Browse data science with Differential geometry and Random matrix theory

Hau-Tieng Wu, MD., PhD.
Mathematics and Statistical Science
Duke University

Modern Data is Massive
Structure is the key

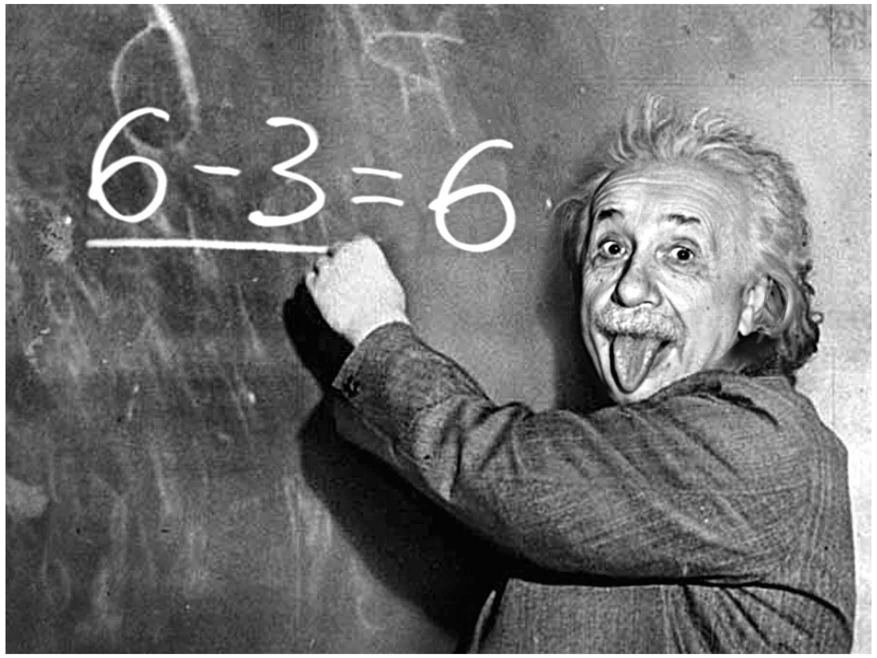
(from internet)

- Large-volume
- High dimensional
- Noisy
- ~~Non stationary~~

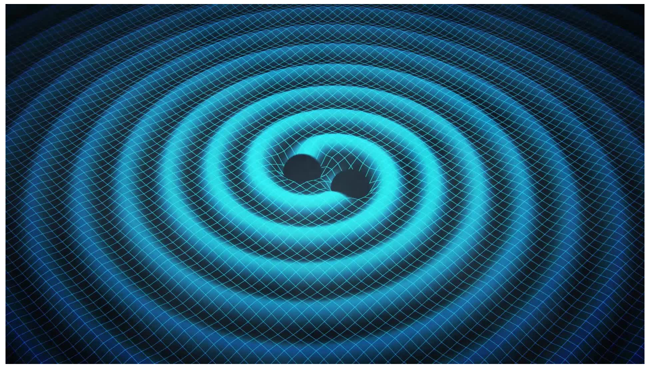
Structure = knowledge!

From data to data science
model and analysis

$F=ma$

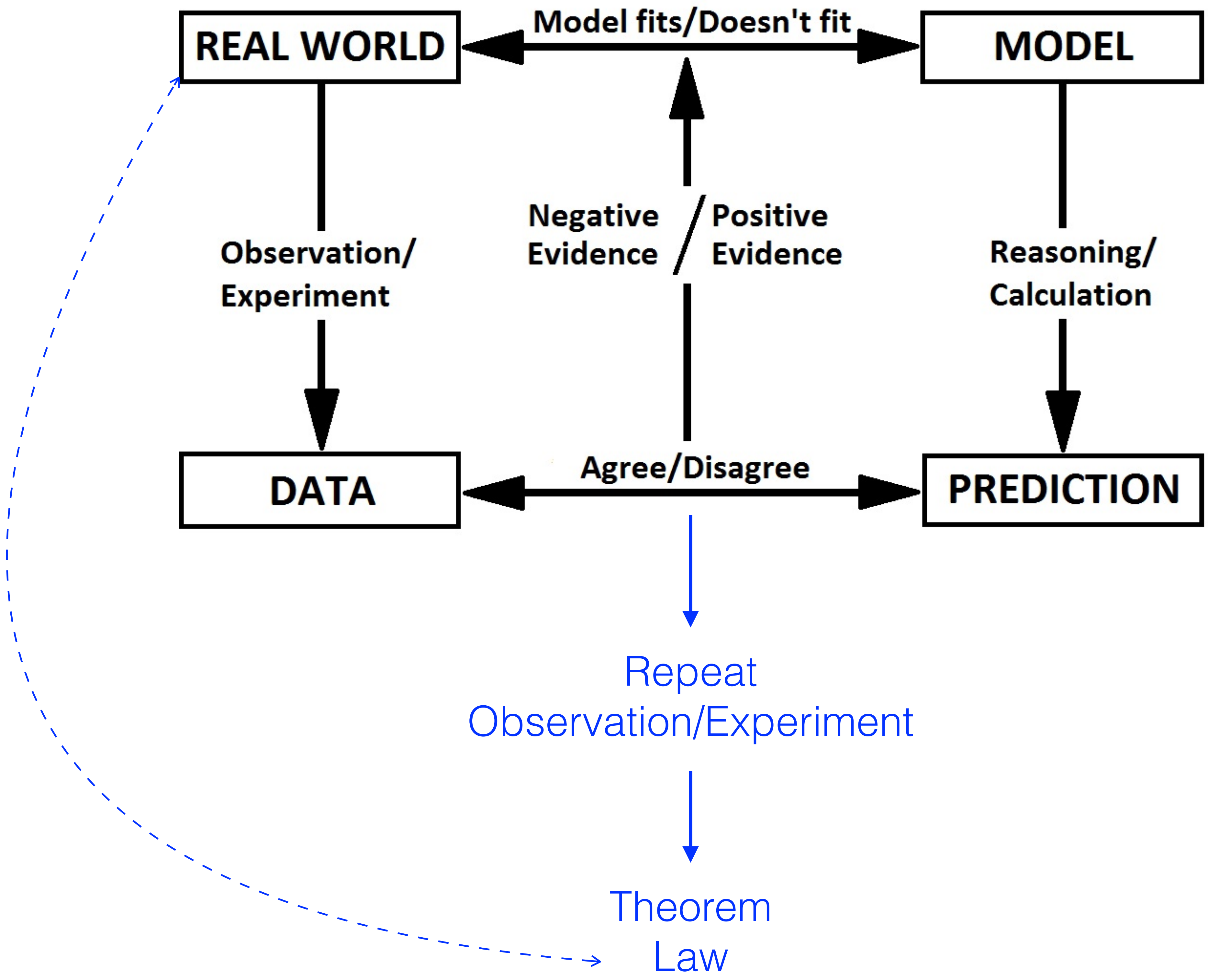


$E=mc^2$

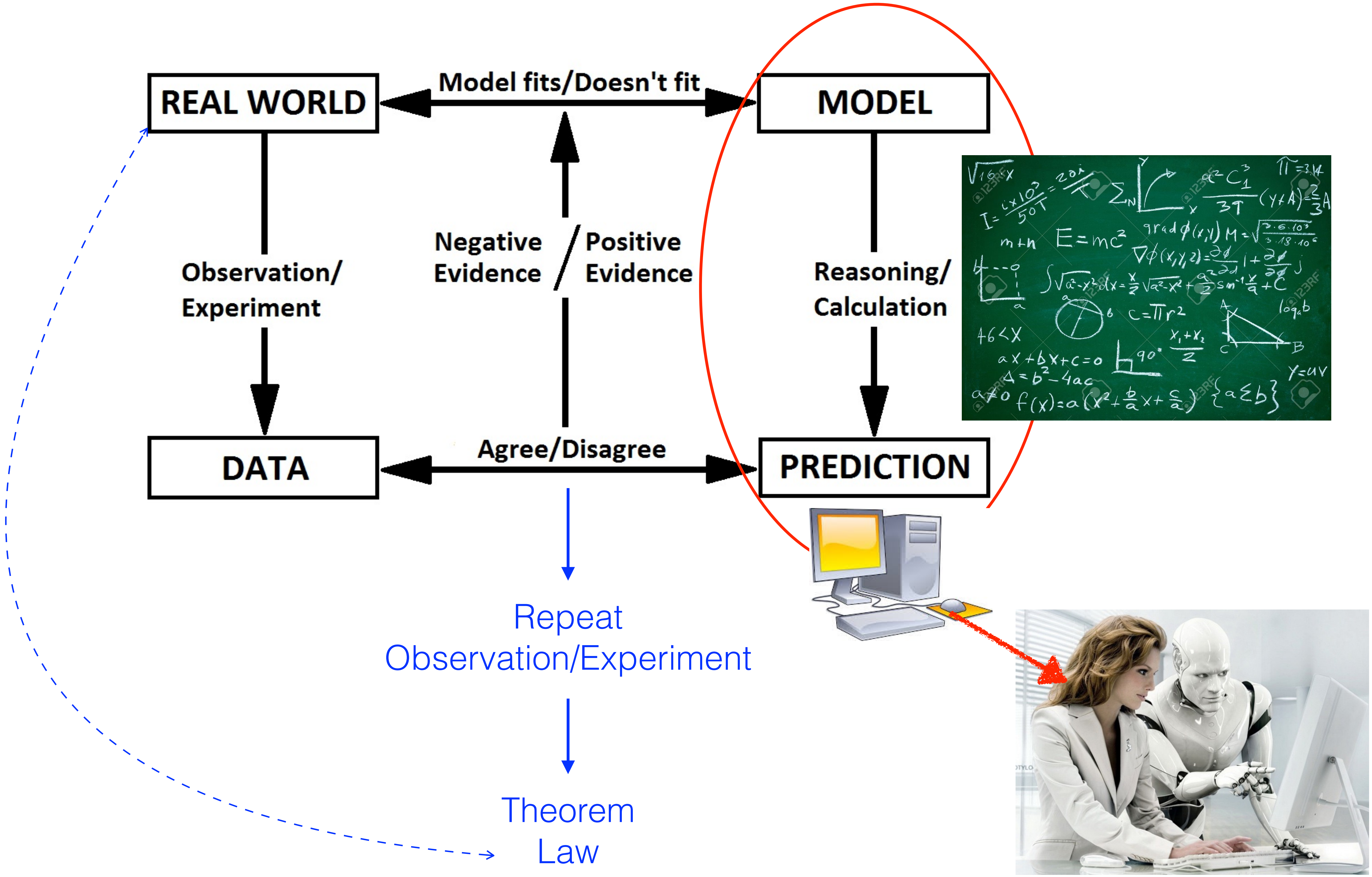


Find the structure under the datasets and use it

Not anything new but scientific argument!!!



Science + computer leads to AI



- One motivative problem
- Manifold learning — algorithm
- Manifold learning — theory
- Random matrix theory
- Toward manifold + RMT
- Some more…

High frequency time series is everywhere in healthcare

Operation Room

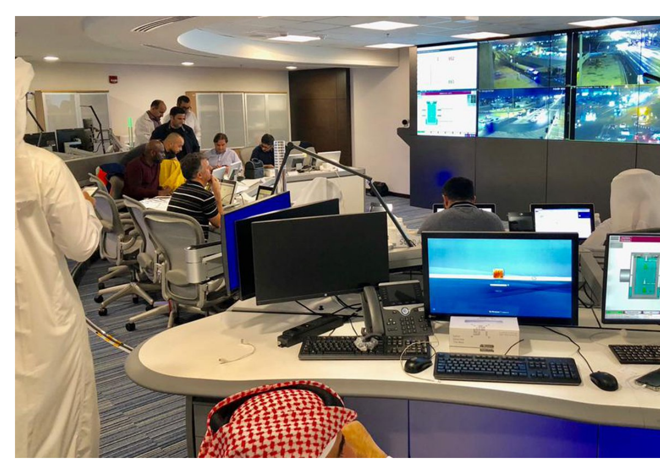
Intensive Care unit

Holter system

Ambulance System

fMRI

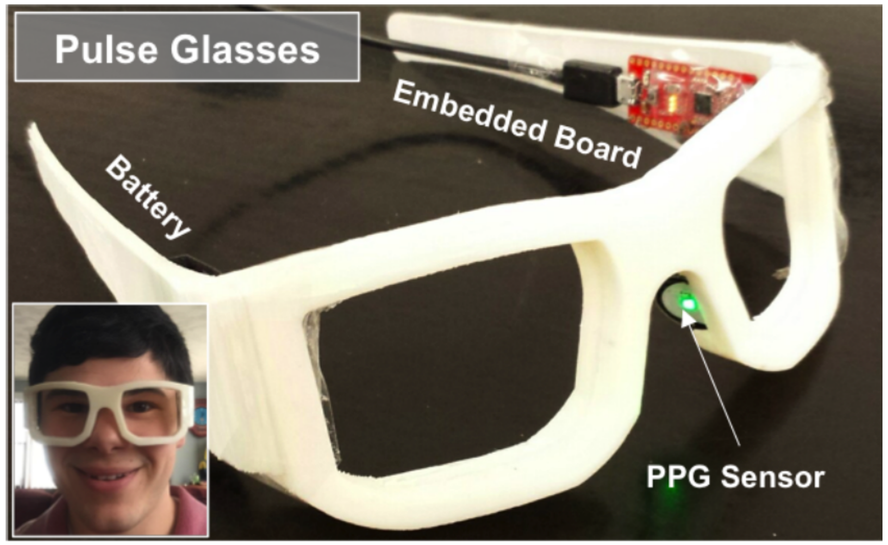
EP/Cath Room

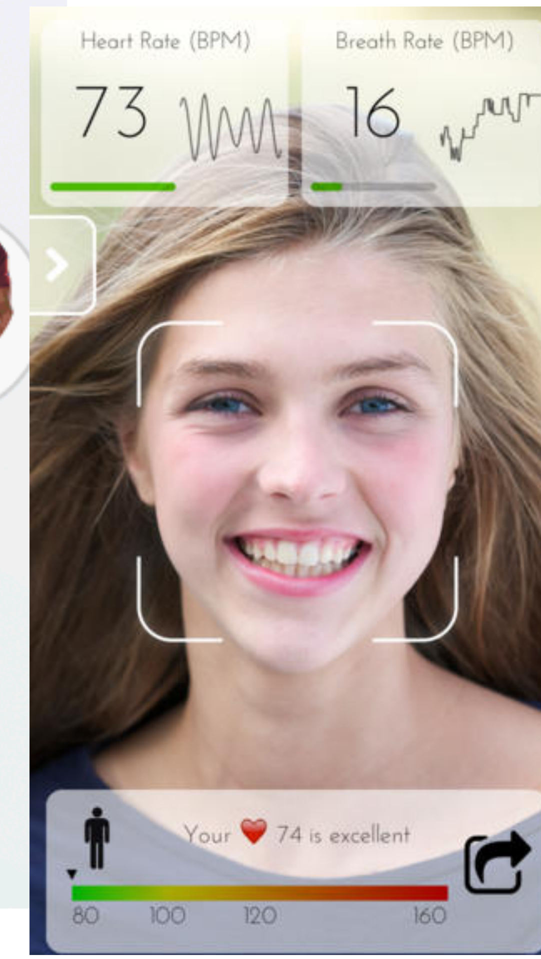


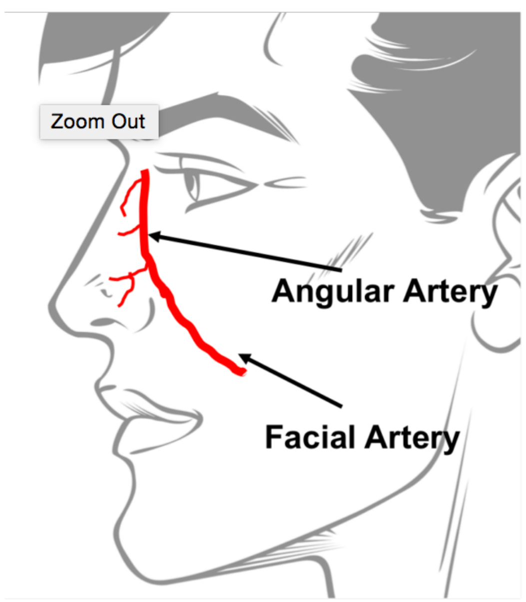
Emergency Room

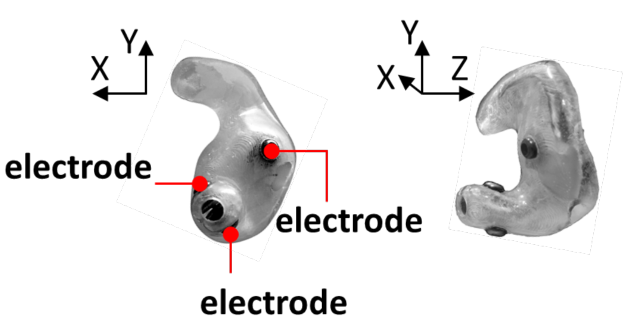
Sleep Lab

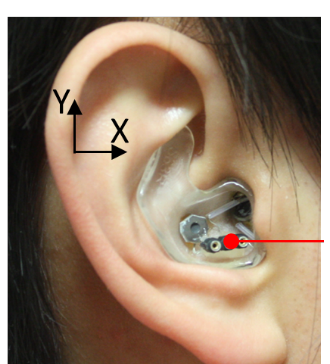
Also everywhere outside hospital

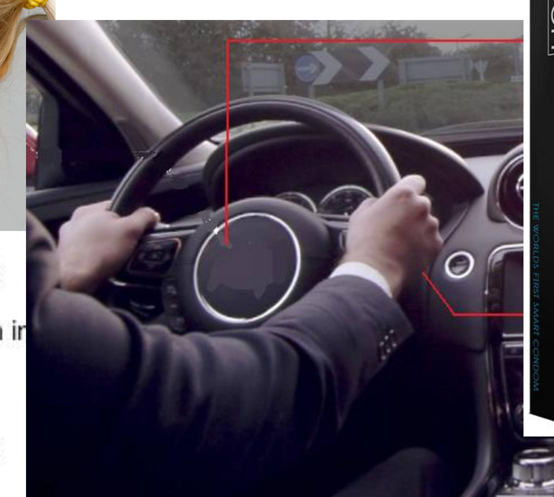






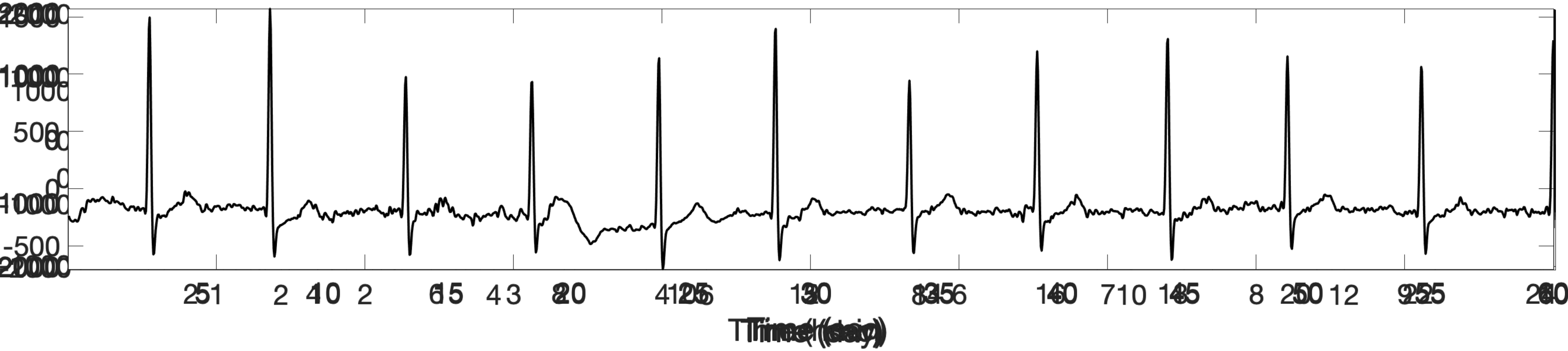






Motivative clinical application :

How to visualize ultra-long signals, like ECG in ICU?



14 days ECG = 120,960 segments to read
Good luck! :)

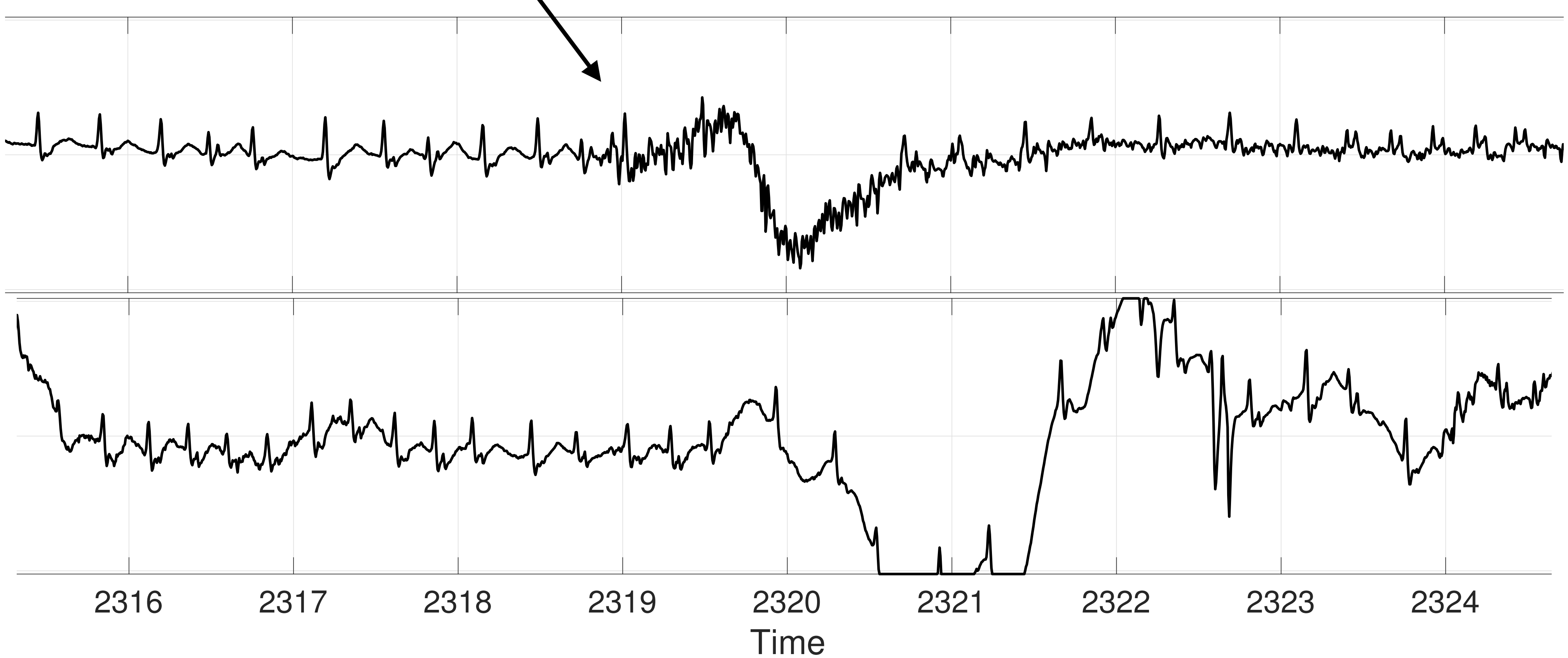
Q: How to summarize/reduce the dimension?

A typical example

- Large-volume
- High dimensional
- Noisy
- Non-stationary

14 days ECG ~ 10^6 beats
3,000 cases ~ $3x10^9$ beats
Only ECG. More channels?

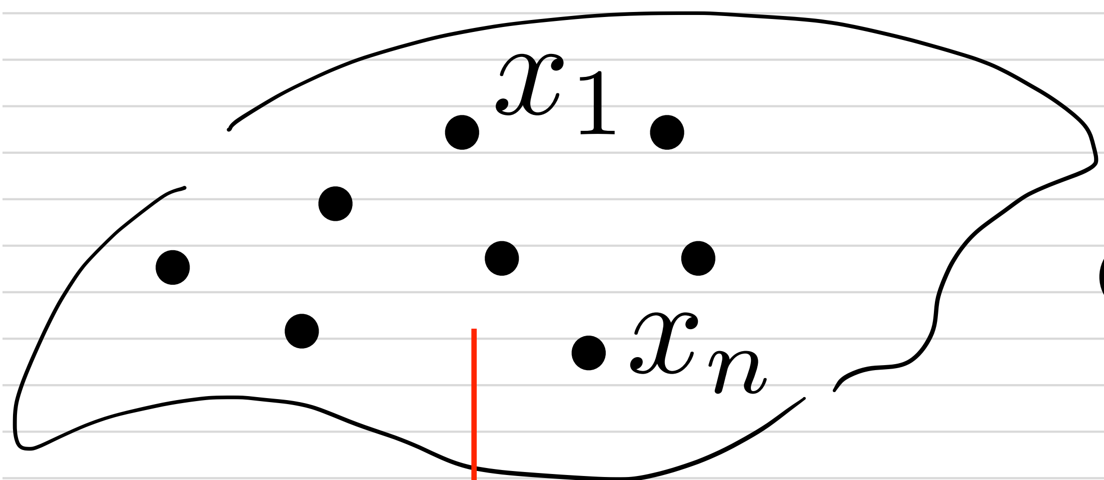
Sampling rate = 1000Hz
Each beat ~ 1000 dim



- One motivative problem
- Manifold learning — algorithm
- Manifold learning — theory
- Random matrix theory
- Toward manifold + RMT
- Some more…

General manifold learning problem

Complicated data structure
Modeled by a low dimensional manifold



$\subset \mathbb{R}^p$ High dimensional data

Data sampled
from manifold

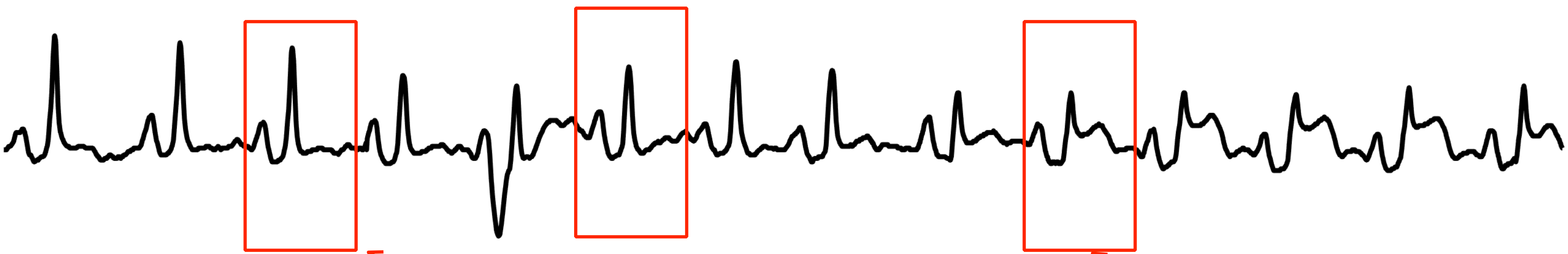
$$\left\{ \begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix} \right\} \subset \mathbb{R}^p$$

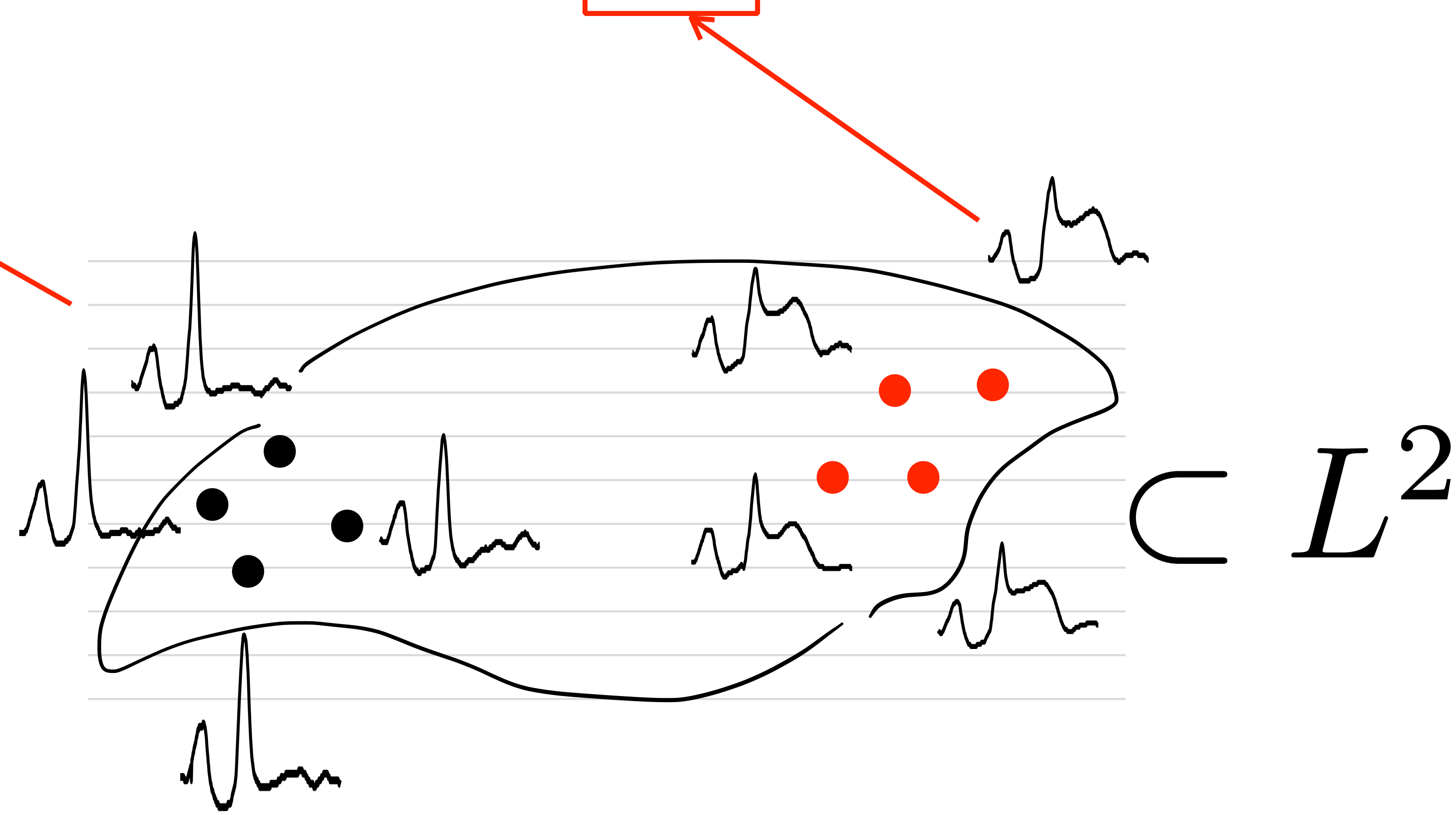
High dimensional
data set (usually noisy)

Can be more general, like bundle, metric space ….
But we will focus on this challenging enough case

*Lin & Malik & **W.**, 2019, arXiv*

Back to the ECG example
Wave-shape manifold model



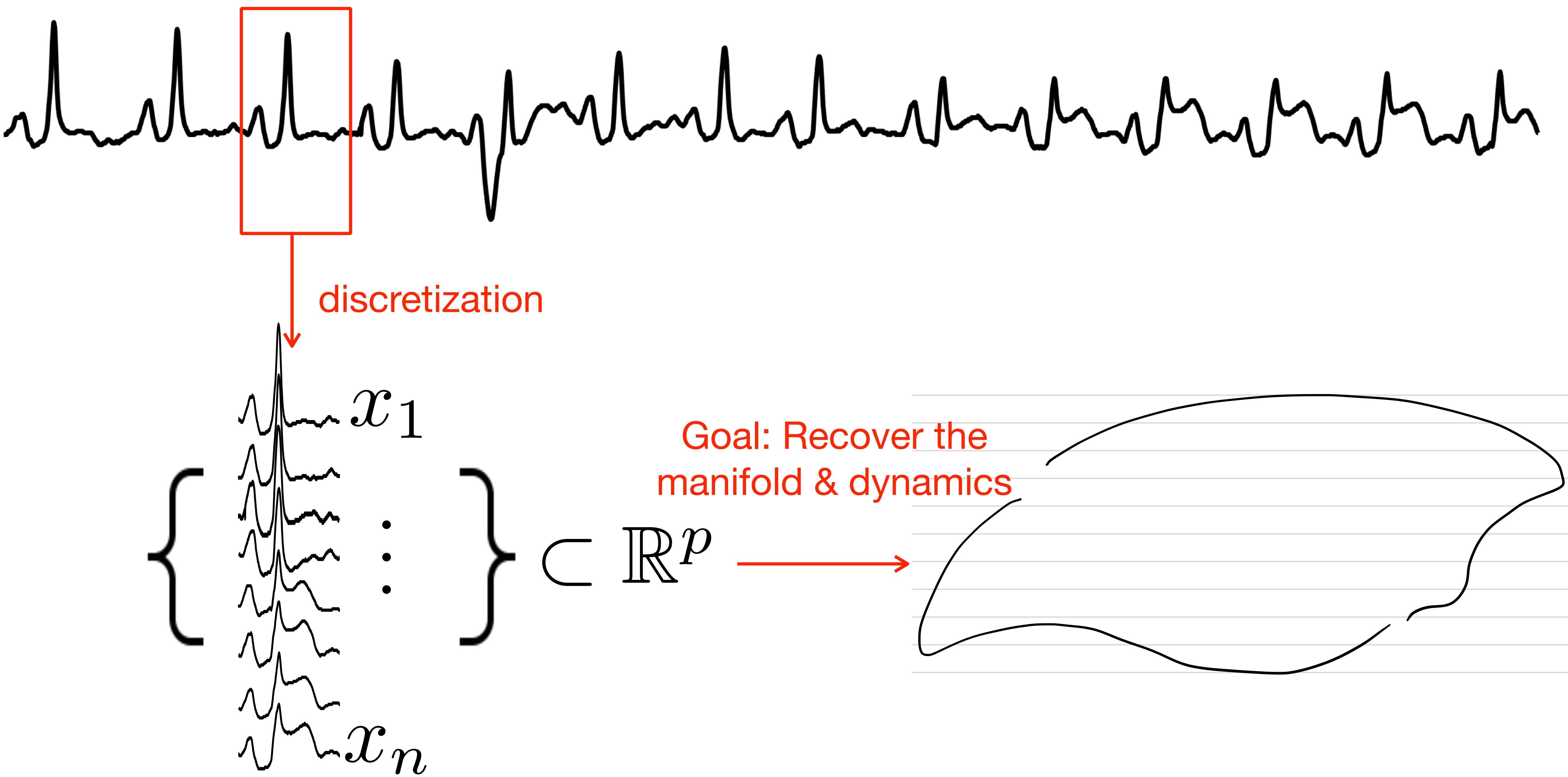


- **Physiological facts:**
  there is a variation but it
  does not go crazy or random

- Model/parametrize wave-shapes
  by a low-dim manifold

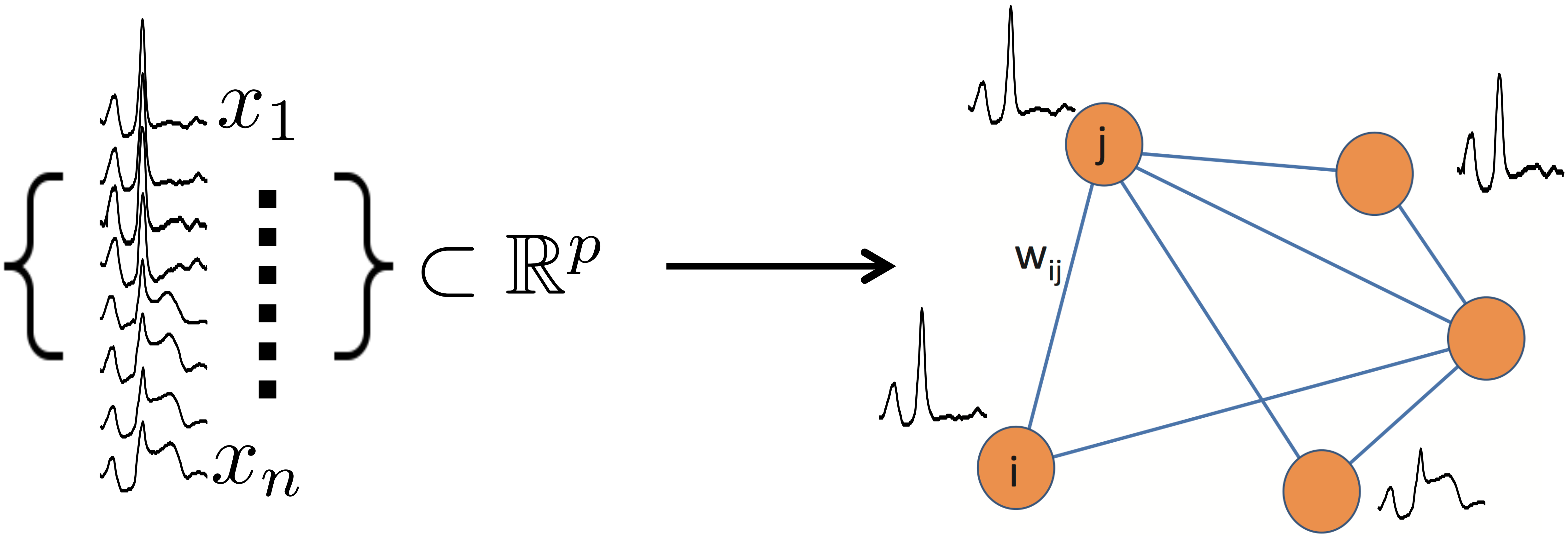
Back to the ECG example

Data preparation step



High dimensional
data set/point cloud

Construct affinity graph

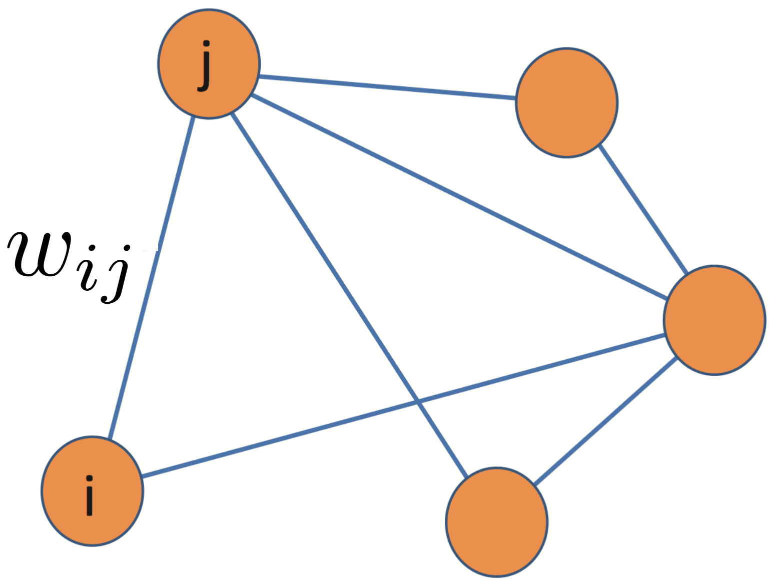


$$\text{affinity}(\text{beat}_i, \text{beat}_j) = e^{-\|\text{beat}_i - \text{beat}_j\|_{L^2}/\epsilon} =: w_{ij}$$

Similar beats have smaller distance & larger affinity

Can consider more complicated metric & kernel
But we will focus on this simple case

Construct graph Laplacian (GL)



*n*x*n* affinity matrix

$$W(i,j) = \begin{cases} w_{ij} & (i,j) \in \mathbb{E} \\ 0 & \text{otherwise} \end{cases}$$

*n*x*n* diagonal degree matrix

$$D(i,i) = \sum_{k=1}^{n} w_{ij}$$

The normalized graph Laplacian

$$I - D^{-1}W$$

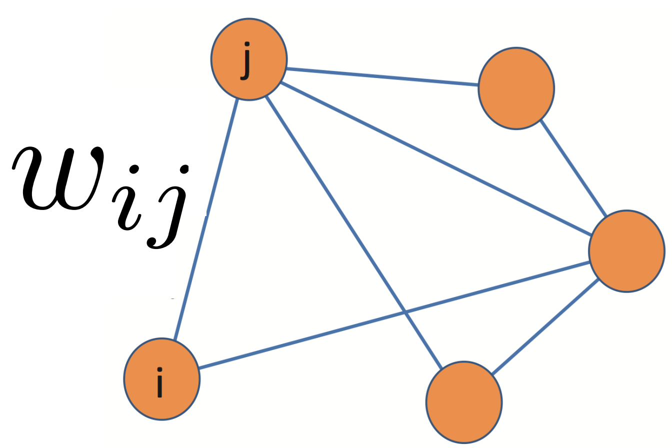
Transition matrix!
Random walk on the graph

$$D^{-1}Wv(i) = \frac{\sum_{j=1}^{n} w_{ij} v_j}{\sum_{j=1}^{n} w_{ij}} = \sum_{j=1}^{n} \left[\frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}} \right] v_j$$

Diffusion map (DM)

Coifman & Lafon, ACHA 2006

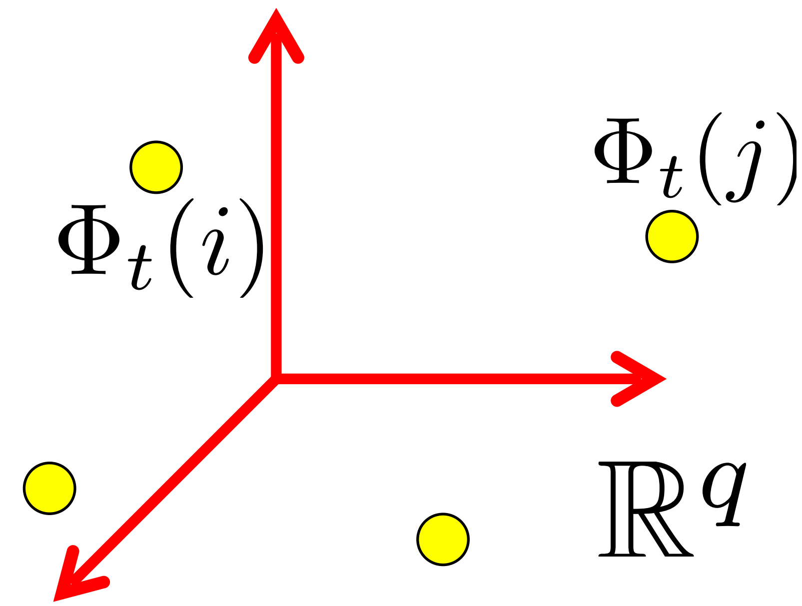
affinity graph



Take t>0

Diffusion map Φ_t

$$\Phi_t(i) = [e^{-t\lambda_l} u_l(i)]_{l=2}^{q+1} \in \mathbb{R}^q$$



$$I - D^{-1}W = \sum_{l=1}^{n} \lambda_l u_l v_l^\top$$

Eigendecomposition

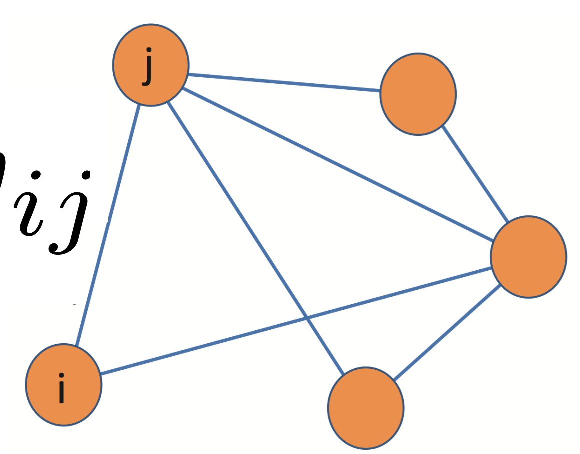
- Visualize high-dim data
- Dimensional reduction
- Recover nonlinear geometry
- Robust to noise! (later)

Diffusion map (DM)

Coifman & Lafon, ACHA 2006

affinity graph

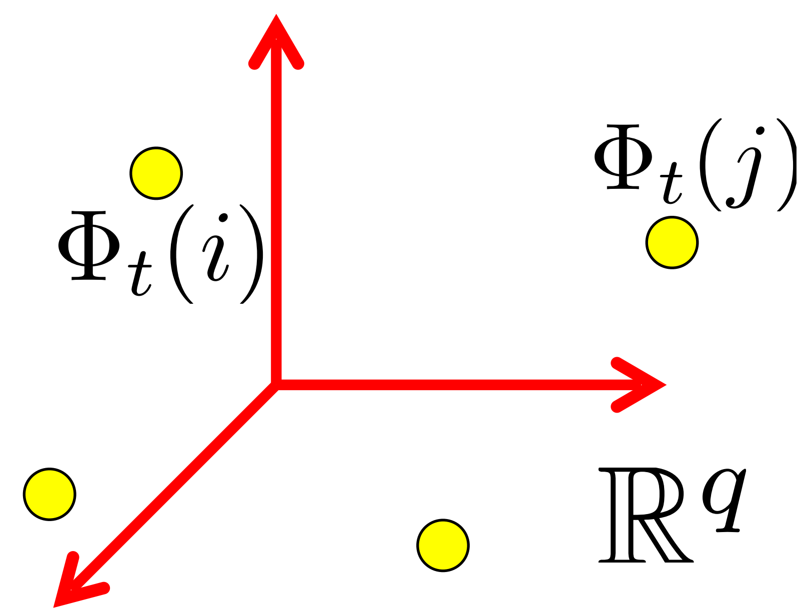
w_{ij}



Take t>0

Diffusion map Φ_t

$$\Phi_t(i) = [e^{-t\lambda_l} u_l(i)]_{l=2}^{q+1} \in \mathbb{R}^q$$



- Large-volume
- High dimensional
- Noisy
- ~~Non-stationarity~~

- Visualize high-dim data
- Dimensional reduction
- Recover nonlinear geometry
- Robust to noise! (later)

Big data v.s. computation

*Chao & **W.**, 2019 arXiv*
*Chao & Lin & **W.**, 2020 biorXiv*

1. Eigendecomposition is expensive. $O(n^{2.89})$
2. Existing solutions – (1) kNN; (2) Nystrom; (3) randomized.

kNN:
1. Good if the data is clean, with theoretical supports.
2. But not robust to noise.

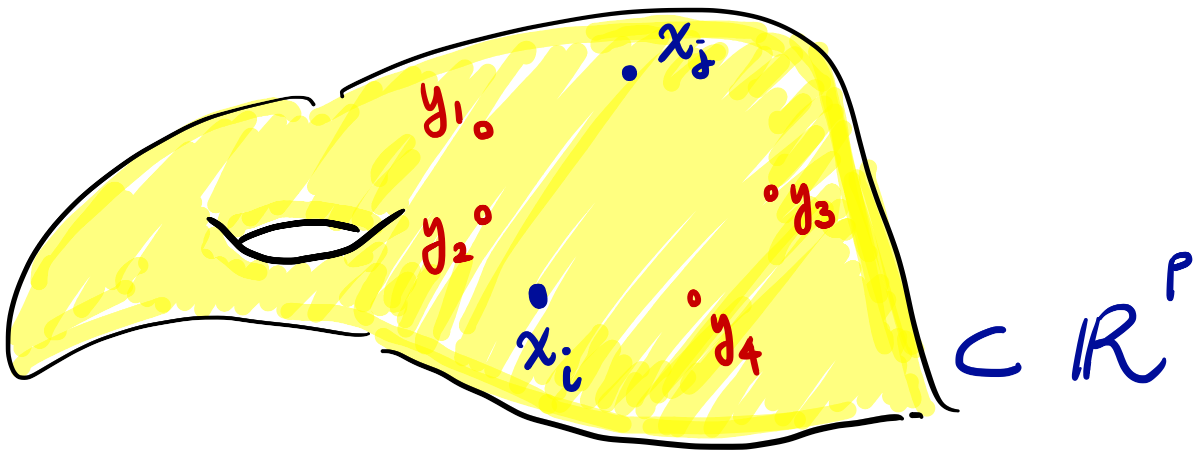
Nystrom:
1. Loss geometric information.
2. Good for spectral clustering (a lot of applications)

randomization:
1. Do you like to hear "with high probability" when you are seeing a doctor?
2. Under theoretical exploration for geometric information retrieval …

Chao & **W.**, 2019 arXiv
Chao & Lin & **W.**, 2020 biorXiv

Our solution — Roseland

RObust & ScalablE LANdmark Diffusion

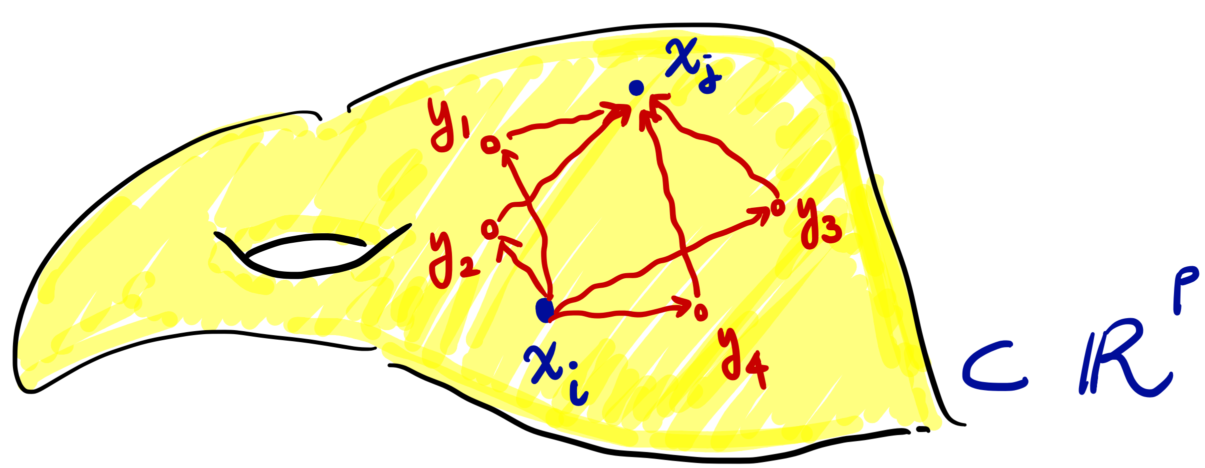


$$\{y_j\}_{j=1}^{m} \subset M$$

Landmark set

$$W_{ij} := \sum_{k=1}^{m} e^{-\|x_i - y_k\|_2/\epsilon} e^{-\|y_k - x_j\|_2/\epsilon}$$

Geometric interpretation — "landmark constraint" diffusion



Chao & **W.**, 2019 arXiv
Chao & Lin & **W.**, 2020 biorXiv

Algorithm & complexity analysis

(Step 1) $W^{(r)} \in \mathbb{R}^{n\times m}, \quad W^{(r)}_{ij} = W_{ij}$

(Step 2) $D^{(\mathrm{R})}_{ii} := e_i^\top W^{(\mathrm{r})}(W^{(\mathrm{r})})^\top \mathbf{1}$

(Step 3) $(D^{(\mathrm{R})})^{-1/2} W^{(\mathrm{r})} = U\Lambda V^\top$ (SVD)

(Step 4) $\bar{U} := (D^{(\mathrm{R})})^{-1/2} U$ (Eigenvectors for DM)

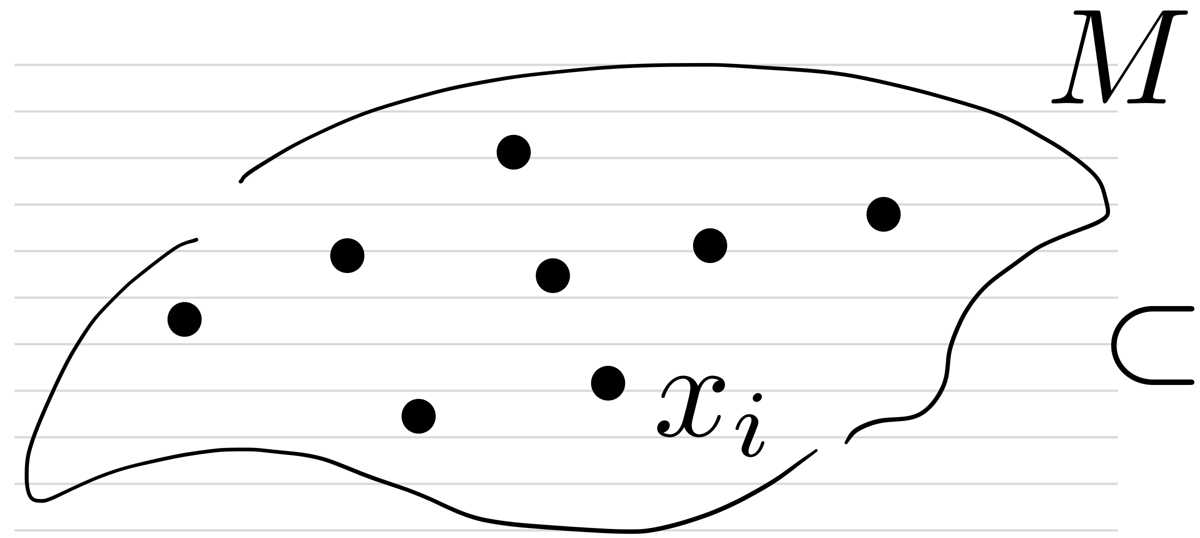
$$m = n^\beta$$

Roseland	$O(n^{1+2\beta})$
kNN-DM	$O(n^{2+\epsilon})$
Nystrom	$O(n^{1+\beta} + n^{3\beta})$

- One motivative problem
- Manifold learning — algorithm
- Manifold learning — theory
- Random matrix theory
- Toward manifold + RMT
- Some more…

Manifold setup

$$(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X}$$



$$\subset \mathbb{R}^p$$

M be a d-dimensional smooth, closed and connected Riemannian manifold isometrically embedded in $\mathbb{R}^p$ through $\iota : M \to \mathbb{R}^p$.

We should be able to sample everywhere of M; that is, the sampling density of *X* :

p_X satisfies $p_X \in \mathcal{C}^4(M^d)$ and $0 < \inf_{x \in M^d} p_X(x) \leq \sup_{x \in M^d} p_X(x)$.

Spectral convergence

Theorem
(Dunson & Wu & **W.** 2019 arXiv)

1. Suppose the kernel is Gaussian
2. Suppose λ_i is simple.
3. $\epsilon = \epsilon(n)$ so that $\epsilon \to 0$ and $\frac{\sqrt{-\log \epsilon}+\sqrt{\log n}}{\sqrt{n}\epsilon^{d/2}} \to 0$, as $n \to \infty$
4. Fix $K \in \mathbb{N}$
5. Assume $\sqrt{\epsilon} \leq \mathcal{K}_1 \min\left\{\left(\frac{\min(\Gamma_K,1)}{\mathcal{K}_2+\lambda_K^{d/2+5}}\right)^2, \frac{1}{(2+\lambda_K^{d+1})^2}\right\}$.

When ϵ is sufficiently small, $\exists\{a_n\}$ s.t. with probability $\geq 1-n^{-2}$, for all $i<K$, we have

$$\|a_n\phi_{\epsilon,n,i} - \phi_i\|_{L^\infty} = \mathcal{O}(\epsilon^{1/2}) + \mathcal{O}\left(\frac{\sqrt{-\log \epsilon}+\sqrt{\log n}}{\sqrt{n}\epsilon^{d+3/2}}\right)$$

$$|\lambda_{\epsilon,n,i} - \lambda_i| = \mathcal{O}(\epsilon^{3/4}) + \mathcal{O}\left(\frac{\sqrt{-\log \epsilon}+\sqrt{\log n}}{\sqrt{n}\epsilon^{d+5/2}}\right)$$

$\Delta\phi_l = -\lambda_l\phi_l$
$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$

Terrible rate… Numerically it converges faster. How to improve?

Finite spectral embedding

"Almost isometric embedding" via finite eigenfunctions

Theorem (Portegies CPAM 2015)

$$\Delta\phi_l = -\lambda_l \phi_l$$
$$0 = \lambda_0 < \lambda_1 \le \lambda_2 \le \dots$$
$$\varepsilon > 0: \text{ tolerable error}$$

then, $\exists t_0 = t_0(d, K, i, \varepsilon)$ such that $\forall 0 < t < t_0$, $\exists N_E = N_E(d, K, i, V, \varepsilon, t)$ such that if $N > N_E$, the spectral embedding

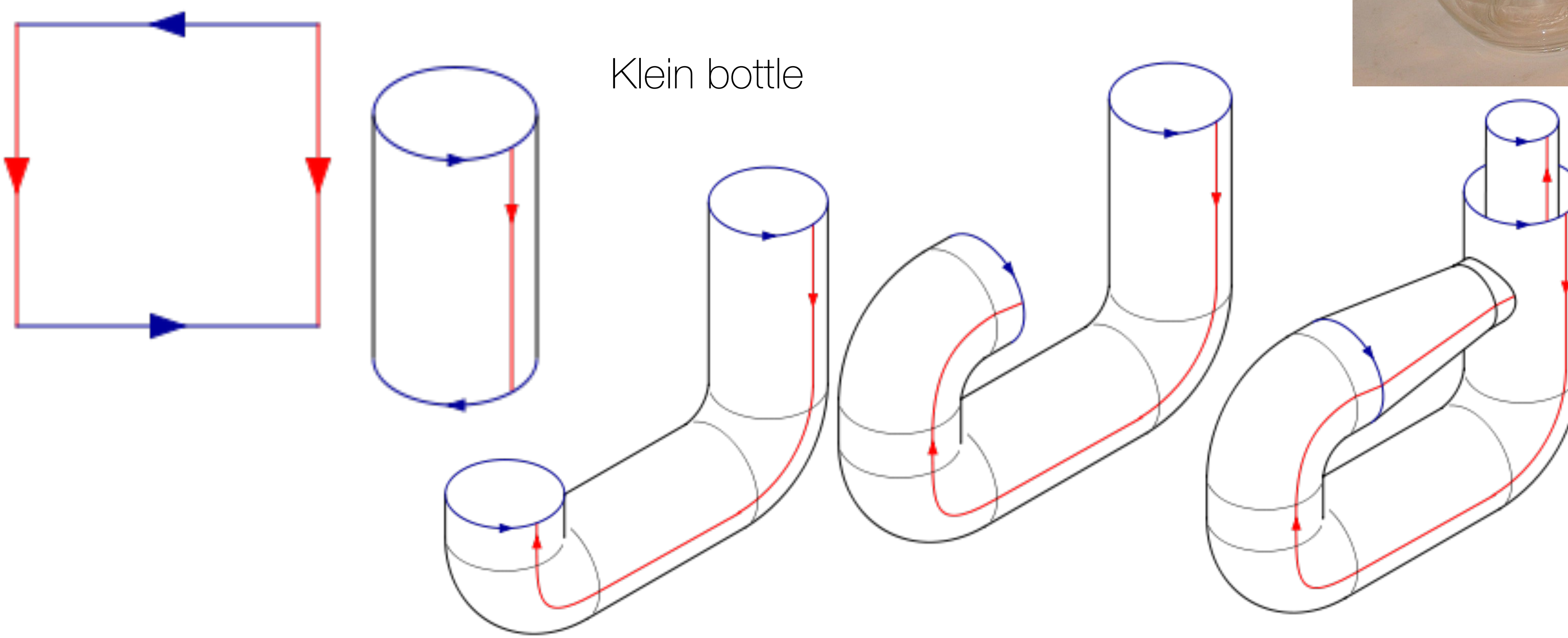
$$x \mapsto 2t^{(d+2)/4}\sqrt{2}(4\pi)^{d/4} \begin{bmatrix} e^{-\lambda_1 t}\phi_1(x) & \dots & e^{-\lambda_N t}\phi_N(x) \end{bmatrix}^\top$$

is almost isometric with the error controlled by ε

Have we solved problems? **NO!**

Q1: **Guarantee if dimension reduction can be achieved?**
by nonlinear embedding algorithms, like diffusion maps, LLE, ISOMAP, t-SNE, etc…, in the sense that the information is preserved geometrically/topologically?

Q2: **Is it possible to faithfully visualize the data?**

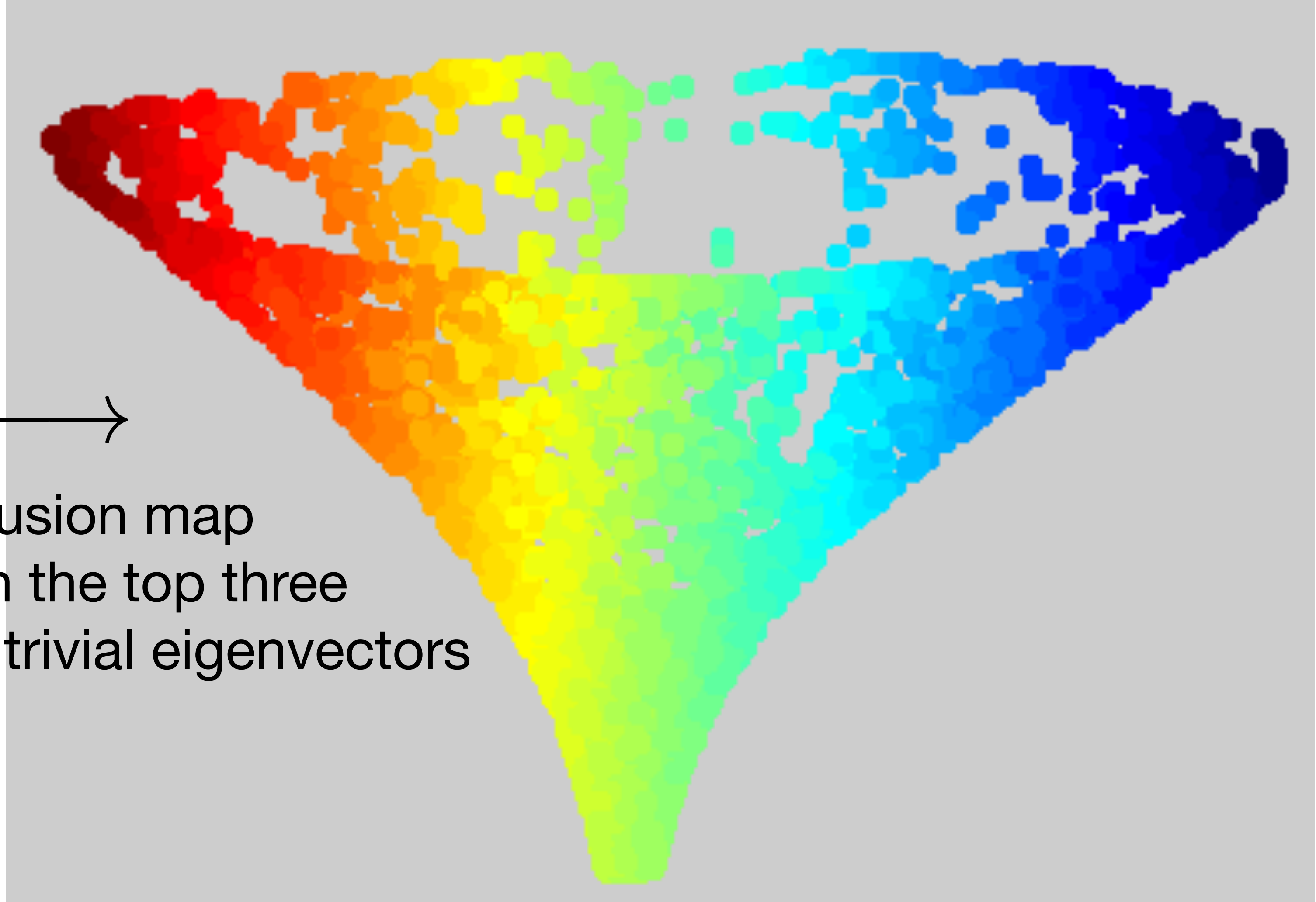


Klein bottle

DM of the Klein bottle

$$\left\{\Psi\left(\begin{bmatrix}u_i\\ v_i\end{bmatrix}\right)\right\}_{i=1}^{n} \subset \mathbb{R}^4 \longmapsto$$

Diffusion map
with the top three
nontrivial eigenvectors



No matter what algorithm you use or how hard you try,
you cannot visualize the Klein bottle in 3-dim Euclidean space.

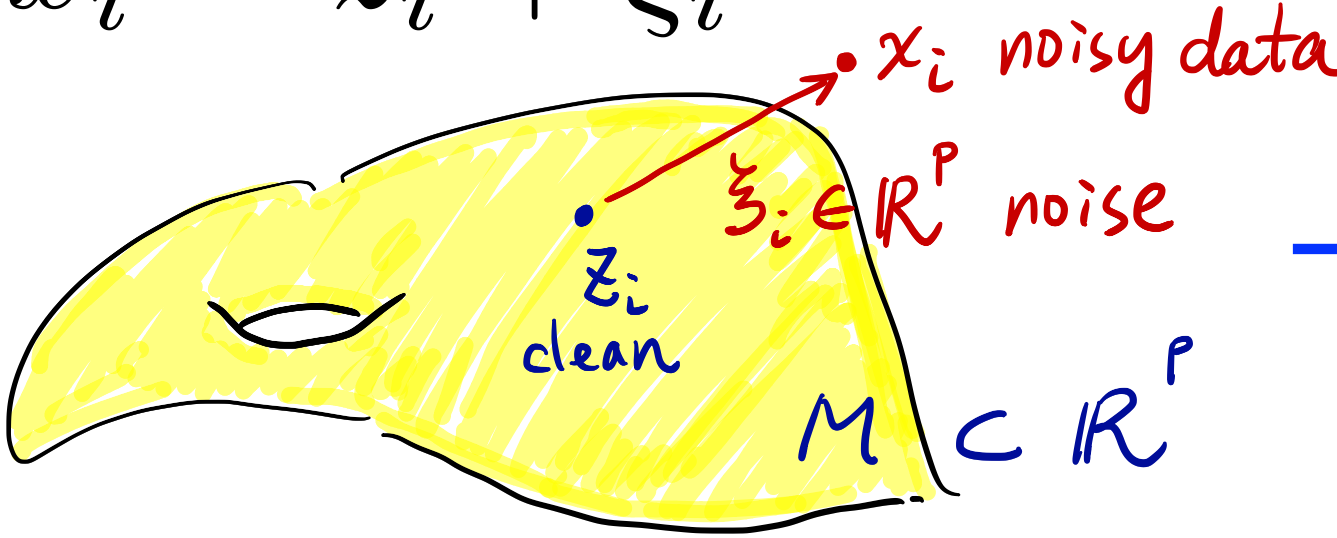
Data analysis cannot break the topological constraint!

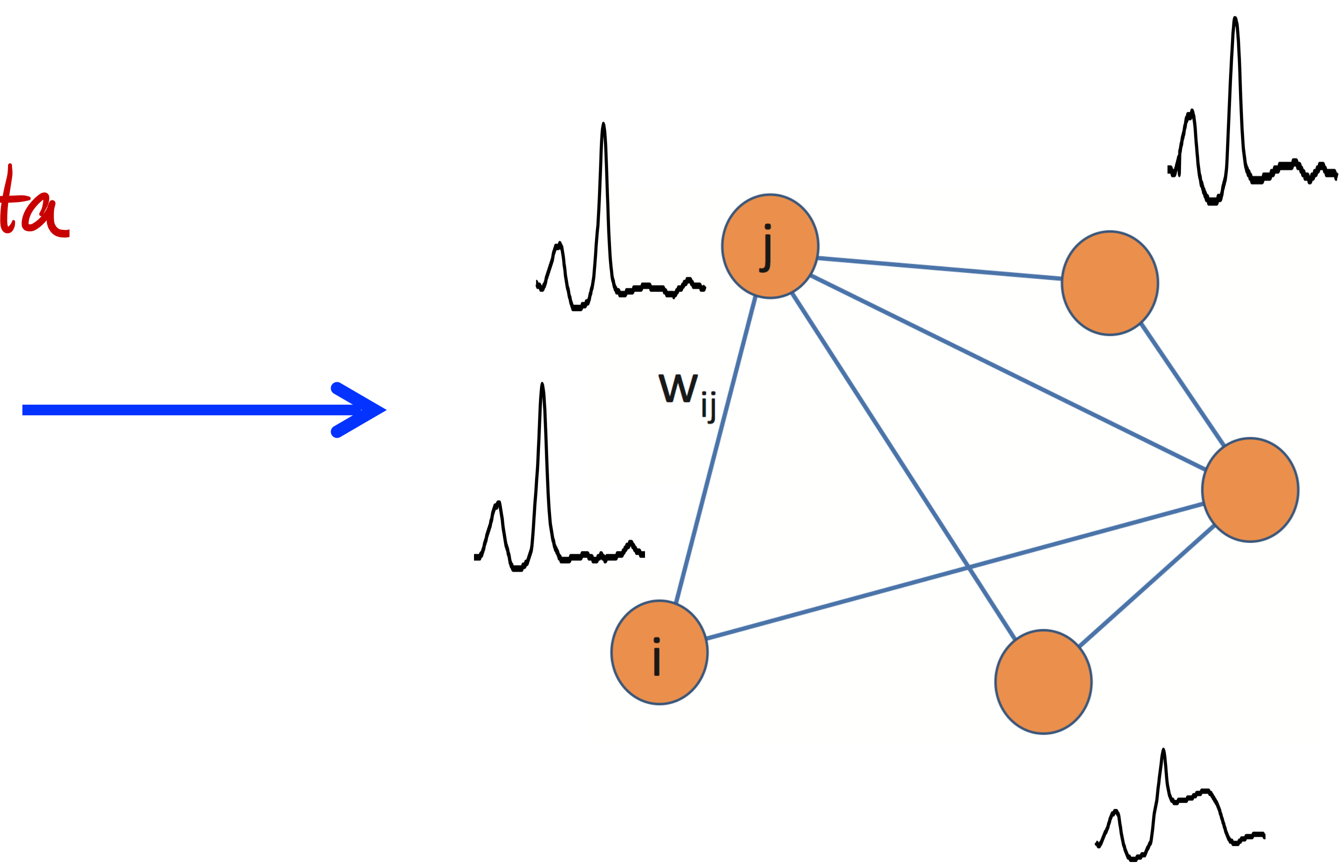
- One motivative problem
- Manifold learning — algorithm
- Manifold learning — theory
- Random matrix theory
- Toward manifold + RMT
- Some more…

Whole spectrum is left unanswered

Where is the promised random matrix theory?

$$x_i = z_i + \xi_i$$



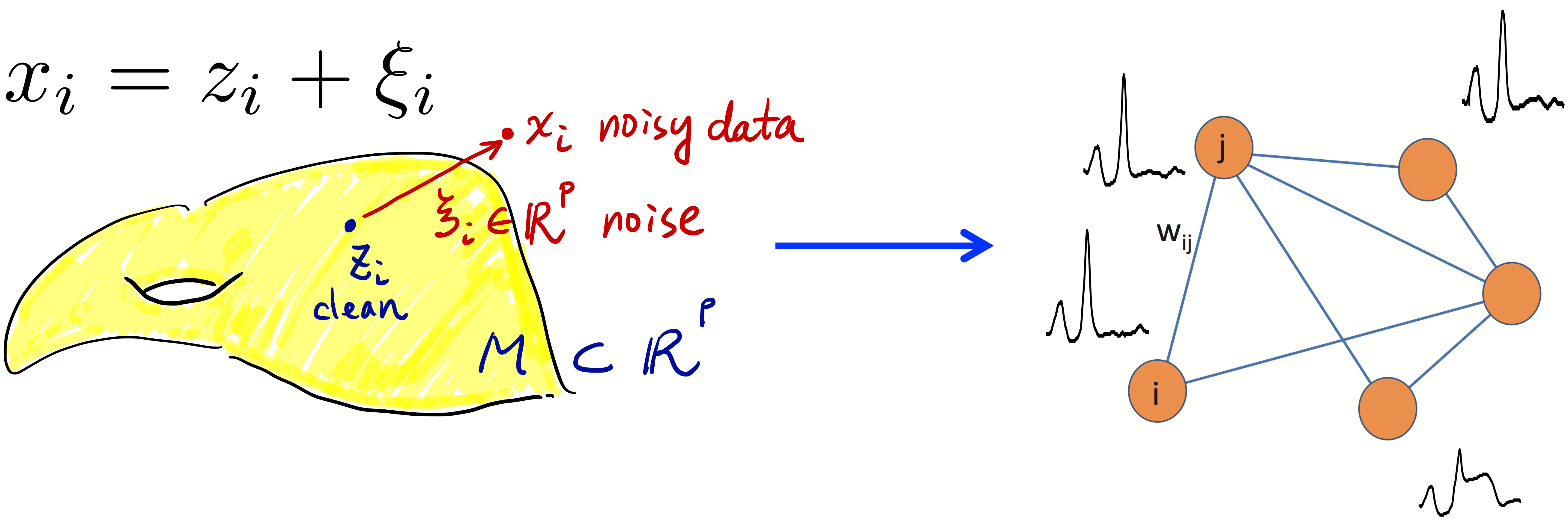


Q1: how does the spectrum of GL looks like from noisy data?

Whole spectrum is left unanswered

Where is the promised random matrix theory?

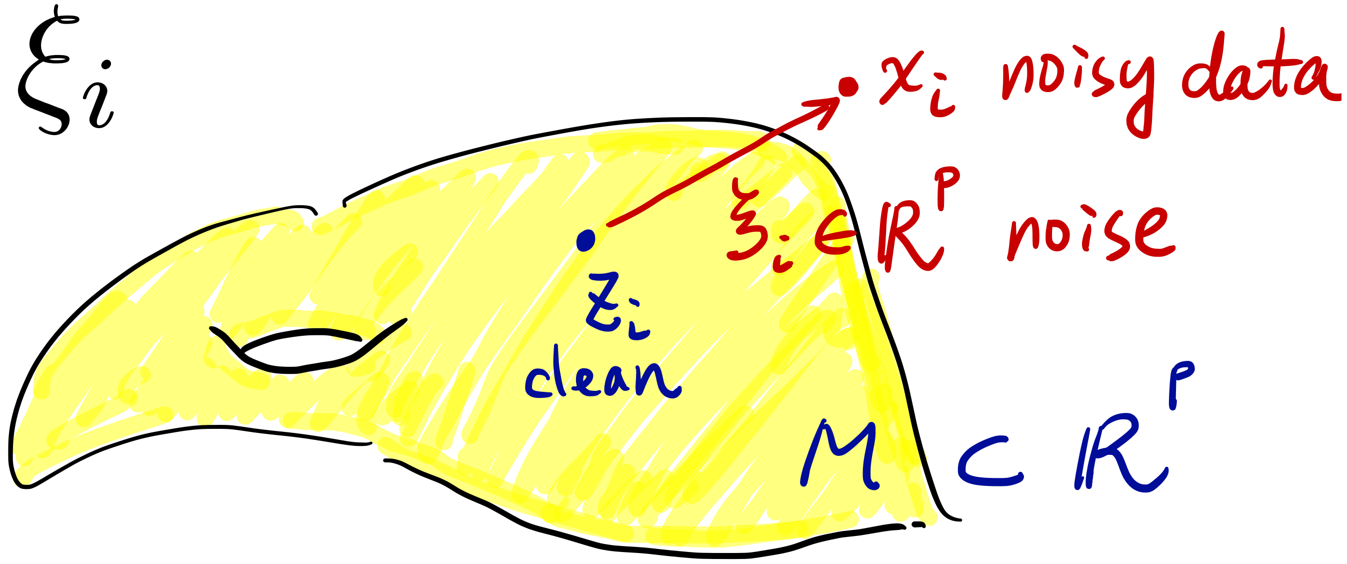
$$x_i = z_i + \xi_i$$



Q1: how does the spectrum of GL looks like from noisy data?

What is the problem? Everywhere!

$$x_i = z_i + \xi_i$$



$$\|x_i - x_j\|_2^2 = \|z_i - z_j\|_2^2 + \|\xi_i - \xi_j\|_2^2 + 2(z_i - z_j)^\top(\xi_i - \xi_j)$$

Q2: how can we find the "true neighbors"? True similarity?
How to compare objects when the data is noisy?

(Ding & **W.** 2020, submitted)

A simplified model

i.i.d. ~~sub~~-Gaussian sequences of $\mathcal{Y} := \{\mathbf{y}_i\}_{i=1}^n \in \mathbb{R}^p$ satisfies

$$\mathbb{E}(\mathbf{y}_i) = \mathbf{0}, \ \operatorname{cov}(\mathbf{y}_i) = \mathbf{I}_p.$$

$$\Sigma = \operatorname{diag}\{\lambda + 1, 1, \cdots, 1\}.$$

$\lambda := \lambda(n) \geq 0$, and when $\lambda > 0$ we consider

$$\lambda \asymp n^\alpha, \ 0 \leq \alpha < \infty.$$

The random point cloud is $\mathcal{X} := \{\mathbf{x}_i\}_{I=1}^n$, where

$$\mathbf{x}_i = \Sigma^{1/2}\mathbf{y}_i.$$

(Ding & **W.** 2020, submitted)

"High dimensional" noise

Assume that for some constant $0 < \gamma \leq 1$, we have

$$\gamma \leq c_n := \frac{n}{p} \leq \gamma^{-1}.$$

NOTE: 1-dim linear manifold

$\mathbf{x}_i^{(1)} = \sqrt{\lambda}\mathbf{x}_i \circ \mathbf{e}_1$, $i = 1, \ldots, n$ be the *clean* cloud points

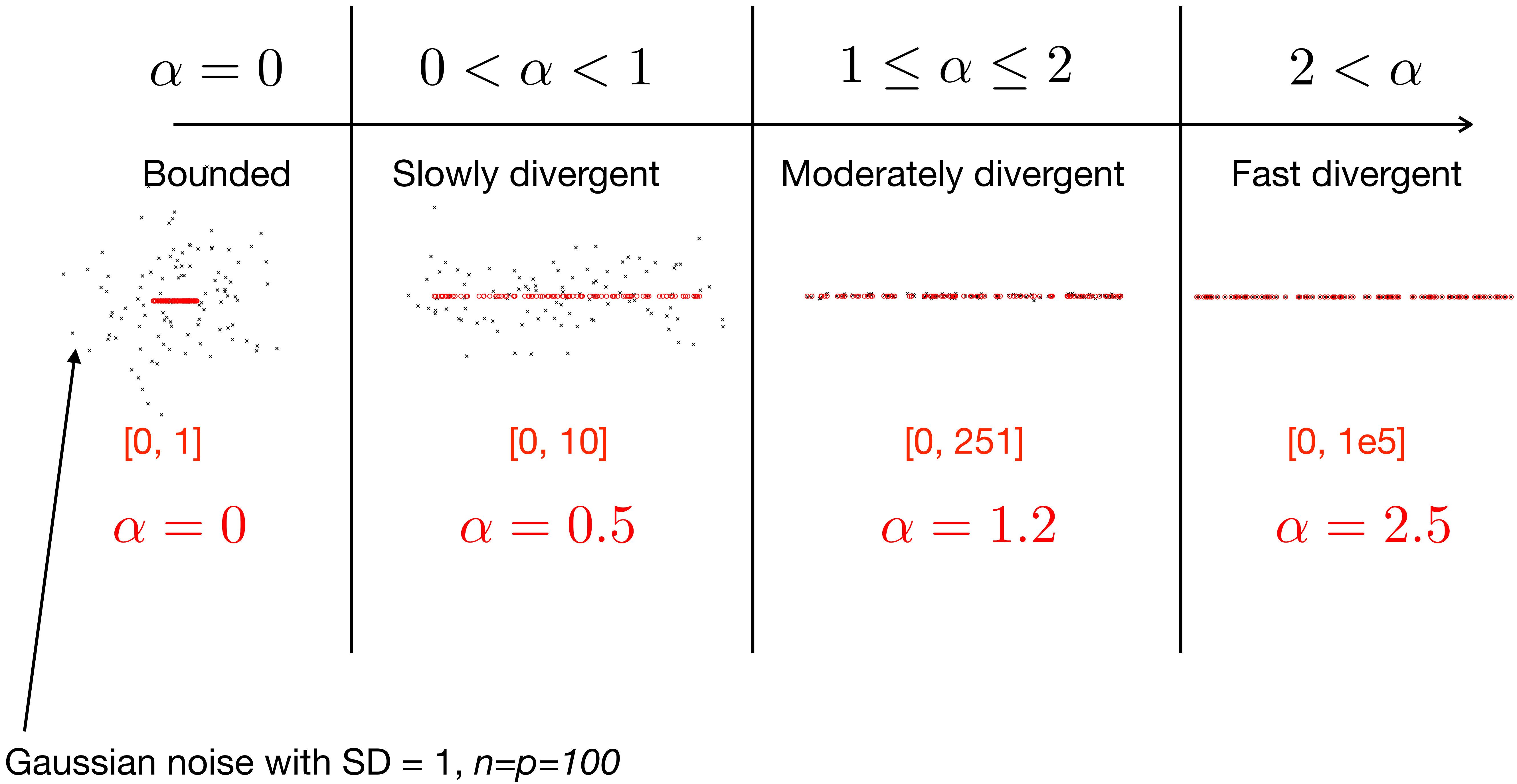
Clean signal

Noisy signal $$\mathbf{x}_i = \mathbf{x}_i^{(1)} + \mathbf{y}_i$$

Noise

(Ding & **W.** 2020, submitted)

Visual illustration / regimes



Question to ask…

Recall the *affinity/kernel matrix*:

$$\mathbf{W}(i,j) = \exp\left(-\upsilon \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{h}\right), \; 1 \leq i, j \leq n,$$

$h \equiv h(n)$: chosen bandwidth

$\upsilon > 0$: chosen parameter

Set $h = p$ ➡ $\mathbf{W}(i,j) = \exp\left(-\upsilon \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{p}\right)$

(Ding & **W.** 2020, submitted)

Question to ask…

$$\mathbf{x}_i^{(1)} = \sqrt{\lambda}\mathbf{x}_i \circ \mathbf{e}_1,\ i = 1, \ldots, n \text{ be the } \textit{clean} \text{ cloud points}$$

$$\mathbf{W}_1(i,j) = \exp\left(-\upsilon \frac{\|\mathbf{x}_i^{(1)} - \mathbf{x}_j^{(1)}\|^2}{p}\right),\ 1 \leq i, j \leq n,$$

From clean manifold (linear 1-dim)

What's the spectral relationship between $\mathbf{W}$ and $\mathbf{W}_1$?

Key observation & setup

Clean signal

Noisy signal $\mathbf{x}_i = \mathbf{x}_i^{(1)} + \mathbf{y}_i$

Noise

$$\mathbf{W} = \mathbf{W}_1 \circ \mathbf{W}_y$$

Classical result (pure noise, null)

For some small $\zeta > 0$ and $\xi > 0$, with probability at least $1 - O(n^{-1/2-\zeta})$, when n is sufficiently large, we have

$$|\mathbf{W}_y - (c_1 p^{-1}\mathbf{Y}^\top \mathbf{Y} + c_2 \mathbf{I}_n + R_3)\| \leq n^{-\xi},$$

Classical result (null case)

By the Taylor expansion at 2, when $i \neq j$,

$$\mathbf{W}_y(i,j) = f(2) + f'(2)\left[\mathbf{O}_y(i,j) - 2\mathbf{P}_y(i,j)\right] + \frac{f''(2)}{2}\left[\mathbf{O}_y(i,j) - 2\mathbf{P}_y(i,j)\right]^2 + \frac{f'''(\xi(i,j))}{6}\left[\mathbf{O}_y(i,j) - 2\mathbf{P}_y(i,j)\right]^3,$$

$$\mathbf{O}_y(i,j) = (1-\delta_{ij})\left(\frac{\|\mathbf{y}_i\|_2^2 + \|\mathbf{y}_j\|_2^2}{p} - 2\right)$$

$$\mathbf{P}_y(i,j) = (1-\delta_{ij})\frac{\mathbf{y}_i^\top \mathbf{y}_j}{p}$$

$$\mathbf{W}_y = f(2)\mathbf{1}\mathbf{1}^\top - \frac{2f'(2)}{p}\mathbf{Y}^\top\mathbf{Y} + f'(2)\mathbf{O}_y + 2f'(2)\left(\frac{1}{p}\mathrm{diag}(\|\mathbf{y}_1\|^2, \ldots, \|\mathbf{y}_n\|^2) - 1\right) + \frac{f''(2)}{2}\mathbf{H}_y + \frac{f'''(\xi(i,j))}{6}\mathbf{Q}_y + \varsigma(\lambda)\mathbf{I}$$

Classical result (null case)

$$\mathbf{W}_y = f(2)\mathbf{1}\mathbf{1}^\top - \frac{2f'(2)}{p}\mathbf{Y}^\top\mathbf{Y} + f'(2)\mathbf{O}_y + 2f'(2)\left(\frac{1}{p}\mathrm{diag}(\|\mathbf{y}_1\|^2, \dots, \|\mathbf{y}_n\|^2) - 1\right)$$
$$+ \frac{f''(2)}{2}\mathbf{H}_y + \frac{f'''(\xi(i,j))}{6}\mathbf{Q}_y + \varsigma(\lambda)\mathbf{I}$$

Higher order & small

$$\mathrm{Sh}_1(2) := f'(2)[\mathbf{1}\Phi^\top + \Phi\mathbf{1}^\top]$$

$$\Phi = (\phi_1, \dots, \phi_n) \text{ with } \phi_i = \frac{1}{p}\|\mathbf{y}_i\|_2^2 - 1$$

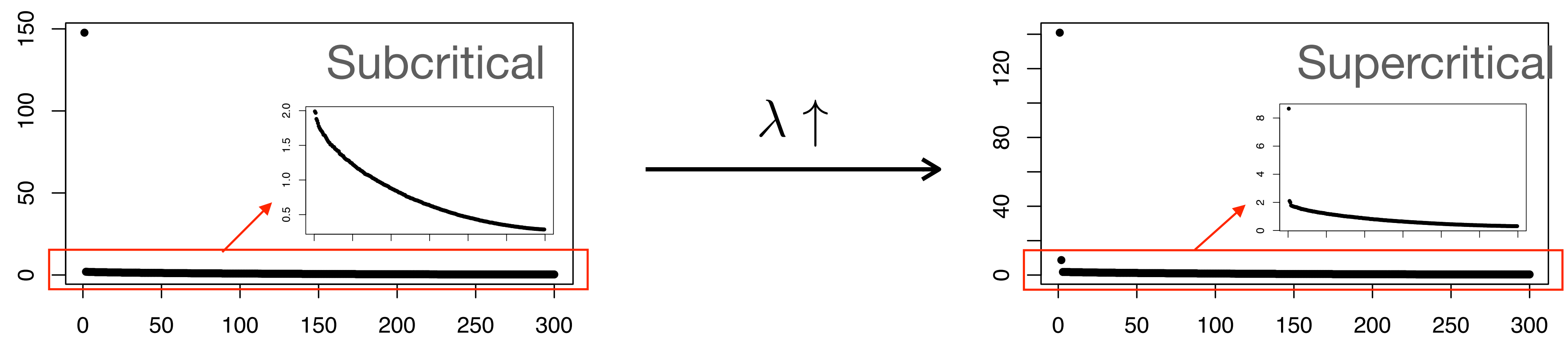
$$\mathrm{Sh}_2(2) := \frac{f''(2)}{2}\left[\mathbf{1}(\Phi \circ \Phi)^\top + (\Phi \circ \Phi)\mathbf{1}^\top + 2\Phi\Phi^\top + 4\frac{p+1}{p^2}\mathbf{1}\mathbf{1}^\top\right]$$

$$\mathbf{W}_y \approx \frac{2f'(2)}{p}\mathbf{Y}^\top\mathbf{Y}$$ up to a low/fixed rank perturbation with high probability!

The random matrix theory results kick in & help !

(Ding & **W.** 2020, submitted)

Theorem: transition phenomena



With probability $\geq 1 - O(n^{-1/2})$, when $0 < \lambda \leq \sqrt{c_n}$,

$$|\lambda_{i+3}(\mathbf{W}) - \gamma_{\nu_0}(i)| \leq \begin{cases} Cn^{-1/9+2\vartheta}, & 1 \leq i \leq C_1 n^{5/6+3\vartheta/2}; \\ Cn^{1/12+\theta} i^{-1/3}, & C_1 n^{5/6+3\vartheta/2} \leq i \leq n/2. \end{cases}.$$

Similar results hold for $i \geq n/2$.

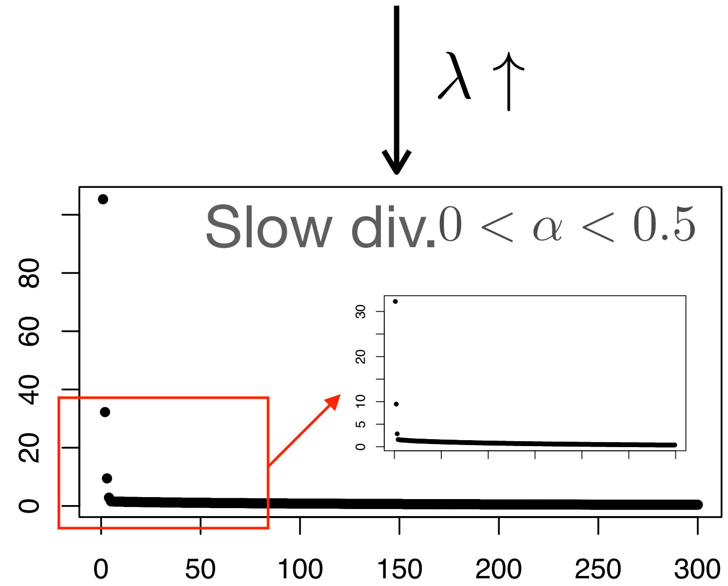
With probability $\geq 1 - O(n^{-1/2})$, when $\alpha = 0$ and $\lambda > \sqrt{c_n}$,

$$|\lambda_{i+4}(\mathbf{W}) - \gamma_{\nu_0}(i)| \leq \begin{cases} Cn^{-1/9+2\vartheta}, & 1 \leq i \leq C_1 n^{5/6+3\vartheta/2}; \\ Cn^{1/12+\theta} i^{-1/3}, & C_1 n^{5/6+3\vartheta/2} \leq i \leq n/2. \end{cases}.$$

Similar results hold for $i \geq n/2$.

(Ding & **W.** 2020, submitted)

Theorem: transition phenomena

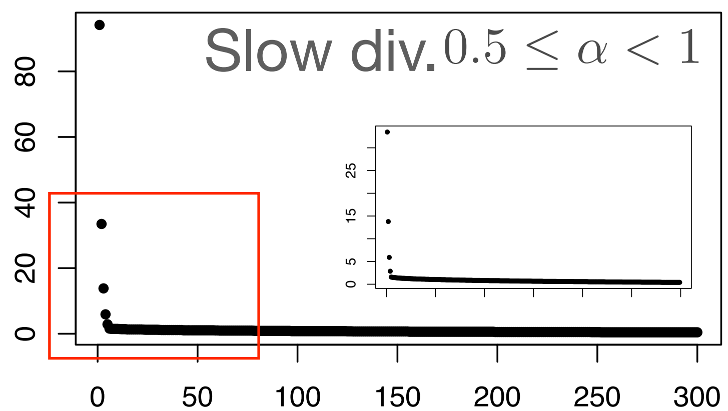


With probability $\geq 1 - O(n^{-1/2})$, when $0 < \alpha < 0.5$,

$$|\lambda_{i+4}(\mathbf{W}) - \gamma_{\nu_0}(i)| \leq \begin{cases} C\max\{n^{-1/9+2\vartheta}, n^{\theta}\frac{\lambda}{\sqrt{n}}\}, & 1 \leq i \leq C_1 n^{5/6+3\vartheta/2}; \\ C\max\{n^{1/12+\theta} i^{-1/3}, n^{\theta}\frac{\lambda}{\sqrt{n}}\}, & C_1 n^{5/6+3\vartheta/2} \leq i \leq n/2. \end{cases}$$

Similar results hold for $i \geq n/2$.

Theorem: transition phenomena



$\lambda \uparrow$

When $0.5 \le \alpha < 1$, denote $d \equiv d(\alpha) := \left\lceil \frac{1}{1-\alpha} \right\rceil + 1$. $\exists \mathsf{K}$ s.t. $5 \le \mathsf{K} \le C2^d$ so that with high probability, for all $1 \le i \le n - \mathsf{K}$, we have that

$$|\lambda_{i+\mathsf{K}}(\mathbf{W}) - \gamma_{\nu_0}(i)| \le C \max\left\{ i^{-1/3} n^{-2/3}, p^{\mathcal{B}(\alpha)}, \frac{\lambda}{p} \right\},$$

where $\mathcal{B}(\alpha) := (\alpha - 1)\left(\left\lceil \frac{1}{1-\alpha} \right\rceil + 1\right) + 1 < 0$.

(Ding & **W.** 2020, submitted)

Theorem: transition phenomena

For some constant $\mathsf{T} \leq \mathsf{C}_\alpha = \begin{cases} C\log n, & \alpha = 1; \\ \min\{Cn^{\alpha-1}, n\}, & 1 < \alpha \leq 2 \end{cases}$, with high probability,

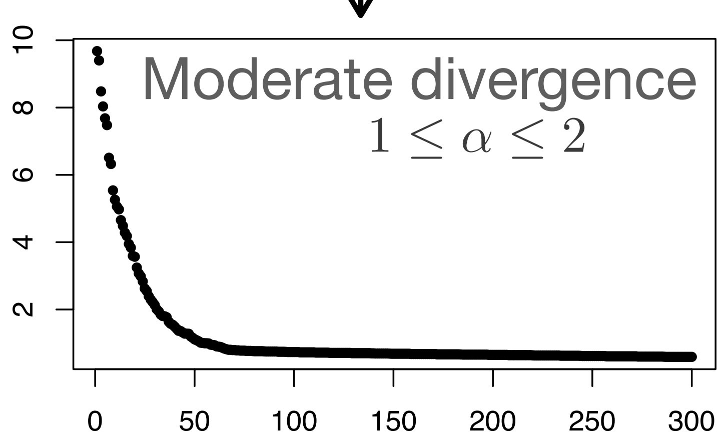
$$\sup_{1 \leq i \leq \mathsf{T}} \left| \lambda_i\left(\frac{1}{n}\mathbf{W}\right) - \lambda_i\left(\frac{1}{n}\mathbf{W}_{a_1}\right) \right| \leq n^{-\alpha/2},$$

where $\mathbf{W}_{a_1} := \exp(-2\upsilon)\mathbf{W}_1 + (1 - \exp(-2\upsilon))\mathbf{I}_n$ Moreover,

$$\sup_{z \in \mathcal{D}} |m_{\mathbf{W}}(z) - m_{\mathbf{W}_{a_2}}(z)| \prec \frac{1}{\sqrt{n}\eta^2},$$

where $\mathbf{W}_{a_2} := \left(\frac{2\upsilon\exp(-2\upsilon)}{p}\mathbf{Y}^\top\mathbf{Y} + 2\upsilon\exp(-4\upsilon)\mathbf{I}_n\right) \circ \mathbf{W}_1$, $\mathcal{D} \equiv \mathcal{D}(1/4, \mathsf{a}) := \left\{z = E + \mathrm{i}\eta : \mathsf{a} \leq E \leq \frac{1}{\mathsf{a}},\ n^{-1/4+\mathsf{a}} \leq \eta \leq \frac{1}{\mathsf{a}}\right\}$ and $0 < \mathsf{a} < 1$ is a small constant.

$\downarrow$ $\lambda \uparrow$



Theorem: transition phenomena

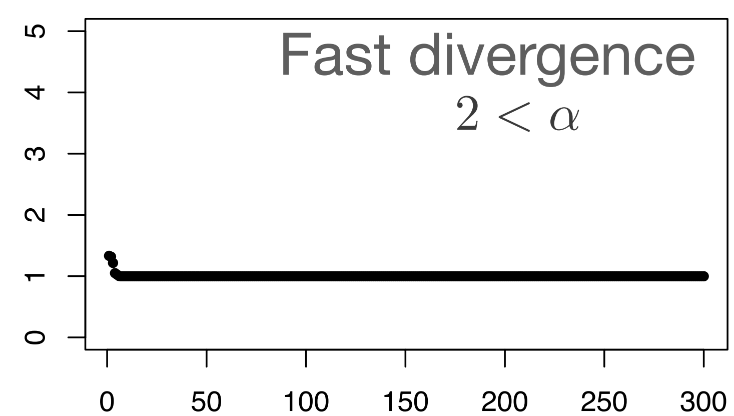
When $\alpha > 2$,

$$\sup_{1 \leq i \leq n} |\lambda_i(\mathbf{W}) - \lambda_i(\mathbf{W}_{a_1})| \prec n^{\frac{2-\alpha}{2}}.$$

Moreover, for any given constant $t \in (0,1)$, if $\alpha > \frac{2}{t}+1$, we have with probability $\geq 1 - n^{1-t(\alpha-1)/2}$,

$$\sup_{1 \leq i \leq n} |\lambda_i(\mathbf{W}) - 1| \leq n \exp(-v(\lambda/p)^{1-t}).$$

$\lambda \uparrow$



- One motivative problem
- Manifold learning — algorithm
- Manifold learning — theory
- Random matrix theory
- Toward manifold + RMT
- Some more…

High dimensional noise model

(Chao & **W.** 2019, arXiv)

Fix a compact smooth d-dim Riemannian manifold M, and assume D is the smallest dimension of the Euclidean space that M can be isometrically embedded into via ι.

Assume $q = q(n) \asymp n$ when $n \to \infty$

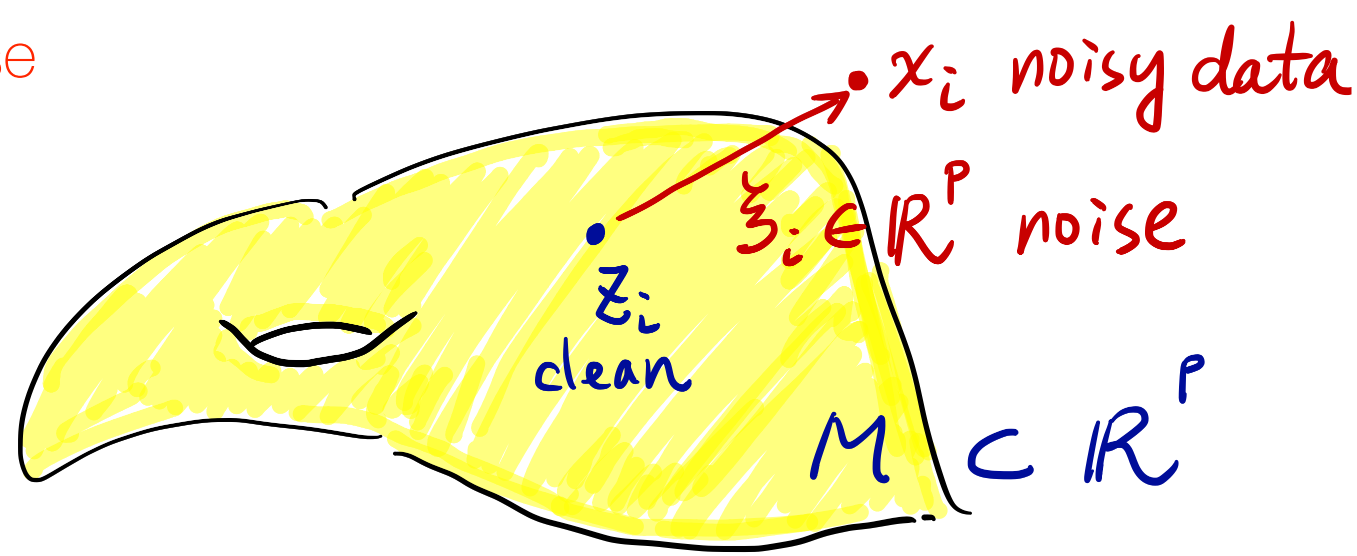
When q is sufficiently large, fix an isometric embedding $\bar{\iota}_q : \mathbb{R}^D \to \mathbb{R}^q$ so that $\bar{\iota}_q(e_i) \in \mathbb{R}^q$ satisfies $|u_i(k)| = 1/\sqrt{q} + O(1/q)$ for $i = 1, \ldots, D$ and $k = 1, \ldots, q$.

Assume x_i is i.i.d. sampled from $\bar{\iota}_q \circ \iota(M)$.

Clean manifold data

$$\tilde{x}_i = x_i + \xi_i$$

Technical condition for non-Gaussian noise
Needed for Gaussian approximation
Chernozhukov & Chetverikov & Kato, AoS 2013



High dimensional noise model

(Chao & **W.** 2019, arXiv)

$$\tilde{x}_i = x_i + \xi_i \qquad \mathbb{E}\xi_i = 0,\ \mathbb{E}\xi_i\xi_i^\top = \Sigma \in \mathbb{R}^{q\times q}$$

$$\|\Sigma\|_2 \le \sigma_q^2$$

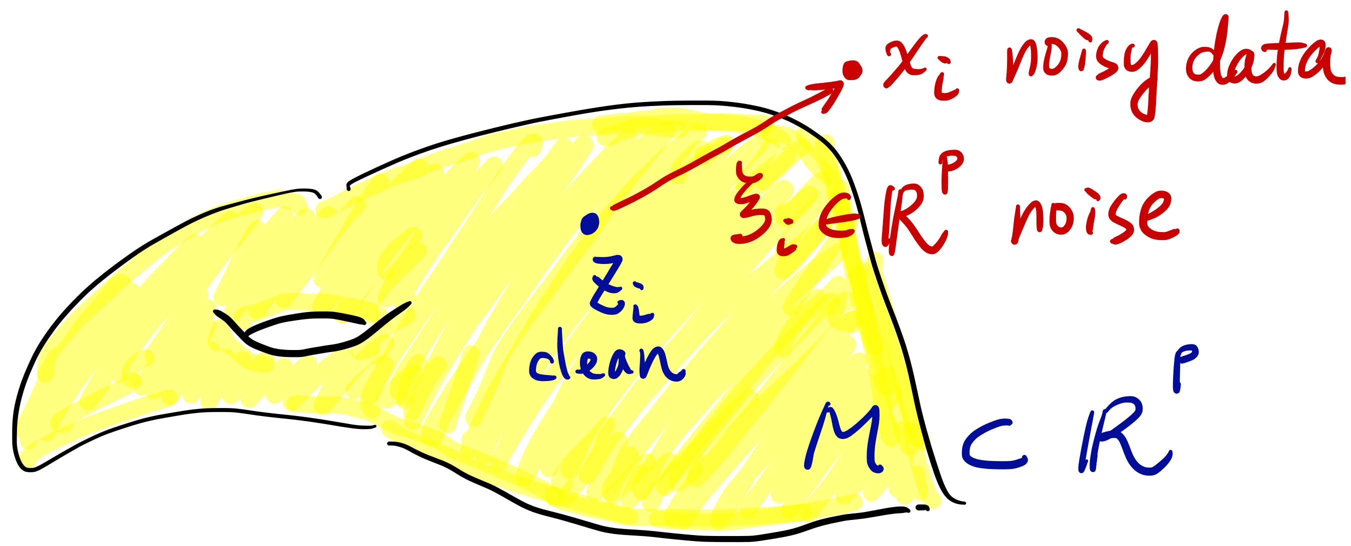
Assume for all convex 1-Lipschitz function f,

$$\mathbb{P}(|f(\xi_i) - m_{f(\xi_i)}| > t) \le 2\exp(-c_i t^2),$$

where $m_{f(\xi_i)}$ is the median of $f(\xi_i)$ and $c_i > 0$.

1. $c_1 \le \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}\xi_i(l)^2 \le C_1$ and $\max_{k=1,2} \frac{1}{q}\sum_{l=1}^q \mathbb{E}\big[\frac{|\xi_i(l)|^{2+k}}{B_q^k}\big] + \mathbb{E}\big[\frac{\exp(|\xi_i(l)|}{B_q})\big] \le$
   4. Moreover, $\frac{1}{q}B_q^2(\log(qn^2))^7 \le C_2 q^{-c_2}$.

2. $c_1 \le \frac{1}{n}\sum_{i=1}^n \mathbb{E}(\xi_i(l))^2 \le C_1$ and $\max_{k=1,2} \frac{1}{q}\sum_{i=1}^q \mathbb{E}\big[\frac{|\xi_i(l)|^{2+k}}{B_q^k}\big] + \mathbb{E}\big[(\max_{i=1,\ldots,n} \frac{|\xi_i(l)|}{B_q})\big] \le$
   4. Moreover, $\frac{1}{q}B_q^4(\log(qn^2))^7 \le C_2 q^{-c_2}$.



Robust to noise

Theorem (Chao & **W.** 2020)

1. For Gaussian noise, assume

$$\delta_q := \sigma_q \sqrt{\log n^2}\,[\sigma_q \sqrt{q} + K] \to 0$$

2. For non-Gaussian noise, assume $\sup_{i,j} \sqrt{(\sigma_q^2)/c_{ij}} \sqrt{\log n^2} \to 0$, and set

$$\delta_q := \sigma_q \sqrt{\log n^2} \left[\sup_{i,j} \sqrt{c_{ij}^{-1}}\,(\sigma_q \sqrt{q} \vee 1) + \sqrt{D} K \right]$$

3. Fix $q' \in \mathbb{N}$ and $t > 0$.
4. W and $\widetilde{W}$: affinity matrices from clean and noisy datasets respectively.
5. $\Phi L^t \in \mathbb{R}^{n \times q'}$ and $\widetilde{\Phi} \widetilde{L}^t \in \mathbb{R}^{n \times q'}$: DM from clean and noisy datasets respectively.

Then, when ϵ is sufficiently small,

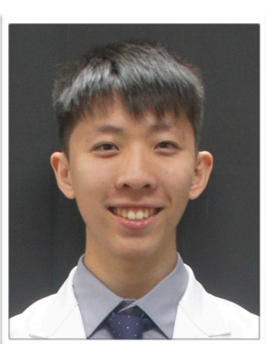
$$\|\Phi O L^t - \widetilde{\Phi} \widetilde{L}^t\|_F = \mathcal{O}_P \left(\frac{\delta_q \sqrt{q' \lambda_2^{2t}}}{\epsilon^{3d/4+1}} \right) .$$

where $O \in \mathbb{R}^{q' \times q'}$ is an orthogonal matrix, and λ_2 are the largest non-trivial eigenvalue from the clean.

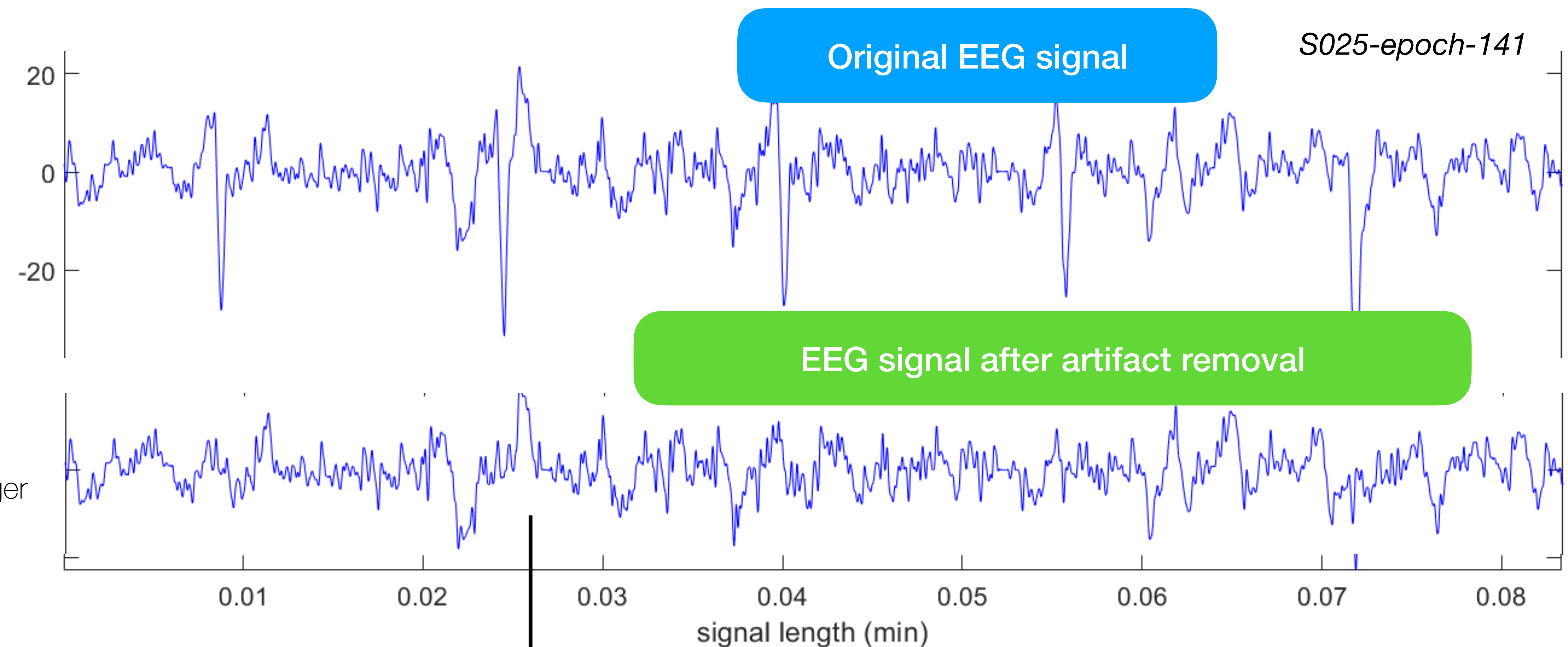
- One motivative problem
- Manifold learning — algorithm
- Manifold learning — theory
- Random matrix theory
- Toward manifold + RMT
- Some more…

One more application

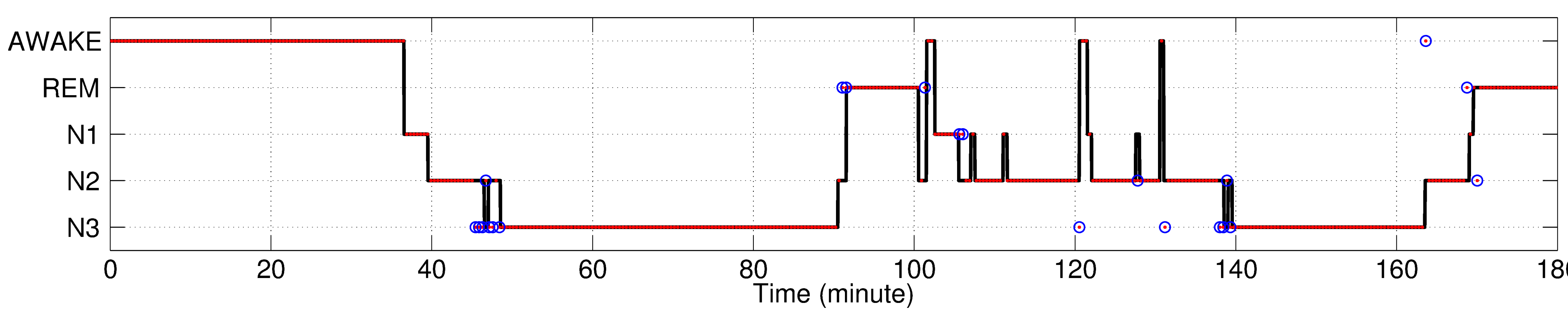
Caroline Lustenberger
Neuroscience
ETH



邱能態
Medicine
NYMU



Apply the automatic annotation



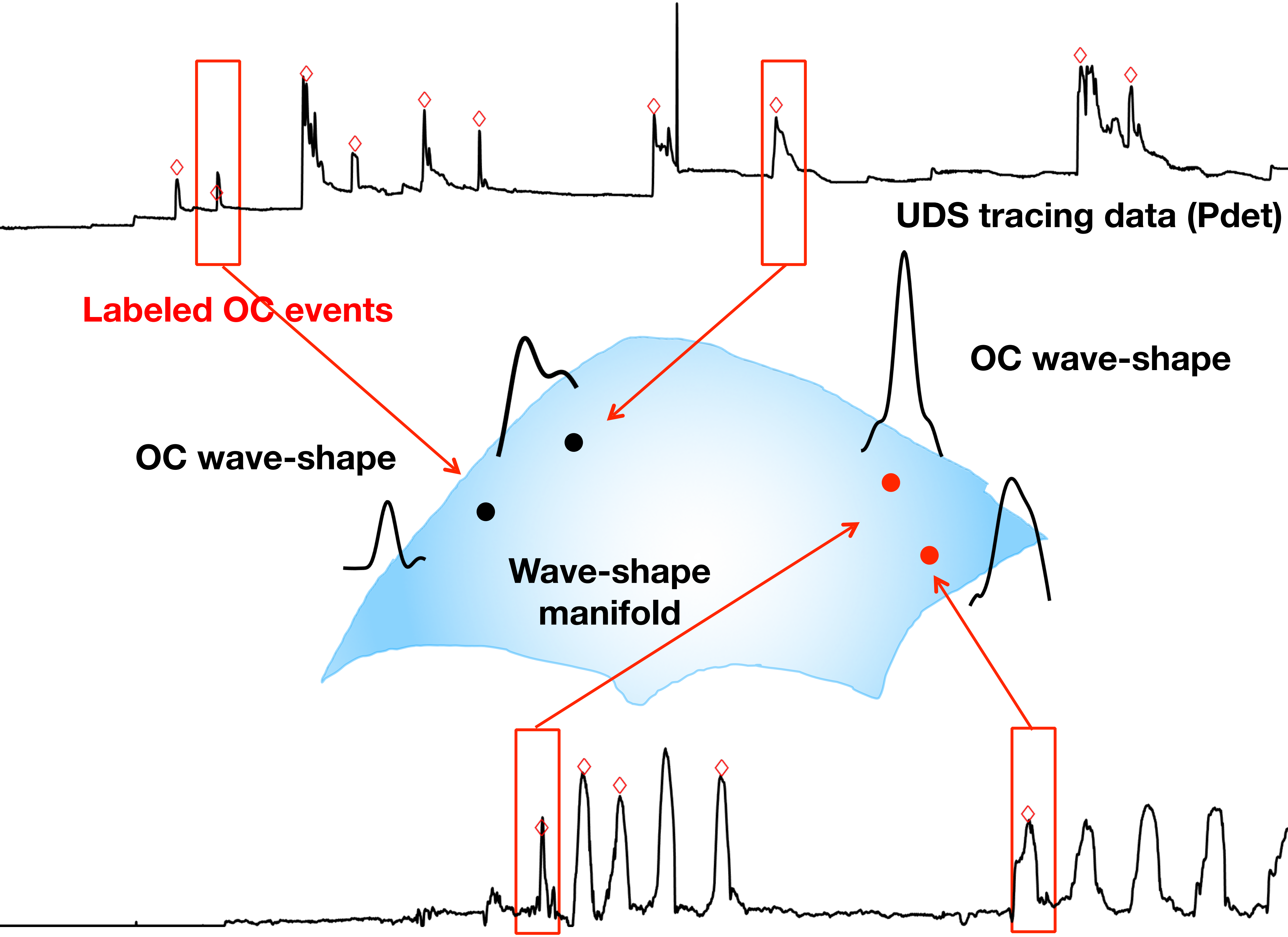
Wang & Cahill & Nelson & **W.** & Estrada, NAU 2020

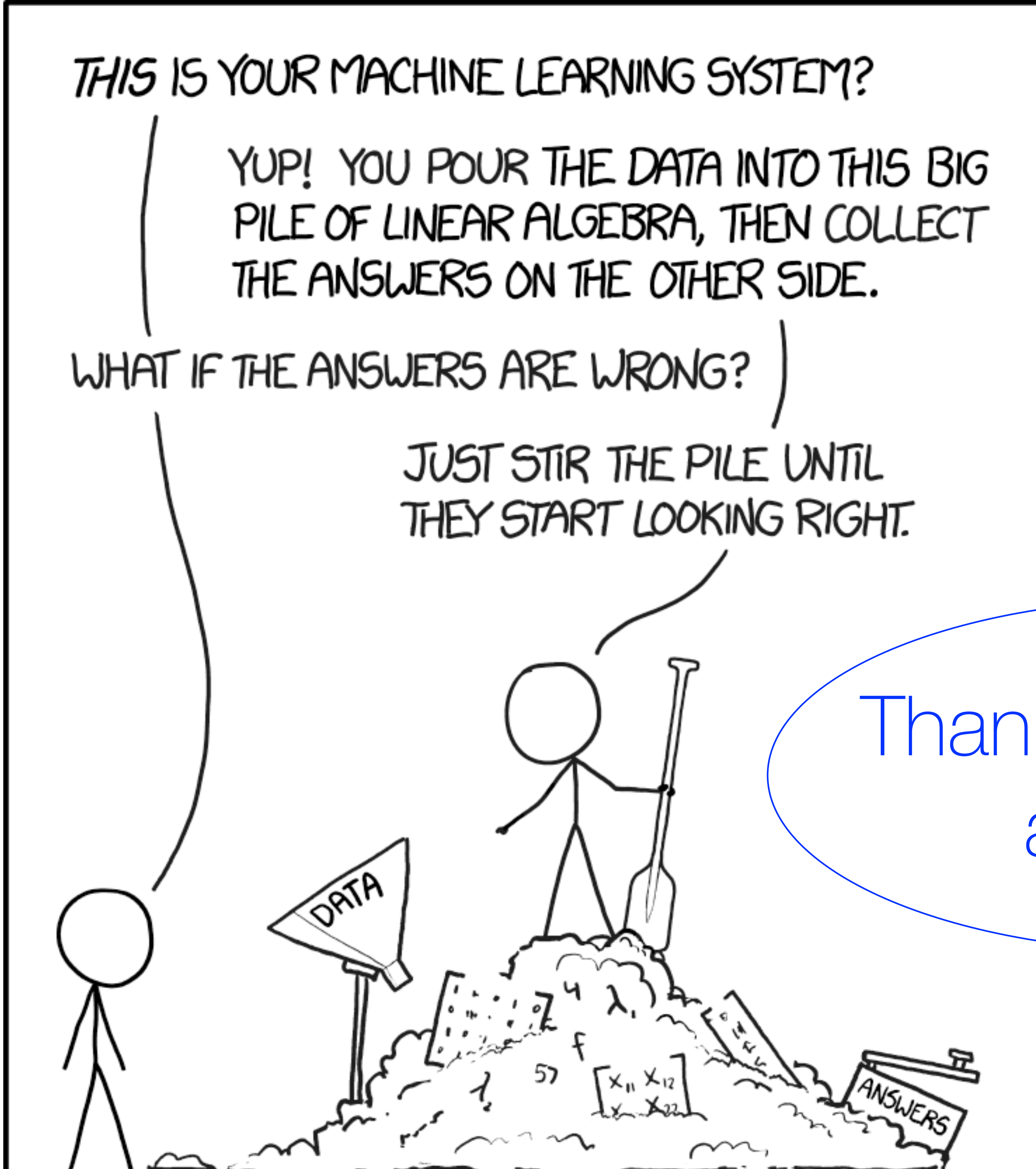
Urodynamics analysis

Scott Wang
Boston Children

Carlos Estrada
Boston Children

Caleb Nelson
Boston Children





https://xkcd.com/1838/

Thank you for your attention!