

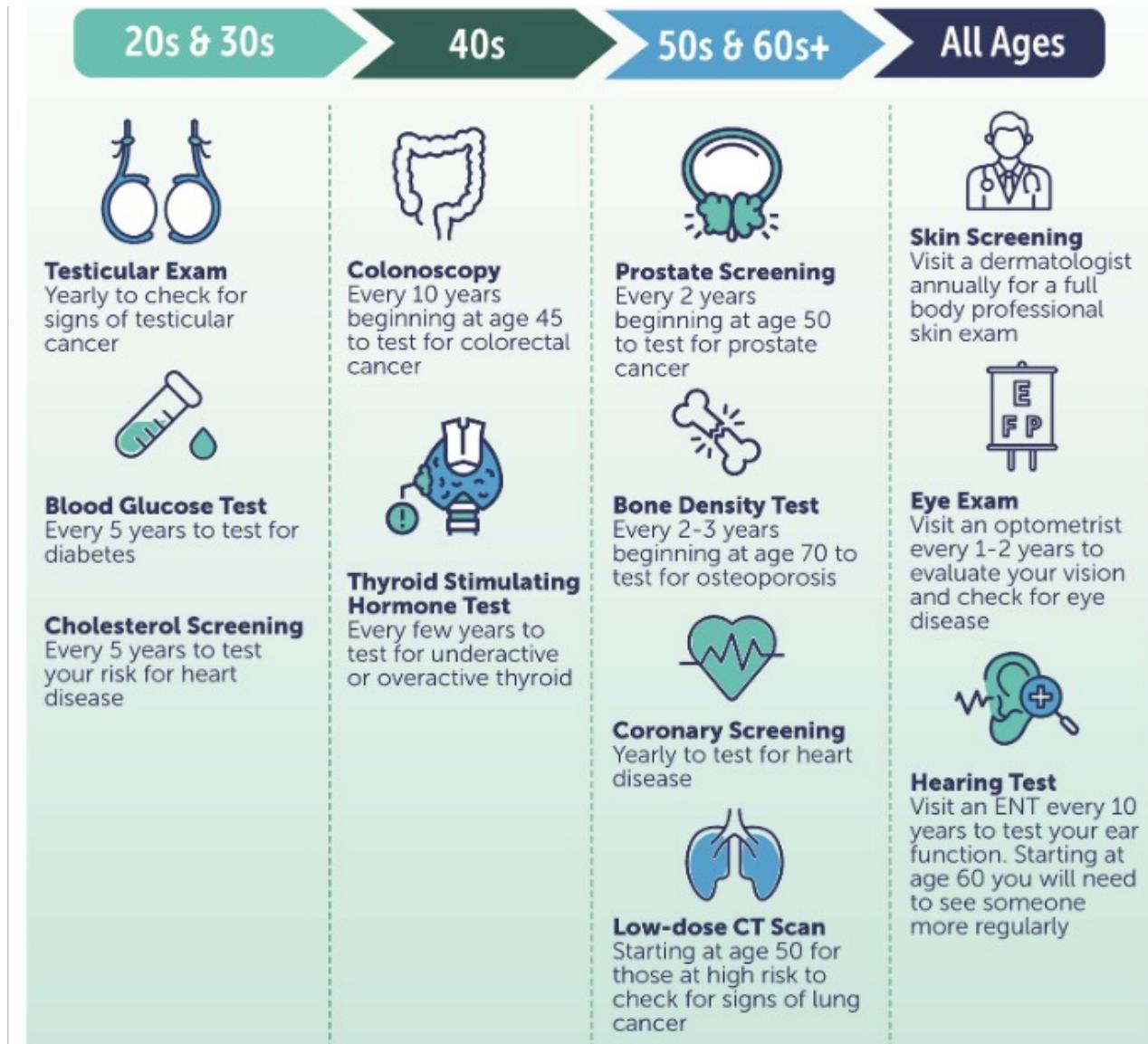
# Low-rank models for multiple non-Euclidean datasets

Kipoong Kim

Department of Statistics, Changwon National University

April 18, 2025

# Men's health screening checklist

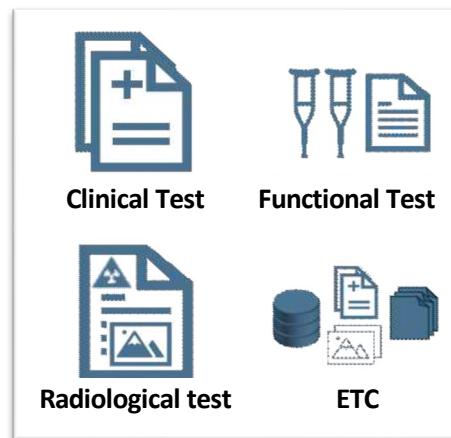


# Datasets with different characteristics

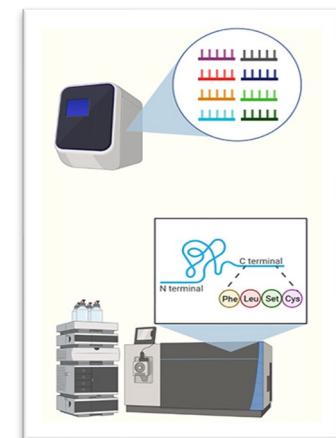
Clinical Outcomes

A	B	C	D	E	F	G	H	I
DATE	DAY	BRAIN	LUNGS	HEART	SYSTOLIC	DIASTOLIC	CELSIUS	PULSE
11/1/2020	Sunday	5	5	5	123	82	36.6	172
11/2/2020	Monday	5	5	5	119	78	36.6	179
11/3/2020	Tuesday	5	5	5	111	80	36.6	84
11/4/2020	Wednesday	5	5	5	120	80	36.6	162
11/5/2020	Thursday	5	4	5	120	80	36.6	52
11/6/2020	Friday	5	5	5	125	81	36.6	80
11/7/2020	Saturday	2	4	5	90	56	37.2	95
11/8/2020	Sunday	2	2	3	101	68	37.4	171
11/9/2020	Monday	5	4	4	147	95	37.6	76
11/10/2020	Tuesday	5	3	4	199	133	37.7	151
11/11/2020	Wednesday	4	2	3	97	70	37.8	154
11/12/2020	Thursday	4	3	4	193	125	38.3	140
11/13/2020	Friday	2	1	2	114	74	38.4	134
11/14/2020	Saturday	2	1	3	207	151	38.5	102

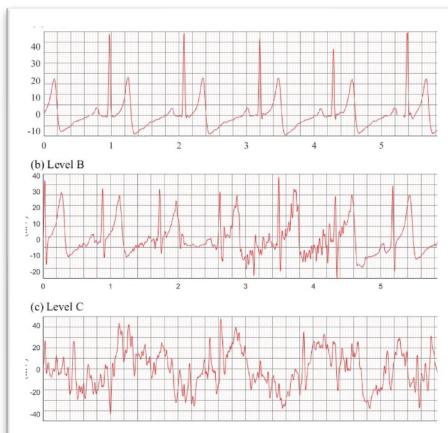
Electrical Health Records



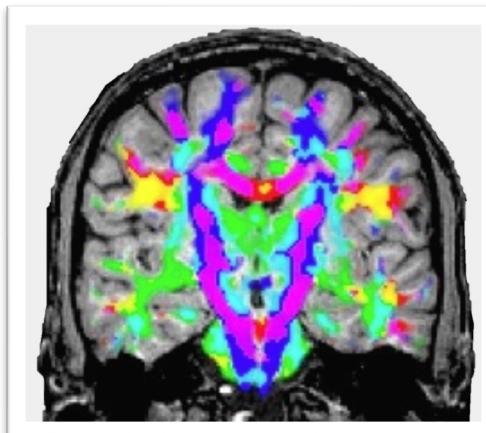
Sequencing data



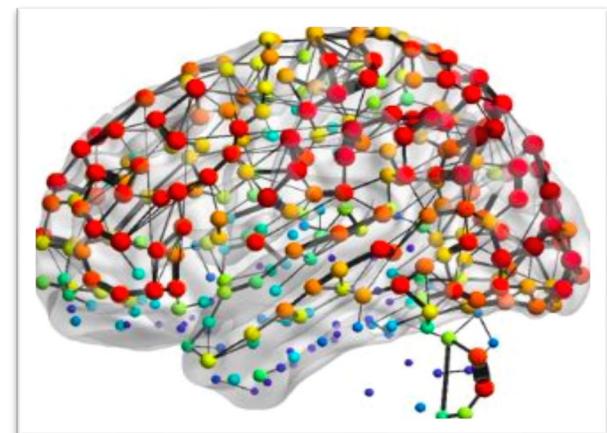
Electrocardiogram (ECG)



Medical Imaging data

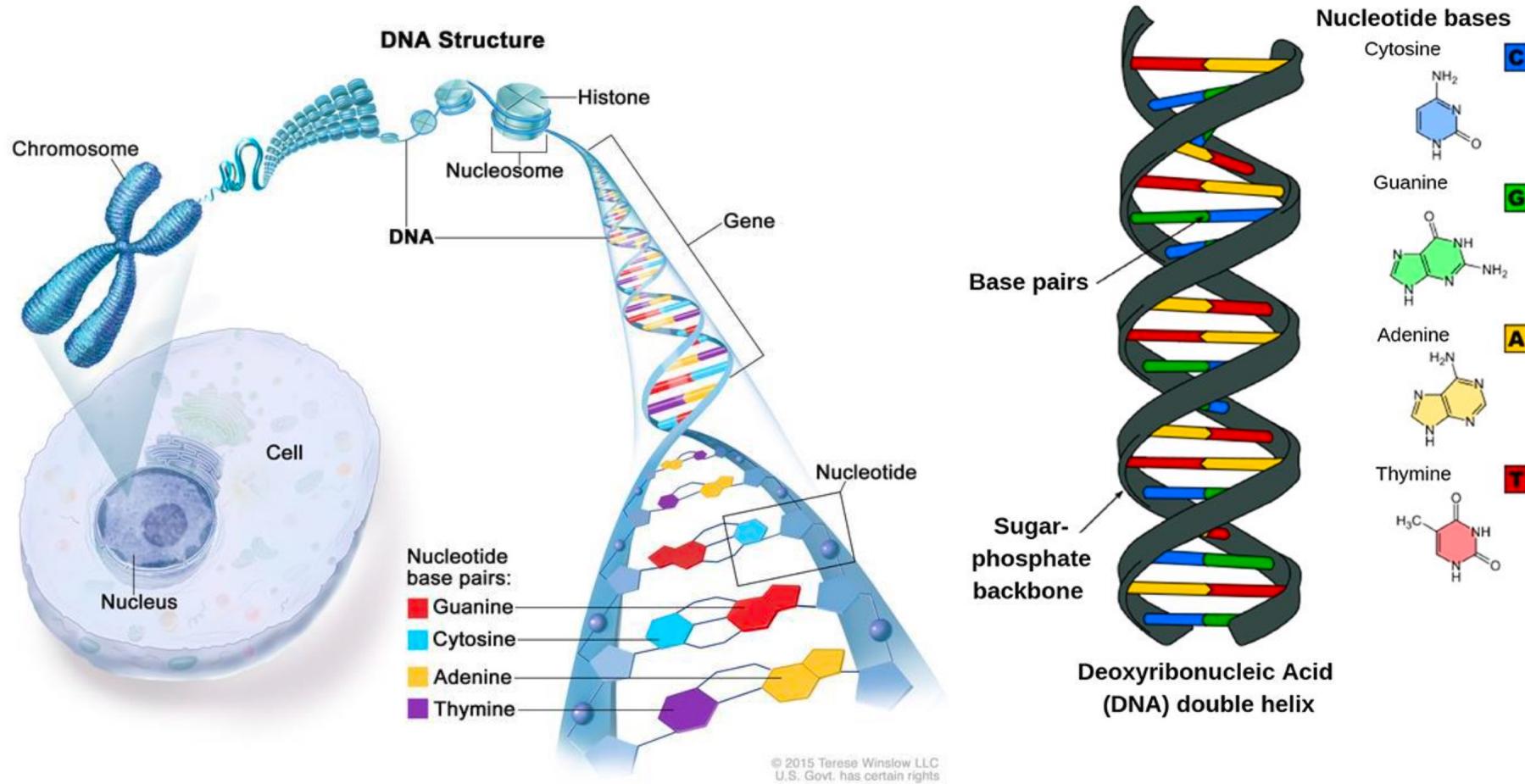


Connectome



# Genomic data

## □ DNA structure



# Genomic data

## □ Human Genome Project

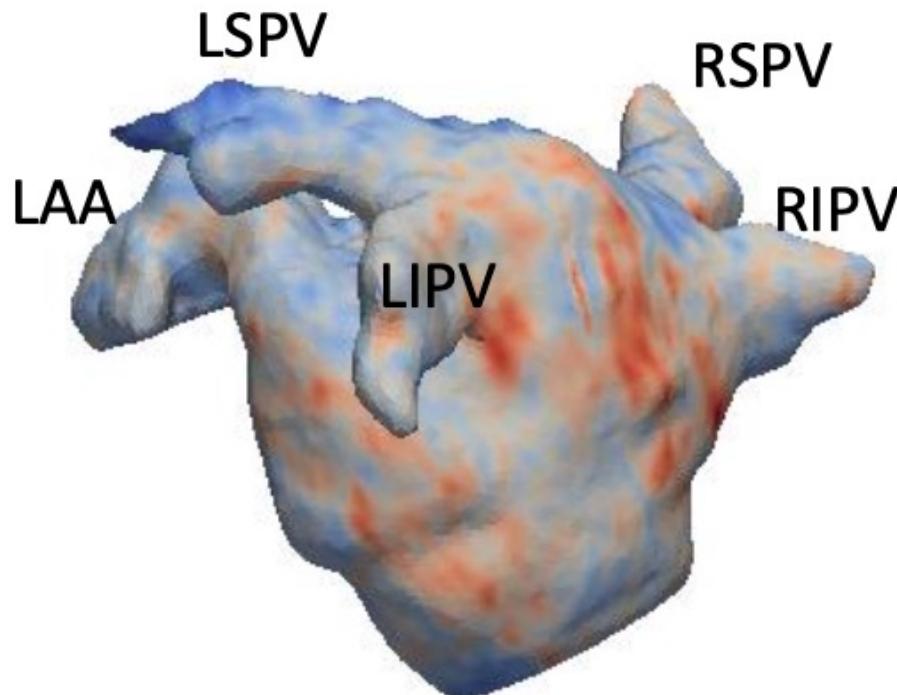
- Genes: ~40,000 개
- DNA: ~5,000,000 개
- Cost of Whole Genome Sequencing (WGS): 1,000\$

## □ Features of genomic data

- Low sample high-dimensional
- Strong correlation
- Genetic network among genes
- Population structure (white vs asian)

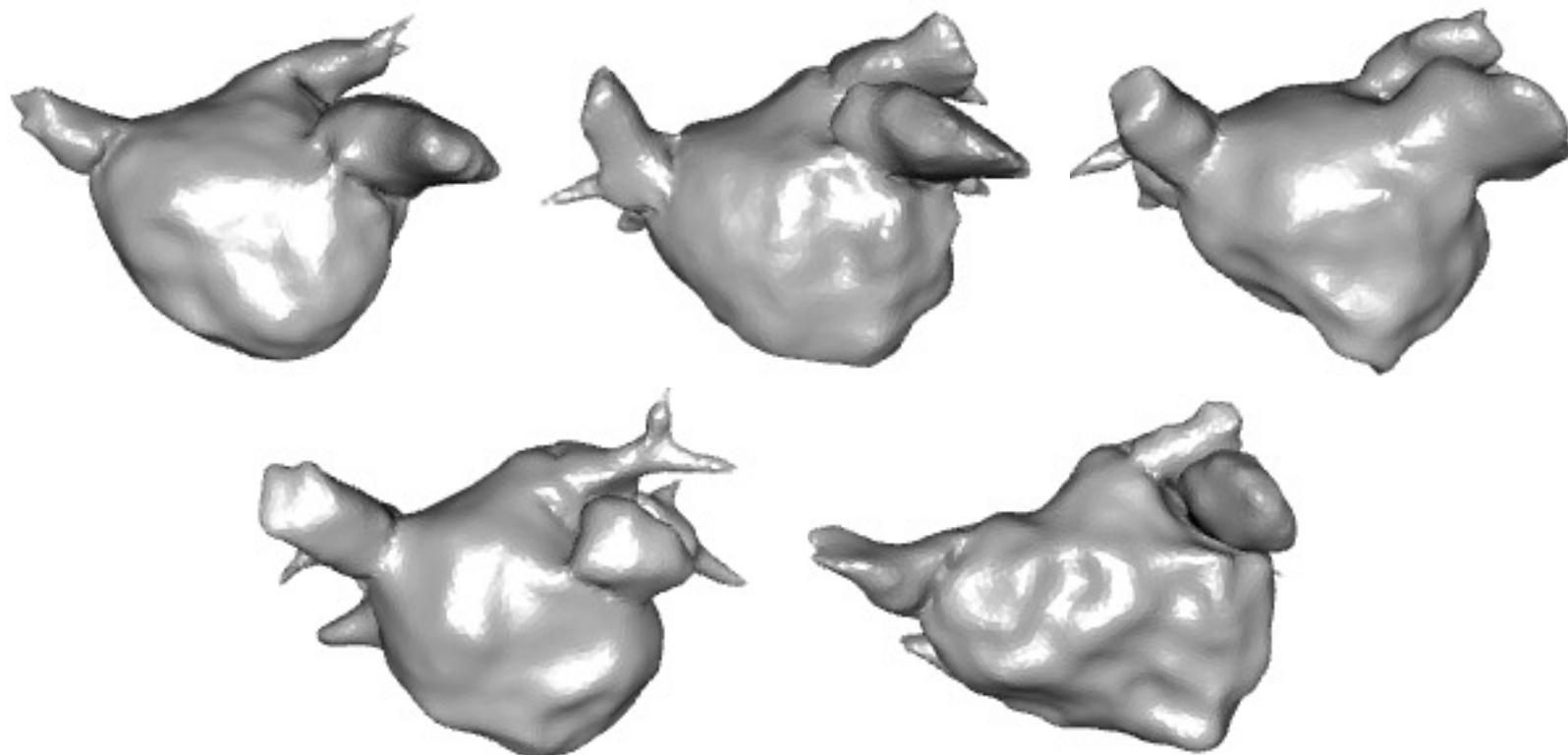
# CT / MRI: 3d point cloud data

- 3 dimensional point cloud data (e.g. Left Atrium of heart)
  - Additionally, voltage, wall thickness, activation time, etc. are collected at each point



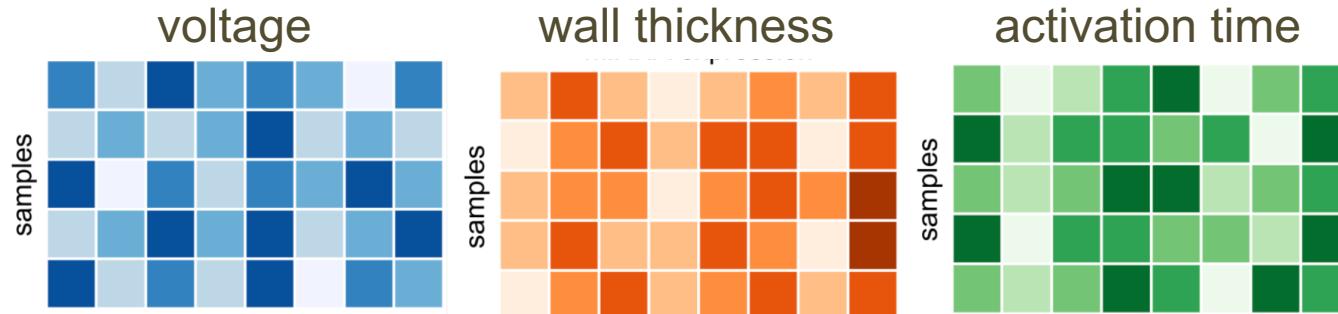
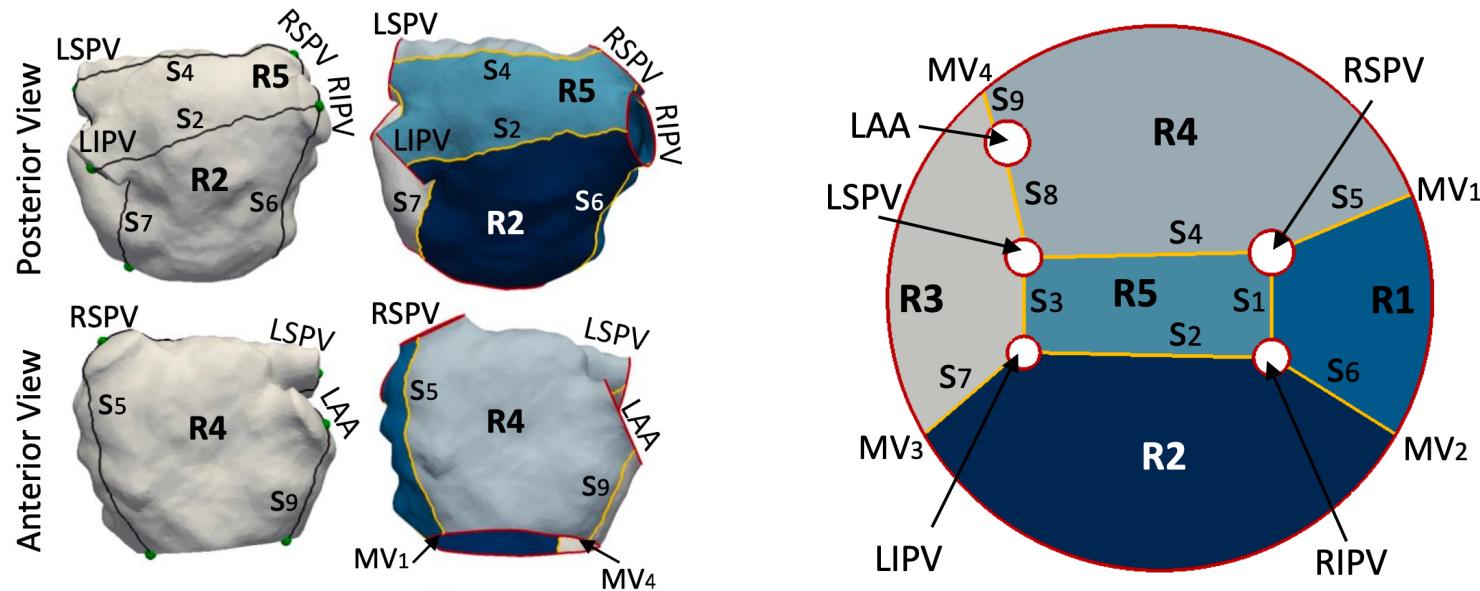
# CT / MRI: 3d point cloud data

- Issue: LA shape differs among individuals.



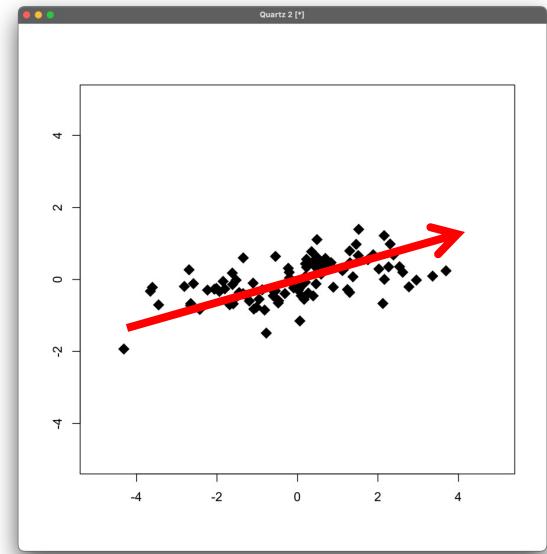
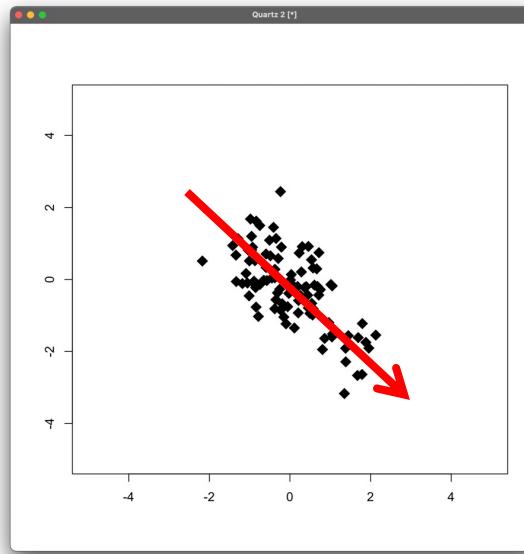
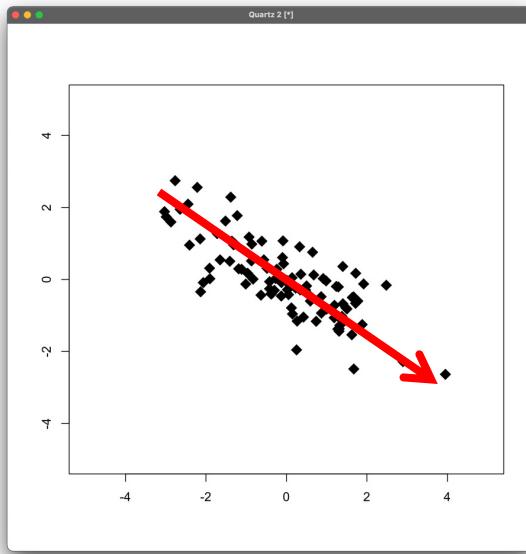
# CT / MRI: 3d point cloud data

- Registration procedure using flattening



# Multi-source data integration

- Our future work of interest:
  - structural decomposition for multiple (non-)Euclidean datasets:



# Multi-source data integration

- Joint and Individual Variation Explained (JIVE) model for Euclidean data

$$\begin{bmatrix} \mathbf{X}_{(1)}^T \\ \vdots \\ \mathbf{X}_{(D)}^T \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{\mu}_{(1)} \\ \vdots \\ \boldsymbol{\mu}_{(D)} \end{bmatrix}}_{\text{Intercept}} \mathbf{1}_n^T + \underbrace{\begin{bmatrix} \mathbf{V}_{(1)} \\ \vdots \\ \mathbf{V}_{(D)} \end{bmatrix}}_{\text{Joint}} \mathbf{U}_{(0)}^T + \underbrace{\begin{bmatrix} \mathbf{A}_{(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{A}_{(D)} \end{bmatrix}}_{\text{Individual}} \begin{bmatrix} \mathbf{U}_{(1)}^T \\ \vdots \\ \mathbf{U}_{(D)}^T \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{E}_{(1)} \\ \vdots \\ \mathbf{E}_{(D)} \end{bmatrix}}_{\text{Error}}$$

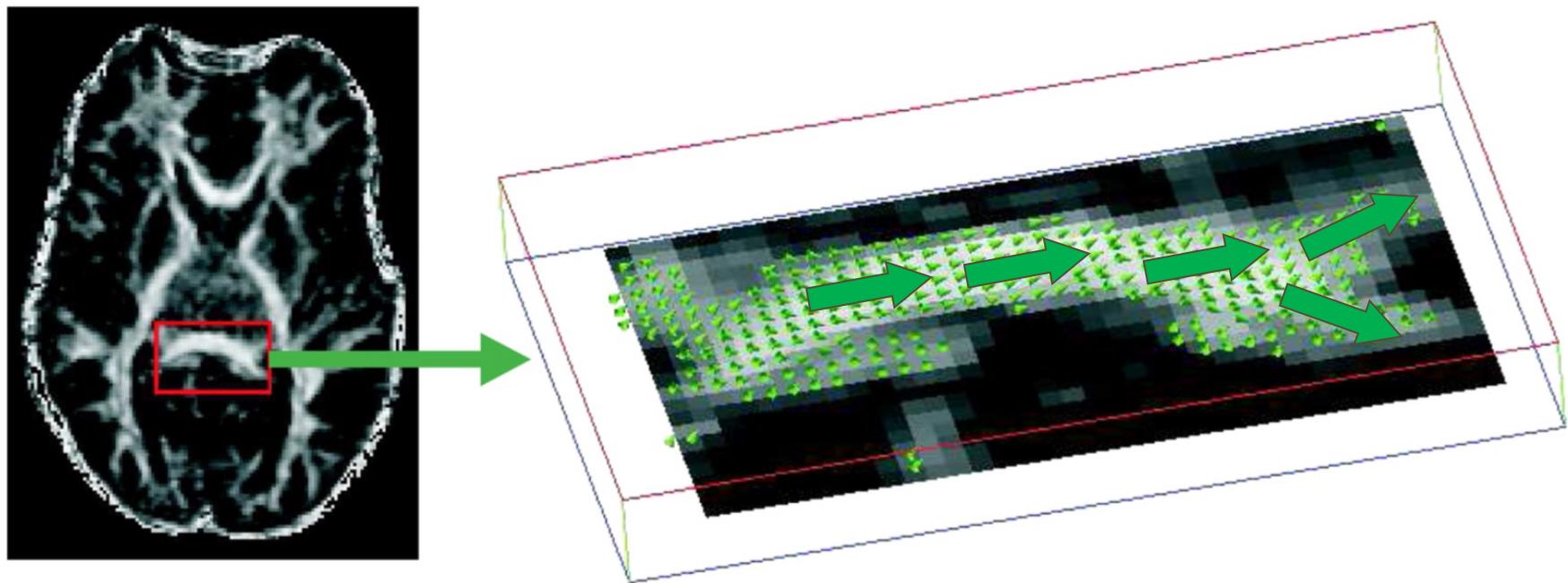
- For the  $d$ -th source case:

$$\mathbf{X}_{(d)} = \mathbf{1}\boldsymbol{\mu}_{(d)}^T + \mathbf{U}_{(0)}\mathbf{V}_{(d)}^T + \mathbf{U}_{(d)}\mathbf{A}_{(d)}^T + \mathbf{E}_{(d)},$$

where  $\mathbf{U}_{(0)} \in \mathbb{R}^{n \times r_0}$ ,  $\mathbf{V}_{(d)} \in \mathbb{R}^{p \times r_0}$ ,  $\mathbf{U}_{(d)} \in \mathbb{R}^{n \times r_d}$ ,  $\mathbf{A}_{(d)} \in \mathbb{R}^{p \times r_d}$

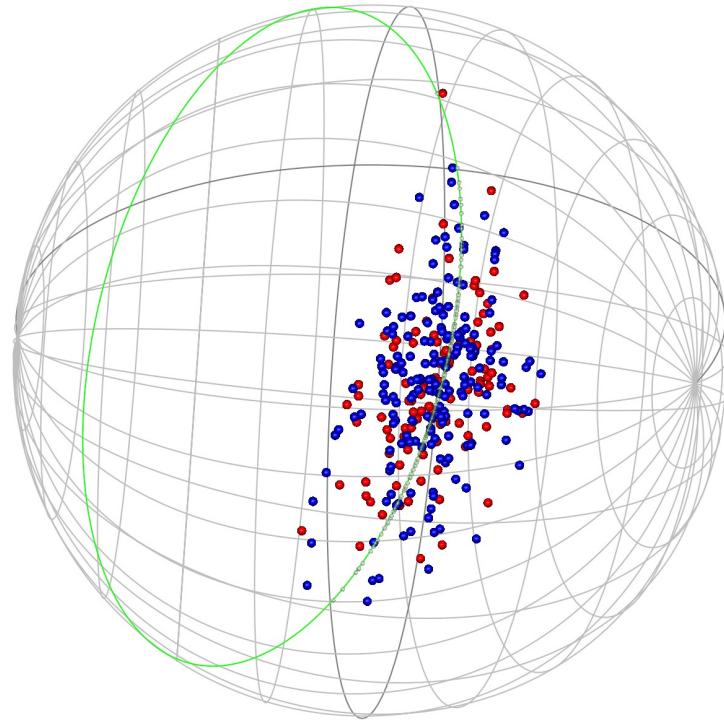
# CT / MRI: DTI data

- Diffusion Tensor Imaging (DTI) data from MRI



# CT / MRI: DTI data

- The movement of water molecules is represented by a  $3 \times 3$  diffusion tensor, from which a principal 3D direction or 4D direction can be extracted.
  - Spherical data



# Spherical data is on a vector space?

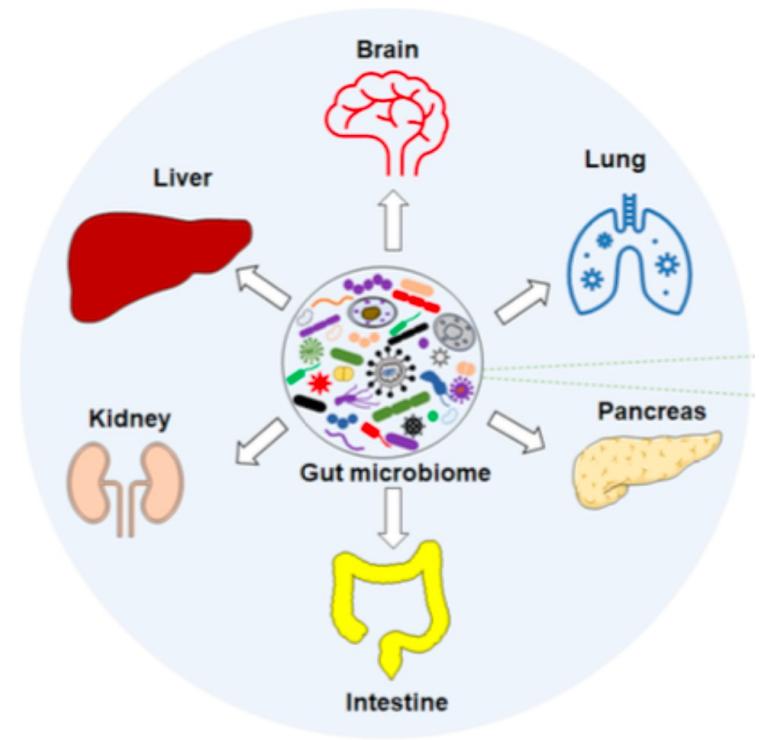
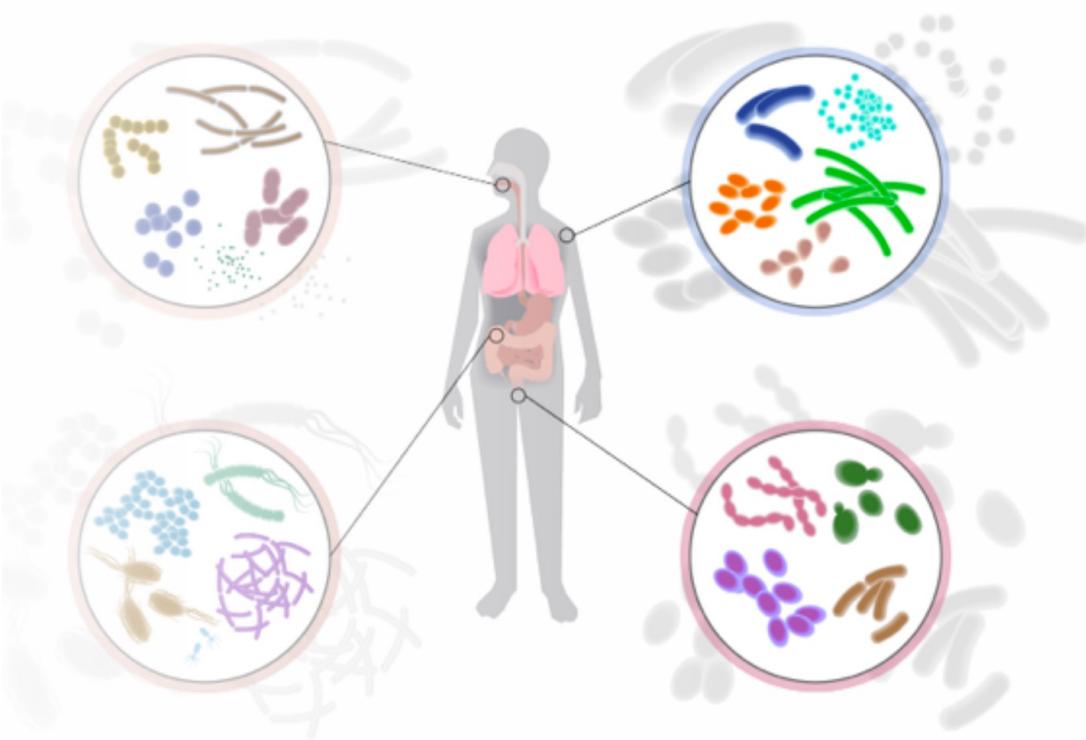
□ For  $\mathbf{x}, \mathbf{y} \in S^{p-1} = \{\mathbf{x} \in \mathbb{R}^p: \|\mathbf{x}\|_2 = 1\}$

➤  $\mathbf{x} + \mathbf{y} \notin S^{p-1}$

➤  $c \cdot \mathbf{x} \notin S^{p-1}, \quad \text{for } c \in \mathbb{R} \setminus \{1\}$

➤  $\mathbf{x} - \mathbf{y} \notin S^{p-1}$

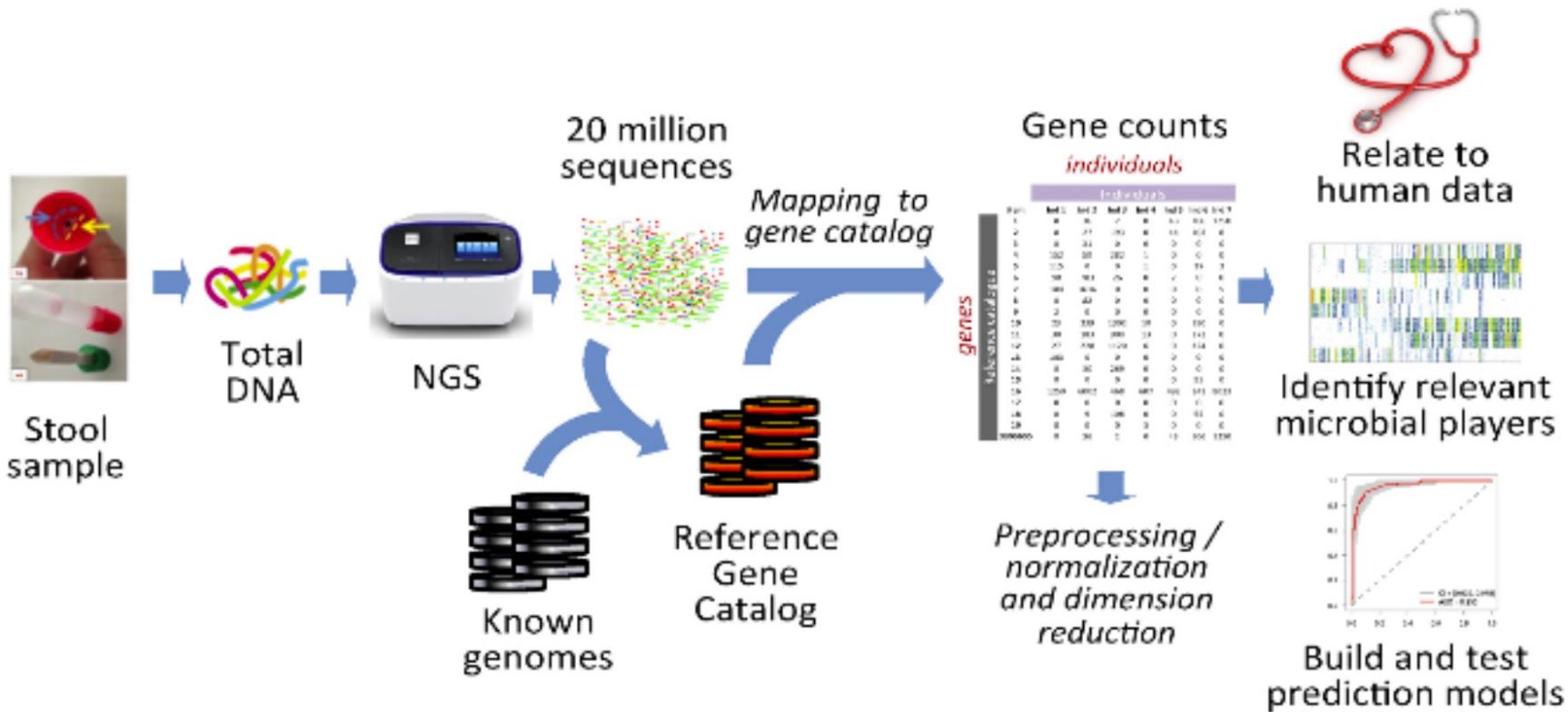
# Microbiome data



# Microbiome data

## □ Data Extraction Process

- Sampling → DNA Extraction → PCR Amplification & Library Preparation  
→ Sequencing & Library Mapping → Microbiome Count Data



# Microbiome data

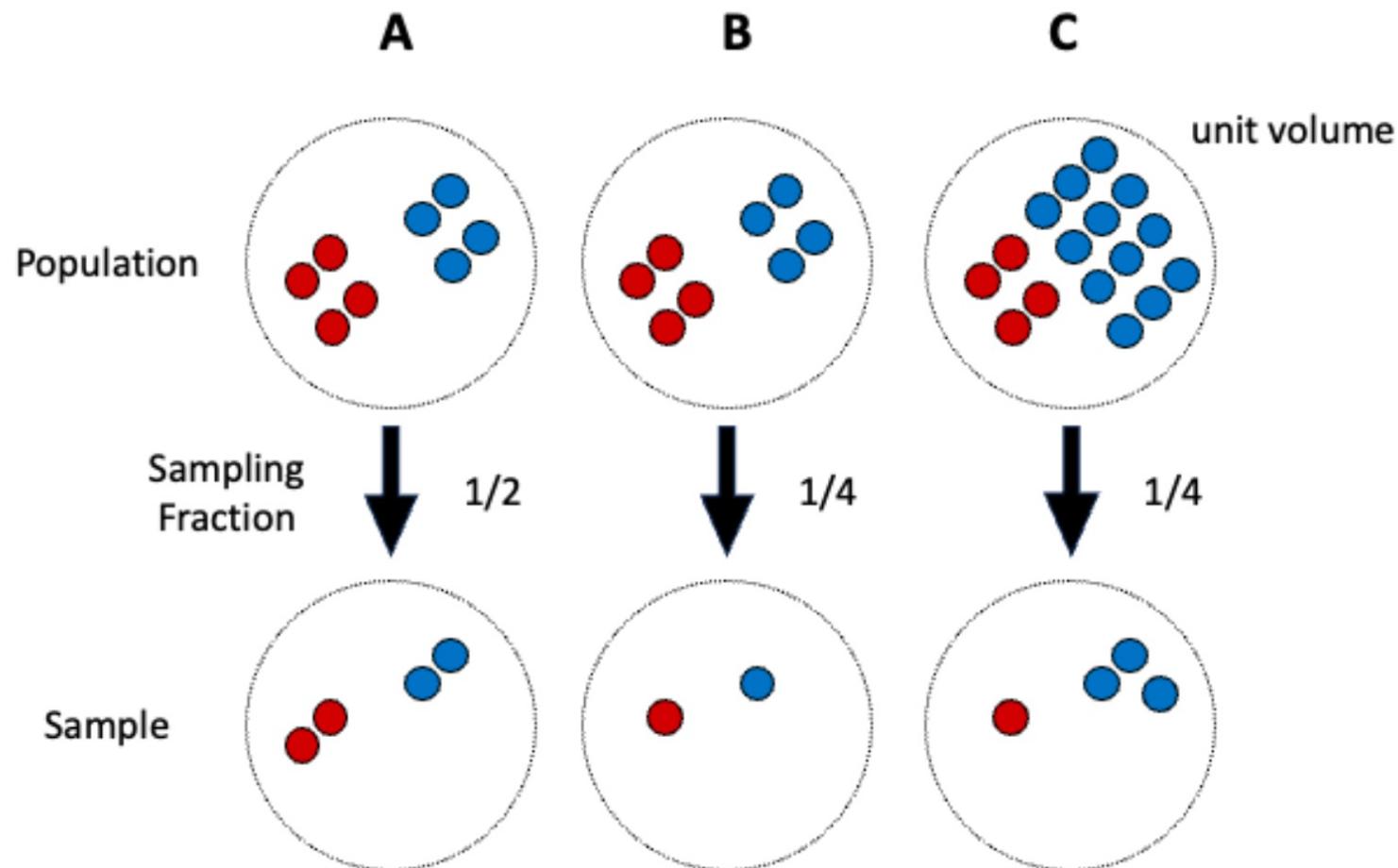


Figure: A vs B: different library size; A vs C: different sampling fraction

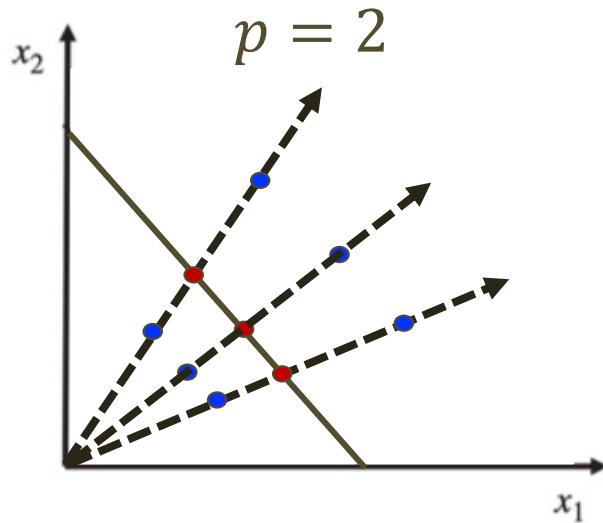
# Microbiome compositional data

## □ Conversion of count data to compositional data

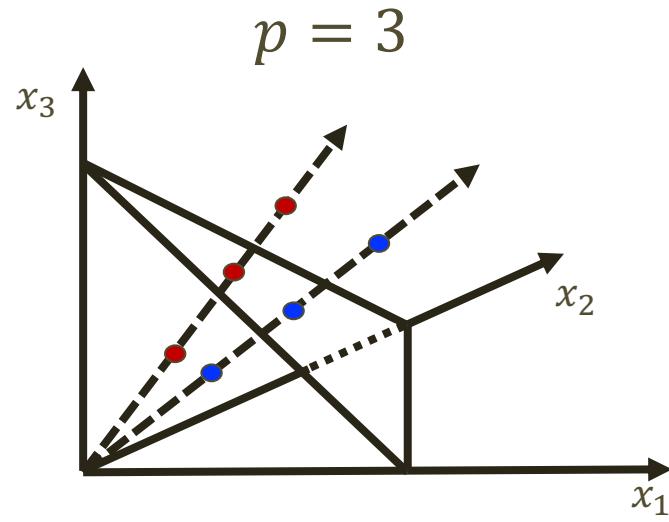
➤ Closure operator:  $\text{cls}(x_1, \dots, x_p) = \left[ \frac{x_1}{\sum x_j}, \dots, \frac{x_p}{\sum x_j} \right]$

➤ Compositional space

$$\mathcal{C}^{p-1} = \{ \mathbf{x} = [x_1, \dots, x_p] \in \mathbb{R}^p : x_1 + \dots + x_p = 1, x_j > 0 \forall j \}$$



$p = 2$

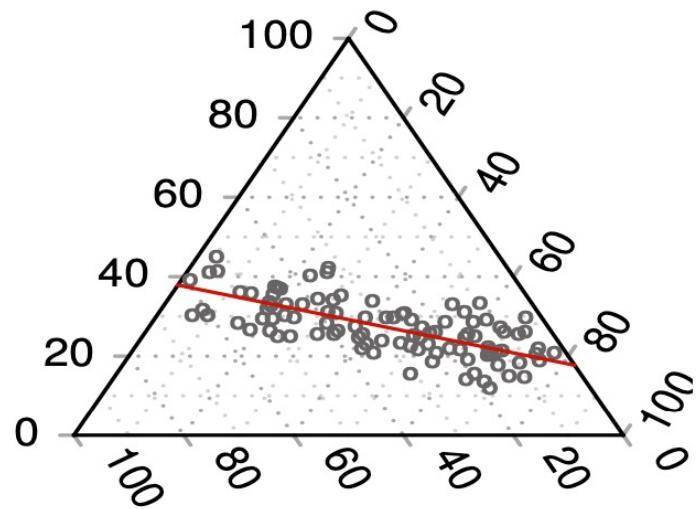


$p = 3$

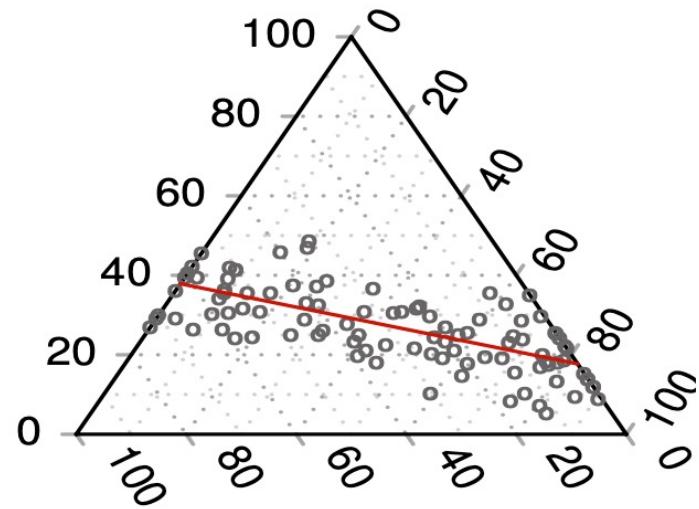
# Microbiome compositional data

- Simulated data example when  $p = 3$

High signal-to-noise ratio



Low signal-to-noise ratio



# Compositional data is on a vector space?

□ For  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^{p-1}$

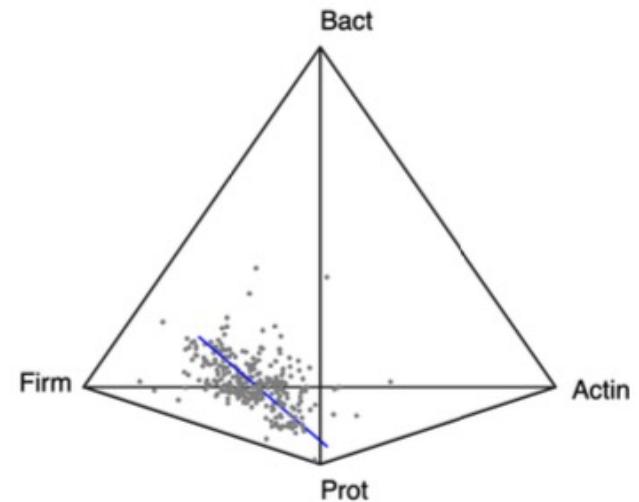
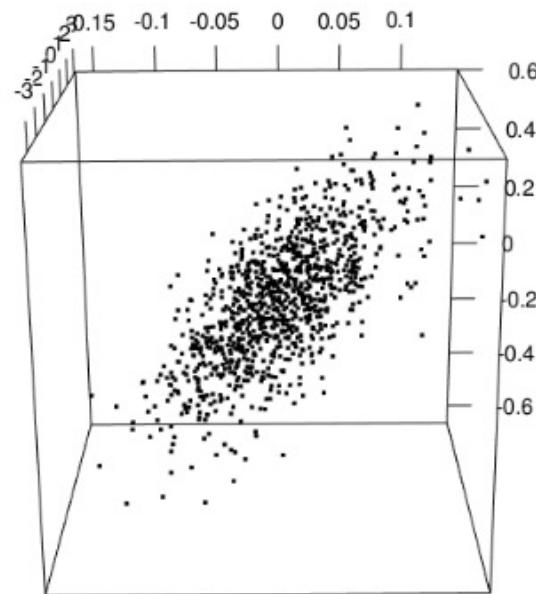
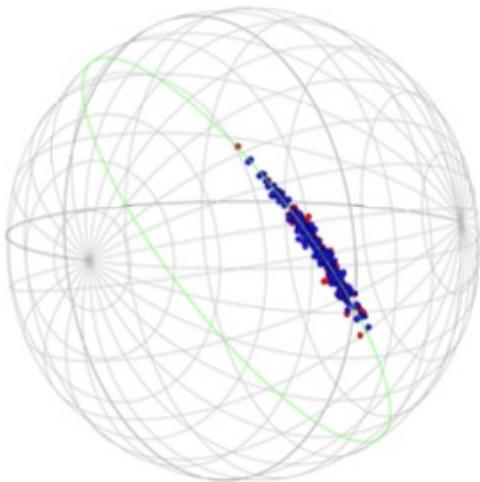
➤  $\mathbf{x} + \mathbf{y} \notin \mathbb{C}^{p-1}$

➤  $c \cdot \mathbf{x} \notin \mathbb{C}^{p-1}$ , for  $c \in \mathbb{R} \setminus \{1\}$

➤  $\mathbf{x} - \mathbf{y} \notin \mathbb{C}^{p-1}$

# Non-Euclidean data integration

- Our future work of interest:
  - structural decomposition for multiple (non-)Euclidean datasets:



# Non-Euclidean data integration

- JIVE model for non-Euclidean case

- ▶  $X_{ij}^{(d)} \sim$  Exponential family with the natural parameter  $\theta_{ij}^{(d)}$
- ▶  $g(\mathbb{E}X_{ij}^{(d)}) = \theta_{ij}^{(d)} = \boldsymbol{\mu}_j^{(d)} + \mathbf{u}_i^{(0)T} \mathbf{v}_j^{(d)} + \mathbf{u}_i^{(d)T} \mathbf{a}_j^{(d)}$
- ▶ Matrix-version

$$\boldsymbol{\Theta}_{(d)} = \underbrace{\mathbf{1}_n \boldsymbol{\mu}_{(d)}^T}_{\text{Intercept}} + \underbrace{\mathbf{U}_{(0)} \mathbf{V}_{(d)}^T}_{\text{Joint}} + \underbrace{\mathbf{U}_{(d)} \mathbf{A}_{(d)}^T}_{\text{Individual}},$$

where  $\mathbf{U}_{(0)} \in \mathbb{R}^{n \times r_0}$ ,  $\mathbf{V}_{(d)} \in \mathbb{R}^{p \times r_0}$ ,  $\mathbf{U}_{(d)} \in \mathbb{R}^{n \times r_d}$ ,  $\mathbf{A}_{(d)} \in \mathbb{R}^{p \times r_d}$

- ▶ Estimate each of  $\mathbf{U}_{(0)}$ ,  $\mathbf{V}_{(d)}$ ,  $\mathbf{U}_{(d)}$  and  $\mathbf{A}_{(d)}$  with others fixed

# Our research topics in non-Euclidean data

- PCA for zero-inflated compositional data
- GLM for spherical responses

# Aitchison geometry for compositional data

## □ New vector operations for compositional data

➤ Perturbation

$$\mathbf{x} \oplus \mathbf{y} = \left[ \frac{x_1 y_1}{\sum x_j y_j}, \dots, \frac{x_p y_p}{\sum x_j y_j} \right] = \text{cls}[x_1 y_1, \dots, x_p y_p]$$

➤ Powering

$$\alpha \odot \mathbf{y} = \left[ \frac{x_1^\alpha}{\sum x_j^\alpha}, \dots, \frac{x_p^\alpha}{\sum x_j^\alpha} \right] = \text{cls}[x_1^\alpha, \dots, x_p^\alpha]$$

➤ Subtraction

$$\mathbf{x} \ominus \mathbf{y} = \mathbf{x} \oplus (-1 \odot \mathbf{y})$$

➤ Distance

$$d_A(\mathbf{x}, \mathbf{y})^2 = \|\mathbf{x} \ominus \mathbf{y}\|_A = \frac{1}{2p} \sum_i \sum_j \left[ \log \frac{x_i}{y_i} - \log \frac{y_i}{x_i} \right]^2$$

# Log-ratio transformations

- Centered Log-Ratio (CLR) transformation:

$$clr(\mathbf{x}) = \log x_j - \frac{1}{p} \sum_{j=1}^p \log x_j$$

- Isometric Log-Ratio (ILR) transformation

$$ilr(\mathbf{x}) = \mathbf{H}_p clr(\mathbf{x})$$

➤  $\mathbf{H}_p$  is the  $(p-1) \times p$  Helmert sub-matrix (Dryden & Mardia, 1998) of which the  $j$ -th row is given by  $(h_j, \dots, h_j, -j h_j, 0, \dots, 0)$  and  $h_j = [j(j+1)]^{-1/2}$

- Isometry of the CLR & ILR transformation

$$d_A(x, y) = \|clr(x) - clr(y)\|_2 = \|ilr(x) - ilr(y)\|_2$$

# Log-ratio PCA

- Log-ratio PCA (Aitchison, 1983) was proposed to cope with both linear and curved data patterns.

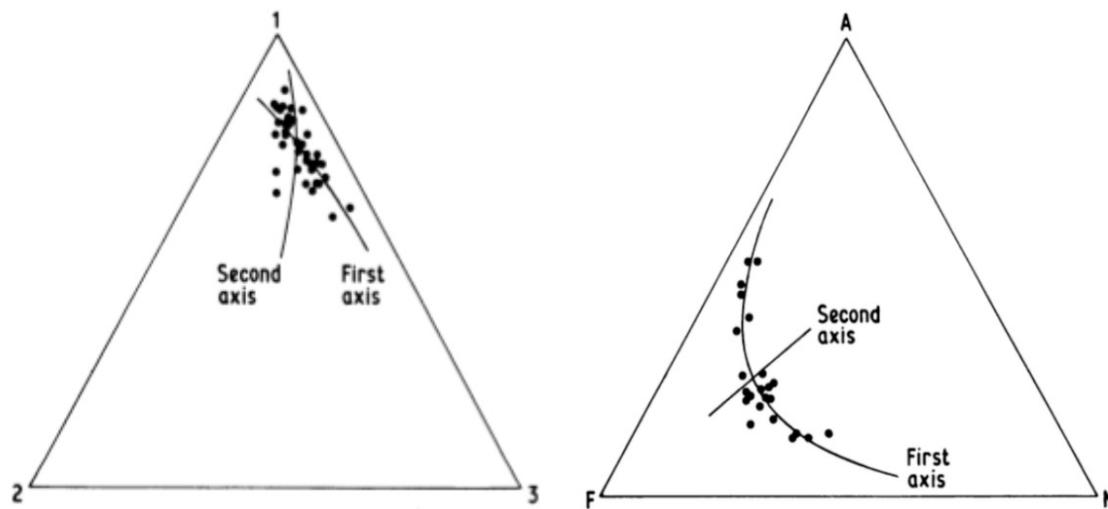


Figure: Ternary diagram with log-ratio principal axes

## □ Limitation:

- Log-ratio transformation inherently cannot handle zero values.

# Log-ratio PCA with zero replacement

## □ Zero replacement strategies

- Simple replacement:

$$(x_1, x_2, 0) \rightarrow \text{cls}((x_1, x_2, \delta))$$

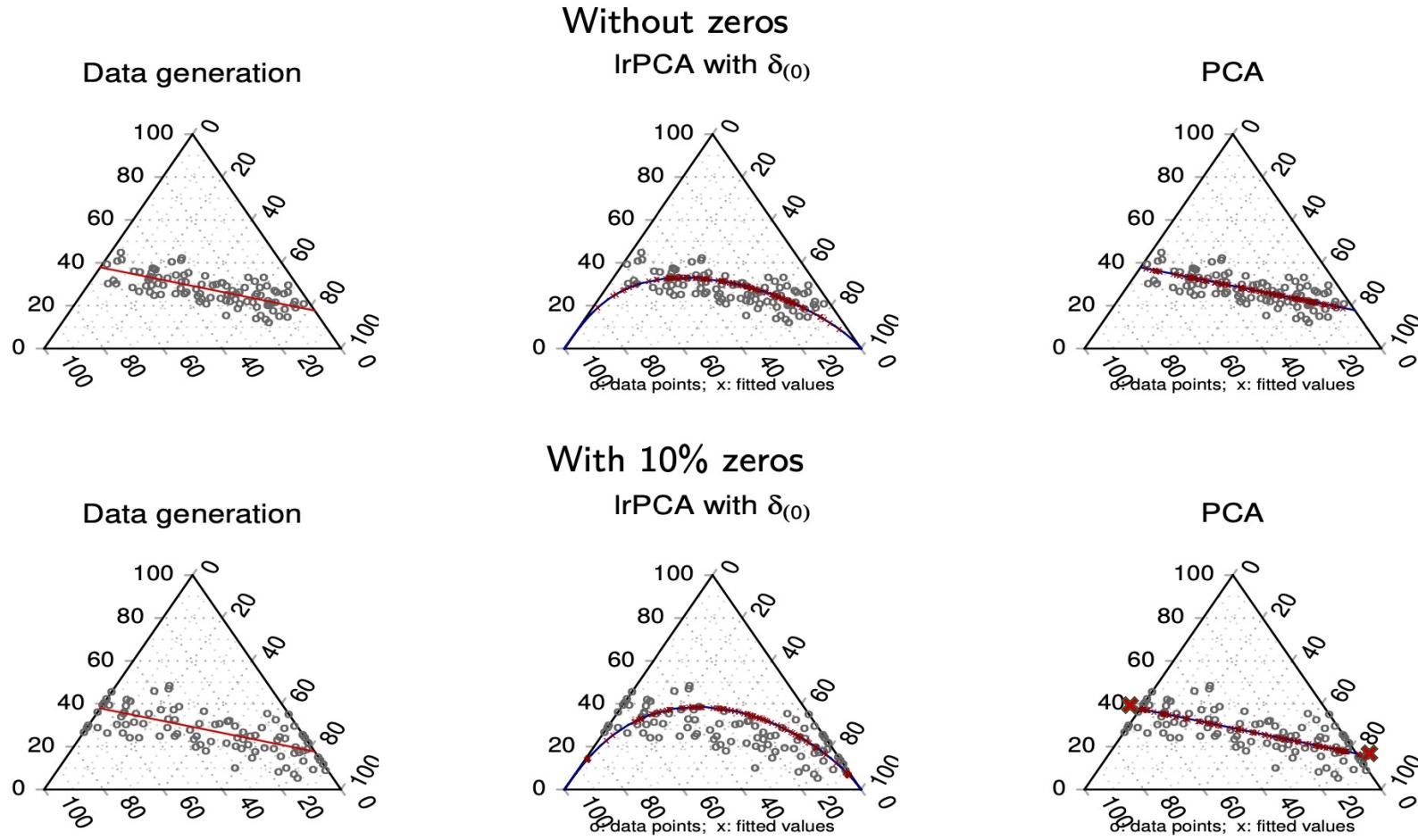
- Additive, Multiplicative, and etc.

## □ Determination of $\delta$

- $\frac{1}{2} \min\{x_j : x_j > 0\}$

# Limitation of log-ratio PCA

- However, the zero inflation may result in the distortion.



# Intuitive approach

- We want to propose a new dimension reduction method that prevents its low-rank reconstructions from being out of the composition space.
- Compositional reconstruction PCA (crPCA)
  - Find the principal directions (classical PCA)
  - Project the principal scores into the compositional space

# Compositional PCA

- Denote the  $i$ -th row of  $\mathbf{A}$  by  $\mathbf{a}_i$  and the  $k$ -th column of  $\mathbf{A}$  by  $A_k$ .
- Global compositional PCA (global CPCCA) problem:

$$(\hat{\mathbf{U}}^{(r)}, \hat{\mathbf{V}}^{(r)}) = \arg \min_{\mathbf{U} \in \mathbb{R}^{n \times r}, \mathbf{V} \in \mathbb{R}^{p \times r}} \left\| \mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T - \mathbf{UV}^T \right\|_F^2,$$

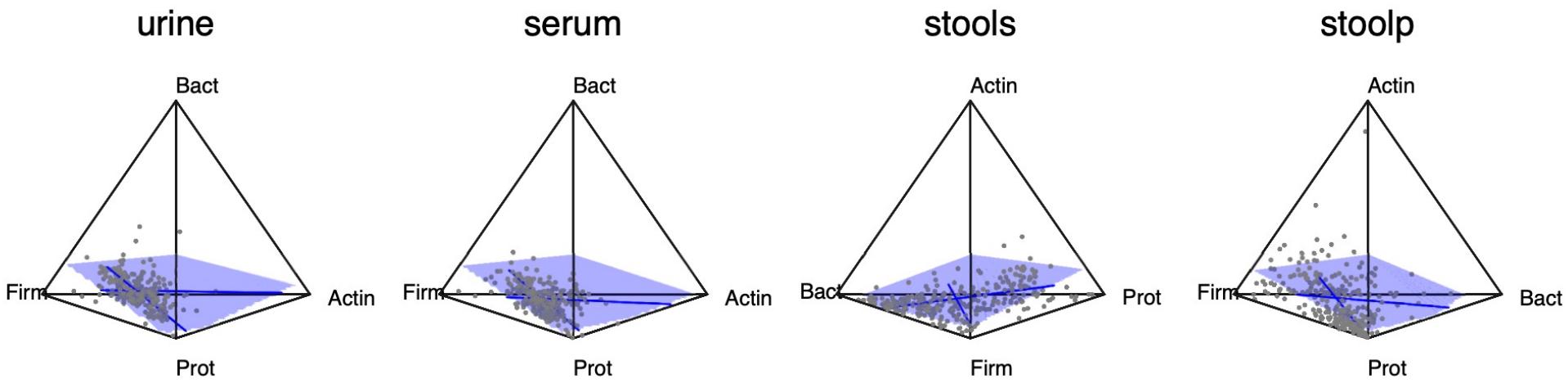
subject to

- $\mathbf{U}$  and  $\mathbf{V}$  have orthogonal and orthonormal columns
- $\boldsymbol{\mu} \in \mathbb{C}^{p-1}$ ,  $\boldsymbol{\mu} + \mathbf{Vu}_i \in \mathbb{C}^{p-1}$  for all  $i = 1, \dots, n$

# Compositional PCA

- Compositional subspace, spanned by  $\{\mathbf{V}_1, \dots, \mathbf{V}_r\}$  at  $\boldsymbol{\mu}$  :

$$\mathbb{CS}_{(\boldsymbol{\mu}; \{\mathbf{V}_1, \dots, \mathbf{V}_r\})} := \mathbb{C}^p \cap \{\boldsymbol{\mu} + c_1 \mathbf{V}_1 + \dots + c_r \mathbf{V}_r : c_1, \dots, c_r \in \mathbb{R}\}$$



# Compositional PCA

□ Sequential estimation procedure:

Rank-1 case:

$$(\hat{\mathbf{U}}, \hat{\mathbf{V}}_1) = \arg \min_{\mathbf{U}_1, \mathbf{V}_1} \|\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T - \mathbf{U}_1 \mathbf{V}_1^T\|_F^2,$$

Rank-2 case:

$$(\hat{\mathbf{U}}, \hat{\mathbf{V}}_2) = \arg \min_{(\mathbf{U}_1, \mathbf{U}_2), \mathbf{V}_2 \perp \hat{\mathbf{V}}_1} \|\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T - \mathbf{U}_1 \hat{\mathbf{V}}_1^T - \mathbf{U}_2 \mathbf{V}_2^T\|_F^2,$$

⋮

Rank- $k$  case:

$$(\hat{\mathbf{U}}, \hat{\mathbf{V}}_k) = \arg \min_{(\mathbf{U}_1, \dots, \mathbf{U}_k), \mathbf{V}_k \perp \hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1}} \|\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T - \mathbf{U}_1 \hat{\mathbf{V}}_1^T - \dots - \mathbf{U}_k \mathbf{V}_k^T\|_F^2,$$

with the appropriate constraints and  $k = 1, \dots, r$ .

# Compositional PCA

- *Compositional PCA (CPCA)*: Given  $\mu, \hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1}$ ,

$$\arg \min_{\mathbf{U}_1, \dots, \mathbf{U}_k, \mathbf{V}_k} \|\mathbf{X} - \mathbf{1}\mu^T - \mathbf{U}_1\hat{\mathbf{V}}_1^T - \dots - \mathbf{U}_{k-1}\hat{\mathbf{V}}_{k-1}^T - \mathbf{U}_k\mathbf{V}_k^T\|_F^2, \quad (2)$$

subject to

- $\mu + \sum_{h=1}^{k-1} u_{ih} \hat{\mathbf{V}}_h + u_{ik} \mathbf{V}_k \in \mathbb{C}^p \quad \forall i$
- $\mathbf{V}_k \perp \mathbf{1}_p, \hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1}$  and  $\|\mathbf{V}_k\|_2 = 1$

- *Approximated CPCA (aCPCA)*: Given  $\mu, (\hat{\mathbf{U}}_1, \hat{\mathbf{V}}_1), \dots, (\hat{\mathbf{U}}_{k-1}, \hat{\mathbf{V}}_{k-1})$ ,

$$\arg \min_{\mathbf{U}_k, \mathbf{V}_k} \|\mathbf{X} - \mathbf{1}\mu^T - \hat{\mathbf{U}}_1\hat{\mathbf{V}}_1^T - \dots - \hat{\mathbf{U}}_{k-1}\hat{\mathbf{V}}_{k-1}^T - \mathbf{U}_k\mathbf{V}_k^T\|_F^2, \quad (3)$$

subject to

- $\mu + \sum_{h=1}^{k-1} \hat{u}_{ih} \hat{\mathbf{V}}_h + u_{ik} \mathbf{V}_k \in \mathbb{C}^p \quad \forall i$
- $\mathbf{V}_k \perp \mathbf{1}_p, \hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1}, \|\mathbf{V}_k\|_2 = 1$

# Algorithm of CPCA

---

## Algorithm 1: Rank- $k$ approximation for CPCA

---

**Input:**  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$  and  $(\hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1})$ .

**Initialize**  $\mathbf{V}_k^{(0)} \perp \mathbf{1}_p$ .

**Repeat** for  $t = 0, 1, 2, \dots$ :

1 U-update: obtain  $\mathbf{u}_i^{(t+1)}$  by (4) with  $\mu = \bar{\mathbf{x}}$  and  $\mathbf{V}_k = \mathbf{V}_k^{(t)}$   $\forall i$ .

2 U-shrinkage:  $\mathbf{u}_i^{(t+1)} \leftarrow (1 - \frac{\gamma}{t+1})\mathbf{u}_i^{(t+1)}$ .

3 V-update: obtain  $\mathbf{V}_k^{(t+1)}$  by (6) with  $\mu = \bar{\mathbf{x}}$  and  $\mathbf{U} = (\mathbf{U}_1^{(t+1)}, \dots, \mathbf{U}_k^{(t+1)})$

4 V-scaling:  $\mathbf{V}_k^{(t+1)} \leftarrow \mathbf{V}_k^{(t+1)} / \|\mathbf{V}_k^{(t+1)}\|_2$ .

**until convergence:**  $\|\mathbf{V}_k^{(t+1)} - \mathbf{V}_k^{(t)}\|_F^2 < \epsilon$ .

**Re-estimation of U:** estimate  $\mathbf{u}_i^{(t+1)}$  without the shrinkage  $\forall i$ .

**Output:**  $(\mathbf{U}_1^{(t+1)}, \dots, \mathbf{U}_k^{(t+1)})$  and  $(\hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1}, \mathbf{V}_k^{(t+1)})$ .

---

# Algorithm of aCPCA

---

## Algorithm 2: Rank- $k$ approximation for aCPCA

---

**Input:**  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ ,  $(\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_{k-1})$  and  $(\hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1})$ .

**Initialize**  $\mathbf{V}_k^{(0)}$ .

**Repeat** for  $t = 0, 1, 2, \dots$ :

1 U-update: obtain  $u_{ik}^{(t+1)}$  by (5) with  $\mathbf{c}_i = \bar{\mathbf{x}} + \sum_{h=1}^{k-1} \hat{u}_{ih} \hat{\mathbf{V}}_h$  and  $\mathbf{V}_k = \mathbf{V}_k^{(t)}$   $\forall i$ .

2 U-shrinkage:  $u_{ik}^{(t+1)} \leftarrow (1 - \frac{\gamma}{t+1}) u_{ik}^{(t+1)}$ .

3 V-update: obtain  $\mathbf{V}_k^{(t+1)}$  by (6) with  $\boldsymbol{\mu} = \bar{\mathbf{x}}$  and  $\mathbf{U} = (\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_{k-1}, \mathbf{U}_k^{(t+1)})$ .

4 V-scaling:  $\mathbf{V}_k^{(t+1)} \leftarrow \mathbf{V}_k^{(t+1)} / \|\mathbf{V}_k^{(t+1)}\|_2$ .

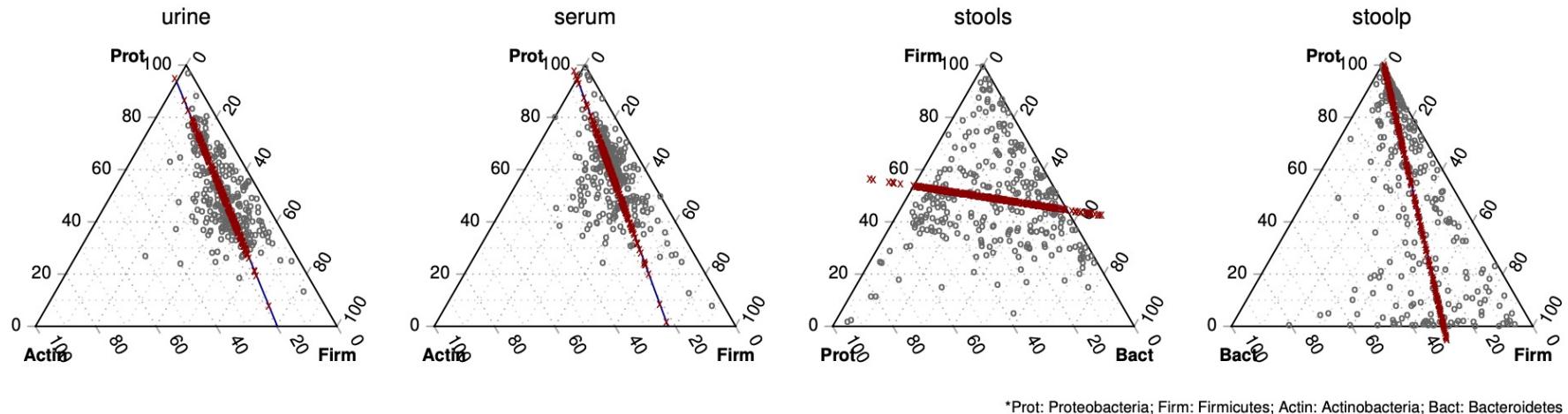
**until convergence:**  $\|\mathbf{V}_k^{(t+1)} - \mathbf{V}_k^{(t)}\|_F^2 < \epsilon$ .

**Re-estimation of  $\mathbf{U}$ :** estimate  $u_{ik}^{(t+1)}$  without the shrinkage  $\forall i$ .

**Output:**  $(\hat{\mathbf{U}}_1, \dots, \hat{\mathbf{U}}_{k-1}, \mathbf{U}_k^{(t+1)})$  and  $(\hat{\mathbf{V}}_1, \dots, \hat{\mathbf{V}}_{k-1}, \mathbf{V}_k^{(t+1)})$ .

---

# Real data example with $p = 3$



# Real data example: rank-1 reconstruction

- In the order of log-ratio PCA, crPCA, and CPCCA,

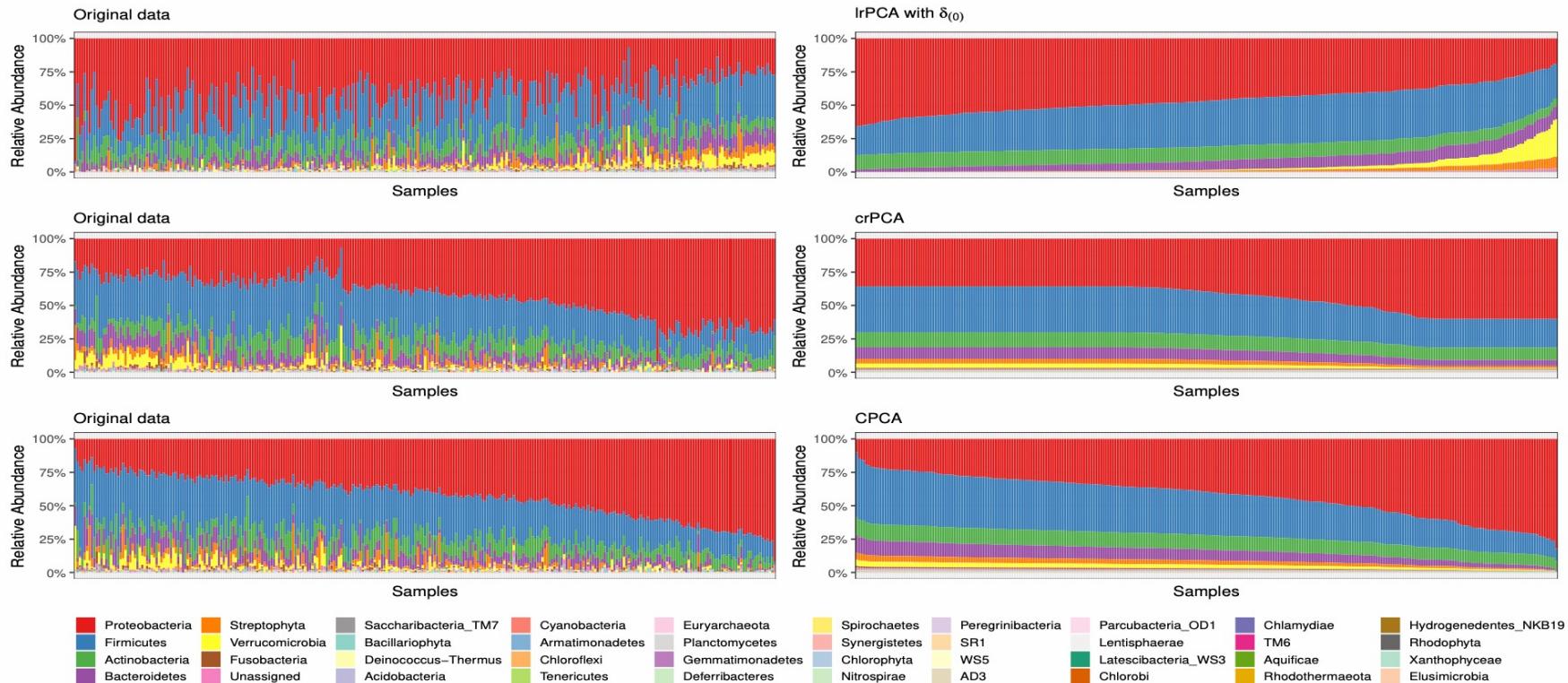


Figure: Left: the original data. Right: reconstructed data. The same sample orders were maintained between left and right panels for each method, based on its estimated first score.

# Application area

- Electrodiagnostic text data (a vector of word frequencies)

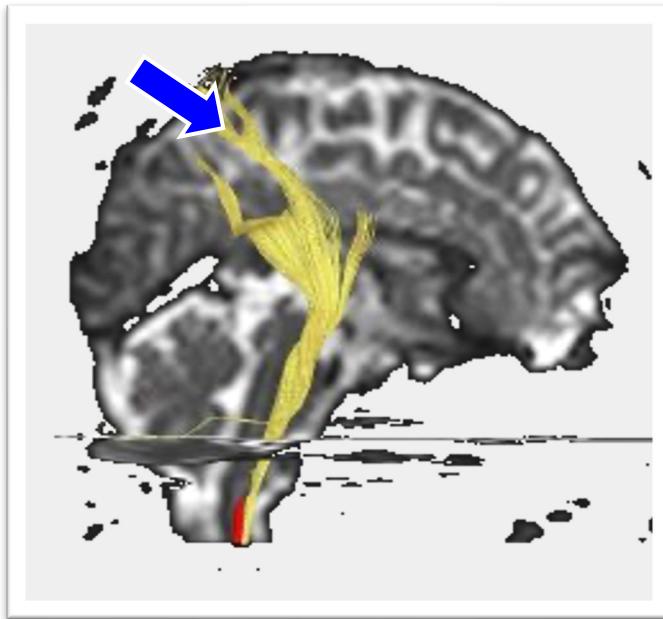
A 12-year old girl with known hyperagglutinability, presented to the emergency department with a 2-week history of headaches and facial weakness. Neurologic examination indicated sensorineural hearing loss on the right side with Weber's test lateralizing to the left, and the Rinne's test demonstrating bone conduction greater than air conduction on the right. Magnetic resonance imaging of the head revealed severe structural defects of the right petrous temporal bone. No indication of cerebral infarction.

# GLM for spherical response

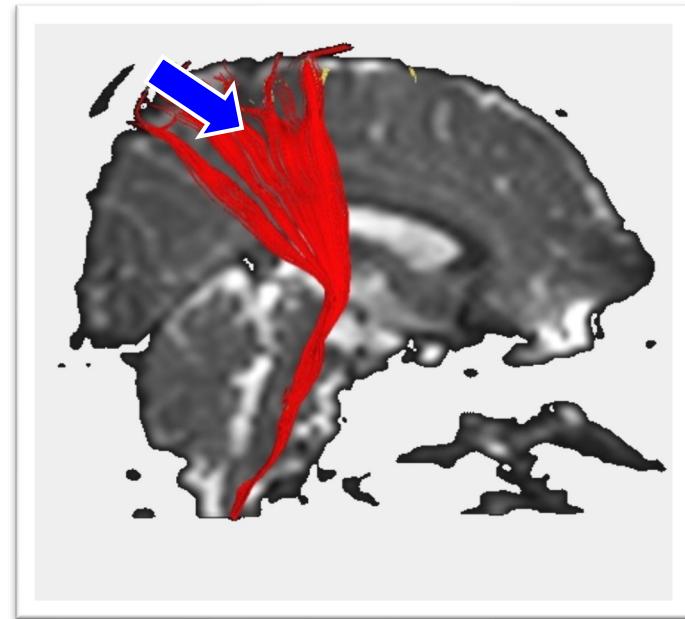
## □ Motivating example

- Two-sample testing problem in DTI data

Patient group



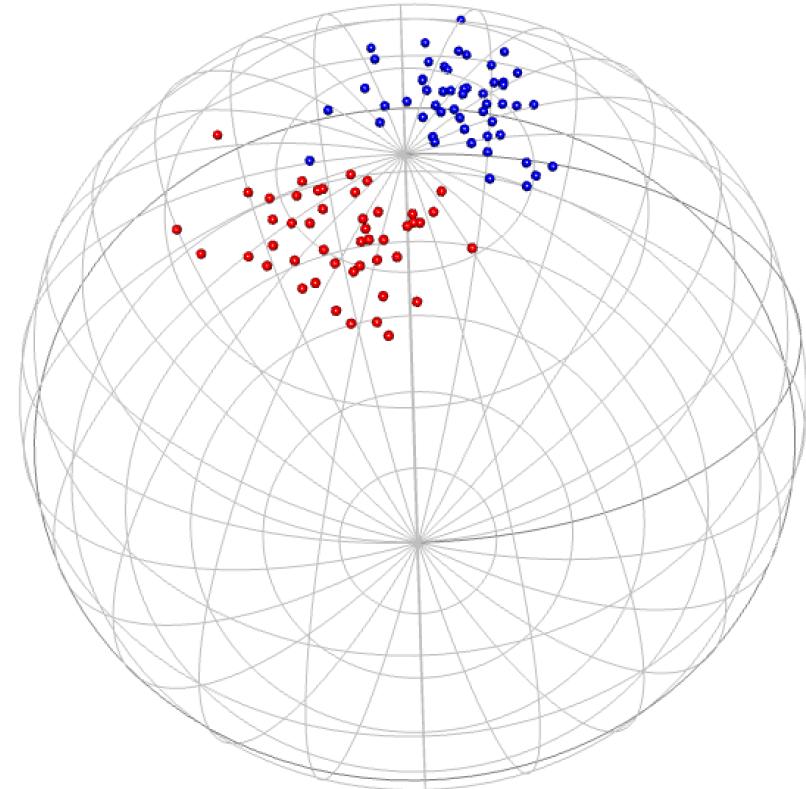
Normal group



# Two-sample testing problem in spherical data

## □ Simulated spherical responses for two-sample test problem

- Blue: patient group
- Red: normal group
  
- Here, we can see that
  - there is location difference
  - but no dispersion difference



# Spherical data and its distribution

- $S^{q-1} := \{x \in \mathbb{R}^q : x_1^2 + \cdots + x_q^2 = 1\}$
- von Mises-Fisher (vMF) distribution

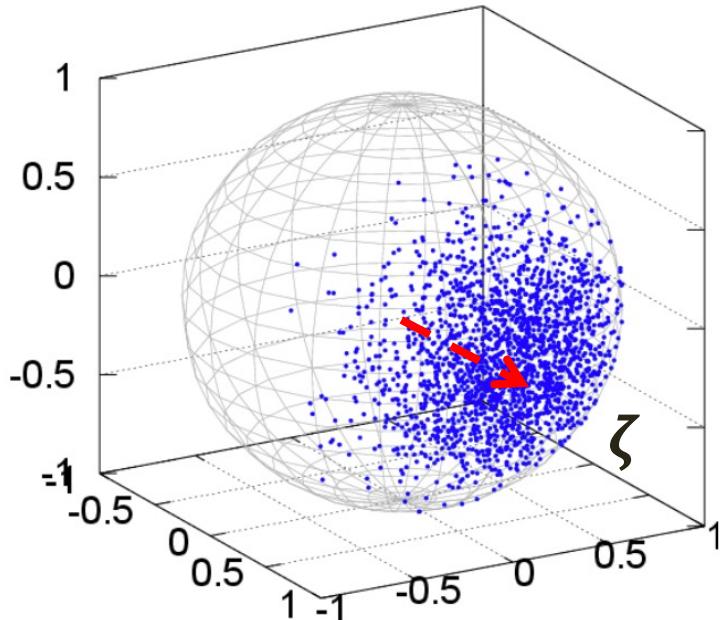
$$f_{vMF}(y; \zeta, \kappa) = C_q(\kappa) \cdot \exp(\kappa \cdot \zeta^T y),$$

where  $C_q(\kappa) = \frac{\kappa^{q/2-1}}{(2\pi)^{q/2} \cdot I_{q/2-1}(\kappa)}$  and  $I_\nu(\cdot)$  is the modified Bessel function of the first kind at order  $\nu$ .

- $\zeta$  : mean direction
- $\kappa$  : concentration parameter

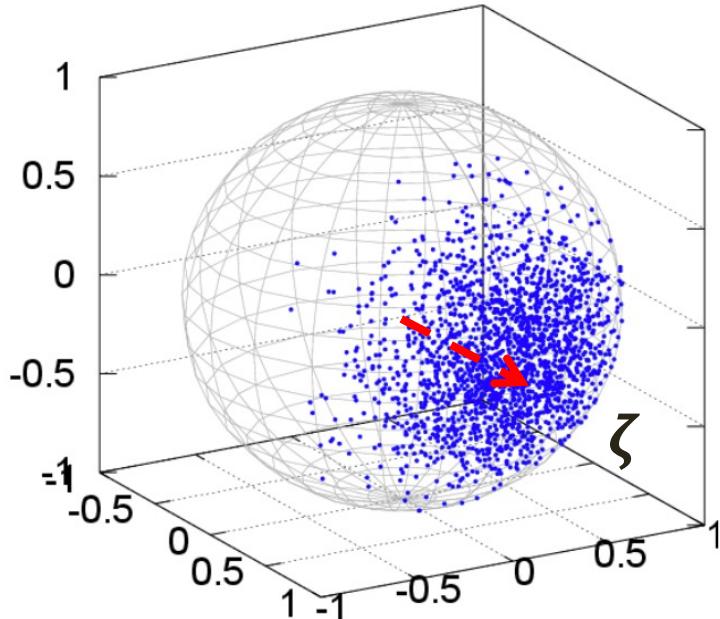
# Spherical data and its distribution

- $q = 3, \zeta = (1,0,0)^T, \kappa = 10$



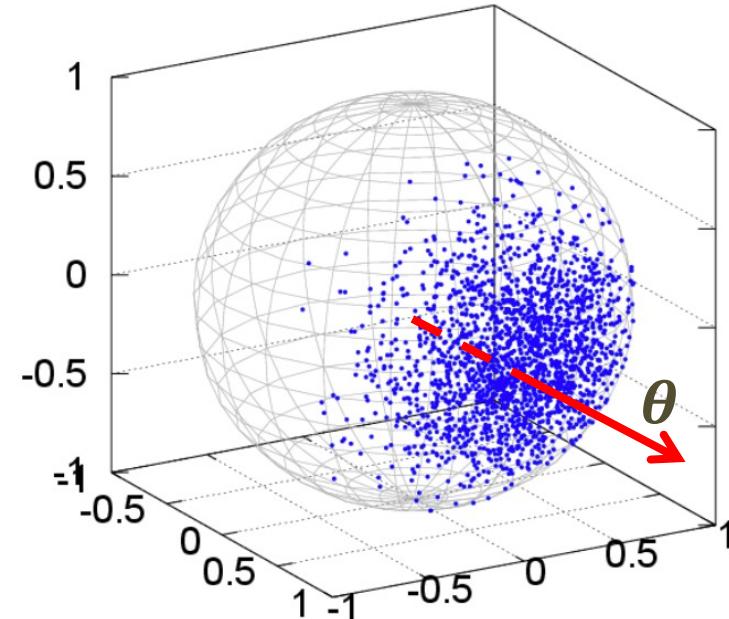
# Spherical data and its distribution

□  $q = 3, \zeta = (1,0,0)^T, \kappa = 10$



□ Reparametrization:  $\theta = \kappa \cdot \zeta$

□  $q = 3, \theta = (100, 0, 0)^T$



# Two-sample testing problem in spherical data

- Then, we can rewrite the pdf of vMF as

$$f_{vMF}(\mathbf{y}; \boldsymbol{\theta}) = C_q(\|\boldsymbol{\theta}\|_2) \cdot \exp(\boldsymbol{\theta}^T \mathbf{y}).$$

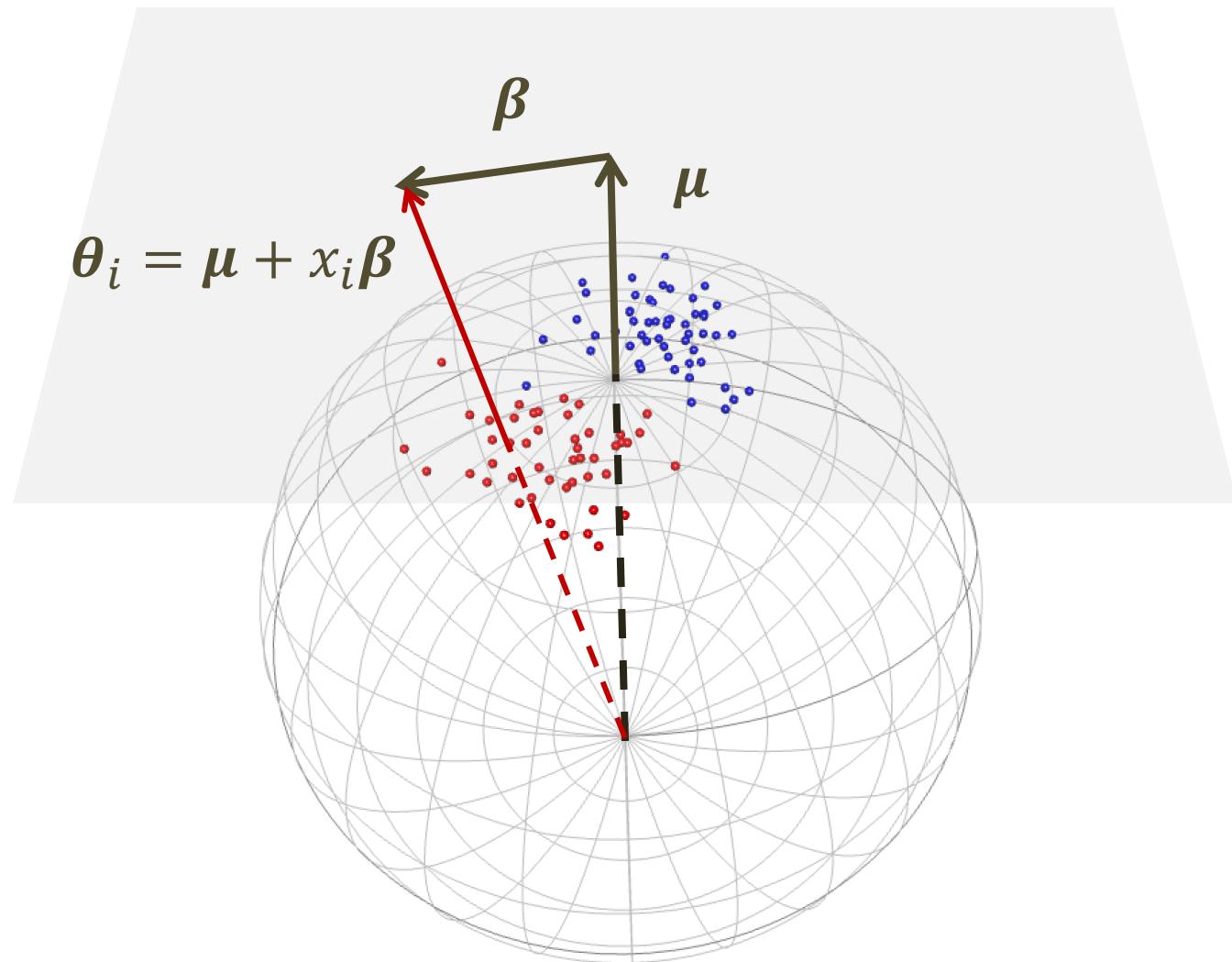
- Generalized linear model framework

$$g( E(\mathbf{y}_i | x_i) ) = \boldsymbol{\theta}_i = \boldsymbol{\mu} + x_i \boldsymbol{\beta}$$

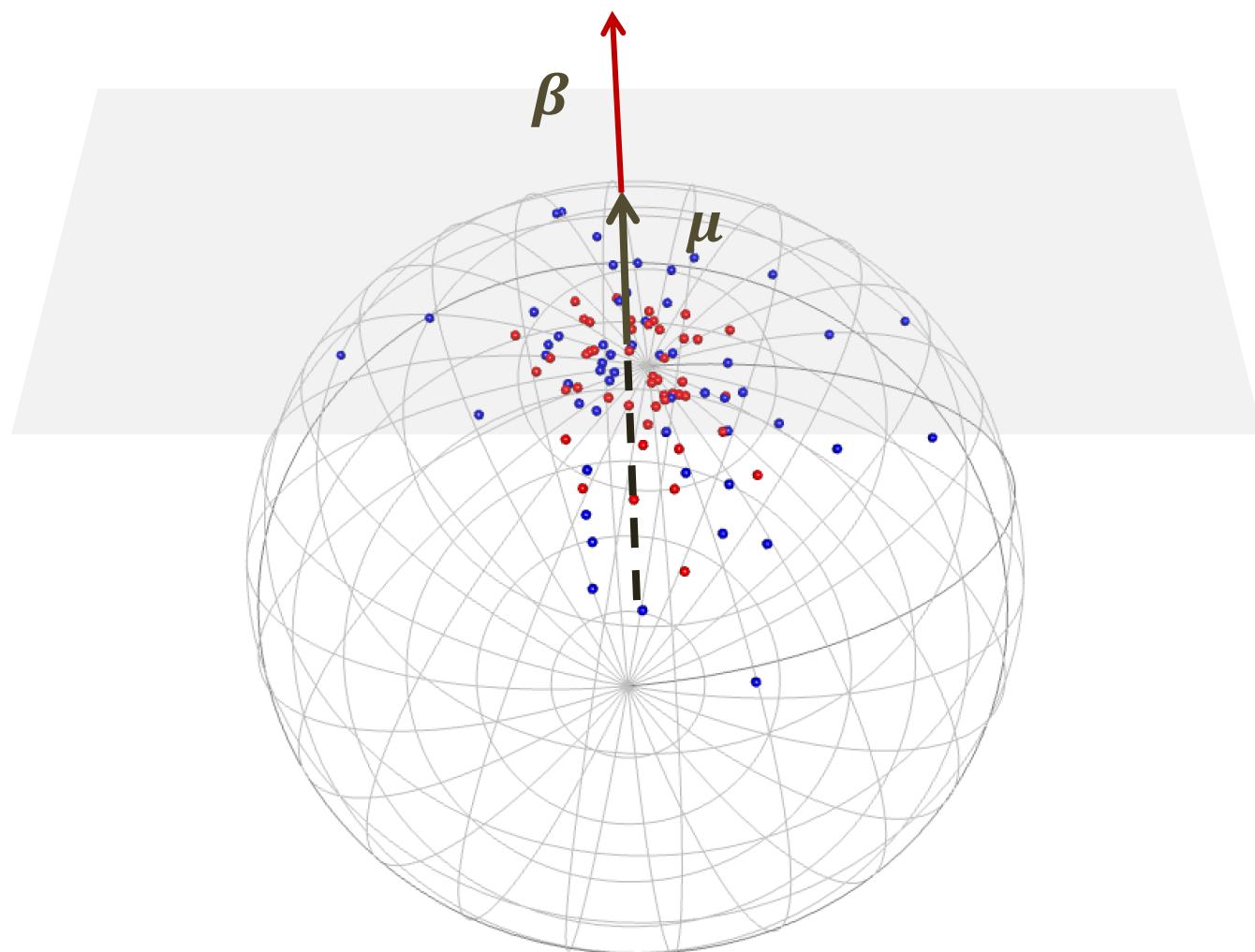
$$\mathbf{y}_i | x_i \sim vMF(\cdot | \boldsymbol{\theta}_i),$$

where  $g(\cdot)$  is a known link function.

# Two-sample problem with location difference



# Two-sample problem with dispersion difference



# Optimization problems w/o orthogonal constraint

- Let  $\tilde{\mathbf{x}}_i := (1, \mathbf{x}_i^T)^T \otimes \mathbf{I}_q \in \mathbb{R}^{(p+1)q \times q}$  and  $\boldsymbol{\beta}^* := (\boldsymbol{\mu}^T, \boldsymbol{\beta}_1^T, \dots, \boldsymbol{\beta}_p^T)^T \in \mathbb{R}^{pq \times 1}$
- Estimation on the unconstrained model

$$\arg \max_{\boldsymbol{\beta}^* \in \mathbb{R}^{pq}} \sum_{i=1}^n \left\{ (\tilde{\mathbf{x}}_i^T \boldsymbol{\beta}^*)^T \mathbf{y}_i + \log C_q(\|\tilde{\mathbf{x}}_i^T \boldsymbol{\beta}^*\|) \right\}.$$

- Estimation on the constrained model

$$\arg \max_{\substack{\boldsymbol{\beta}^* \in \mathbb{R}^{pq} \\ \boldsymbol{\gamma} \in \mathbb{R}^p}} \sum_{i=1}^n \left\{ \boldsymbol{\beta}^{*T} \tilde{\mathbf{x}}_i \mathbf{y}_i - b(\tilde{\mathbf{x}}_i^T \boldsymbol{\beta}^*) \right\} + \sum_{j=1}^p \gamma_j \boldsymbol{\mu}^T \boldsymbol{\beta}_j$$

➤  $\gamma_j$ 's are lagrangian multipliers

# Asymptotic analysis

□ Based on the standard likelihood theory, we have

**Theorem 1.** Let us assume that the assumptions (A1)–(A2) is satisfied.

(i) Under the assumptions (A4), the weak consistency of  $\hat{\beta}^*$  can be obtained as follows:

$$\hat{\beta}^* \xrightarrow{p} \beta_0^*.$$

(ii) Under the assumption (A3), the asymptotic normality of  $\hat{\beta}^*$  can be obtained as follows:

$$F_n^{T/2}(\hat{\beta}^* - \beta_0^*) \xrightarrow{d} N(\mathbf{0}, I_{pq}).$$

➤  $F_n$  is the Fisher information matrix at true  $\beta_0^*$  and  $F^{T/2}$  indicates the transpose of the Cholesky square root matrix for  $F$

# Asymptotic analysis following projection

- Let  $P_{\hat{\mu}} = \frac{\hat{\mu}\hat{\mu}^T}{\|\hat{\mu}\|_2^2}$ ,  $P_{\hat{\mu}^\perp} = \mathbf{I} - P_{\hat{\mu}}$  be the projection matrix onto  $\hat{\mu}$  and its orthogonal complement, respectively.

- Then we have

$$\mathbf{P}_{\hat{\mu}^\perp} \hat{\boldsymbol{\beta}}_j | \hat{\mu} \xrightarrow{d} N(\mathbf{P}_{\mu^\perp} \boldsymbol{\beta}_j, \Sigma_{\mu^\perp})$$

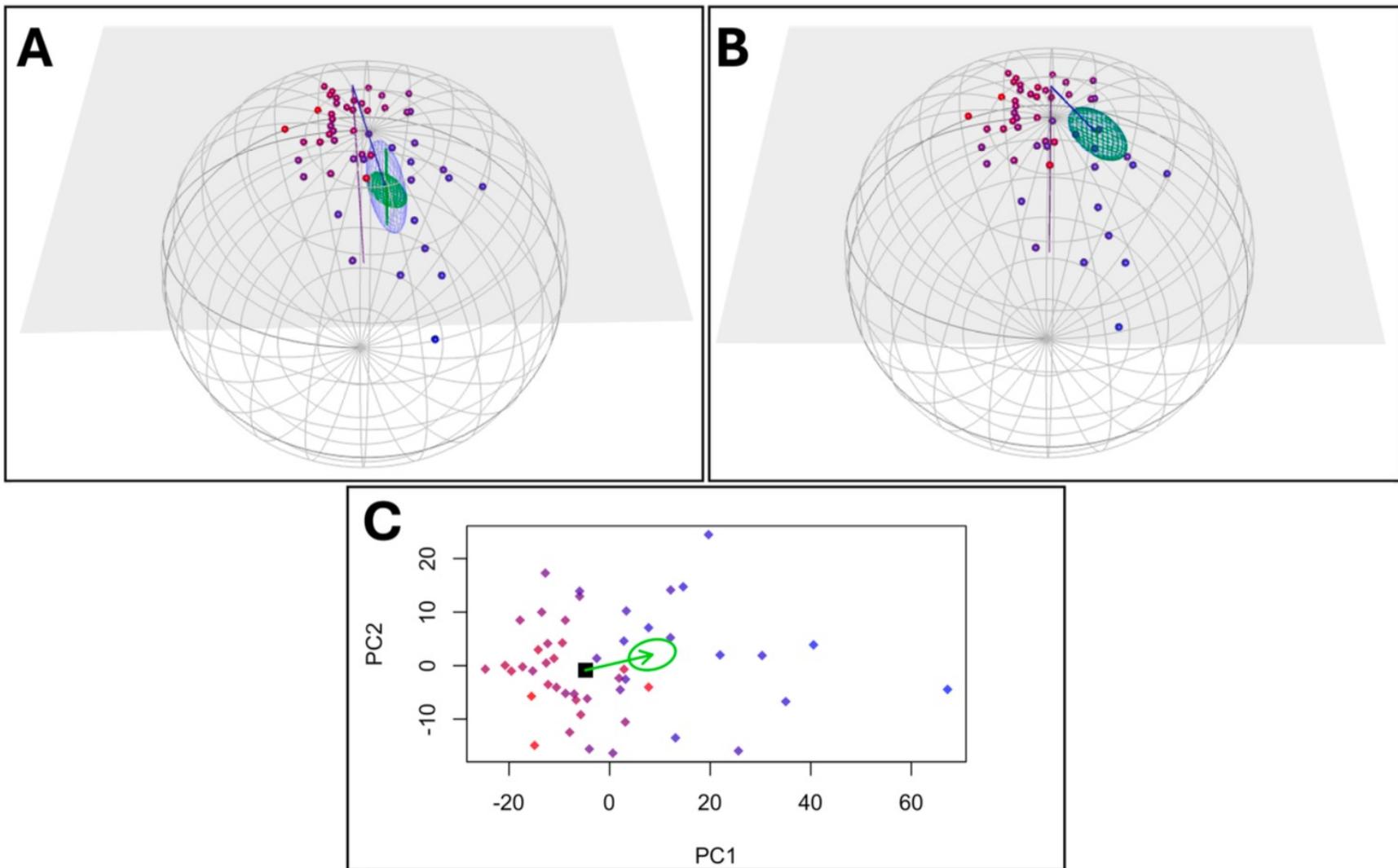
➤  $\Sigma_{\mu^\perp} = P_{\mu^\perp} c_j^T [\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}] c_j P_{\mu^\perp}^T$  and  $c_j$  is the index matrix representing the  $j$ -th covariate

- Similarly, for the  $\mu$ -directional projection, we have the following result

$$\mathbf{P}_{\hat{\mu}} \hat{\boldsymbol{\beta}}_j | \hat{\mu} \xrightarrow{d} N(\mathbf{P}_{\mu} \boldsymbol{\beta}_j, \Sigma_{\mu})$$

➤  $\Sigma_{\mu} = P_{\mu} c_j^T [\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}] c_j P_{\mu}^T$

# Orthogonal constraint: without vs. with



**Thank you for your attention !**