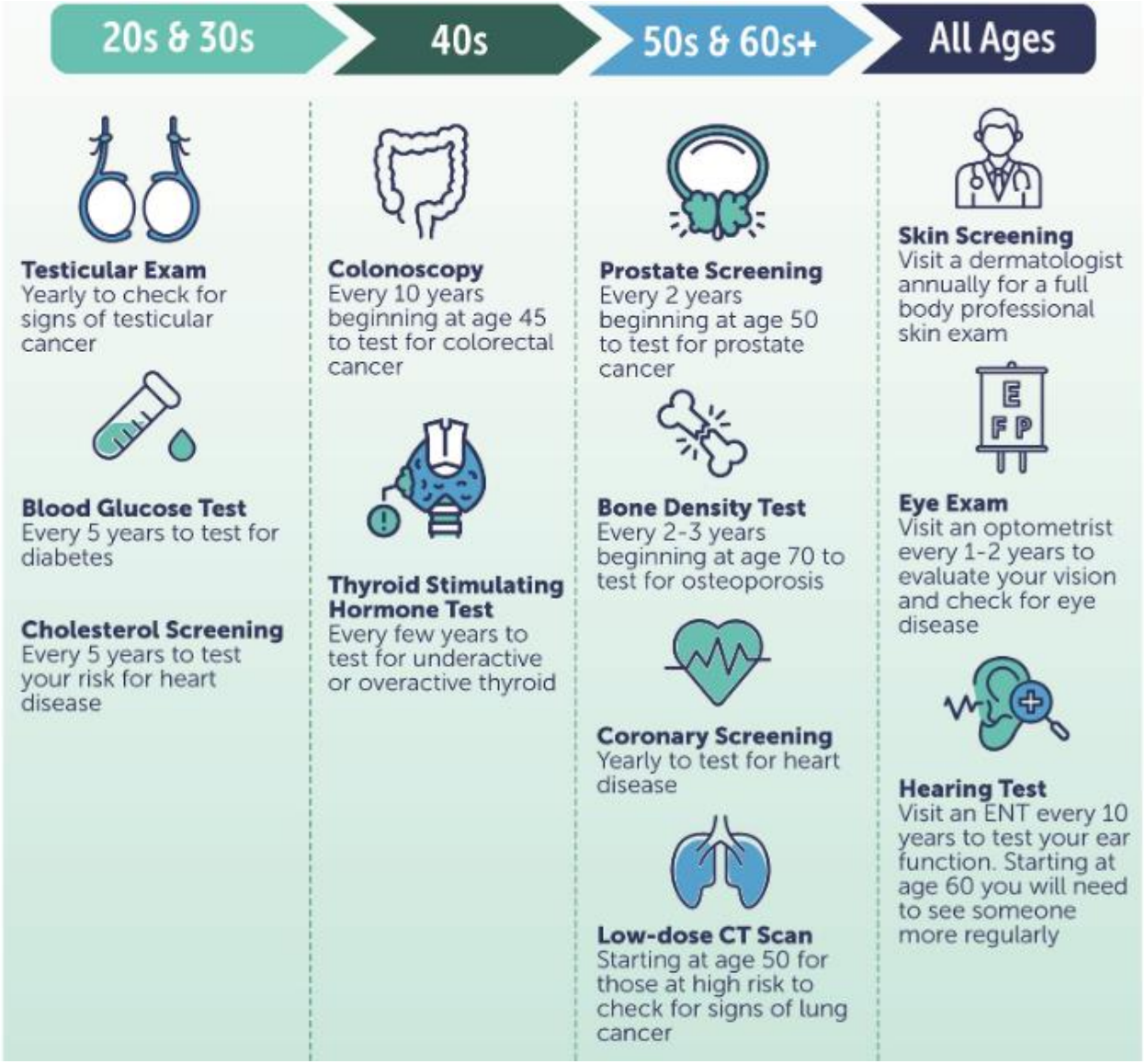
Low-rank models for multiple non-Euclidean datasets

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
1	DATE	DAY	BRAIN	LUNGS	HEART	SYSTOLIC	DIASTOLIC	CELSIUS	PULSE
2	11/1/2020	Sunday	5	5	5	123	82	36.6	172
3	11/2/2020	Monday	5	5	5	119	78	36.6	179
4	11/3/2020	Tuesday	5	5	5	111	80	36.6	84
5	11/4/2020	Wednesday	5	5	5	120	80	36.6	162
6	11/5/2020	Thursday	5	4	5	120	80	36.6	52
7	11/6/2020	Friday	5	5	5	125	81	36.6	80
8	11/7/2020	Saturday	2	4	5	90	56	37.2	95
9	11/8/2020	Sunday	2	2	3	101	68	37.4	171
10	11/9/2020	Monday	5	4	4	147	95	37.6	76
11	11/10/2020	Tuesday	5	3	4	199	133	37.7	151
12	11/11/2020	Wednesday	4	2	3	97	70	37.8	154
13	11/12/2020	Thursday	4	3	4	193	125	38.3	140
14	11/13/2020	Friday	2	1	2	114	74	38.4	134
15	11/14/2020	Saturday	2	1	3	207	151	38.5	102

Electrical Health Records

Clinical Test

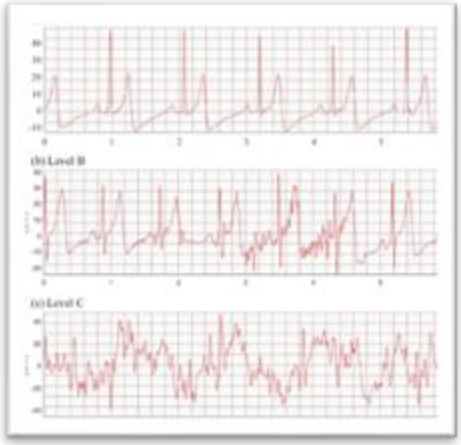
Functional Test

Radiological test

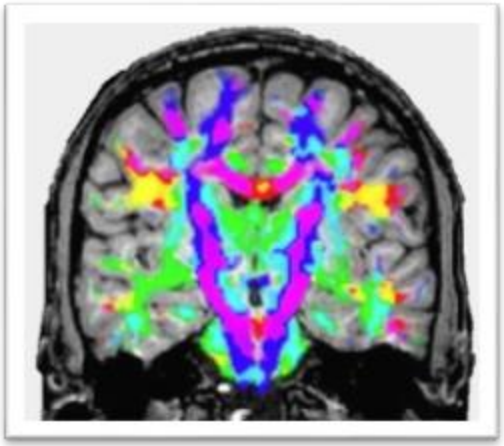
ETC

Sequencing data

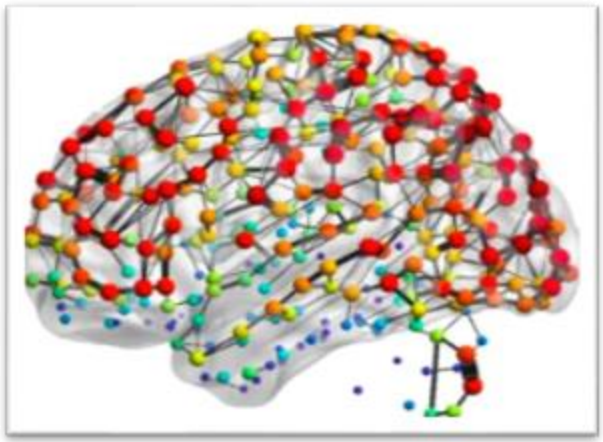
Electrocardiogram (ECG)



Medical Imaging data



Connectome



비유클리드 데이터 (Non-Euclidean data)

- 비유클리드 데이터: 내재된 구조가 유클리드가 아닌 데이터

- k차원 유클리드 공간 (R^k)의 특성:
  - 벡터 덧셈($+$)과 스칼라 곱($\cdot$)에 대해 닫혀있는 벡터 공간
  - 유한 차원 공간

- 비유클리드 데이터의 특성:
  - 벡터 덧셈($+$)과 스칼라 곱($\cdot$)에 대해 닫혀있지 않은 유한 차원 벡터 공간
  - 유한 차원 비벡터 공간
  - 무한 차원 공간

Non-Euclidean data: Image data (pixel)

- A: 고양이 사진
  - $\begin{bmatrix} 100 & 150 \\ 120 & 160 \end{bmatrix}$

- B: 강아지 사진
  - $\begin{bmatrix} 80 & 120 \\ 90 & 140 \end{bmatrix}$

-  +  = $\begin{bmatrix} 180 & 270 \\ 210 & 300 \end{bmatrix}$?

-  x 2 = $\begin{bmatrix} 200 & 300 \\ 240 & 320 \end{bmatrix}$?

Non-Euclidean data: Image data (pixel)

- 유클리드 거리가 실제 유사도를 대표하지 않음.

Non-Euclidean data: Text data

- 거리 개념이 잘 정의되지 않음: (1)
  - A: ChatGPT는 텍스트를 생성한다.
  - B: ChatGPT는 대화를 한다.
  - C: AI는 텍스트를 만든다.

- 거리 개념이 잘 정의되지 않음: (2)
  - A: ChatGPT는 대화를 잘한다.
  - B: 인공지능은 대화를 한다.
  - C: 로봇은 움직인다.
    - d(A, C) >> d(A, B) + d(B, C)

Non-Euclidean data: Network data

- Research topics of interest
  - [가설검정]: 두 네트워크의 구조는 동일하다.
  - [회귀분석]: 시간에 따른 네트워크 구조의 변화

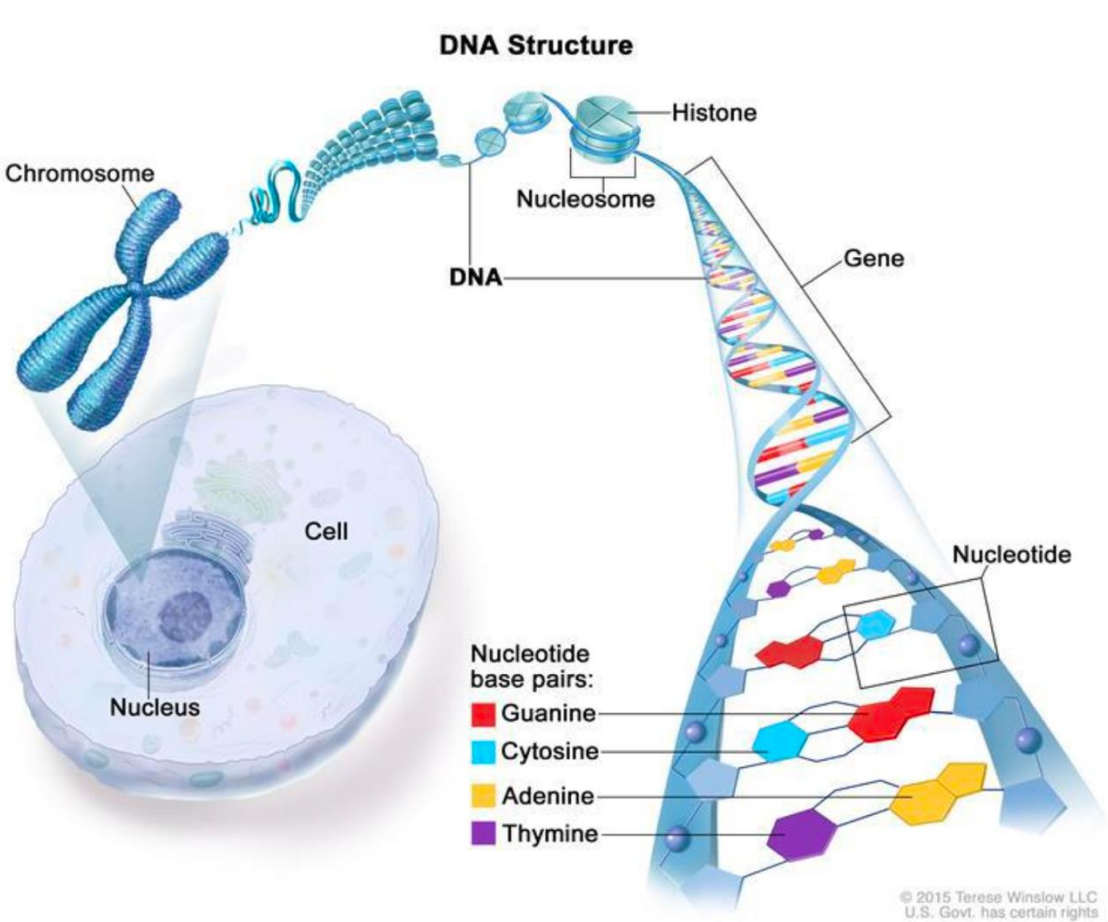
- Key questions:
  - [가설검정]: 네트워크 구조 간 차이를 어떻게 정의할까?
  - [회귀분석]: 평균 네트워크가 어떻게 정의되는가?

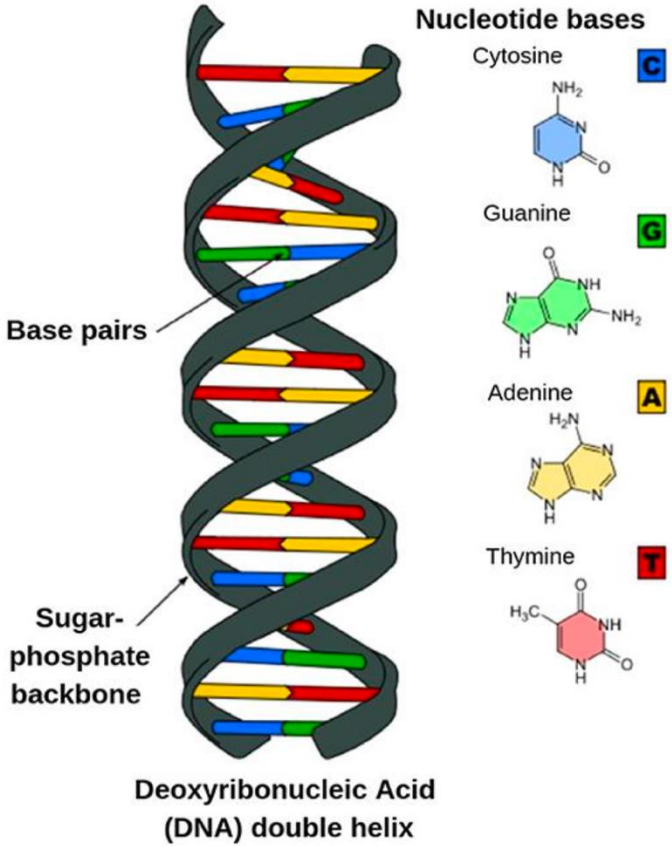
Non-Euclidean data: Compositional data

- Examples
  - 토양 구성 비율
    - 모래 (45.3%), 점토 (30.2%), 유기물 (15.8%), 미네랄 (8.7%)
  - 하루 시간 사용 분석 (전체 24시간)
    - 수면 (8시간), 업무 (9.5시간), 여가 (3.5시간), 기타 (3.0시간)
  - 주식 포트폴리오 구성
    - 엔비디아 (45.9%), 테슬라 (7.2%), TQQQ (34.7%), 비트코인 (12.2%)
  - 문서 내 단어 빈도 (배민리뷰 감성분석)
    - 긍정표현 (65%), 부정표현 (10%), 중립표현 (25%)
  - **장내 미생물 비율**
    - Firmicutes (59.5%), Proteobacteria (35.5%), Bacteroidetes (5%)

Genomic data

- DNA structure



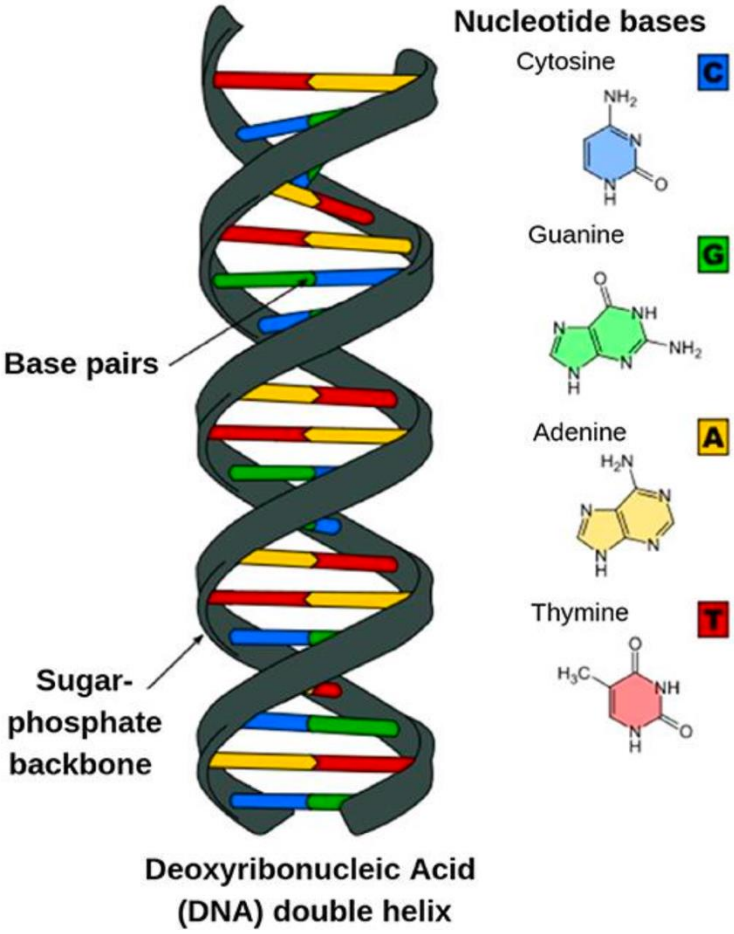


Genomic data

- DNA = 염기구조
  - 첫번째 지역 (base)에서의 염기
    - 사람 A: CC >> '0' 코딩
    - 사람 B: GC >> '1' 코딩
    - 사람 C: GG >> '2' 코딩
    - .....

- 간단하게 말하면,

  DNA 데이터 = {0, 1, 2}로 구성



Genomic data

- Human Genome Project
  - Genes: ~40,000개
  - DNA: ~5,000,000개
  - 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)

Application area of genomic data

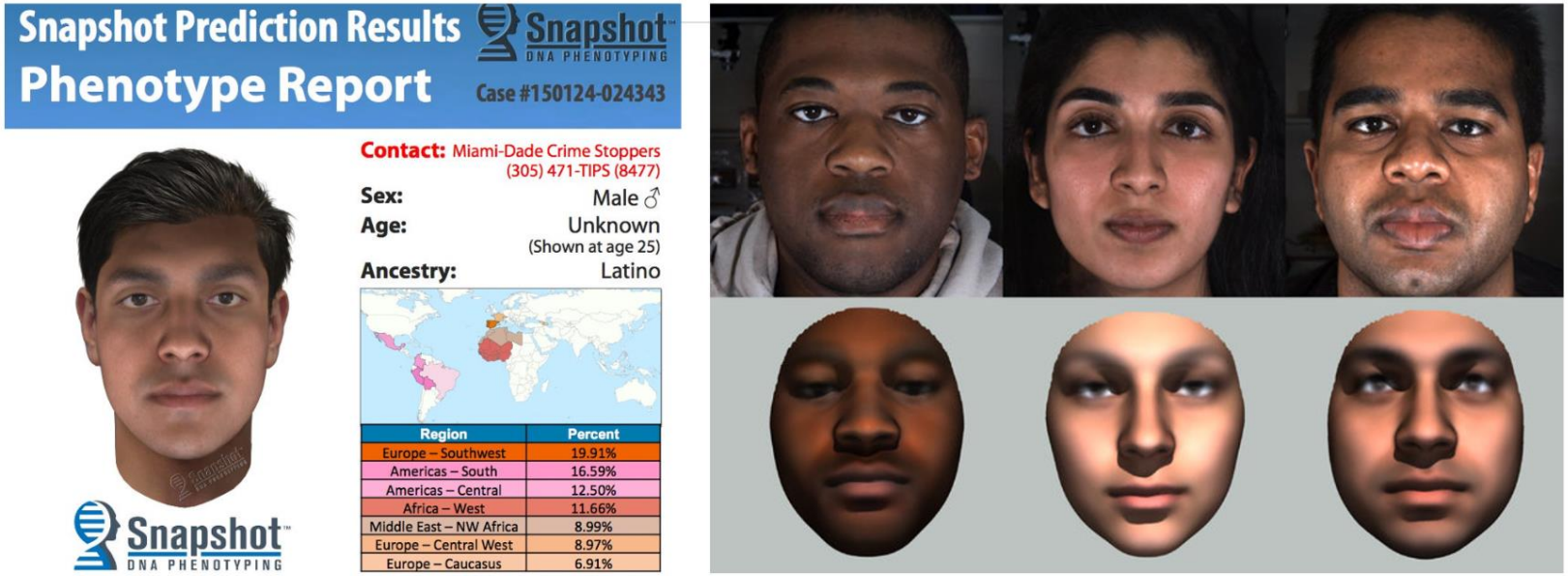
- **Personalized Medicine: disease prediction**

GWAS catalog (J. MacArthur et al., Nucleic acids research 45, D896–D901, 2017)

Application area of genomic data

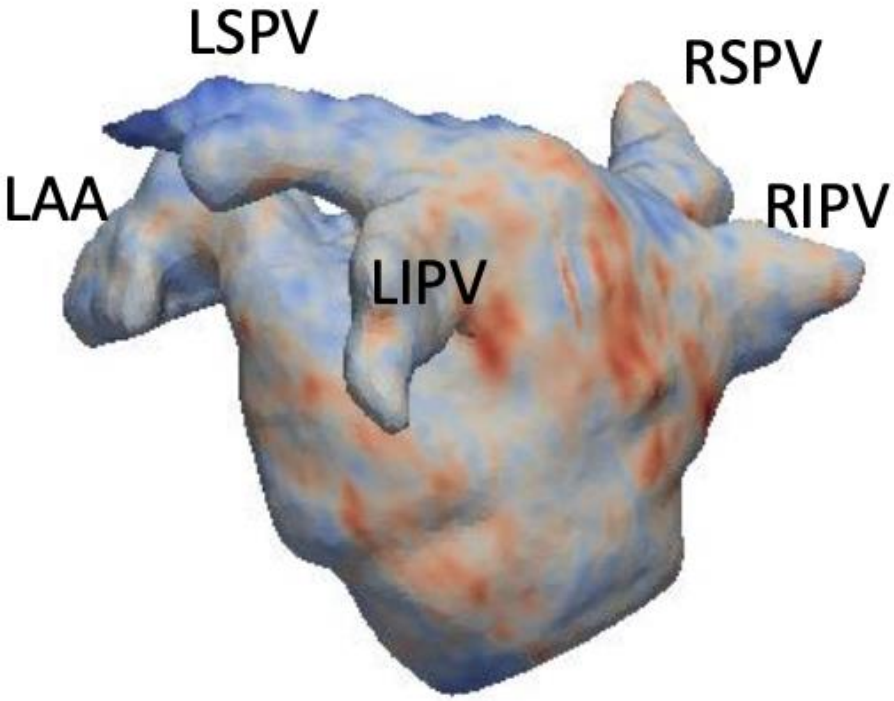
- **DNA phenotyping**

Florida police used a smidgen of DNA to try to fully reconstruct an alleged criminal's face



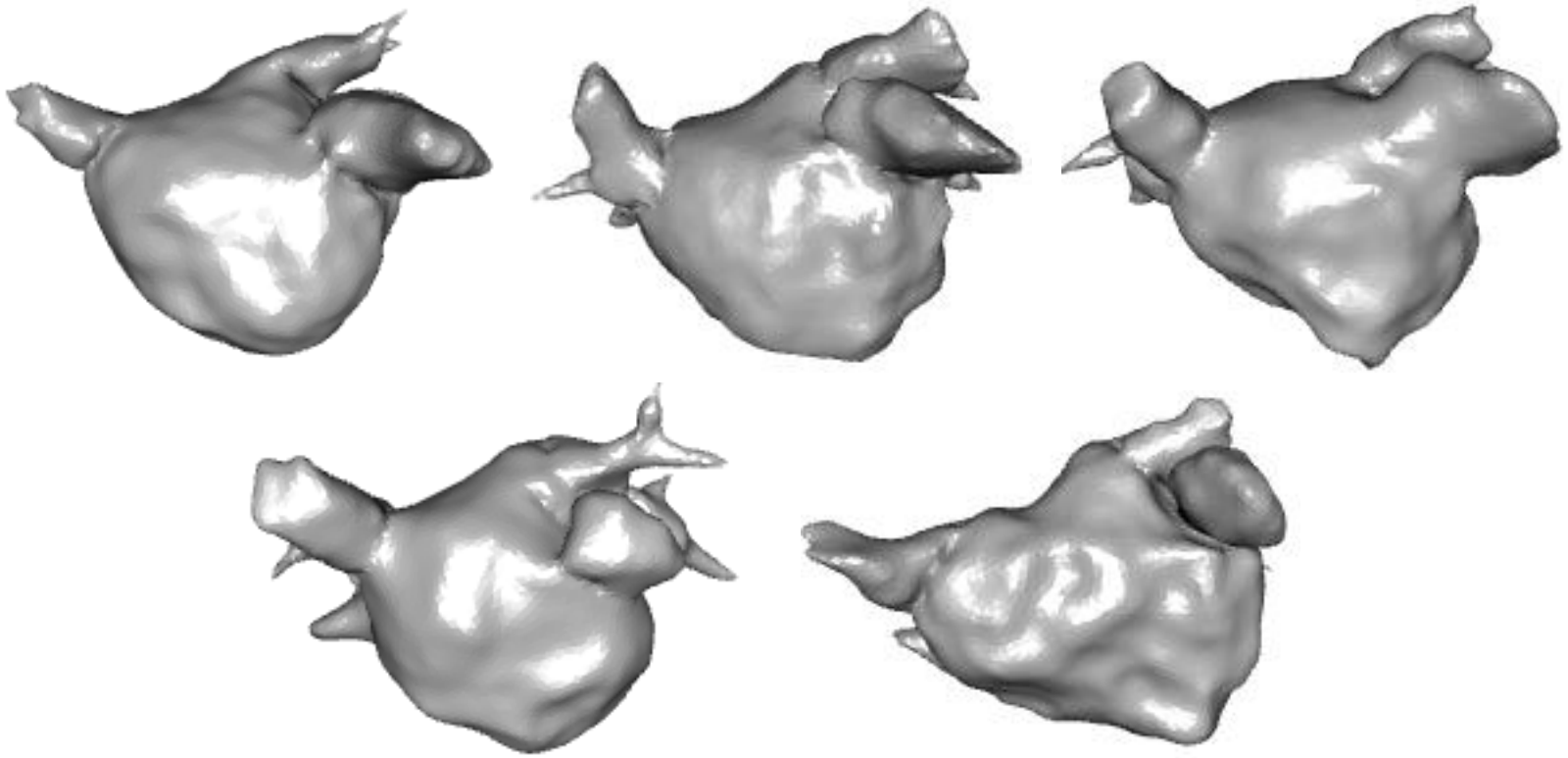
CT / MRI: 3d point cloud data

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



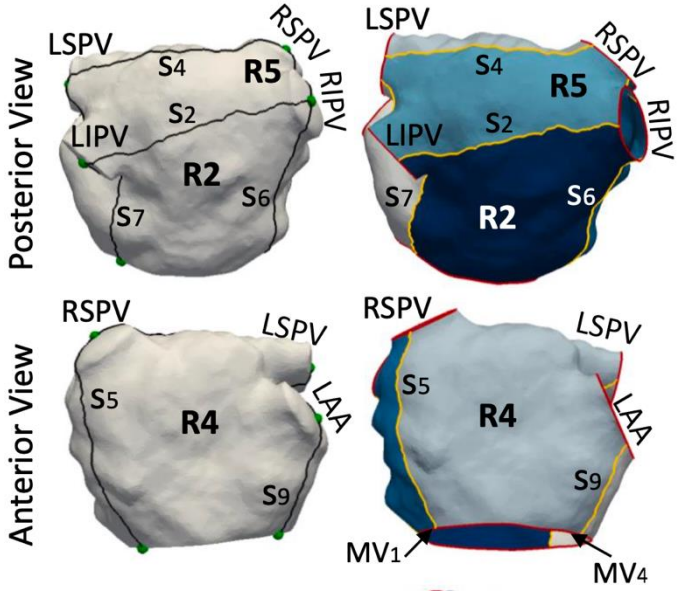
CT / MRI: 3d point cloud data

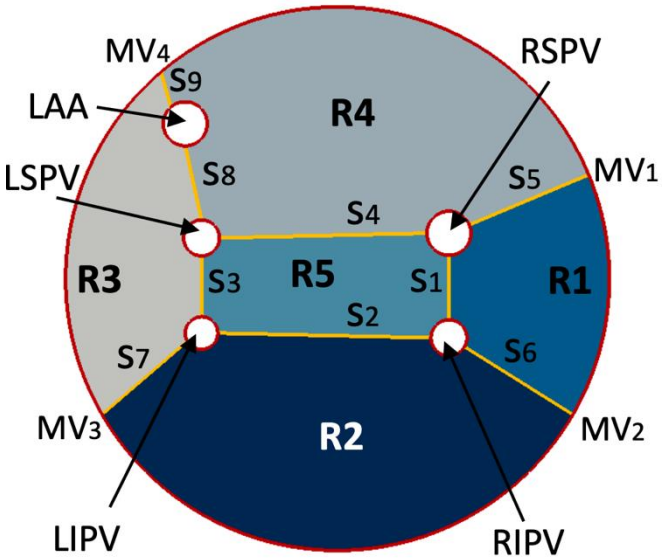
- Issue: LA shape differs among individuals.

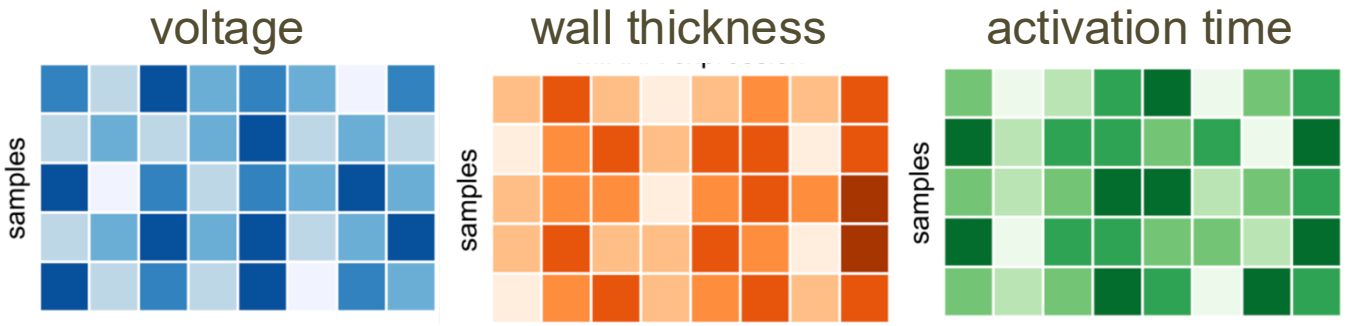


CT / MRI: 3d point cloud data

- Registration procedure using flattening

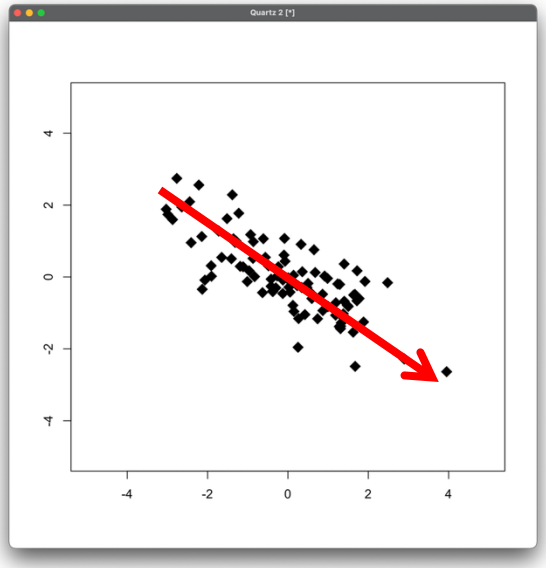


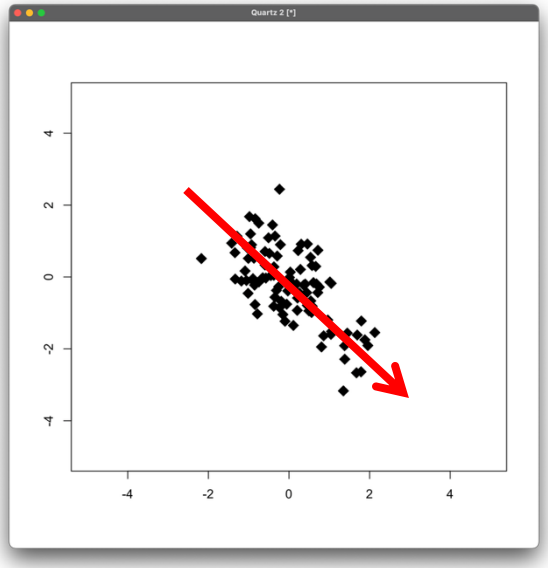


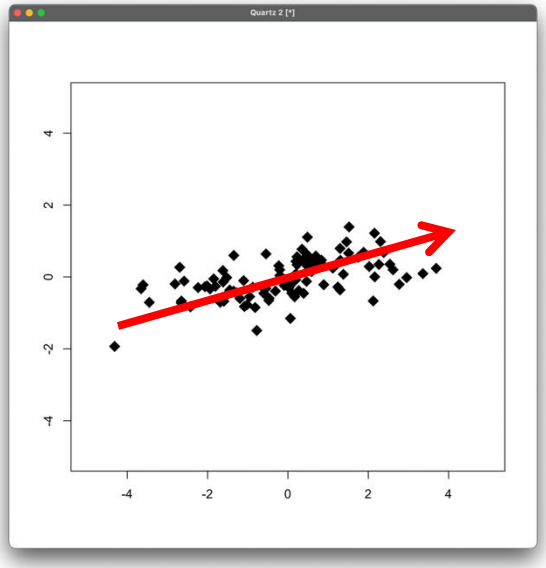


Multi-source data integration

- 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

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

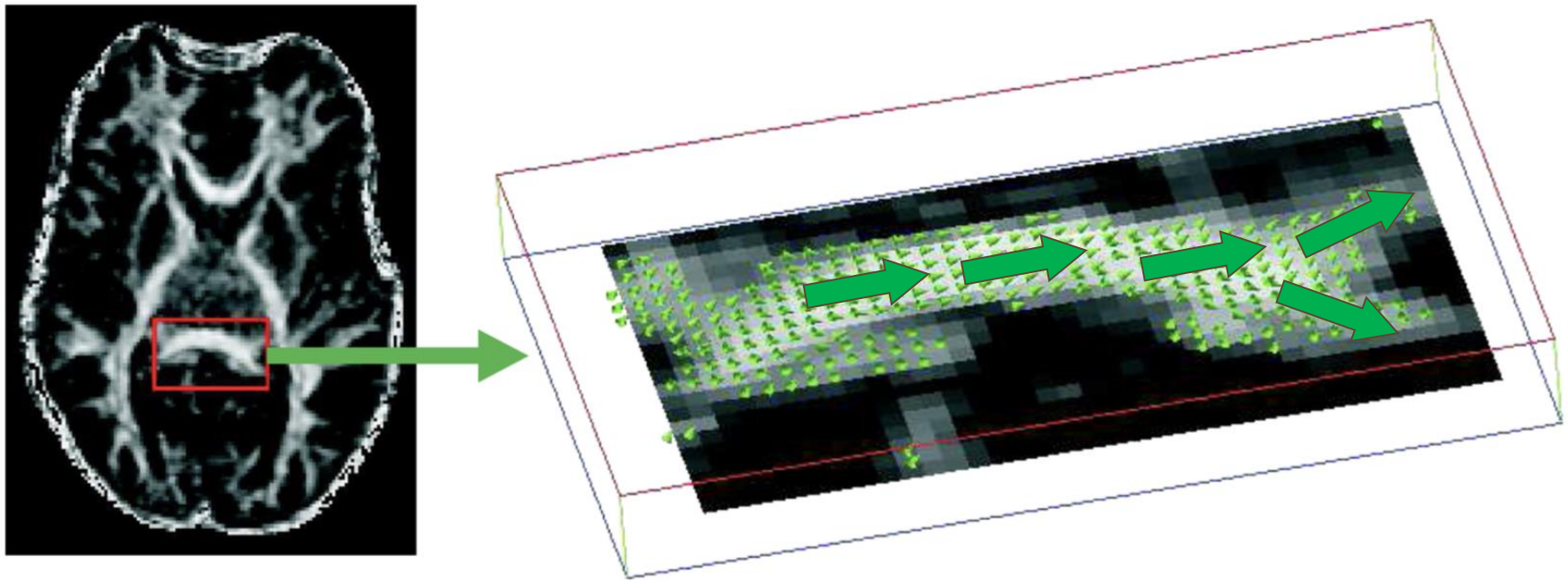
- For the d-th source case:

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

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

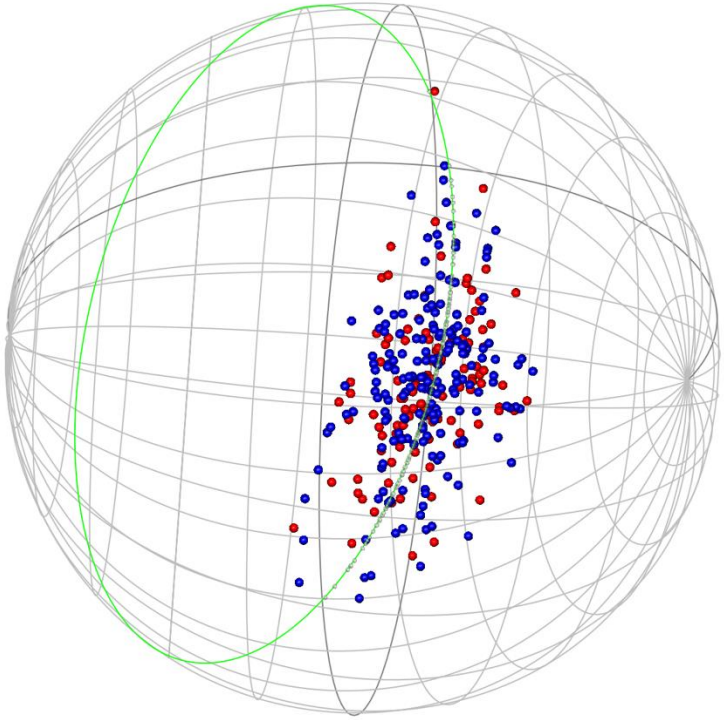
CT / MRI: DTI data

- Diffusion Tensor Imaging (DTI) data from MRI



CT / MRI: DTI data

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



Spherical data is on a vector space?

- For $\mathbf{x}, \boldsymbol{y} \in \mathrm{S}^{p-1} = \{\boldsymbol{x} \in \mathrm{R}^{p} : ||\boldsymbol{x}||_2 = 1\}$
  - $\boldsymbol{x} + \boldsymbol{y} \notin \mathrm{S}^{p-1}$
  - $c \cdot \boldsymbol{x} \notin \mathrm{S}^{p-1}$, for $c \in \mathrm{R} \setminus \{1\}$
  - $\boldsymbol{x} - \boldsymbol{y} \notin \mathrm{S}^{p-1}$

CT / MRI: 3d point cloud data

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2022, VOL. 00, NO. 0, 1–13
https://doi.org/10.1080/10618600.2022.2116445

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Statistical Analysis of Locally Parameterized Shapes

Mohsen Taheri[a] and Jörn Schulz[b]

Department of Mathematics & Physics, University of Stavanger, Stavanger, Norway

ABSTRACT

In statistical shape analysis, the establishment of correspondence and defining shape representation are crucial steps for hypothesis testing to detect and explain local dissimilarities between two groups of objects. Most commonly used shape representations are based on object properties that are either extrinsic or noninvariant to rigid transformation. Shape analysis based on noninvariant properties is biased because the act of alignment is necessary, and shape analysis based on extrinsic properties could be misleading. Besides, a mathematical explanation of the type of dissimilarity, for example, bending, twisting, stretching, etc., is desirable. This work proposes a novel hierarchical shape representation based on invariant and intrinsic properties to detect and explain locational dissimilarities by using local coordinate systems. The proposed shape representation is also superior for shape deformation and simulation. The power of the method is demonstrated on the hypothesis testing of simulated data as well as the left hippocampi of patients with Parkinson's disease and controls. Supplementary materials for this article are available online.

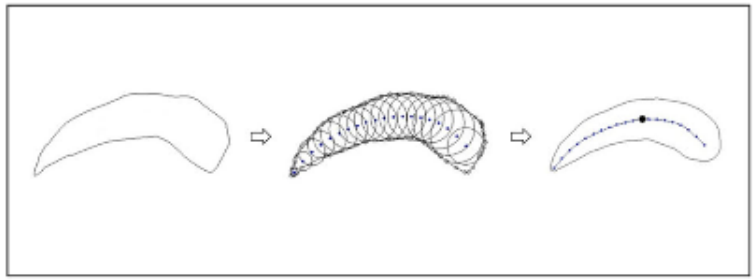
ARTICLE HISTORY
Received August 2021
Accepted July 2022

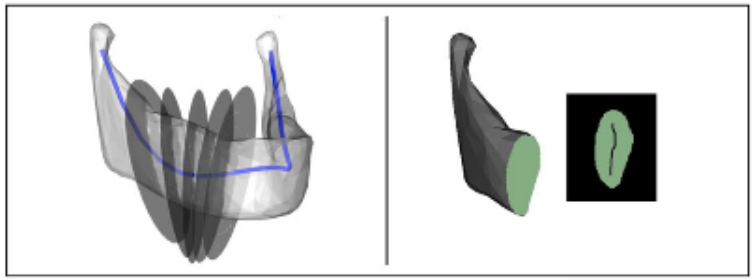
KEYWORDS
Local coordinate system; Local dissimilarity; Parkinson's disease; Shape alignment; Skeletal representation; s-Rep hypothesis testing

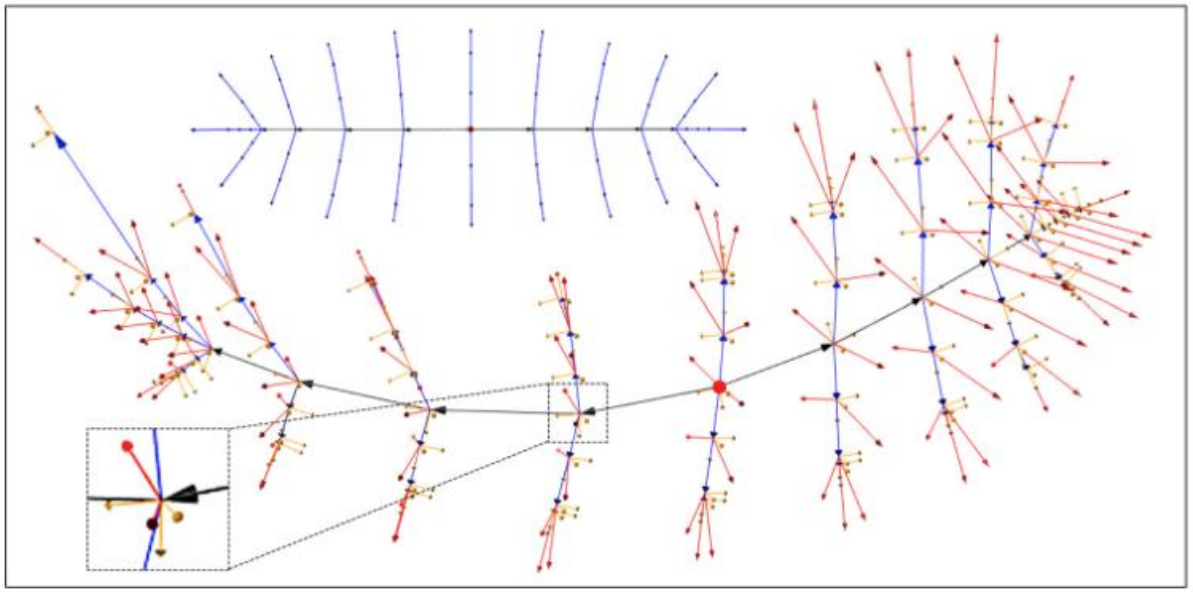
(a) 2D-ellipsoidal (b) 3D-ellipsoidal

CT / MRI: 3d point cloud data

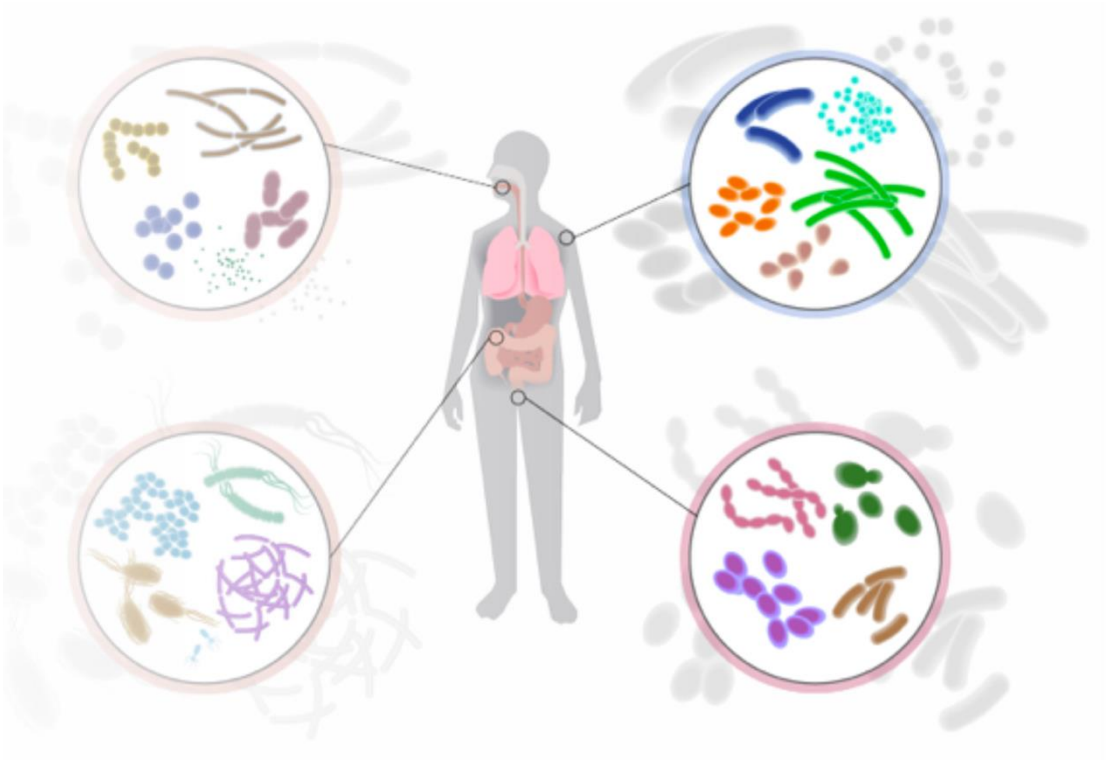
- Shape analysis $\mathbf{x} \in (S^2)^{n_s+n_c} \times \left(SO(3)\right)^{n_f} \times (R_+)^{n_s+n_c+1}$
  - Discrete skeletal representation

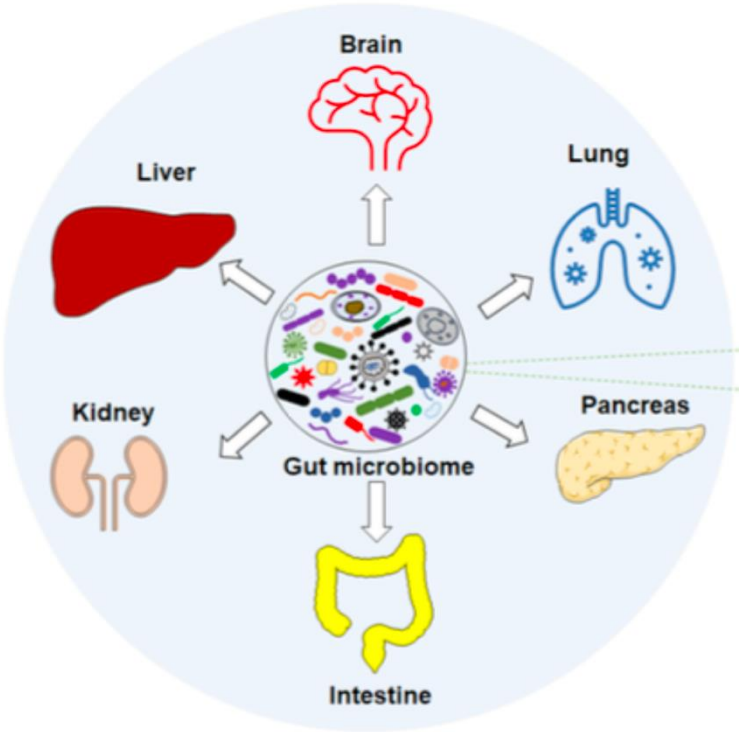






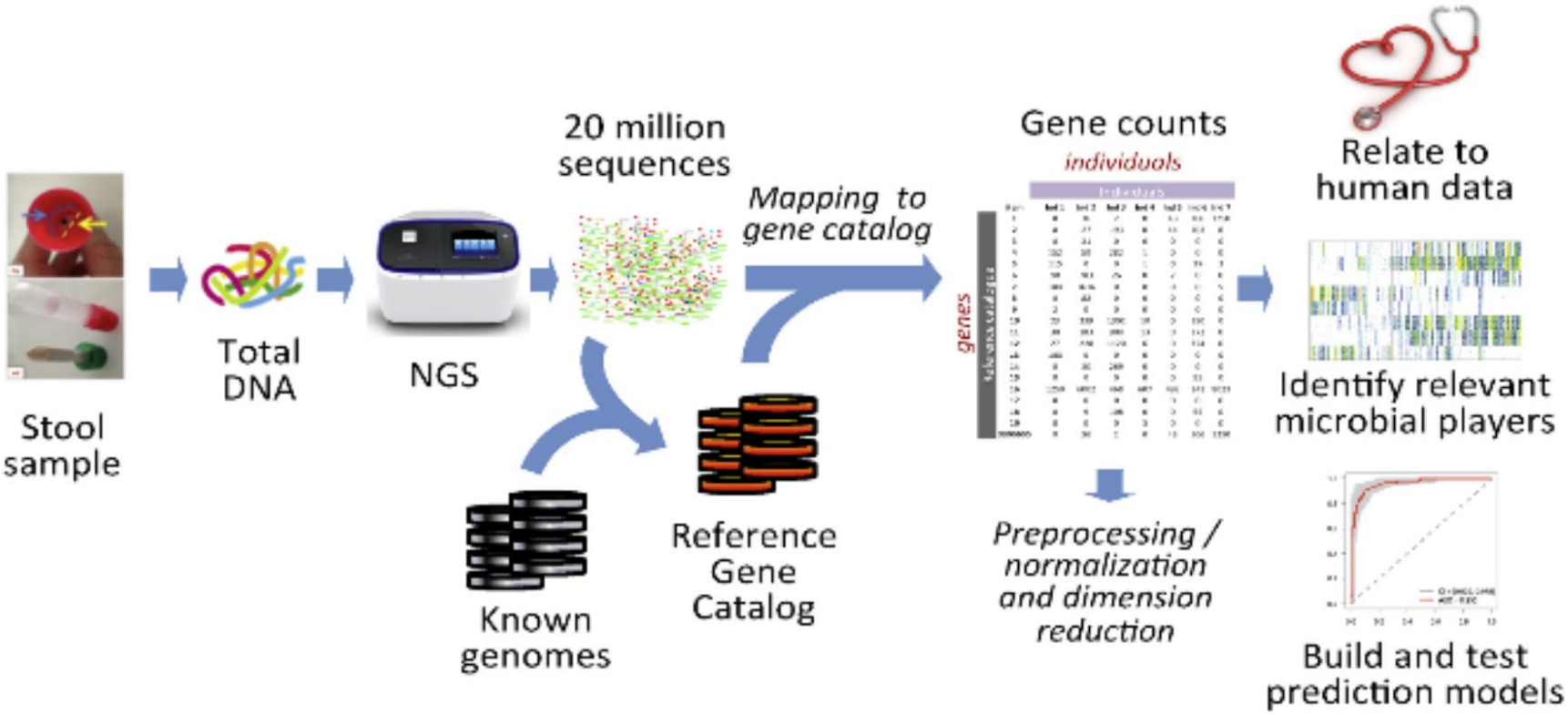
Microbiome data





Microbiome data

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



Microbiome data

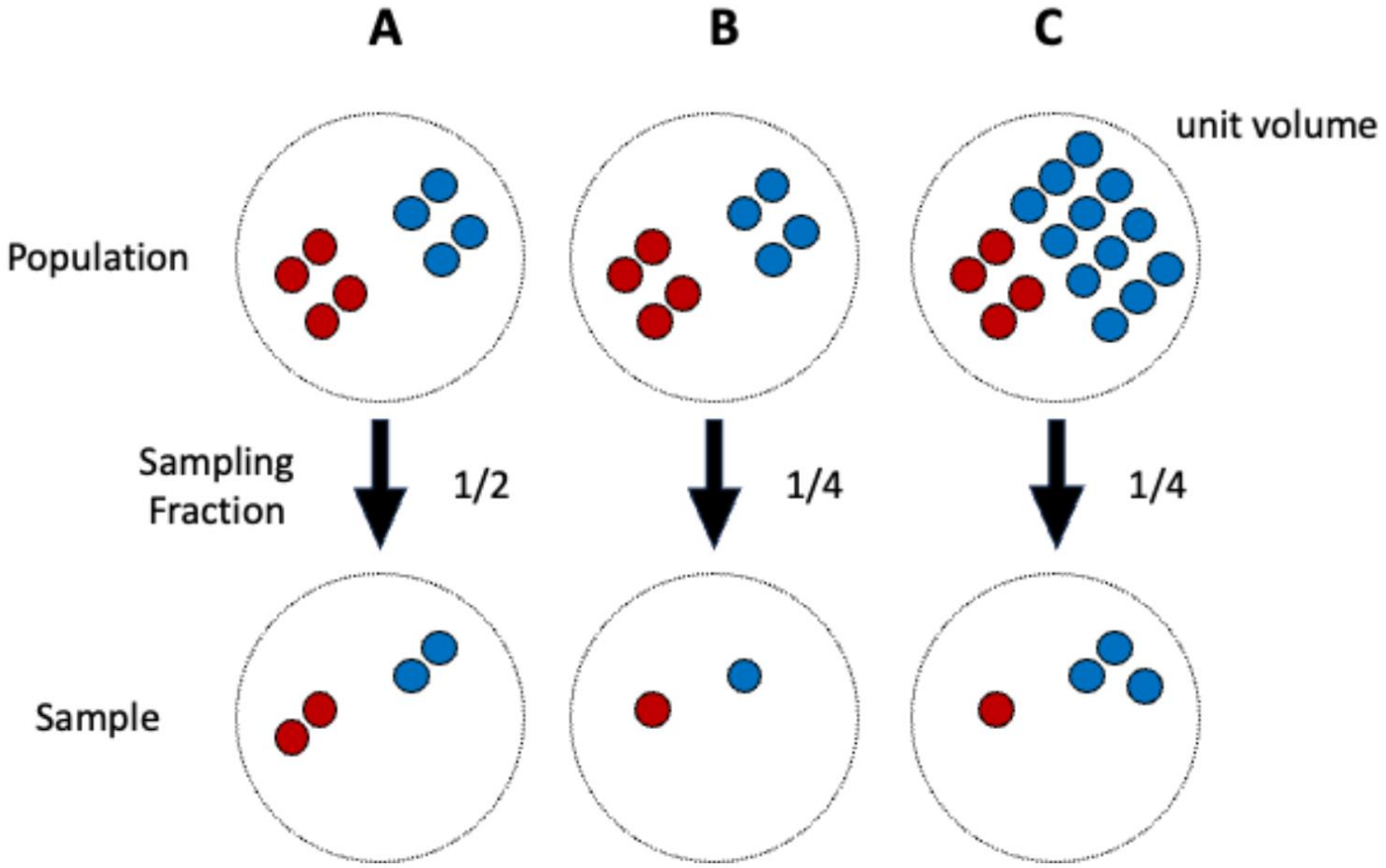
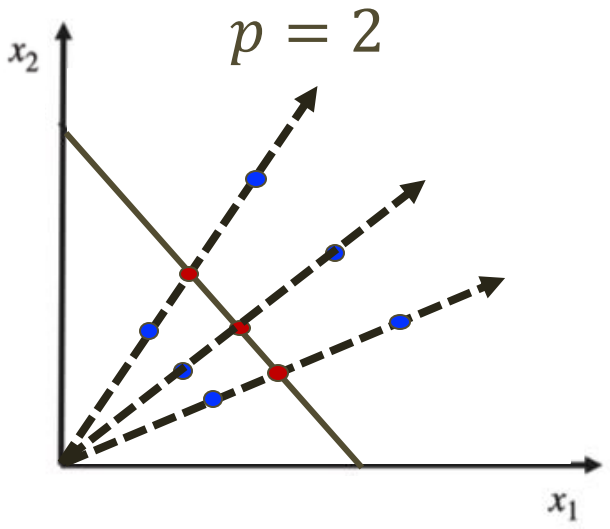


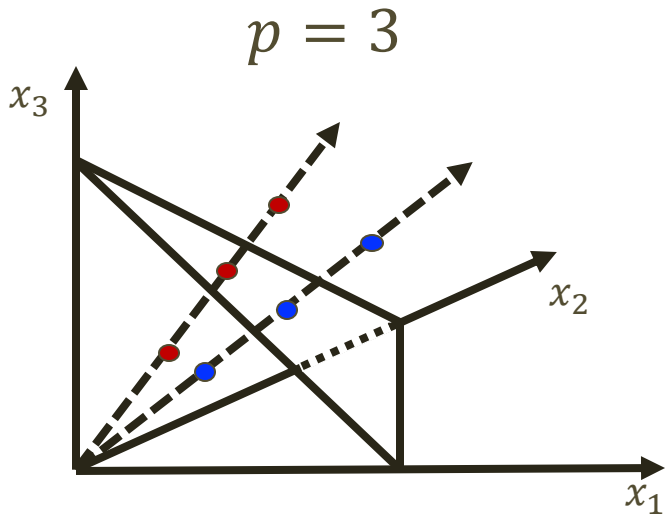
Figure: A vs B: different library size; A vs C: different sampling fraction

Microbiome compositional data

- Conversion of count data to compositional data
  - Closure operator: $cls(x_1, \dots, x_p) = \left[\frac{x_1}{\sum x_j}, \dots, \frac{x_p}{\sum x_j}\right]$
  - Compositional space

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

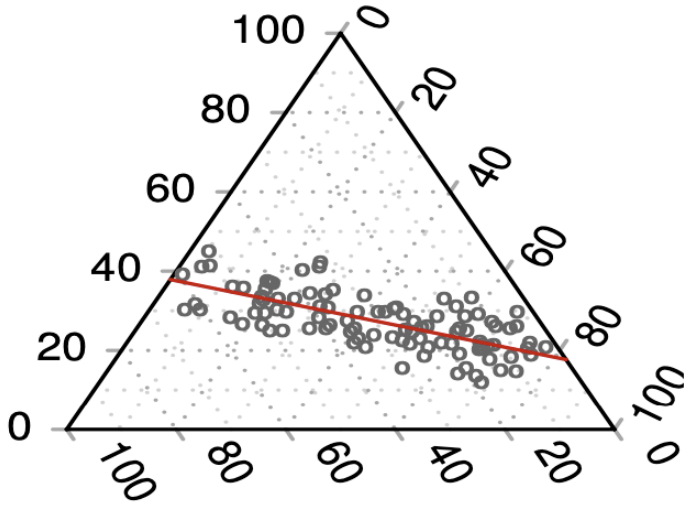




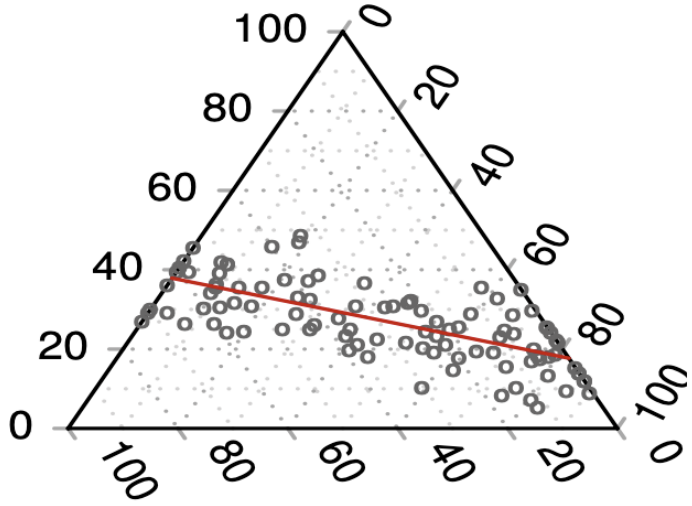
Microbiome compositional data

- Simulated data example when $p = 3$

High signal-to-noise ratio



Low signal-to-noise ratio

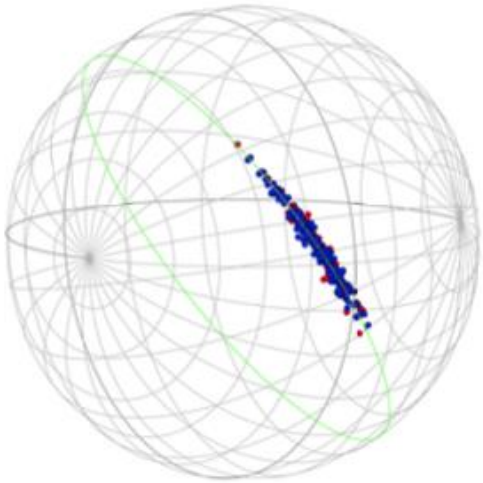


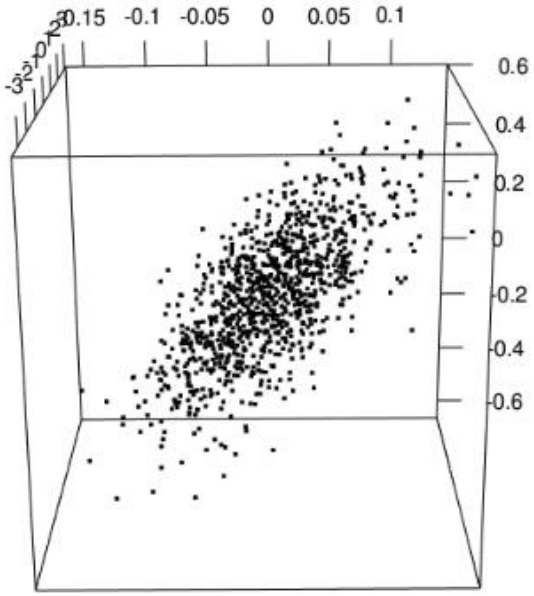
Compositional data is on a vector space?

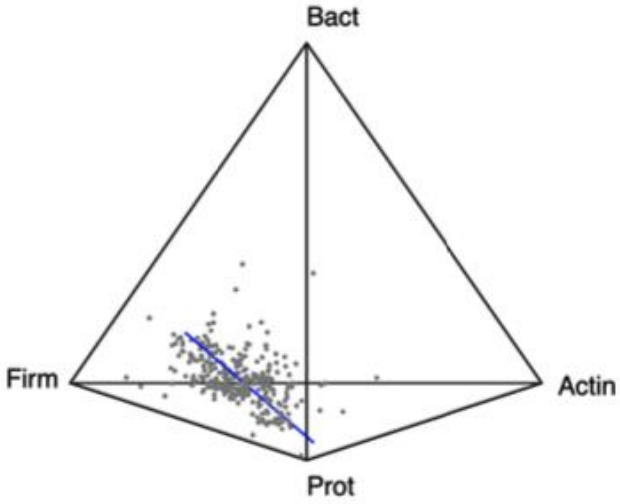
- For $\mathbf{x}, \boldsymbol{y} \in \mathrm{C}^{p-1}$
  - $\boldsymbol{x} + \boldsymbol{y} \notin \mathrm{C}^{p-1}$
  - $c \cdot \boldsymbol{x} \notin \mathrm{C}^{p-1}$, for $c \in \mathrm{R} \setminus \{1\}$
  - $\boldsymbol{x} - \boldsymbol{y} \notin \mathrm{C}^{p-1}$

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

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

  - Powering

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

  - Subtraction

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

  - Distance

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

Log-ratio transformations

- Centered Log-Ratio (CLR) transformation:

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

- Isometric Log-Ratio (ILR) transformation

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

  - $\boldsymbol{H}_p$ is the (p-1) x 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.

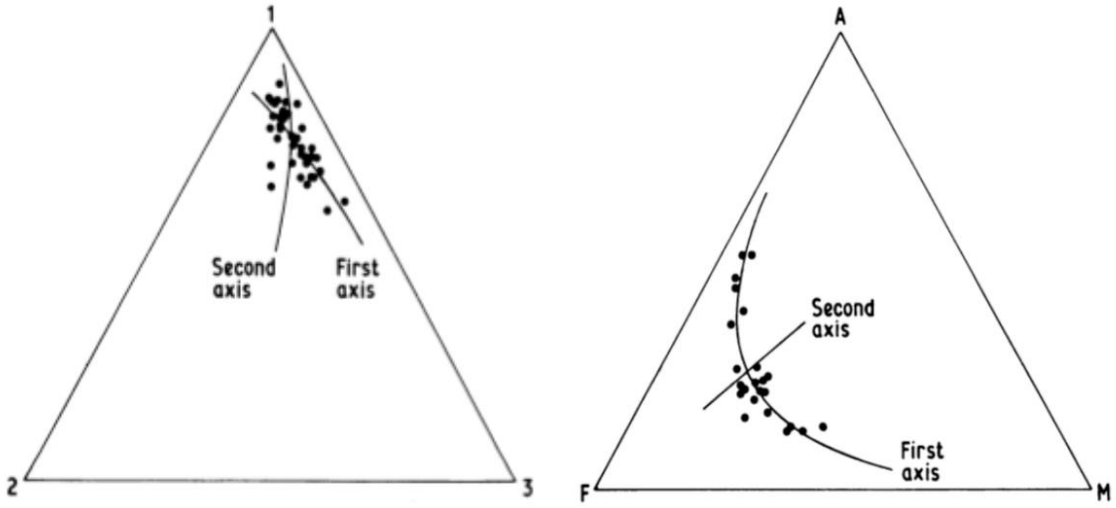


Figure: Ternary diagram with log-ratio principal axes

- Limitation:
  - Log-ratio transformation inherently cannot handle zero values.

Log-ratio PCA with zero replacement

- Zero replacement strategies
  - Simple replacement:

$$(x_1, x_2, 0) \rightarrow cls\big((x_1, x_2, \delta)\big)$$

  - Additive, Multiplicative, and etc.

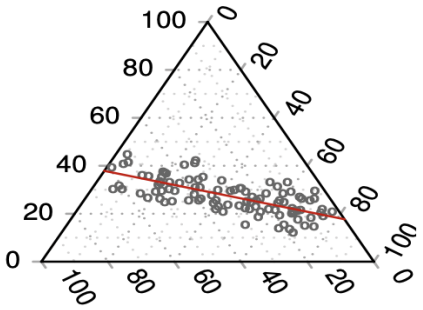
- Determintation of δ
  - $\frac{1}{2}\min\{x_j : x_j > 0\}$

Limitation of log-ratio PCA

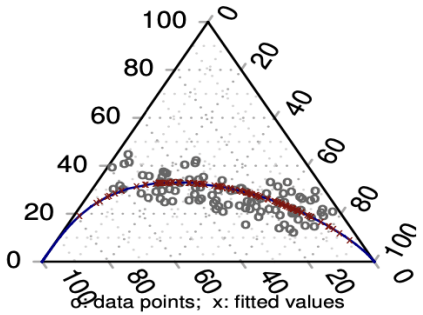
- However, the zero inflation may result in the distortion.

Without zeros

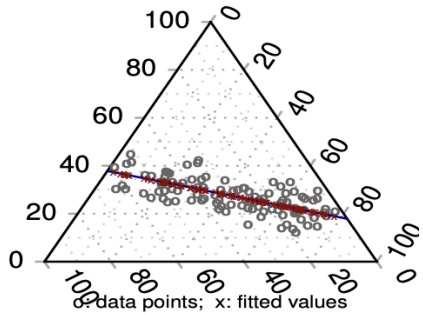
Data generation



lrPCA with $\delta_{(0)}$

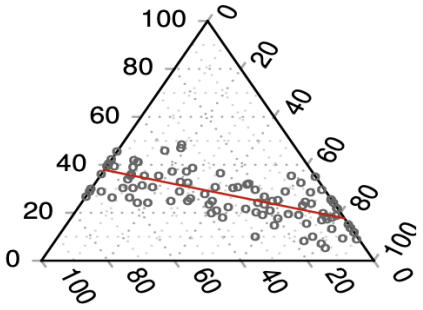


PCA

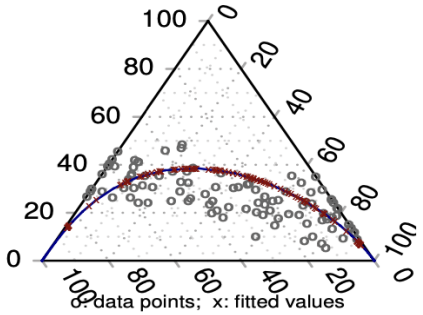


With 10% zeros

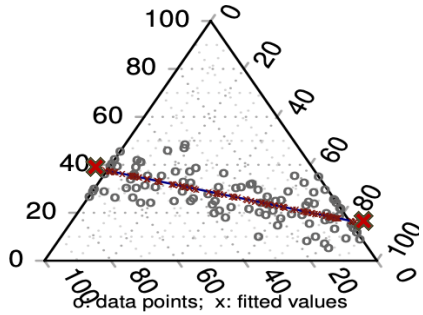
Data generation



lrPCA with $\delta_{(0)}$



PCA



$$\delta_{(0)} = \min\{x_{ij} \in \mathbf{X} : x_{ij} > 0\}$$

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 **A** by $\boldsymbol{a}_i$ and the k-th column of **A** by $\boldsymbol{A}_k$.

- Global compositional PCA (global CPCA) problem:

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

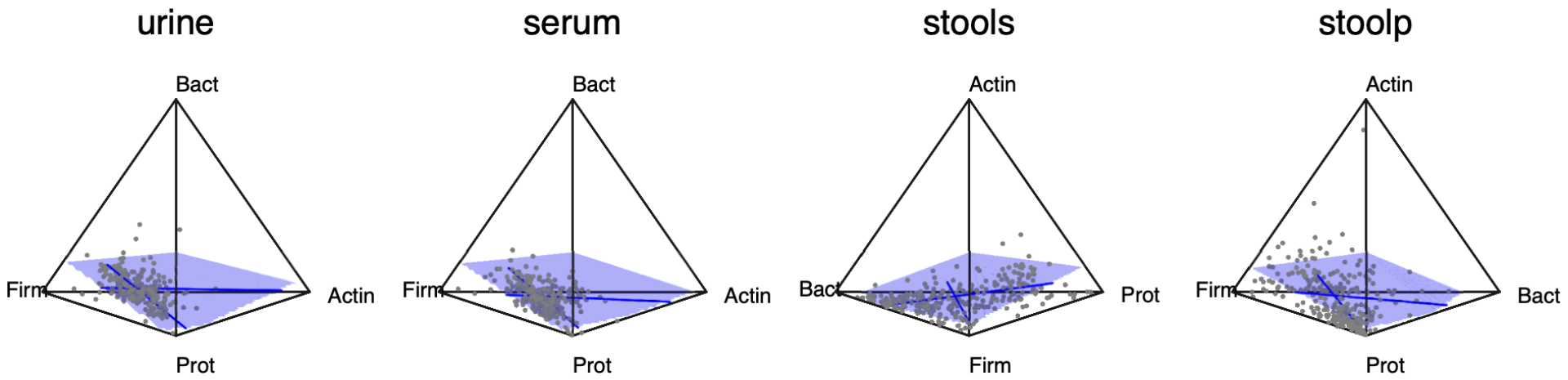
  subject to

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

Compositional PCA

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

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



*Prot: Proteobacteria; Firm: Firmicutes; Actin: Actinobacteria; Bact: Bacteroidetes

Compositional PCA

- Sequential estimation procedure:

  Rank-1 case:

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

  Rank-2 case:

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

$$\vdots$$

  Rank-k case:

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

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

Compositional PCA

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

$$\underset{\mathbf{U}_1,\ldots,\mathbf{U}_k,\mathbf{V}_k}{\arg\min} \|\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T - \mathbf{U}_1\hat{\mathbf{V}}_1^T - \cdots - \mathbf{U}_{k-1}\hat{\mathbf{V}}_{k-1}^T - \mathbf{U}_k\mathbf{V}_k^T\|_F^2, \quad (2)$$

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

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

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

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

Algorithm of CPCA

Algorithm 1: Rank-k approximation for CPCA

Input: $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ and $(\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_{k-1})$.
Initialize $\mathbf{V}_k^{(0)} \perp \mathbf{1}_p$.
Repeat for $t = 0, 1, 2, \ldots$:

1. U-update: obtain $\boldsymbol{u}_i^{(t+1)}$ by (4) with $\boldsymbol{\mu} = \bar{\boldsymbol{x}}$ and $\mathbf{V}_k = \mathbf{V}_k^{(t)}$ $\forall i$.
2. U-shrinkage: $\boldsymbol{u}_i^{(t+1)} \leftarrow (1 - \frac{\gamma}{t+1})\boldsymbol{u}_i^{(t+1)}$.
3. V-update: obtain $\mathbf{V}_k^{(t+1)}$ by (6) with $\boldsymbol{\mu} = \bar{\boldsymbol{x}}$ and $\mathbf{U} = (\mathbf{U}_1^{(t+1)}, \ldots, \mathbf{U}_k^{(t+1)})$
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 $\boldsymbol{u}_i^{(t+1)}$ without the shrinkage $\forall i$.
Output: $(\mathbf{U}_1^{(t+1)}, \ldots, \mathbf{U}_k^{(t+1)})$ and $(\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_{k-1}, \mathbf{V}_k^{(t+1)})$.

Algorithm of aCPCA

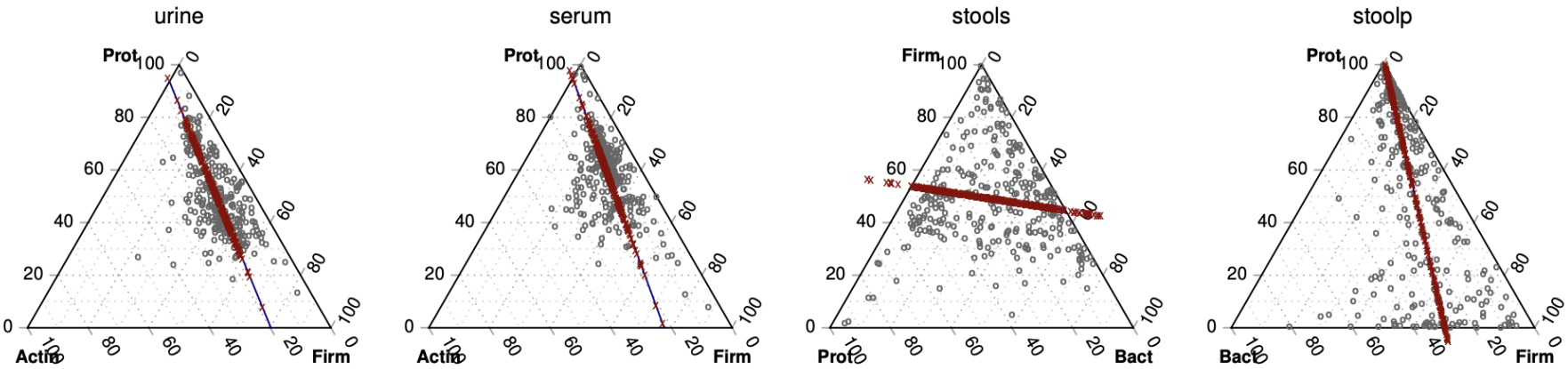
Algorithm 2: Rank-k approximation for aCPCA

Input: $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$, $(\hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_{k-1})$ and $(\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_{k-1})$.
Initialize $\mathbf{V}_k^{(0)}$.
Repeat for $t = 0, 1, 2, \ldots$:

1. U-update: obtain $u_{ik}^{(t+1)}$ by (5) with $\boldsymbol{c}_i = \bar{\boldsymbol{x}} + \sum_{h=1}^{k-1} \hat{u}_{ih} \hat{\mathbf{V}}_h$ 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 $\boldsymbol{\mu} = \bar{\boldsymbol{x}}$ and $\mathbf{U} = (\hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_{k-1}, \mathbf{U}_k^{(t+1)})$.
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: $(\hat{\mathbf{U}}_1, \ldots, \hat{\mathbf{U}}_{k-1}, \mathbf{U}_k^{(t+1)})$ and $(\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_{k-1}, \mathbf{V}_k^{(t+1)})$.

Real data example with $p = 3$



*Prot: Proteobacteria; Firm: Firmicutes; Actin: Actinobacteria; Bact: Bacteroidetes

Real data example: rank-1 reconstruction

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

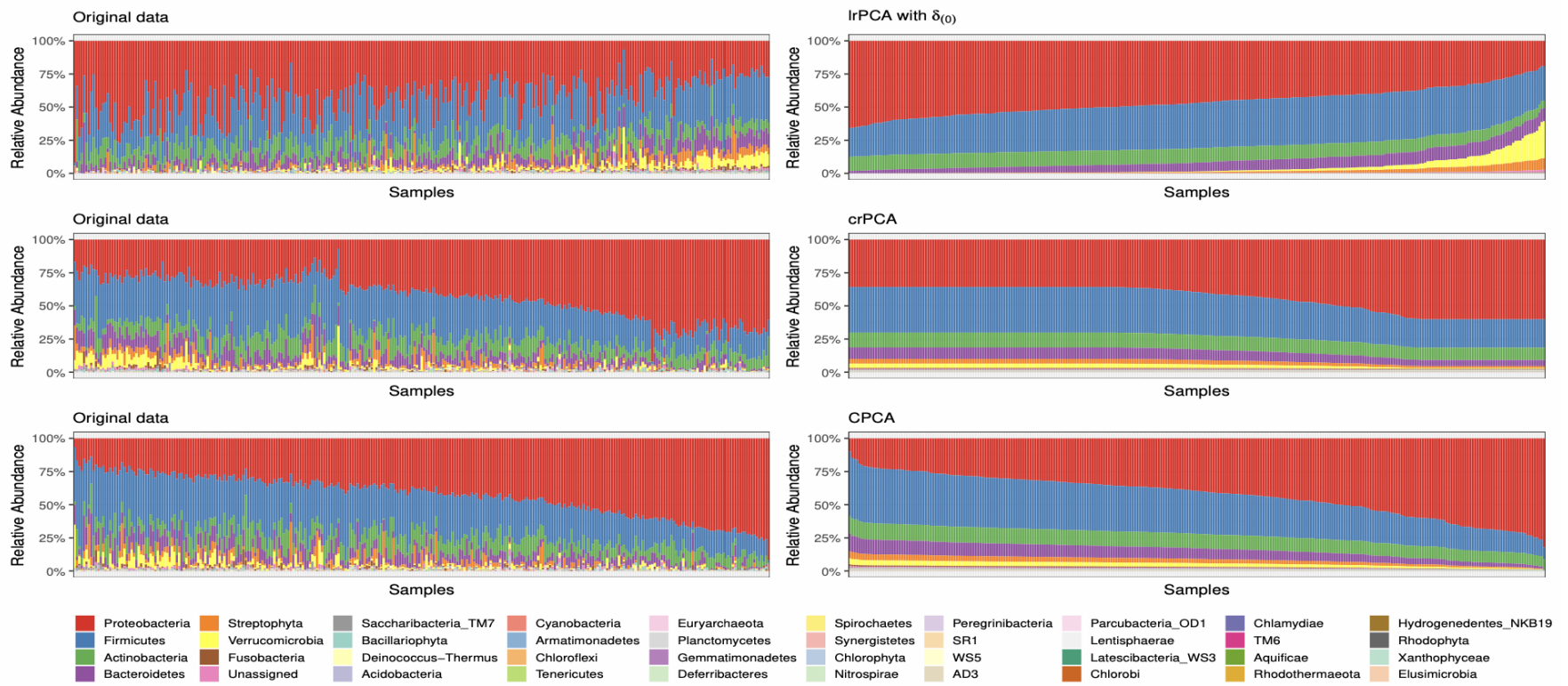


Figure: Left: the original data. Right: reconstructed data. The same sample orders were maintained between left and right panels for each method, based on its estimated first score.

Application area

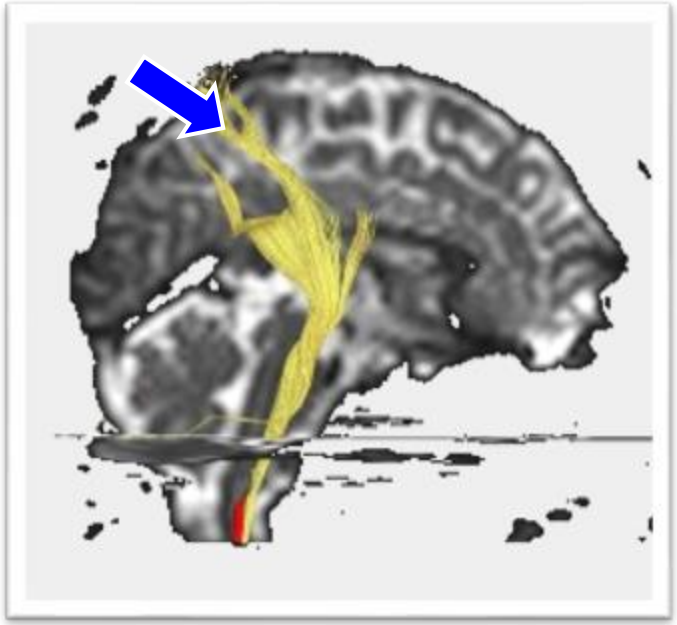
- Electrodiagnostic text data (a vector of word frequencies)

A 12-year old girl with known hyperagglutinability, presented to the emergency department with a 2-week history of headeaches 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.

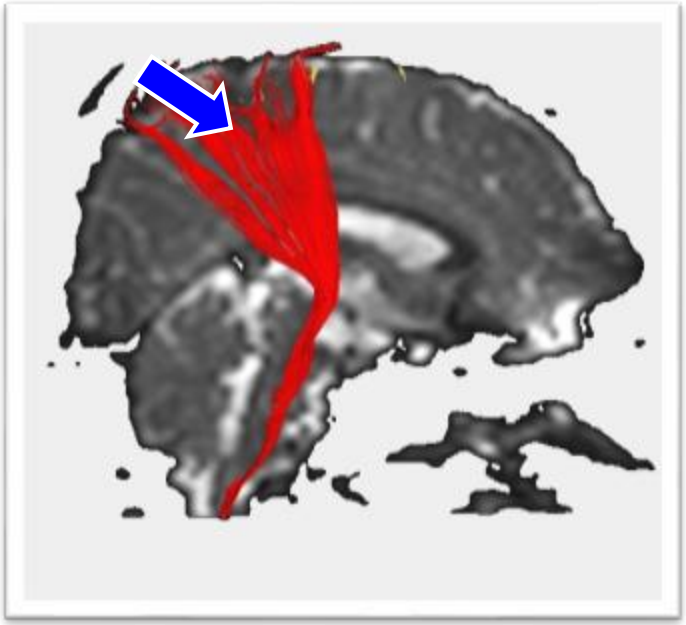
GLM for spherical response

- Motivating example
  - Two-sample testing problem in DTI data

Patient group

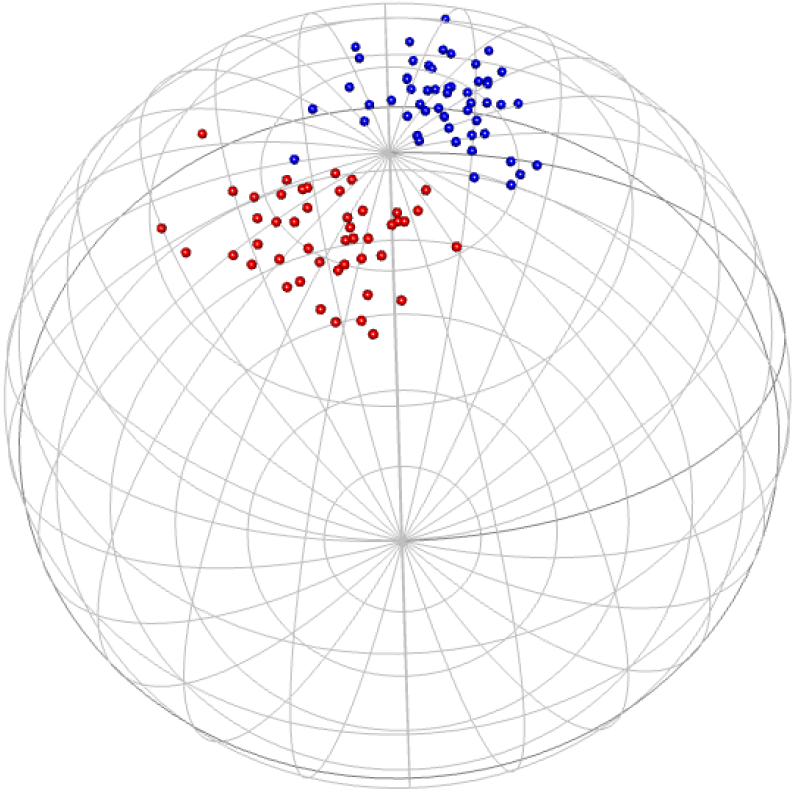


Normal group



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

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

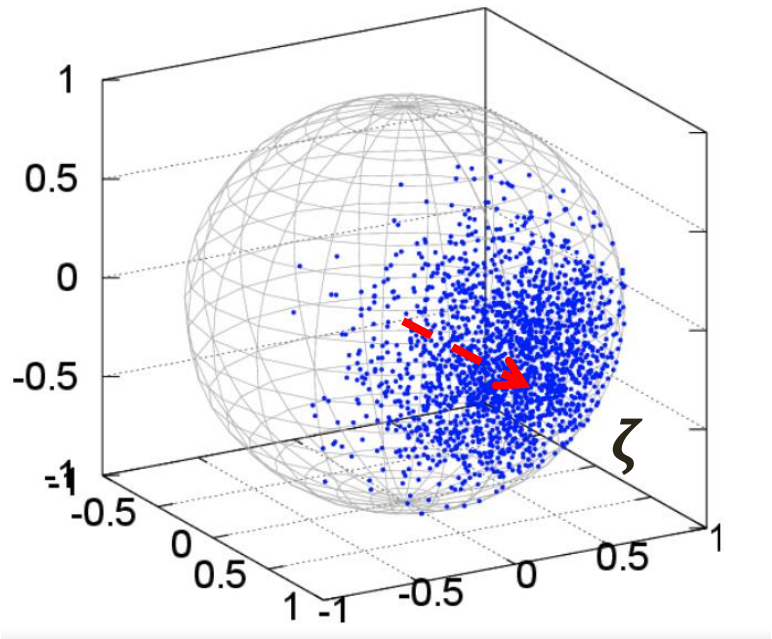
$$f_{vMF}(\boldsymbol{y}; \boldsymbol{\zeta}, \kappa) = \mathrm{C}_q(\kappa) \cdot \exp(\kappa \cdot \boldsymbol{\zeta}^T \boldsymbol{y}),$$

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

  - $\boldsymbol{\zeta}$: mean direction
  - κ : concentration parameter

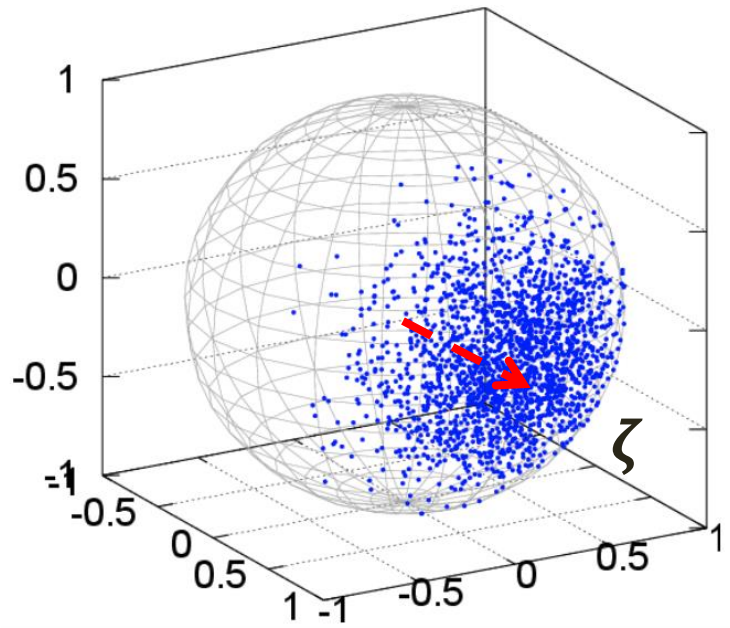
Spherical data and its distribution

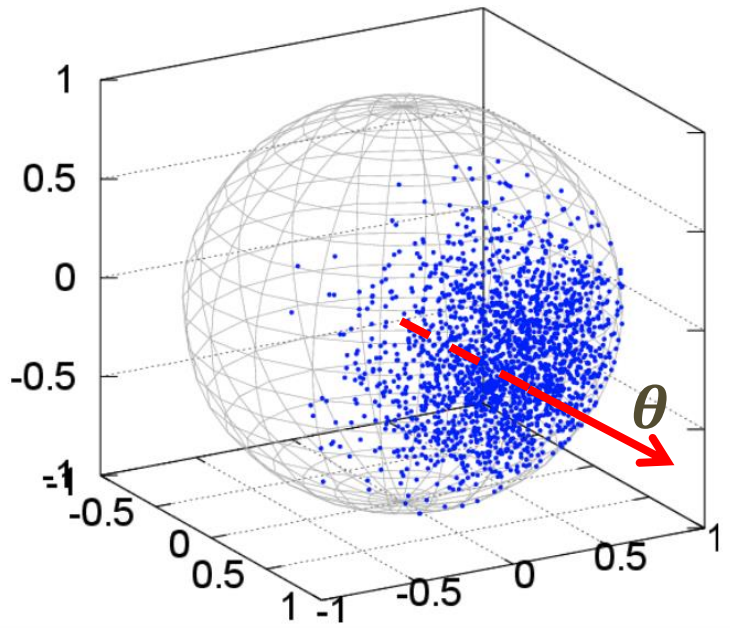
- $q = 3,\ \boldsymbol{\zeta} = (1,0,0)^T,\ \kappa = 10$



Spherical data and its distribution

- $q = 3,\ \boldsymbol{\zeta} = (1,0,0)^T,\ \kappa = 10$
- Reparametrization: $\boldsymbol{\theta} = \kappa \cdot \boldsymbol{\zeta}$
- $\mathbf{q = 3},\ \boldsymbol{\theta} = (100, 0, 0)^T$





Two-sample testing problem in spherical data

- Then, we can rewrite the pdf of vMF as

$$f_{vMF}(\boldsymbol{y}; \boldsymbol{\theta}) = \mathrm{C}_q(||\boldsymbol{\theta}||_2) \cdot \exp\left(\boldsymbol{\theta}^{\boldsymbol{T}} \boldsymbol{y}\right).$$

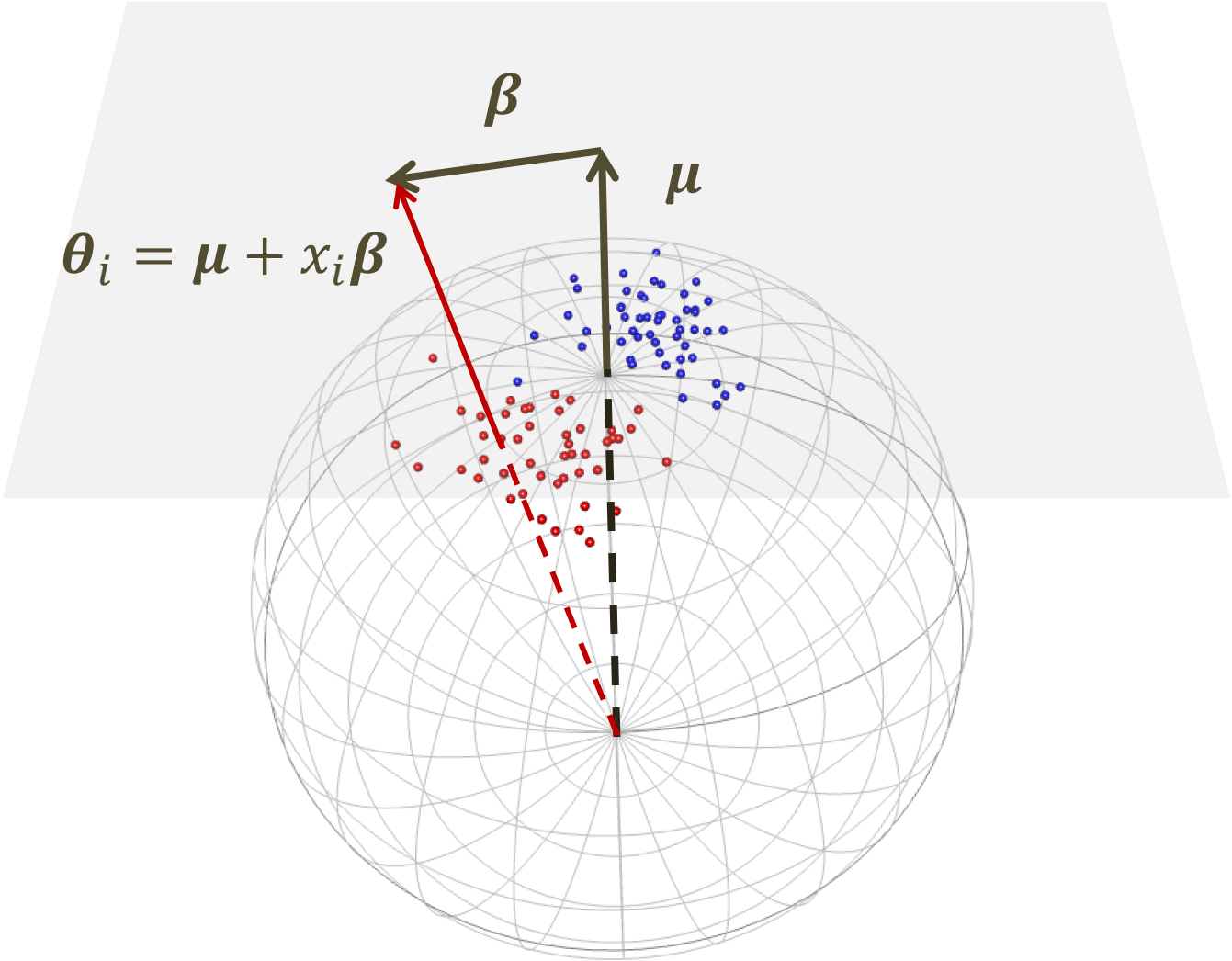
- Generalized linear model framework

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

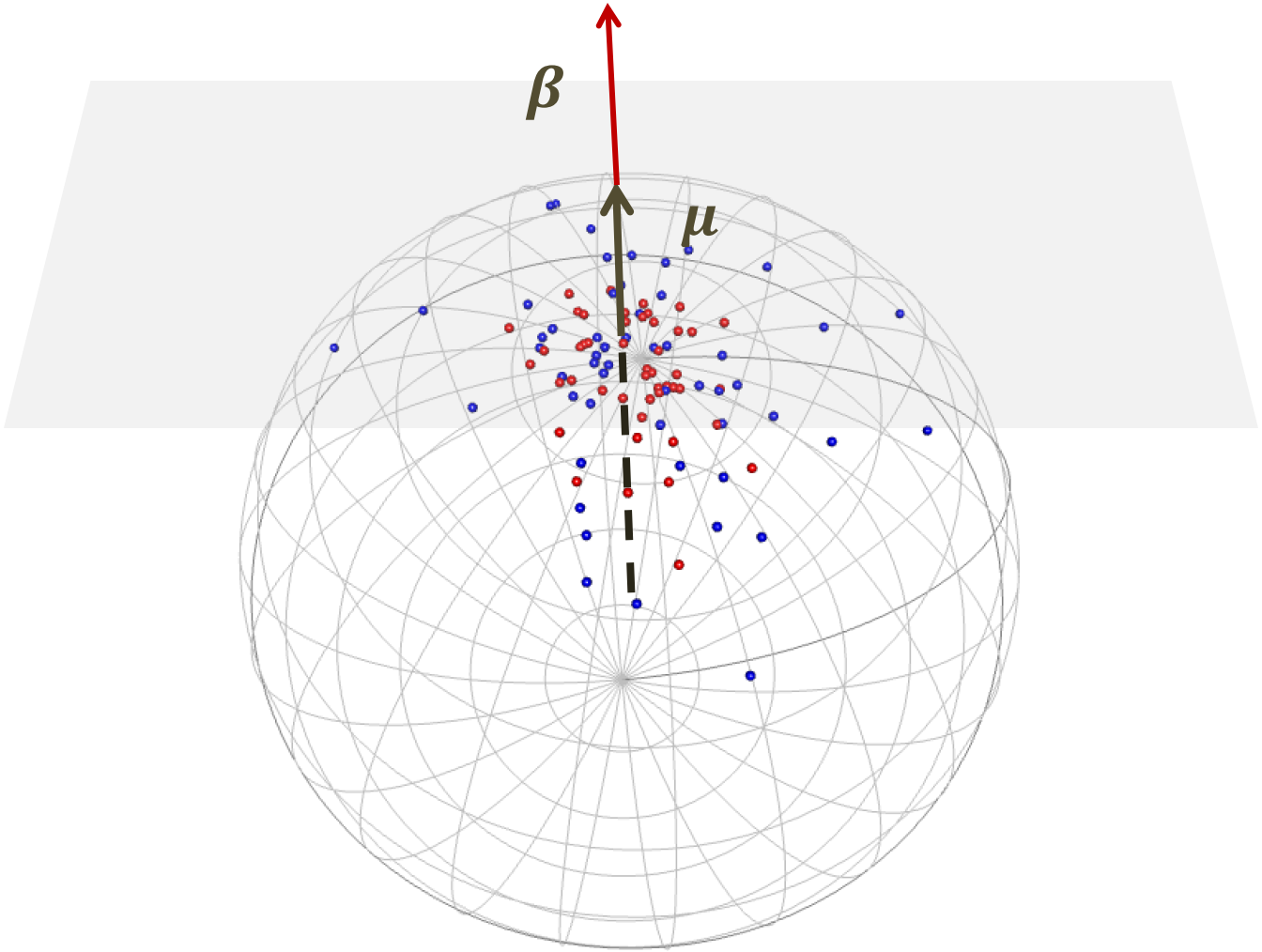
$$\boldsymbol{y}_i |\, x_i \sim vMF(\cdot\,|\boldsymbol{\theta}_i),$$

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 $\tilde{\boldsymbol{x}}_i := \left(1, \boldsymbol{x}_i^T\right)^T \otimes \boldsymbol{I}_q \in \mathrm{R}^{(p+1)q \times q}$ and $\boldsymbol{\beta}^* := \left(\boldsymbol{\mu}^T, \boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_p^T\right)^T \in R^{pq \times 1}$

- Estimation on the unconstrained model

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

- Estimation on the constrained model

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

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

$$\hat{\boldsymbol{\beta}}^* \xrightarrow{p} \boldsymbol{\beta}_0^*.$$

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

$$\boldsymbol{F}_n^{T/2}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}_{pq}).$$

- $\boldsymbol{F}_n$ is the Fisher information matrix at true $\boldsymbol{\beta}_0^*$ and $\boldsymbol{F}^{T/2}$ indicates the transpose of the Cholesky square root matrix for $\boldsymbol{F}$

Asymptotic analysis following projection

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

- Then we have

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

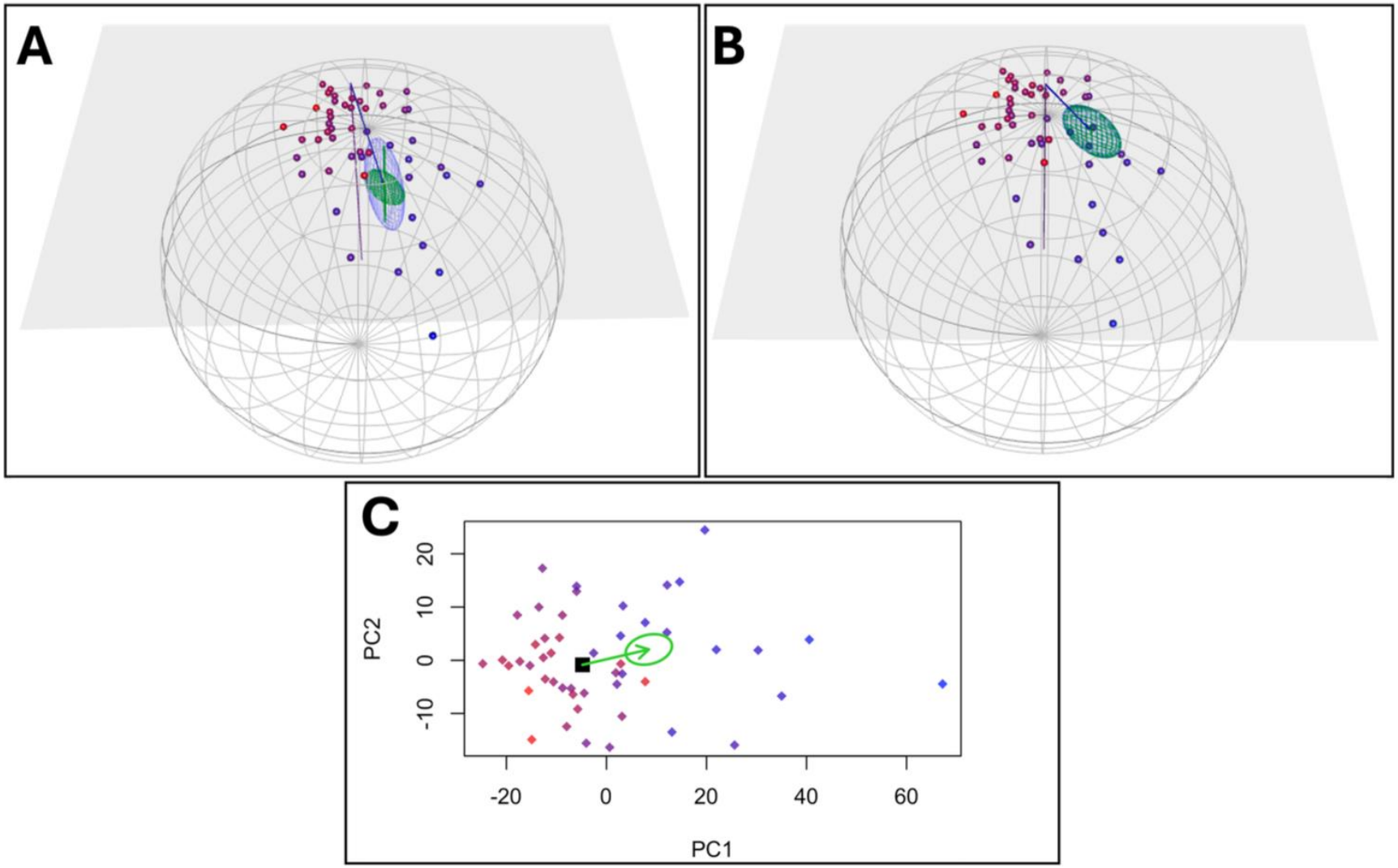
  - $\Sigma_{\mu^\perp} = P_{\mu^\perp} c_j^T [\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}] c_j P_{\mu^\perp}^T$ and c_j is the index matrix representing the j-th covariate

- Similarly, for the μ-directional projection, we have the following result

$$\boldsymbol{P}_{\hat{\boldsymbol{\mu}}}\hat{\boldsymbol{\beta}}_j|\hat{\boldsymbol{\mu}} \xrightarrow{d} N\left(\boldsymbol{P}_{\boldsymbol{\mu}}\boldsymbol{\beta}_j,\ \boldsymbol{\Sigma}_{\boldsymbol{\mu}}\right)$$

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

Orthogonal constraint: without vs. with



Thank you for your attention !