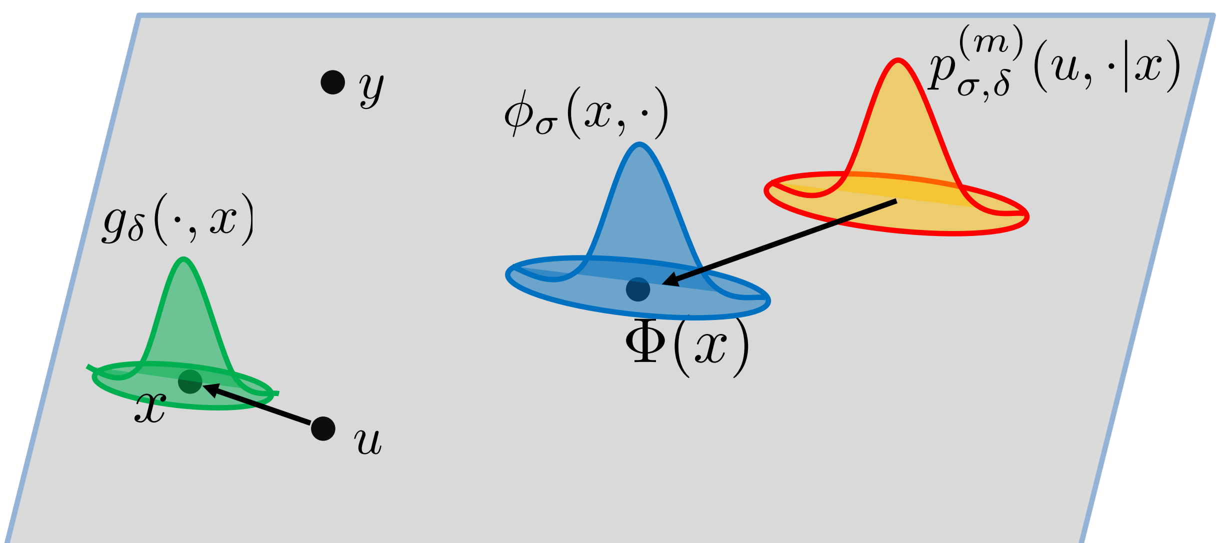
**Lemma 4.2.** *Fix $x \in \mathbb{X}_m$. For any $m > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\lim_{u \to x} \left| p^{(m)}_{\sigma,\delta}(u, y|\mathbb{X}_m) - \phi_\sigma(x, y) \right| = O(\varepsilon).$$

$$\lim_{u \to x} p^{(m)}_{\sigma,\delta}(u, y|\mathbb{X}_m) = p^{(m)}_{\sigma,\delta}(x, y|\mathbb{X}_m) = a_\delta(x, x) p^{(m)}_{\sigma,\delta}(x, y|x) + \sum_{x' \neq x} a_\delta(x, x') p^{(m)}_{\sigma,\delta}(x, y|x')$$

$$= \left( \frac{1}{1 + \sum_{x' \neq x} g_\delta(x, x')} \right) p^{(m)}_{\sigma,\delta}(x, y|x) + \sum_{x' \neq x} \left( \frac{g_\delta(x, x')}{1 + \sum_{x'' \neq x} g_\delta(x, x'')} \right) p^{(m)}_{\sigma,\delta}(x, y|x')$$

$$\Longrightarrow \quad \lim_{u \to x} \left| p^{(m)}_{\sigma,\delta}(u, y|\mathbb{X}_m) - \phi_\sigma(x, y) \right| \le C_\sigma \left| \frac{\sum_{x' \neq x} g_\delta(x, x')}{1 + \sum_{x' \neq x} g_\delta(x, x')} \right| + C_\sigma \sum_{x' \neq x} \left| \frac{g_\delta(x, x')}{1 + \sum_{x'' \neq x} g_\delta(x, x'')} \right|$$



# Arbitrary approximation of known dynamics

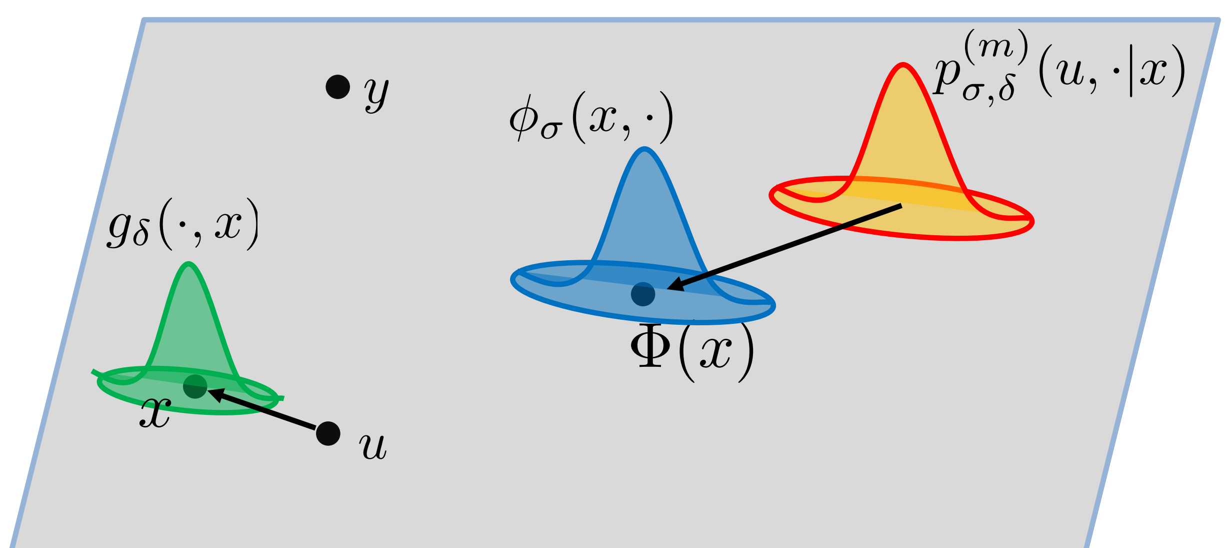
**Lemma 4.2.** *Fix $x \in \mathbb{X}_m$. For any $m > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\lim_{u \to x} \left| p^{(m)}_{\sigma,\delta}(u, y | \mathbb{X}_m) - \phi_\sigma(x, y) \right| = O(\varepsilon).$$

$$\lim_{u \to x} p^{(m)}_{\sigma,\delta}(u, y | \mathbb{X}_m) = p^{(m)}_{\sigma,\delta}(x, y | \mathbb{X}_m) = a_\delta(x, x) p^{(m)}_{\sigma,\delta}(x, y | x) + \sum_{x' \neq x} a_\delta(x, x') p^{(m)}_{\sigma,\delta}(x, y | x')$$

$$= \left( \frac{1}{1 + \sum_{x' \neq x} g_\delta(x, x')} \right) p^{(m)}_{\sigma,\delta}(x, y | x) + \sum_{x' \neq x} \left( \frac{g_\delta(x, x')}{1 + \sum_{x'' \neq x} g_\delta(x, x'')} \right) p^{(m)}_{\sigma,\delta}(x, y | x')$$

$$\Longrightarrow \quad \lim_{u \to x} \left| p^{(m)}_{\sigma,\delta}(u, y | \mathbb{X}_m) - \phi_\sigma(x, y) \right| \leq C_\sigma \left| \frac{\cancelto{\epsilon}{\sum_{x' \neq x} g_\delta(x, x')}}{1 + \sum_{x' \neq x} g_\delta(x, x')} \right| + C_\sigma \sum_{x' \neq x} \left| \frac{\cancelto{\epsilon}{g_\delta(x, x')}}{1 + \sum_{x'' \neq x} g_\delta(x, x'')} \right|$$

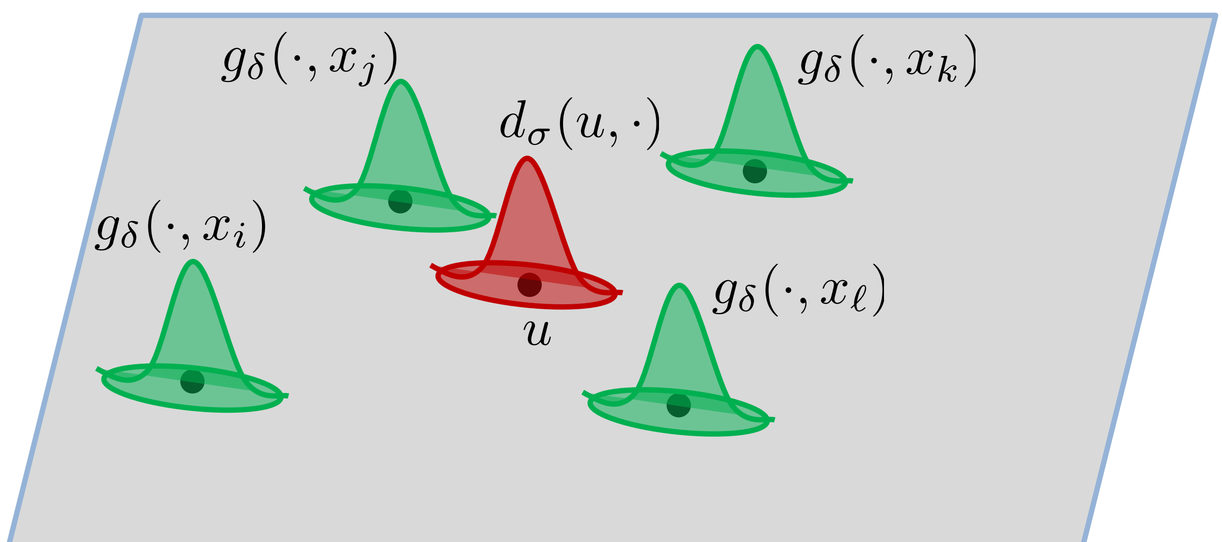


# The pure diffusion limit

**Lemma 4.4.** *Fix $u \notin \mathbb{X}_m$. Then*
$$\lim_{\delta \to 0} p^{(m)}_{\sigma,\delta}(u,y|\mathbb{X}_m) = d_\sigma(u,y)$$

$$\lim_{\delta \to 0} p^{(m)}_{\sigma,\delta}(u,y|x) = d_\sigma(u,y)$$

$$p^{(m)}_{\sigma,\delta}(u,y|x) = \begin{cases} d_\sigma(u,y) & ,m=0 \\ (1-g_\delta(u,x))d_\sigma(u,y) + g_\delta(u,x)\phi_\sigma(x,y) & ,m>0 \end{cases}$$

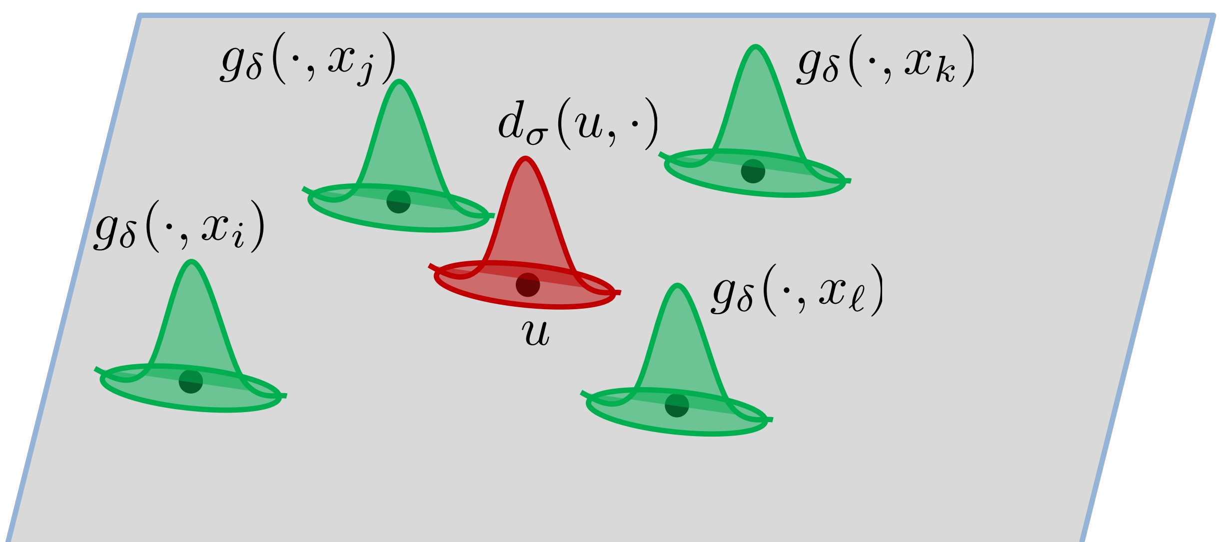


# The pure diffusion limit

**Lemma 4.4.** *Fix $u \notin \mathbb{X}_m$. Then*
$$\lim_{\delta \to 0} p^{(m)}_{\sigma,\delta}(u,y|\mathbb{X}_m) = d_\sigma(u,y)$$

$$\lim_{\delta \to 0} p^{(m)}_{\sigma,\delta}(u,y|x) = d_\sigma(u,y)$$

$$p^{(m)}_{\sigma,\delta}(u,y|x) = \begin{cases} d_\sigma(u,y) & ,m=0 \\ (1-\underbrace{g_\delta(u,x)}_{0})d_\sigma(u,y) + \underbrace{g_\delta(u,x)}_{0}\phi_\sigma(x,y) & ,m>0 \end{cases}$$



# Partial dynamics integral operator

$$(A_{\sigma,\delta} f)(x) = \int p^{(m)}_{\sigma,\delta}(x, y|\mathbb{X}_m) f(y) d\mu(y)$$

**Proposition 4.5.** *Fix* $\varepsilon > 0$*. For any* $\mathbb{X}_m$*, there exists a* $\delta(\varepsilon, \mathbb{X}_m) > 0$ *such that for any continuous* $f$ *and any* $x \in \mathbb{X}_m$*,*

$$\left|(A_{\sigma,\delta} f)(x) - (A_\sigma f)(x)\right| = O(\varepsilon). \tag{4.41}$$

Recall $\quad (A_\sigma f)(x) = \int \phi_\sigma(x, y) f(y) d\mu(y)$

$$\begin{aligned}\left|(A_{\sigma,\delta} f)(x) - (A_\sigma f)(x)\right| &= \left|\int \left(p^{(m)}_{\sigma,\delta}(x, y|\mathbb{X}_m) - \phi_\sigma(x, y)\right) f(y) d\mu(y)\right| \\ &\le \int \left|p^{(m)}_{\sigma,\delta}(x, y|\mathbb{X}_m) - \phi_\sigma(x, y)\right| |f(y)| d\mu(y) \\ &\le C_\sigma \varepsilon \|f\|_1 .\end{aligned}$$

# Unscrambling Time

**Given: TBI Study**
- Control patients
  - 1 visit, 2 sampling protocols
  - HFWB : blood drawn ever 2h over 28h
  - LFWB : blood draw ever 4h over 12h
- TBI patients
  - 3 visits (0, 2, 6 months)
  - LFWB protocol
- Gene expression levels (RNAseq) in each blood sample supplied

**Problem**
- The time stamps of the blood samples were not given
- Each patients' blood samples were permuted differently

**Goal**
- Find the true time stamp on a 24h clock of each blood sample

**Assumptions**
- The circadian rhythm is the dominant driving force of the gene expression levels
- HFWB and LFWB samples for control patients treated as separate data sets
  - Note: this is different to how the NU team treated them. We are losing information with this assumption.

# Unscrambling Time

Time manifold
(interval or circle)

Map from time manifold
to gene expression levels
(higher dimensional space)

Gene expression time
series for circadian genes

$$\left[ x_1, x_2, \ldots \right]$$

Time
blind
data

**Data given**

Gene expression levels
for circadian genes

$$\left\{ \ldots, w_i,\, w_j, \ldots \right\}$$

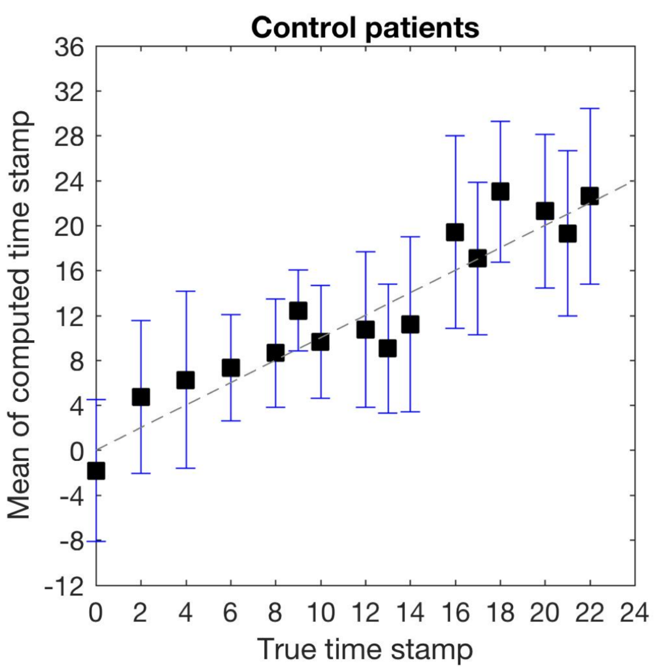
Use Diffusion Maps to find
coordinates $\theta_i$ (time stamp or phase)
(1 eigenfunction to time interval)
(2 eigenfunctions to compute
a phase for circular time manifold)

$$\left\{ \begin{matrix} \ldots, \theta_i,\, \theta_j, \ldots \\ \ldots, w_i,\, w_j, \ldots \end{matrix} \right\}$$

Map diffusion coordinate
to time of day

$$\left\{ \begin{matrix} \ldots, t_i,\, t_j, \ldots \\ \ldots, w_i,\, w_j, \ldots \end{matrix} \right\}$$

# Unscrambling Time



# Summary

- We introduced an integral approximation of the Koopman operator in terms of asymmetric similarity kernels
- We constructed a “homotopy” between the Diffusion Maps operator and the integral Koopman operator, where the homotopy was parameterized by the amount of knowledge we had about the dynamics
  - No knowledge = Diffusion Maps
  - Full knowledge = Integral Koopman operator
- We presented results on the “unscrambling problem” of time series of blood samples where the goal was to correctly order time-blinded and shuffled time series data

Thank you!