Methods for High-dimensional Inference, with Applications to Imaging Genetics

Dustin Pluta

28 Mar 2018

Acknowledgements

- **Gui Xue**, PI and data provider, Beijing Normal University, Center for Brain and Learning Sciences
- **Chuansheng Chen**, UCI, Dept. of Psychology and Social Behavior
- **Hernando Ombao**, KAUST, Dept. of Statistics
- **Zhaoxia Yu**, UCI, Dept. of Statistics
- **Tong Shen**, UCI, Dept. of Statistics (PhD Student)

Overview of Talk

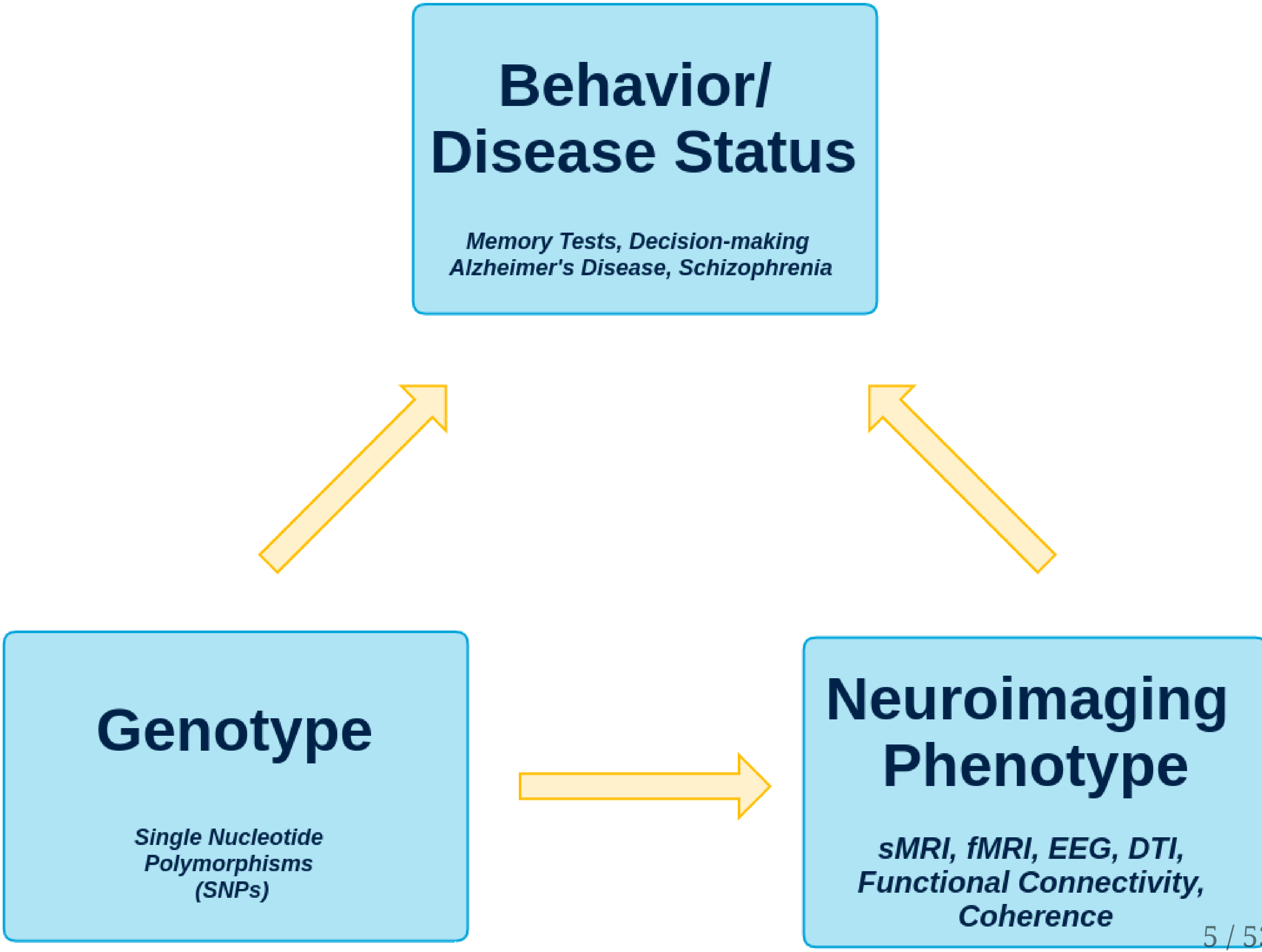
1. **Scientific motivation** from imaging genetics.
2. **Kernel Mantel test** and metric-based association testing.
3. The **adaptive Mantel test** for penalized inference.
4. Application to test heritability of EEG coherence during a working memory task.

Scientific Question

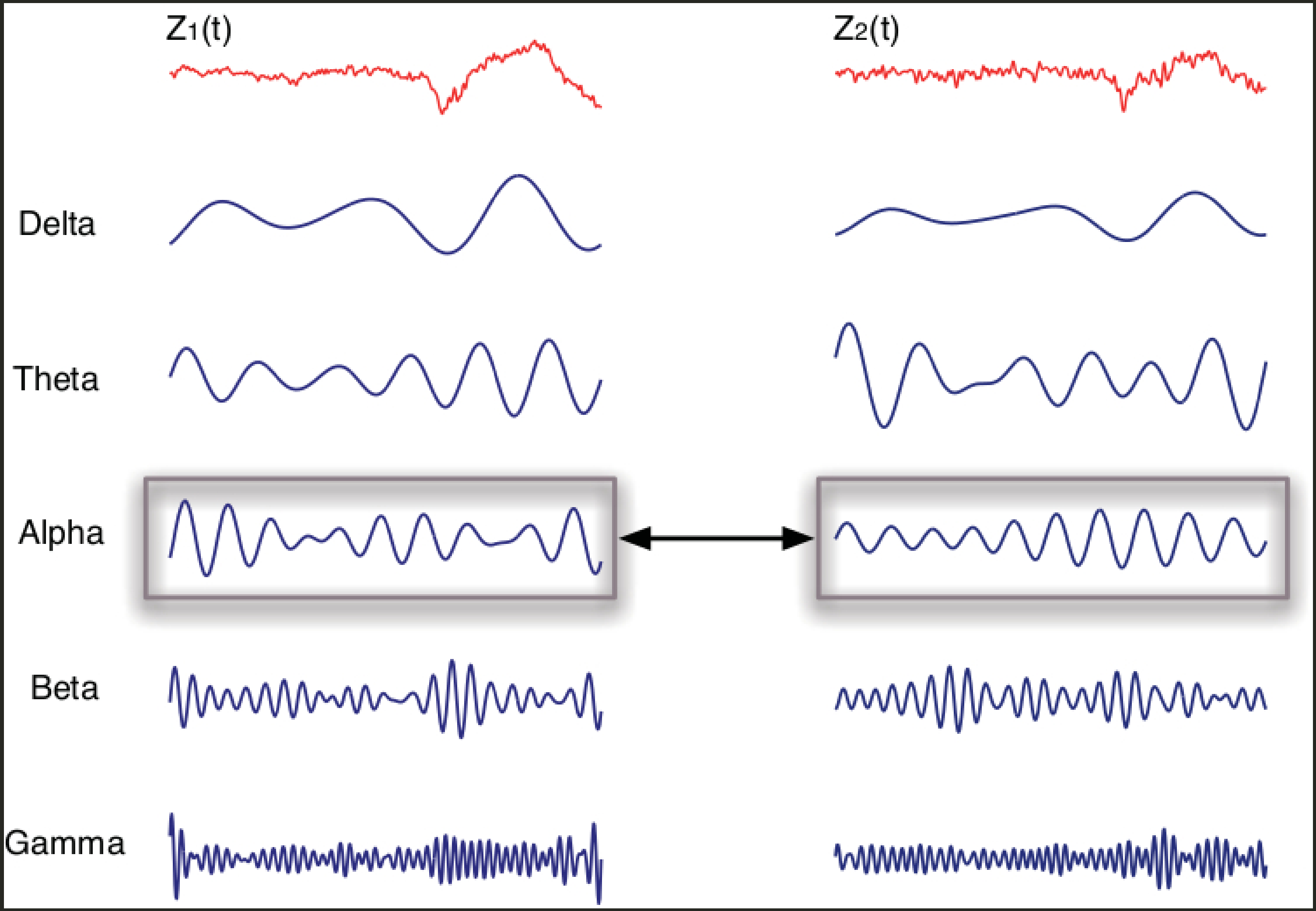
What is the heritability of brain connectivity features (measured from fMRI or EEG)?

Motivating Data

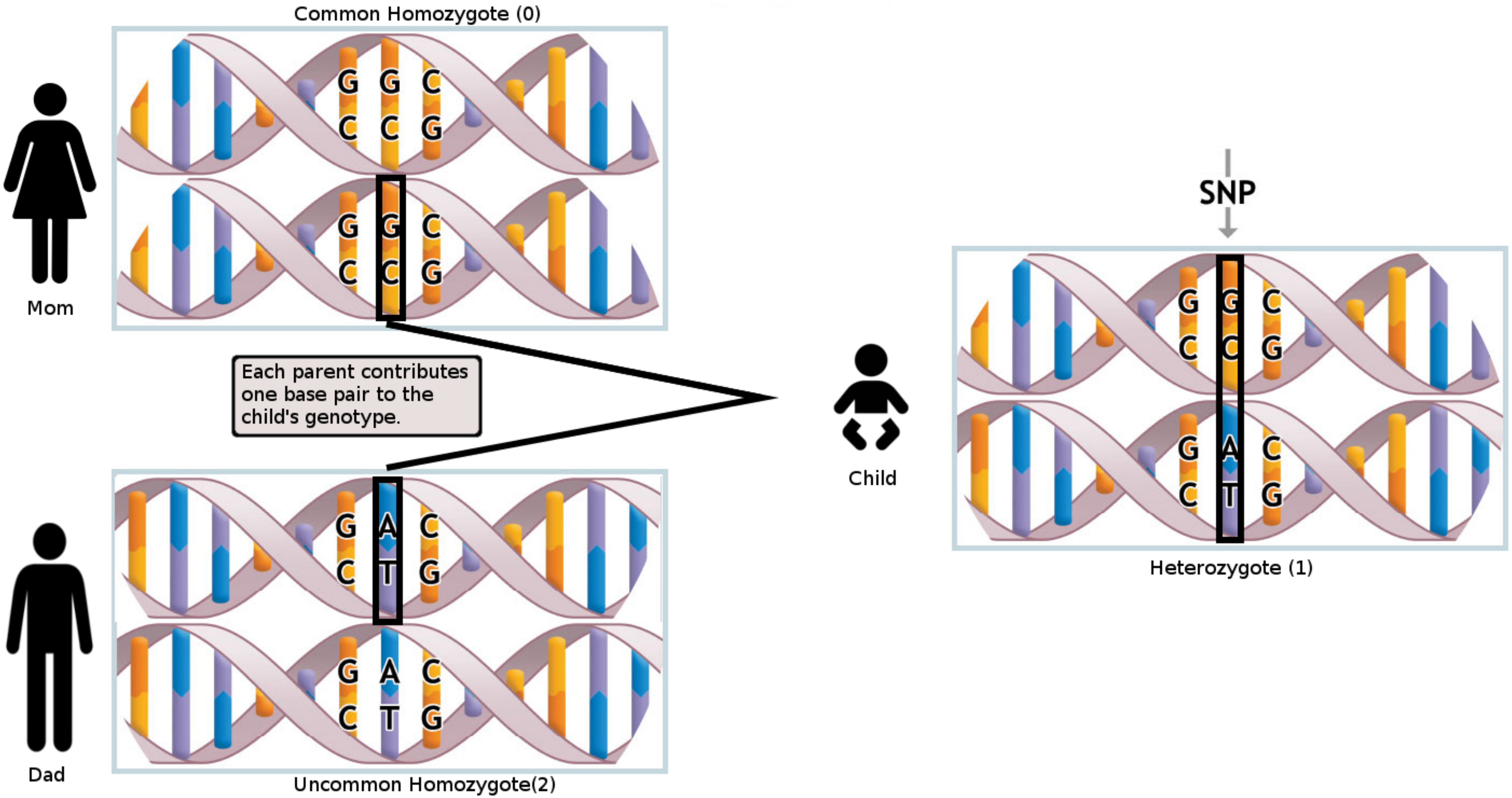
- **350 students** from the Beijing Normal University
- ~10 minute 64 channel **EEG** recording during **Visual Working Memory task**
- **13 SNPs** selected for analysis, previously identified as potential factors for **Alzheimer's disease risk**.


Behavior/
Disease Status
Memory Tests, Decision-making
Alzheimer's Disease, Schizophrenia
Genotype
Single Nucleotide
Polymorphisms
(SNPs)
Neuroimaging
Phenotype
sMRI, fMRI, EEG, DTI,
Functional Connectivity,
Coherence

Functional Connectivity: Coherence



SNPs


Common Homozygote (0)
G G C
C C G
G G C
C C G
Mom
Each parent contributes one base pair to the child's genotype.
G A C
C T G
G A C
C T G
Dad
Uncommon Homozygote(2)
Child
SNP
G G C
C C G
G A C
C T G
Heterozygote (1)

Scientific Motivation

Estimating Heritability with Variance Components Model

- Let X be an $n \times p$ matrix of single nucleotide polymorphism (SNP) data, and Y be an observed scalar phenotype.

$$Y = g + \varepsilon,$$

where $g \sim N(0, \sigma_g^2 G)$, for $G = XX^T/p$, and $\varepsilon \sim N(0, \sigma_\varepsilon^2 I_n)$.

Narrow-sense heritability of the phenotype measured by Y can then be estimated as

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_\varepsilon^2}.$$

Mantel Test and Distance-based Association Testing

Association Testing Methods

- **Mantel's test** (Mantel 1967) uses the inner product of the pairwise distance/similarity matrices from X and Y.
- The **RV coefficient** (Escoufier 1976) uses a test statistic based on the multivariate correlation between X and Y.
- The **distance covariance** (dCov) test (Szekely, Rizzo, Bakirov, 2007) is defined as the covariance of distances between X and Y.
- **Adaptive sum of powered score test** (Xu et. al 2017).

Kernel Mantel Test

- Given **similarity functions** $\mathcal{K}_X : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ and $\mathcal{K}_Y : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, we can form two $n \times n$ **Gram matrices** K and H, where

$$K_{ij} = \mathcal{K}_X(X_i, X_j)$$

$$H_{ij} = \mathcal{K}_Y(Y_i, Y_j).$$

- The **correlation** of these matrices is

$$r(H, K) := \frac{\langle K, H \rangle}{\|K\| \cdot \|H\|},$$

Kernel Mantel Test

How should we test the significance of the correlation?

Mantel's original approach (1967) is to **permute** rows and columns of one of the Gram matrices to generate the reference distribution.

That is, for test statistic

$$T = \langle K, H \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} K_{ij} H_{ij} = \mathrm{tr}(KH),$$

we compute the **permutation P-value** by permuting H to approximate the reference distribution.

Kernel Mantel Test

Similarity with Weighted Inner Products

For two vectors $u, v \in \mathbb{R}^p$, the **weighted inner product** $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ for some positive semi-definite matrix $\mathcal{W}$, is defined as

$$\langle u, v \rangle_{\mathcal{W}} = u^T \mathcal{W} v.$$

The **Mantel Test Statistic** for similarity $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ is

$$T_{\mathcal{W}} = \mathrm{tr}(X \mathcal{W} X^T Y Y^T) = Y^T X \mathcal{W} X^T Y.$$

Kernel Mantel Test

Similarity with Weighted Inner Products

For two vectors $u, v \in \mathbb{R}^p$, the **weighted inner product** $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ for some positive semi-definite matrix $\mathcal{W}$, is defined as

$$\langle u, v \rangle_{\mathcal{W}} = u^T \mathcal{W} v.$$

The **Mantel Test Statistic** for similarity $\langle \cdot, \cdot \rangle_{\mathcal{W}}$ is

$$T_{\mathcal{W}} = \mathrm{tr}(X \mathcal{W} X^T Y Y^T) = Y^T X \mathcal{W} X^T Y.$$

Questions

How does the choice of weight matrix affect the test characteristics?

How should the weight matrix be chosen?

Similarity Measures

Euclidean Inner Product

- Choosing $\mathcal{W} = I_p$ gives

$$K_E = XX^T,$$

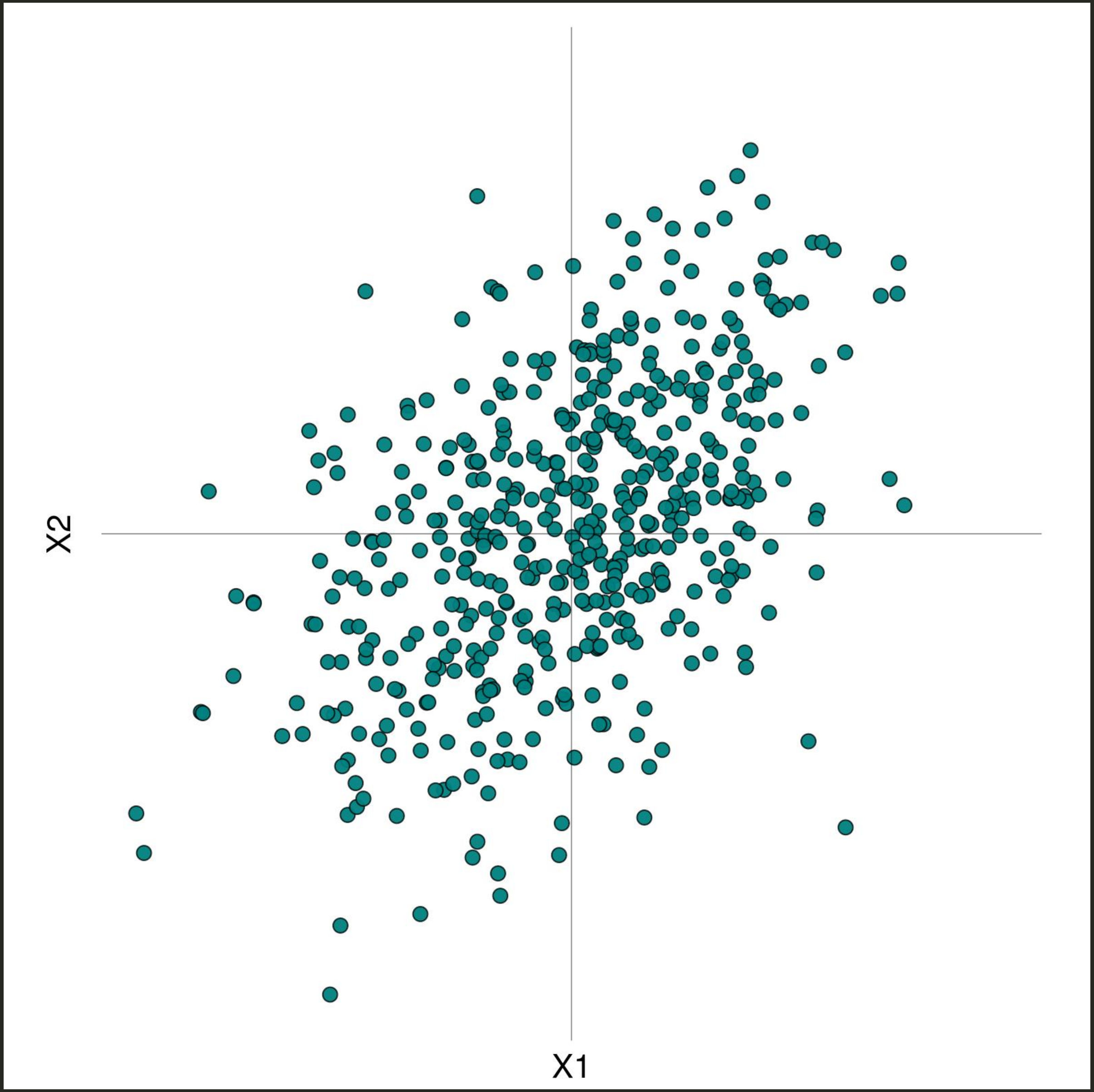
which is the Gram matrix for the standard Euclidean inner product.

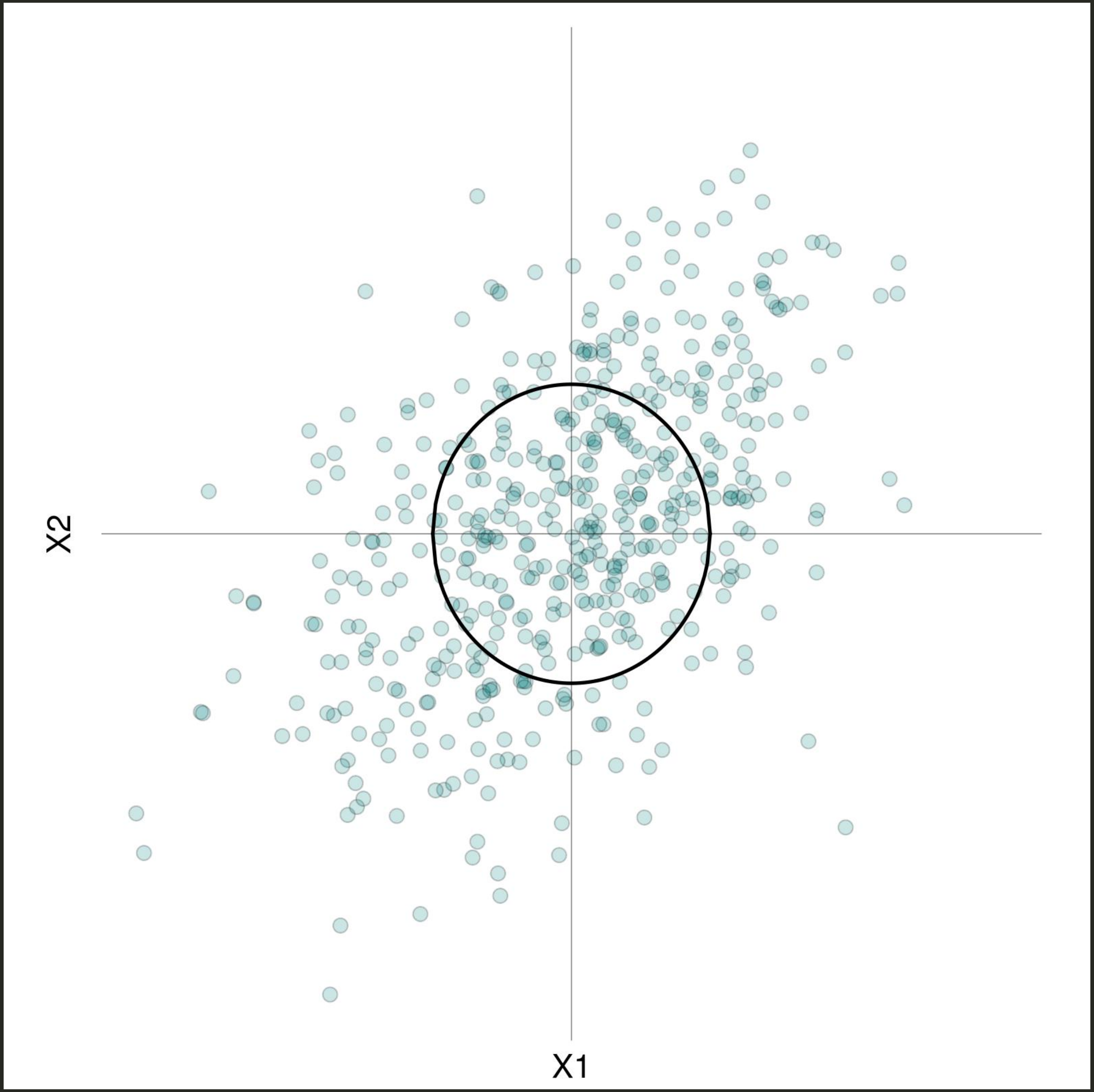
Mahalanobis Inner Product

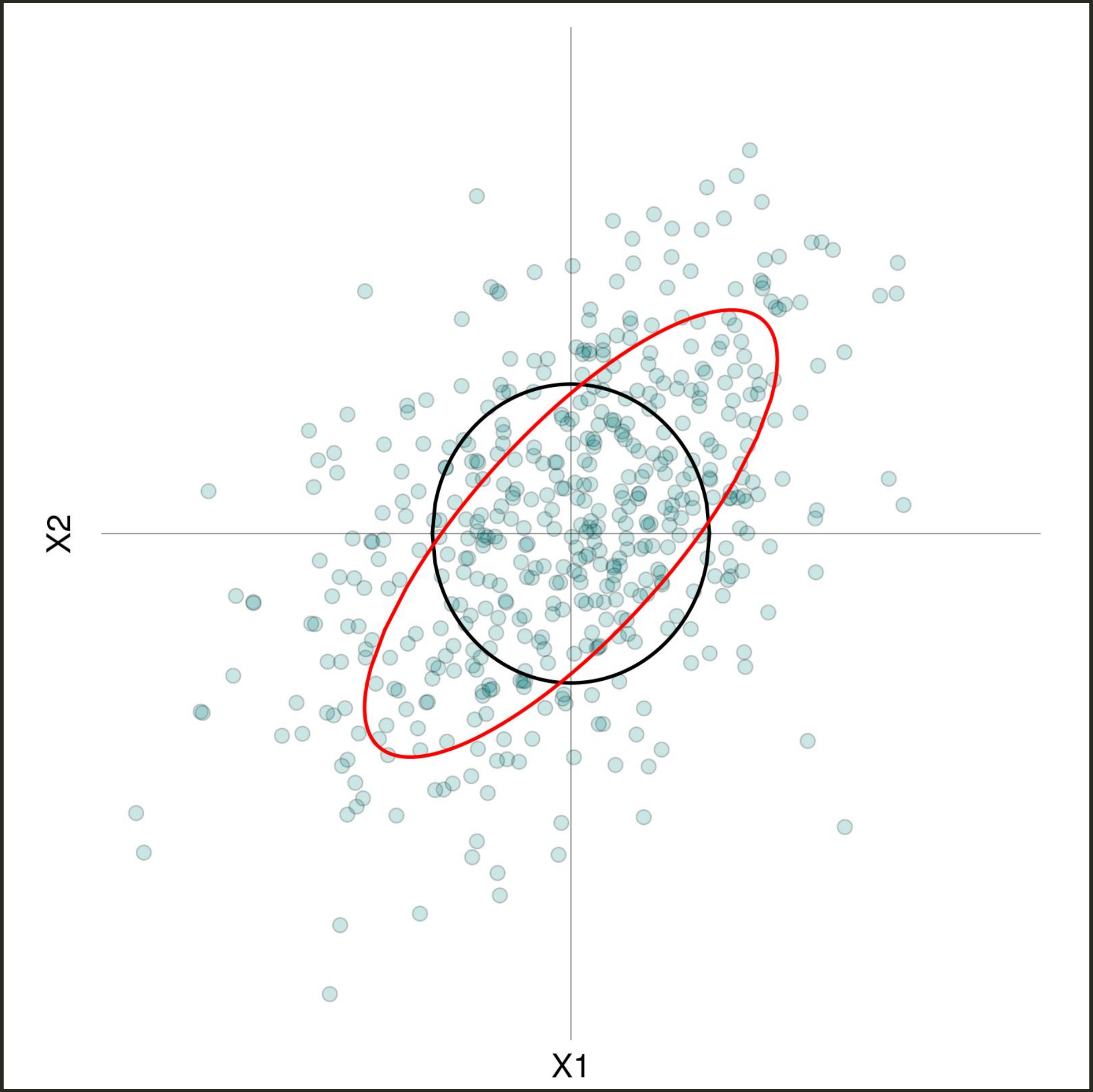
- Choosing $\mathcal{W} = (X^T X)^{-1}$ gives

$$K_M = X(X^T X)^{-1} X^T,$$

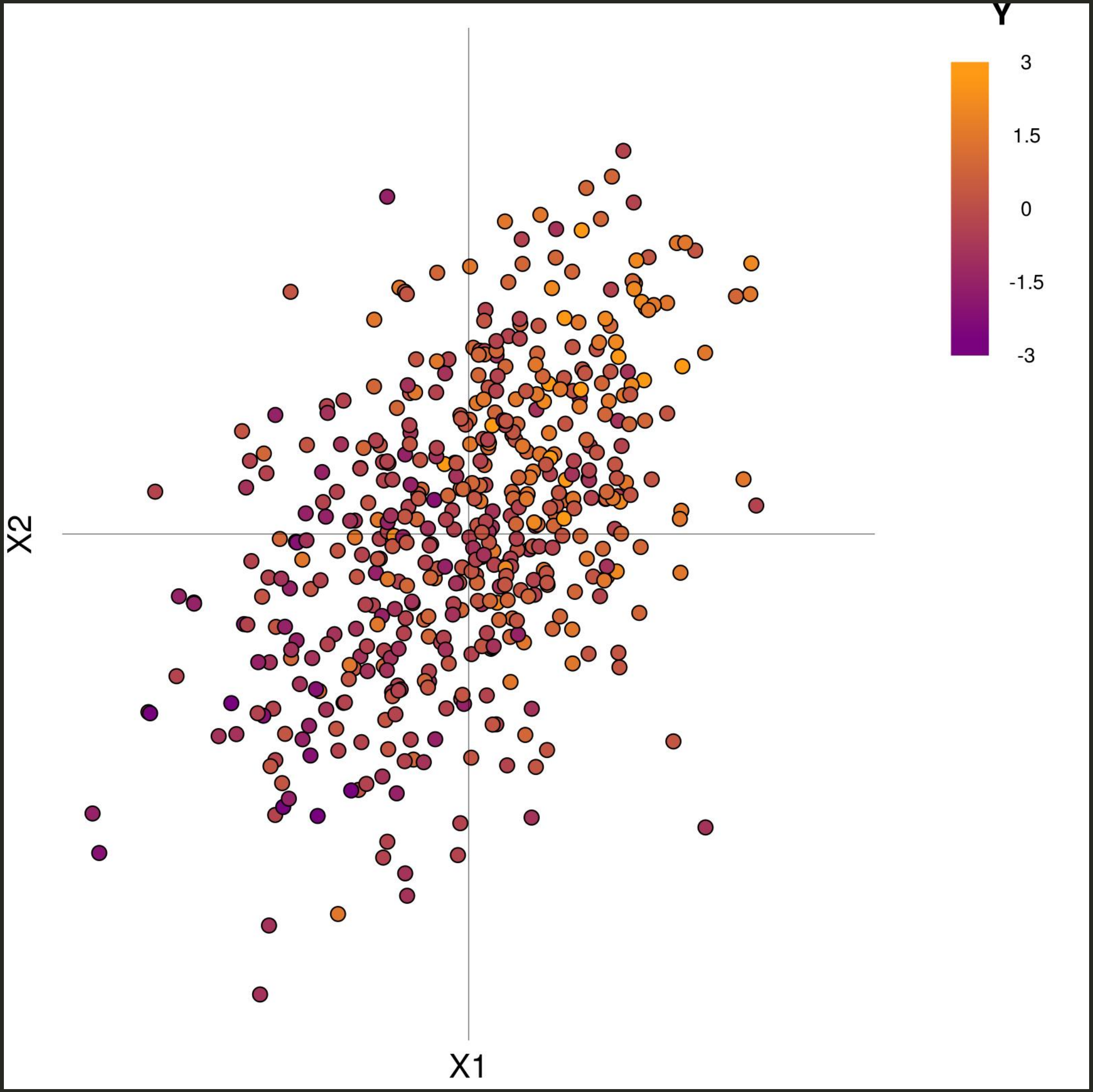
which is a similarity matrix related to the Mahalanobis distance, and is the projection matrix for the column space of X.

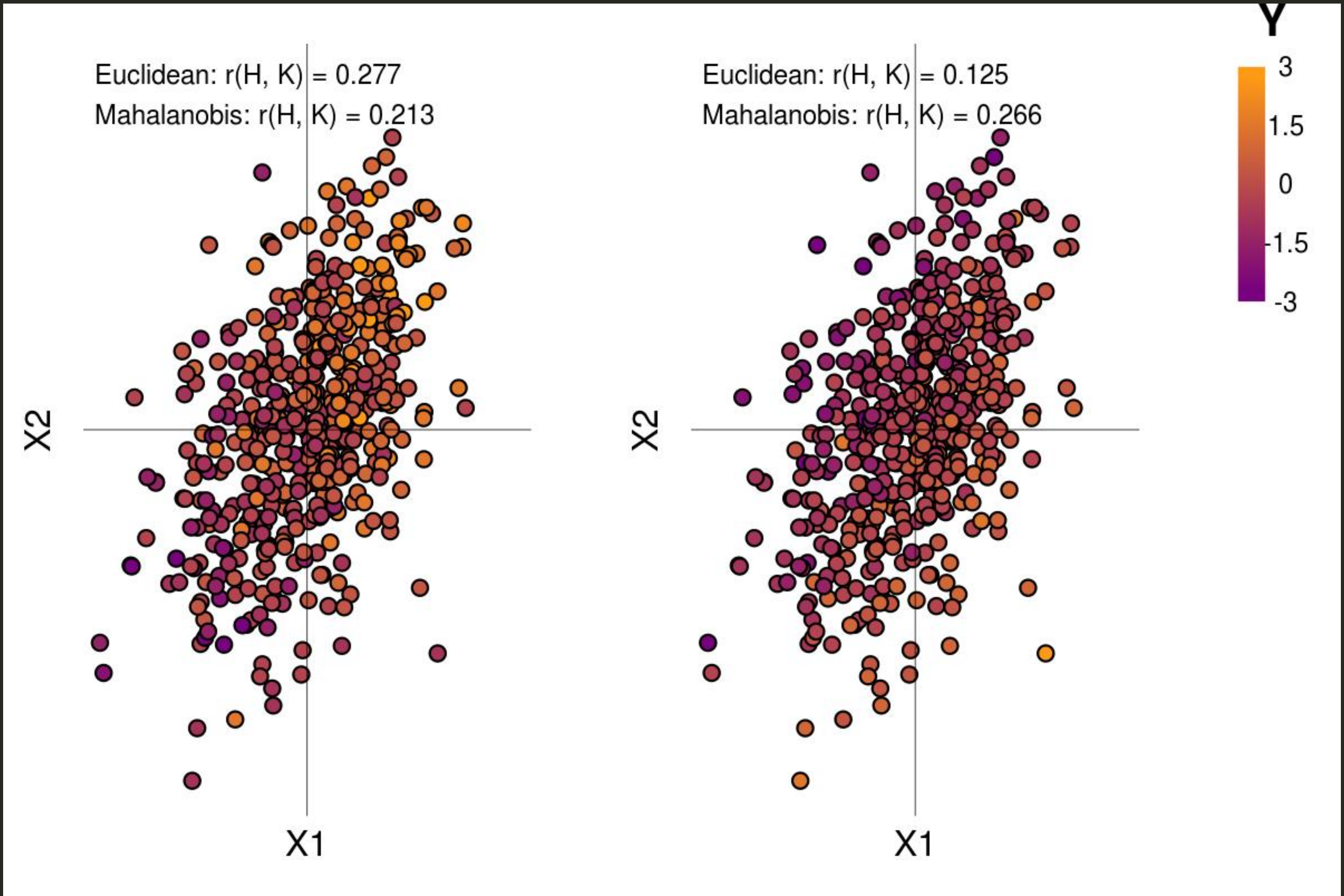

X2
X1


X2
X1


X2
X1

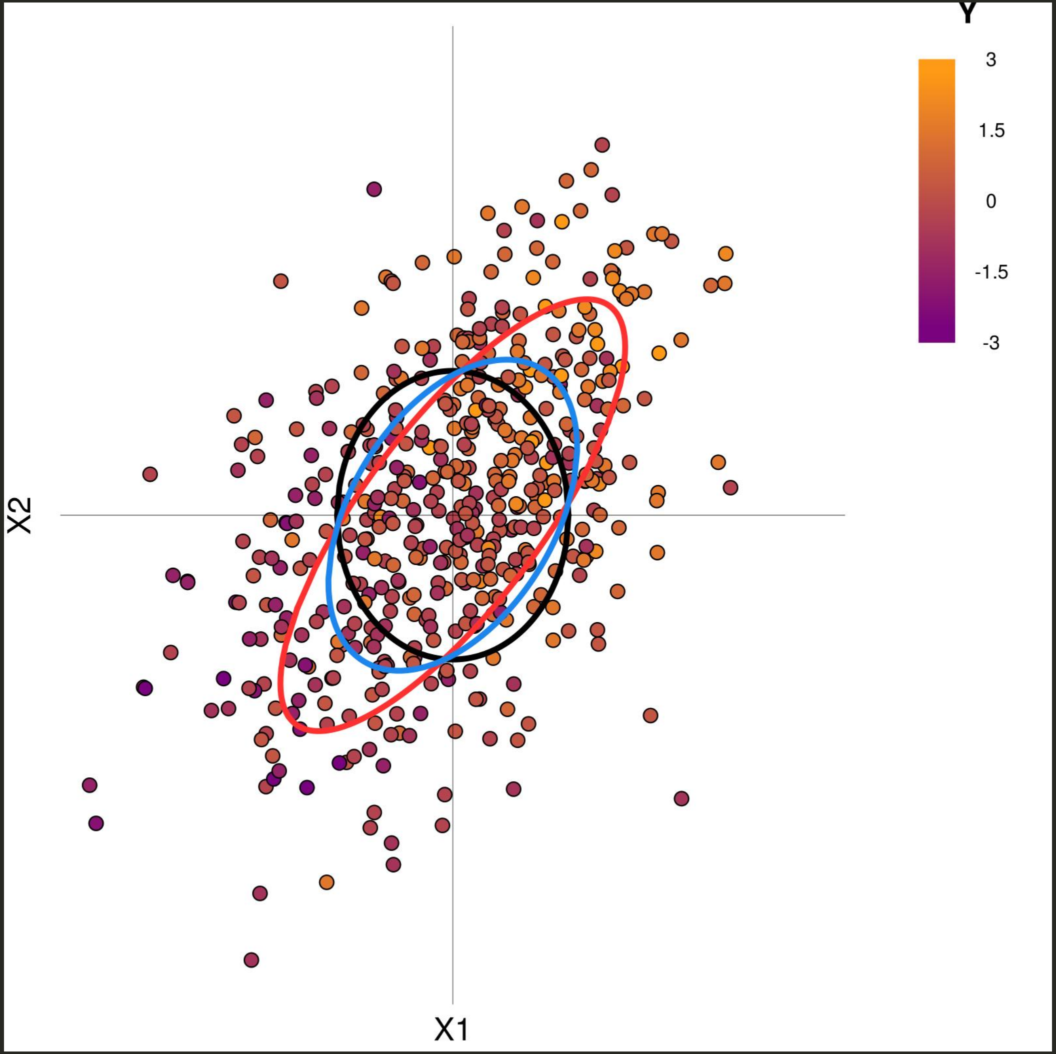
Data generated from variance components model with $\sigma_b^2 = 1$





Left Data generated from variance components model with $\sigma_b^2 = 1$. **Right** Data generated from fixed effects model with $\beta = (0.75, -0.75)$.

Can we compromise between the Mahalanobis and Euclidean metrics?



Weight Matrices

Goal

We want a weight matrix that adjusts for the correlation structure in the data like the Mahalanobis metric, but which is closer to the Euclidean metric.

Weight Matrices

Goal

We want a weight matrix that adjusts for the correlation structure in the data like the Mahalanobis metric, but which is closer to the Euclidean metric.

Solution: Penalize the Mahalanobis weight matrix.

- Let $\lambda \geq 0$. Consider the penalized weight matrix:

$$\mathcal{W}_\lambda = (X^T X + \lambda I_p)^{-1}.$$

- As $\lambda \to \infty$, the penalty term λI_p dominates $X^T X$, and so W_λ tends to a constant diagonal matrix.
- We call $\mathcal{K}(u, v) = u^T \mathcal{W}_\lambda v$ the **ridge kernel**.

Kernel Mantel Test

Weight Matrices

- **Summarizing**, the Euclidean and Mahalanobis inner products are linked by the ridge kernel, where $W_{\lambda=0}$ gives the Mahalanobis metric, and $\lambda \to \infty$ gives the Euclidean metric.

Metric	Gram Matrix
Mahalanobis	$K_M = X(X^TX)^{-1}X^T$
Euclidean	$K_E = XX^T$
Ridge Kernel	$K_\lambda = X(X^TX + \lambda I)^{-1}X^T$

Mantel Test

Contributions

Derive testing properties of the ridge kernel in the Mantel test.

Link the random effects, fixed effects, and ridge regression score tests through the kernel Mantel test framework.

Develop the adaptive Mantel test for simultaneous testing across a range of tuning parameters.

Advantages of Proposed Methods

- Provides an extremely flexible and computationally practical framework for testing a wide variety of relationships between different modalities.
- Simulations show the kernel Mantel test is often more powerful than competing methods.

Kernel Mantel Test

Correlation of Similarities

Assume $\mathrm{rank}(X) = r$ with singular value decomposition $X = U_{n\times r} D_{r\times r} V_{p\times r}^T$, where $\eta_j, j = 1, \cdots, r$ are the squared singular values. Let $H = YY^T$ and $Z = U^T Y$.

Mahalanobis Metric $$r(H, K_M) = \frac{\sum_{j=1}^r z_j^2}{\sqrt{p}\sum_{i=1}^n y_i^2},$$

Euclidean Metric $$r(H, K_E) = \frac{\sum_{j=1}^r \eta_j z_j^2}{\sqrt{\sum_{j=1}^r \eta_j^2 \sum_{i=1}^n y_i^2}},$$

Ridge Similarity $$r(H, K_\lambda) = \frac{\sum_{j=1}^r \frac{\eta_j}{\lambda+\eta_j} z_j^2}{\sqrt{\sum_{j=1}^r \left(\frac{\eta_j}{\eta_j+\lambda}\right)^2 \sum_{i=1}^n y_i^2}}.$$

Kernel Mantel Test

Correlation of Similarities

Assume $\operatorname{rank}(X) = r$ with singular value decomposition $X = U_{n\times r} D_{r\times r} V_{p\times r}^T$, where $\eta_j, j = 1, \cdots, r$ are the squared singular values. Let $H = YY^T$ and $Z = U^T Y$.

Mahalanobis Metric $\quad r(H, K_M) \asymp \sum_{j=1}^r z_j^2 \quad = \operatorname{tr}(HK_M) = T_M,$

Euclidean Metric $\quad r(H, K_E) \asymp \sum_{j=1}^r \eta_j z_j^2 \quad = \operatorname{tr}(HK_E) = T_E,$

Ridge Similarity $\quad r(H, K_\lambda) \asymp \sum_{j=1}^r \frac{\eta_j}{\lambda+\eta_j} z_j^2 \quad = \operatorname{tr}(HK_\lambda) = T_\lambda.$

Kernel Mantel Test

Limiting Relationship

From the previous results, we get the following limiting relationships between the ridge test, and tests for the fixed effects and random effects models.

$$T_{\lambda=0} = T_M$$

$$T_\lambda \asymp \left\{ \lambda \sum_{j=1}^{r} \frac{\eta_j}{\lambda + \eta_j} z_j^2 \right\} \quad \overset{\lambda\to\infty}{\to} \quad T_E$$

Similarly, for the matrix correlations

$$r(H, K_{\lambda=0}) = r(H, K_M)$$

$$\lim_{\lambda\to\infty} r(H, K_\lambda) = r(H, K_E)$$

Kernel Mantel Test

Linear Model Definitions

Model Name	Definition
Fixed	$Y \sim N(X\beta, \sigma_\varepsilon^2 I_N)$
Random	$Y \sim N(0, \sigma_b^2 G + \sigma_\varepsilon^2 I_N), \quad G = XX^T/p$
Ridge	$Y \sim N(X\beta, \sigma_\varepsilon^2 I_N), \qquad \|\beta\|_2^2 < c(\lambda)$

Linear Model Score Tests

Model	Score Stat.	Equivalent Stat.	Null Distribution
Fixed	$S = Z^T D(D^T D)^{-1} D^T Z$	$T_M = \mathrm{tr}(K_M H)$	$c_1 \chi_p^2$
Random	$S = Z^T D D^T Z$	$T_E = \mathrm{tr}(K_E H)$	$c_2 \sum_{j=1}^{r} \eta_j \chi_1^2$
Ridge	$S = Z^T D(D^T D + \lambda I_p)^{-1} D^T Z$	$T_\lambda = \mathrm{tr}(K_\lambda H)$	$c_3 \sum_{j=1}^{r} \frac{\eta_j}{\lambda + \eta_j} \chi_1^2$

Kernel Mantel Test

Geometric Interpretation

Consider $Z = U^T Y$, as the projection of Y into the column space of X.

1. The *Random Effects* model tests the **weighted Euclidean norm** of Z, where the jth component is weighted by the jth eigenvalue η_j.

2. The *Fixed Effects* model tests the **Euclidean norm** of Z

3. The *Ridge Penalization* **weights the Euclidean norm of Z proportional to the eigenvalues**, but these weights are now **flattened** by a factor of $(\lambda + \eta_j)^{-1}$.

Kernel Mantel Test

Proportion of Variance Explained

$$R^2(X, Y) = \sqrt{p} \cdot r(H, K_M)$$

Heritability

For large n and assuming that $\text{rank}(X) = p$,

$$\hat{h}^2_{MOM} = \frac{tr(HG) - n}{tr(G^2) - n} \approx p\sqrt{\frac{tr(H^2)}{\text{tr}(K_E^2)}} r(H, K_E) \in [r(H, K_E), \sqrt{p} \cdot r(H, K_E)]$$

Goal

We want to **use the ridge kernel in the Mantel test** to improve power in high-dimensional settings.

Goal

We want to **use the ridge kernel in the Mantel test** to improve power in high-dimensional settings.

Problem

However, **choosing a good penalty term for inference can be difficult,** since we must control the type I error.

Goal

We want to **use the ridge kernel in the Mantel test** to improve power in high-dimensional settings.

Problem

However, **choosing a good penalty term for inference can be difficult,** since we must control the type I error.

Idea

To simultaneously test a set of tuning parameters, use the **minimum** ***P*-value** across all parameters as the test statistic.

Adaptive Mantel Test

Idea

To simultaneously test a set of tuning parameters, use the **minimum** ***P*-value** across all parameters as the test statistic.

Algorithm

- **Input:**
  - $X, n \times p$ covariates, column centered and scaled
  - $Y, n \times 1$ response, centered and scaled
  - $\left\{\left(\mathcal{K}_m^{\mathrm{X}}, \mathcal{K}_m^{\mathrm{Y}}\right)\right\}, m = 1, \cdots, M$
- **Output:** P_{ADA} = adaptive Mantel P-value for global test of significant association

Adaptive Mantel Test

Algorithm 1: Adaptive Mantel Algorithm

Input: $X, Y, \Lambda = \{\lambda_j\}_{j=1}^m$.

Output: Adaptive Mantel P-value.

$H^{(0)} \leftarrow YY^T$

for $j = 1, \ldots, m$ **do**

$\quad K_{\lambda_j} \leftarrow X(X^TX + \lambda I_p)^{-1}X^T$

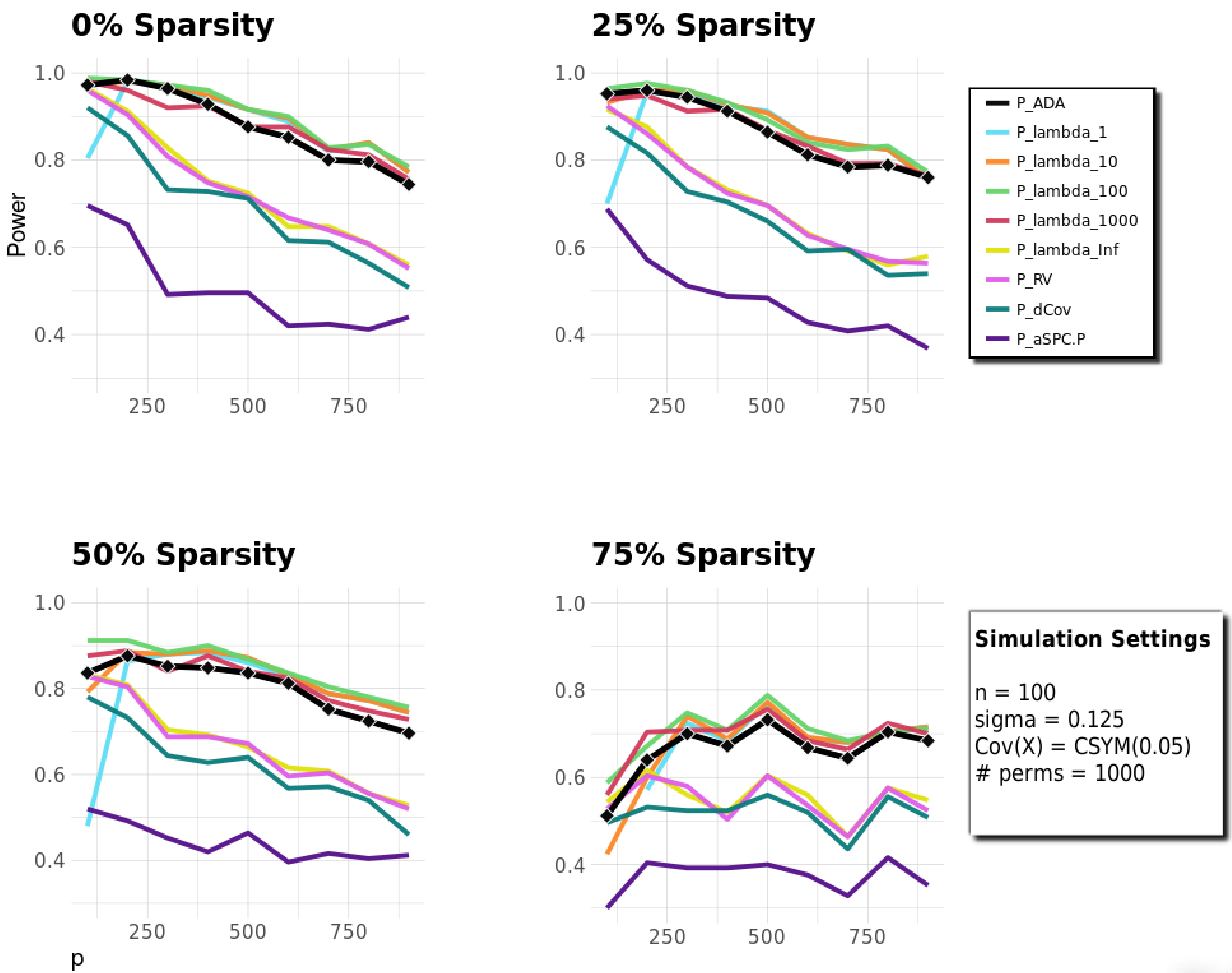
Generate B permutations of $H^{(0)}$, labeled $H^{(b)}, b = 1, \ldots, B$

$T_j^{(b)} \leftarrow \mathtt{tr}(H^{(b)}K_{\lambda_j}), \quad \forall b = 0, \ldots, B; j = 1, \ldots, m$

$P_j^{(b)} \leftarrow \frac{1}{B+1} \sum_{b'=0}^{B} \mathbb{1}(T_j^{(b)} \geq T_j^{(b')}), \quad \forall b = 0, \ldots, B; j = 1, \ldots, m$

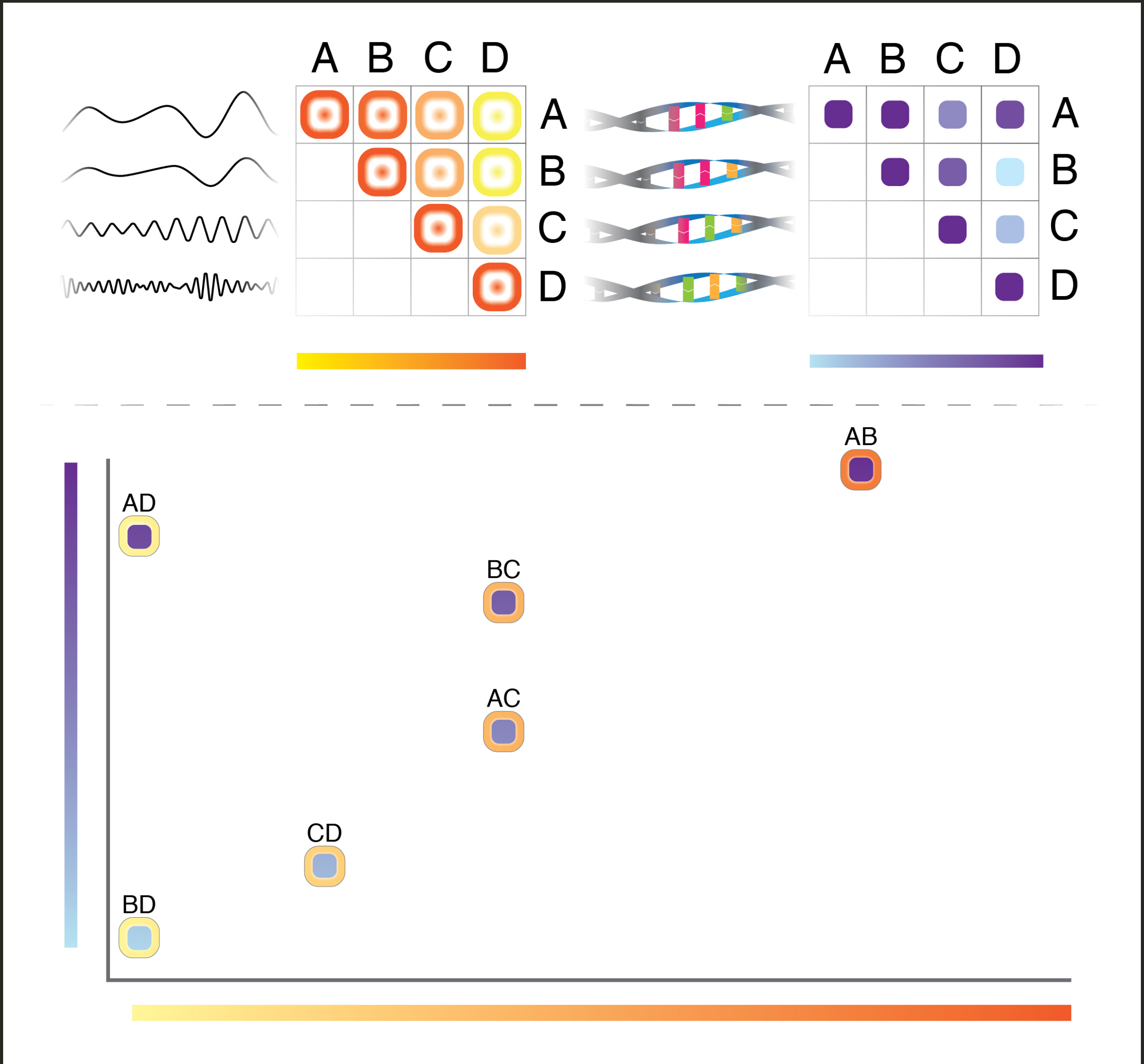
$P^{(b)} \leftarrow \min_j P_j^{(b)}, \quad \forall b = 0, \ldots, B$

$P_{ADA} \leftarrow \frac{1}{B+1} \sum_{b=0}^{B} \mathbb{1}(P^{(0)} \leq P^{(b)})$

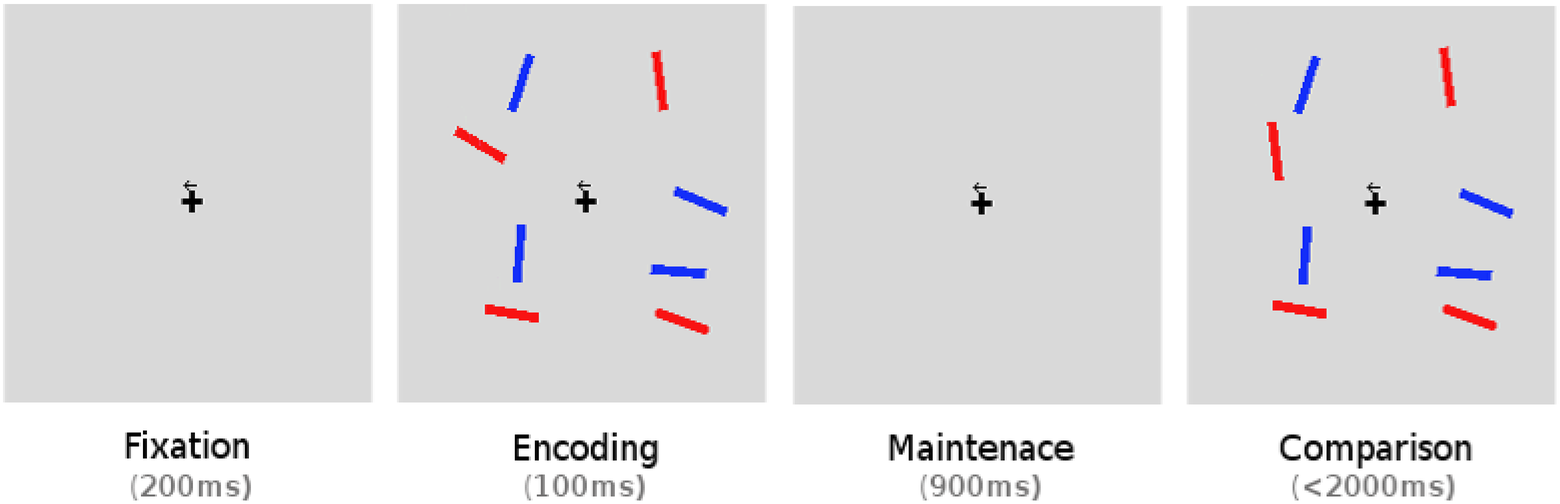

0% Sparsity
25% Sparsity
50% Sparsity
75% Sparsity
Power
p
1.0
0.8
0.6
0.4
250
500
750
P_ADA
P_lambda_1
P_lambda_10
P_lambda_100
P_lambda_1000
P_lambda_Inf
P_RV
P_dCov
P_aSPC.P
Simulation Settings
n = 100
sigma = 0.125
Cov(X) = CSYM(0.05)
# perms = 1000

Application Example

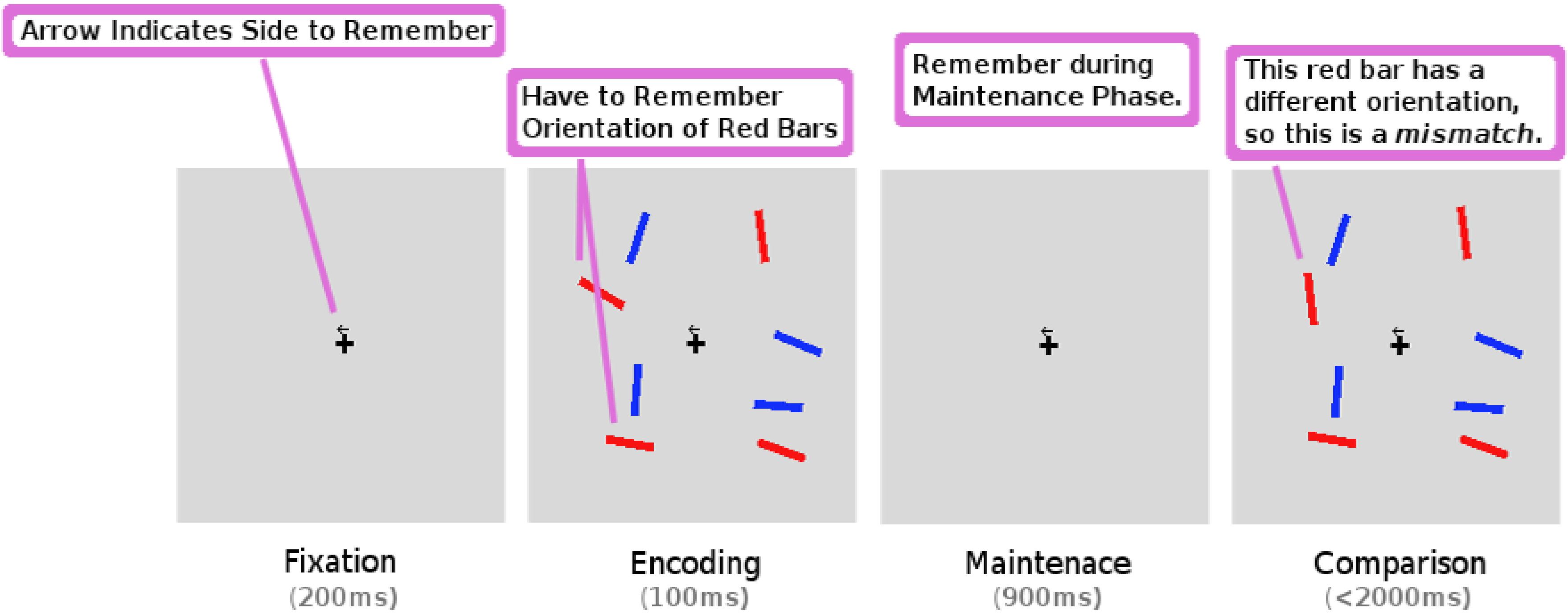
- 350 Subjects from the BNU data set
- ~10 minute 64 channel EEG recording during VWM task
    - Preprocessed according to standard pipeline
    - Coherence measures for each channel pair was calculated by the FFT, and grouped into five frequency bands (in Hz): $\delta\ (1-4), \theta\ (4-8), \alpha\ (8-16), \beta\ (16-32), \gamma\ (32+)$
- 13 SNPs selected for analysis, previously identified as potential factors for Alzheimer's disease risk
    - All 13 SNPs passed standard quality control checks


A B C D
A
B
C
D
A B C D
A
B
C
D
AB
AD
BC
AC
CD
BD

Visual Working Memory Experiment



Visual Working Memory Experiment



Related Previous Results

- *Sauseng et al. (2005)* also found that **theta** and **alpha coherence** play a significant role in “top-down” control during working memory tasks
- *Jiang et al. (2005)* A study of EEG power and coherence in patients with mild cognitive impairment revealed differences in **theta, alpha, and beta band** power between MCI and normal controls during working memory tasks.
- *Vogler et al. (2014)* **Estimated SNP-based heritability for working memory performance** to be in the range of $h^2 = 0.31$ to $0.41 (P = 0.0008)$.
- *Dai et al. (2017)* found topological reorginzation of EEG cortical **connectivity in the theta and alpha bands** during working memory tasks.

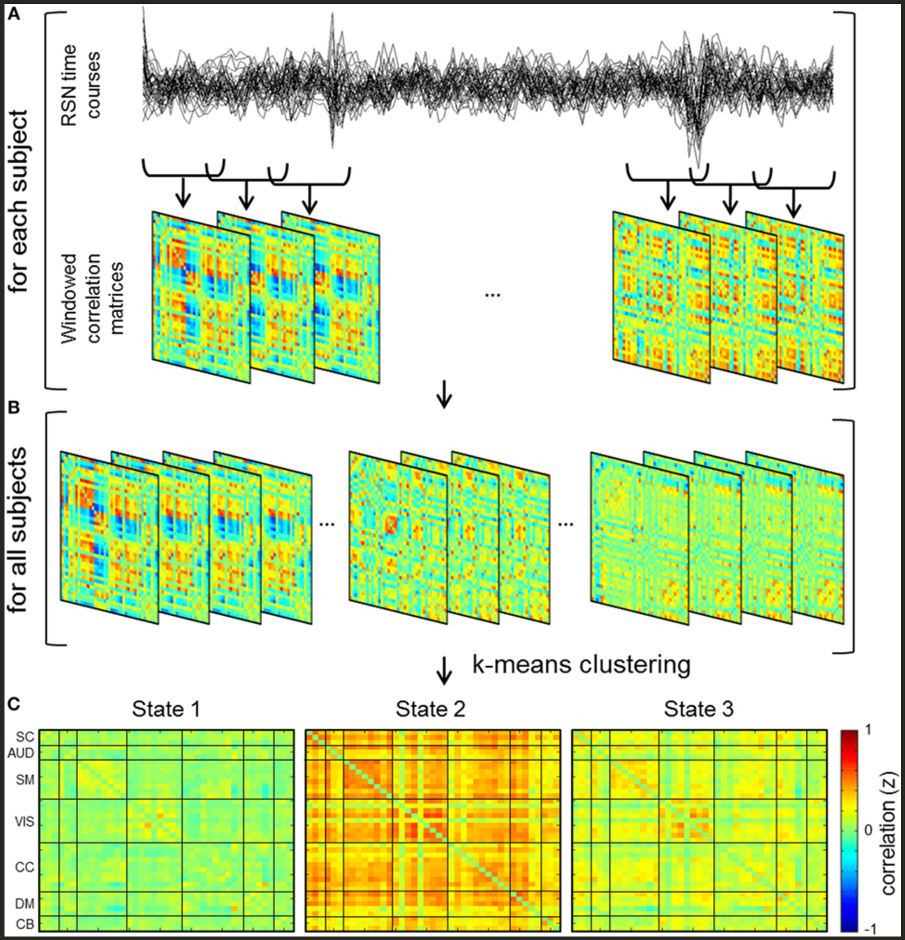
Adaptive Mantel Test Results

- Results of adaptive Mantel test for association of AD SNPs and EEG Coherence at particular frequency bands
- Used L_2 similarity for SNPs, and ridge kernel similarity for coherence, with penalty terms $\Lambda = \{0.5, 1, 5, 10, 100, 1000, \infty\}$

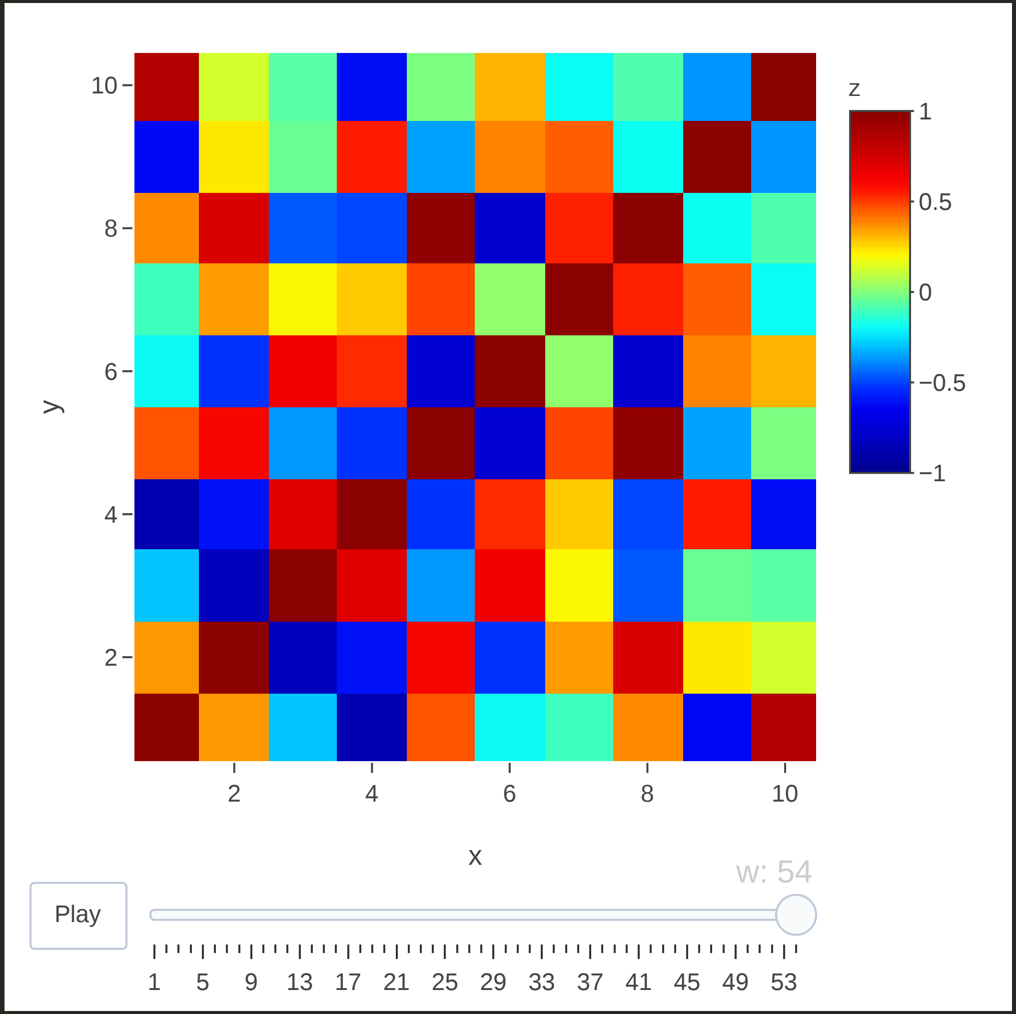
Band	Channels	$P-$ value
β	All	0.619
β	Frontal	0.517
α	All	0.075
α	Frontal	0.381
θ	All	0.416
θ	Frontal	0.081
δ	All	0.015
δ	Frontal	0.088

Future Work: Methods for Dynamic Connectivity Analysis

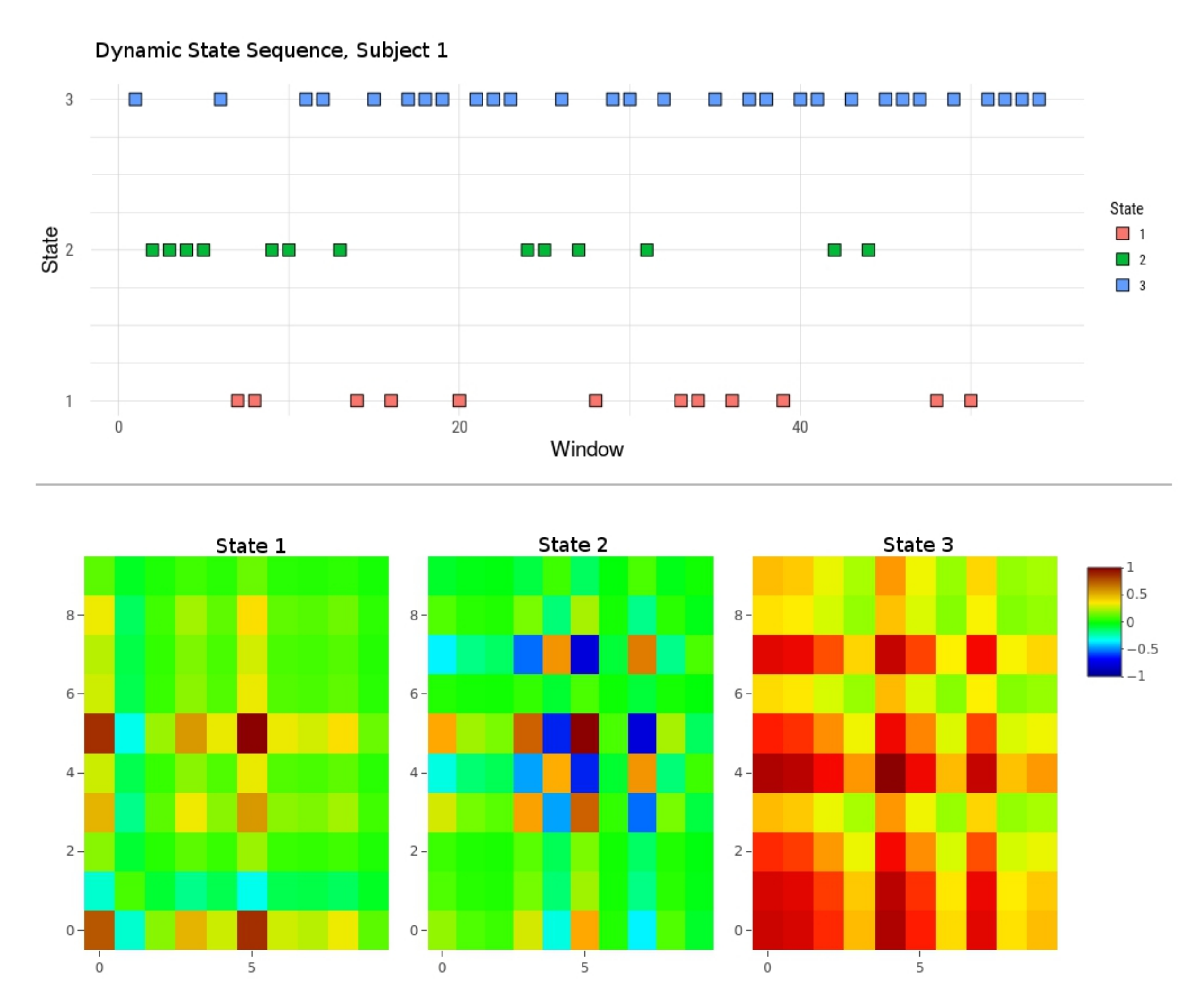
Dynamic Connectivity



Subj. 2, Win. Length = 10s



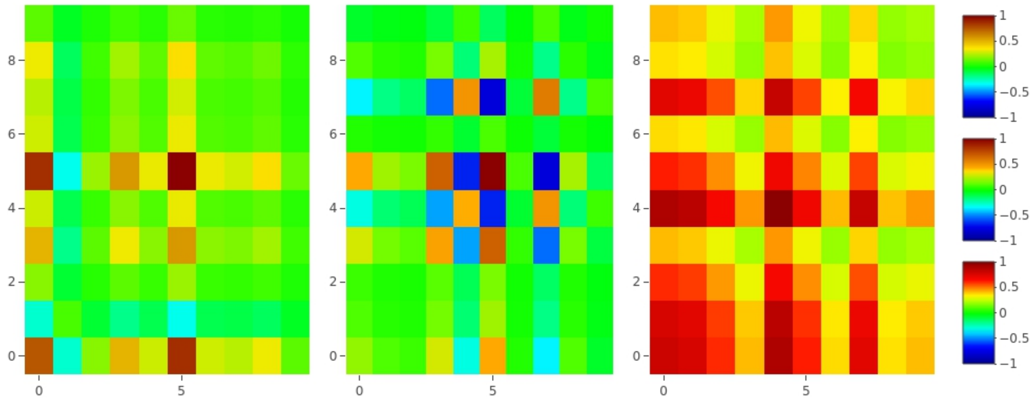
Dynamic Connectivity



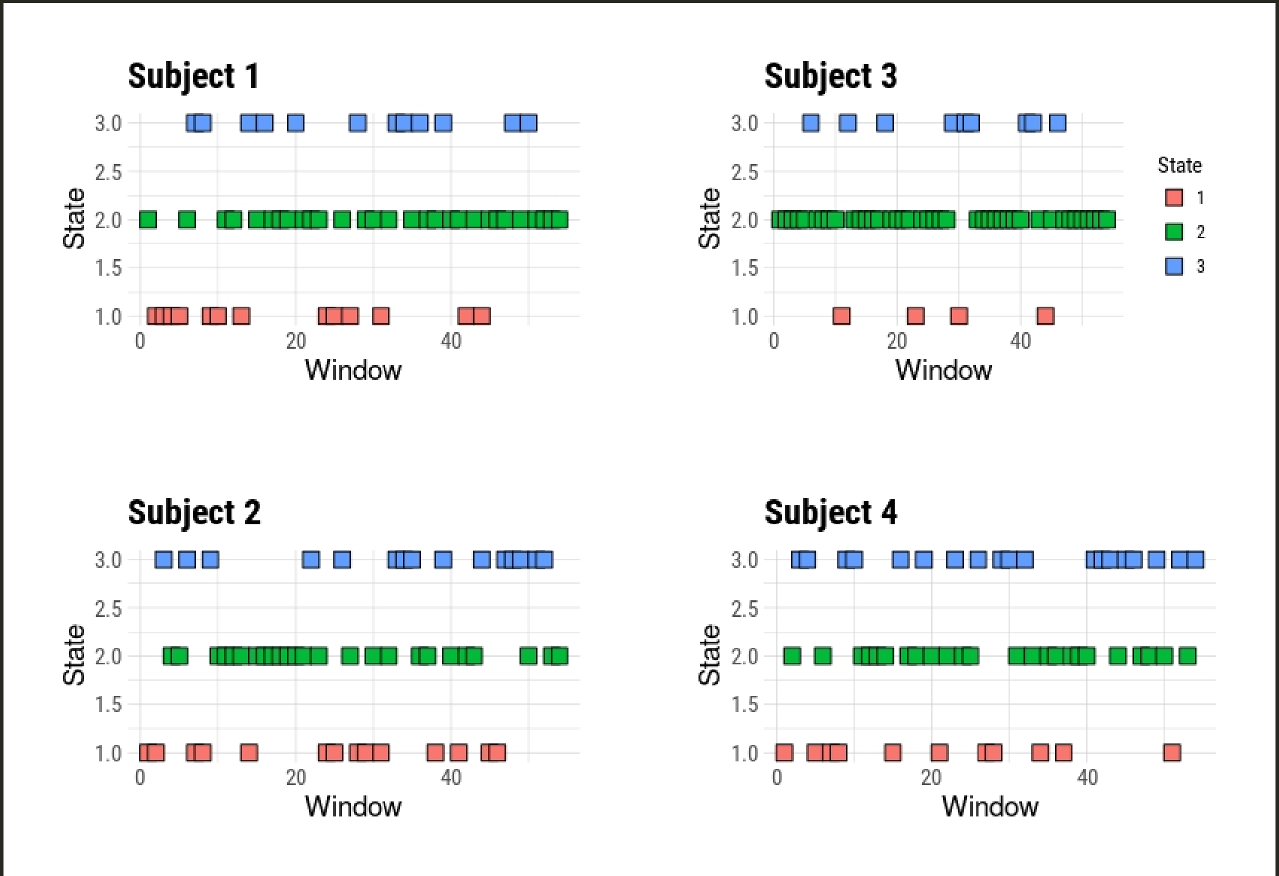
Dynamic Connectivity

Global transition probabilities.

FROM \ TO	State 1	State 2	State 3
State 1	0.195	0.496	0.309
State 2	0.174	0.549	0.277
State 3	0.182	0.507	0.311



Dynamic Connectivity



Dynamic Connectivity

Scientific Questions

- How do behavioral outcomes correlate with dynamic connectivity patterns?
- Are the states linked to particular experimental settings?
- How can one interpret the state matrices?
- Can subjects be classified into similar dynamic connectivity profiles by some set of covariates?

Dynamic Connectivity

Statistical Challenges

- Choice of window size,
- Estimated states are highly sensitive to modeling choices,
- The concatenated k-means method is somewhat *ad hoc* for estimating common states across subjects,
- The number of states needs to be chosen.
- Can (or should) we try to encourage temporal continuity in state estimation?
- How best to reduce dimension when estimating states?

Dynamic Connectivity

Some Ideas

- Choice of window size $\Rightarrow$ Phase Coherence (gives an "instantaneous" measure of connectivity)
- Instead of k-means $\Rightarrow$ Dirichlet process mixture model or Classification Trees
- Temporal continuity $\Rightarrow$ model based estimation of state sequence
- Higher Order SVD can be used instead of k-means to cluster while still respecting matrix structure
- Dimension reduction $\Rightarrow$ HOSVD, PC reduction, shrinkage methods, manifold regression models

Dynamic Connectivity

Cabral et al. (2018)

DFC for fMRI with Time-varying BOLD Phase Coherence

Calculate BOLD Phase Coherence Connectivity to obtain a $q \times q \times T$ dFC matrix.

- Computing Phase Coherence:
    - Estimate phase $\theta(j, t)$ of BOLD signals in area j at time t using Hilbert transform
    - Given phase of the BOLD signals, phase coherence between areas q and q' at time t (denoted $dFC(q, q', t)$) is defined as

$$dFC(q, q', t) = \cos(\theta(q, t) - \theta(q', t)).$$

- *Leading Eigenvector* of each $dFC(t)$ is used to capture connectivity structures.

Dynamic Connectivity

Studying FC Dynamics

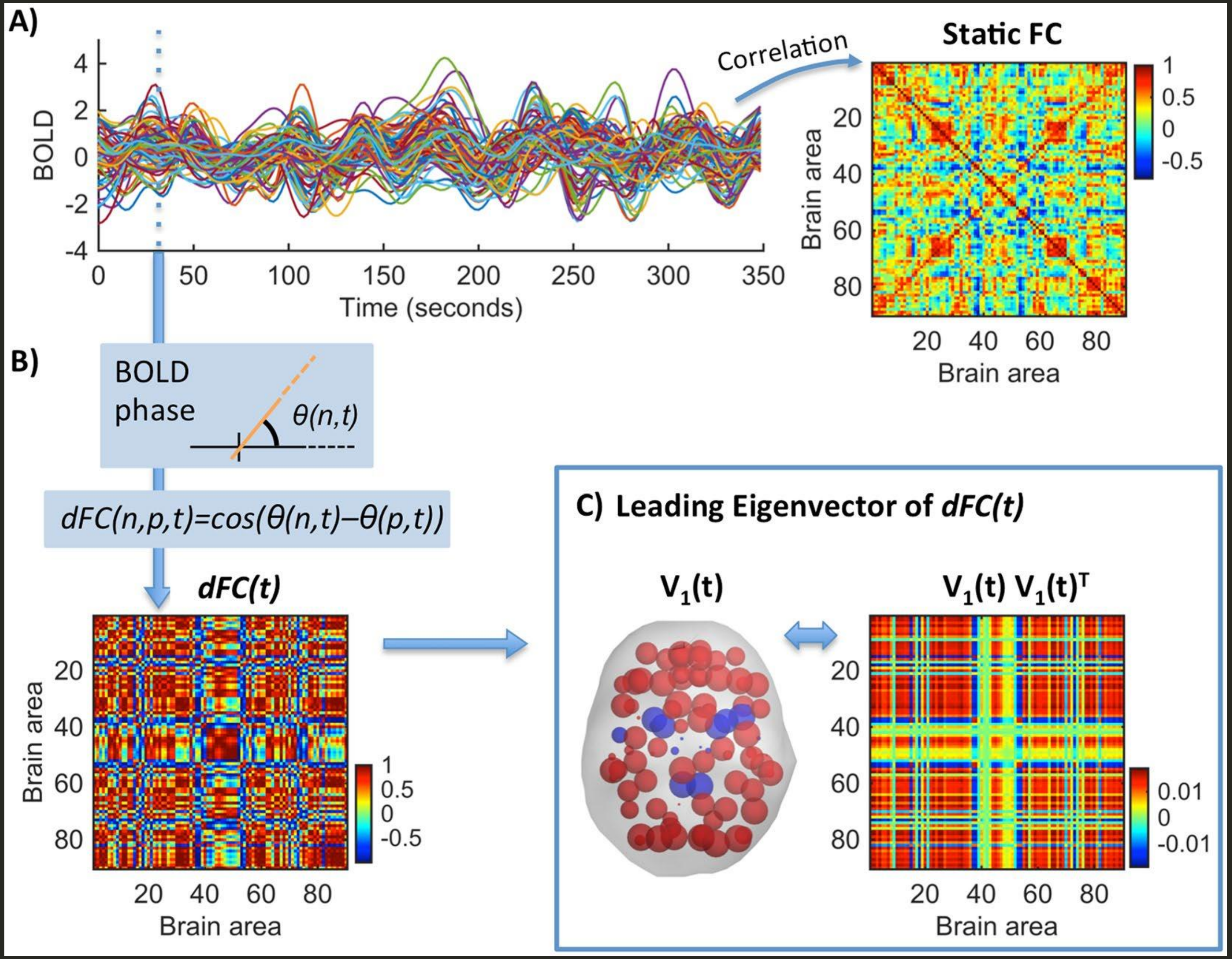
- Compute time-versus-time matrix to represent functional connectivity dynamics (FCD), where each $FCD(t_x, t_y)$ measures the resemblance between the dFC at times t_x and t_y using Pearson correlation or cosine similarity:

$$FCD(t_x, t_y) = \frac{\langle V_1(t_x), V_1(t_y) \rangle}{\|V_1(t_x)\| \|V_1(t_y)\|} \in [-1, 1].$$

- *FC States:* detect discrete FC states with k-means clustering on the leading eigenvectors $V_1(t)$ across all time points and subjects.

- *Between group Comparisons:* Group differences are tested using permutation-based paired t-test

Dynamic Connectivity



Thanks!

Links

- **Adaptive Mantel Test Paper:** arxiv.org/pdf/1712.07270.pdf
- **Slides and References:** github.com/dspluta/Presentations/
- **Adaptive Mantel R Package:** github.com/dspluta/adamant

References

- Wu M., et al. Kernel Machine SNP-set Testing under Multiple Candidate Kernels. Genetic Epideomiology. 2013. 37(3): 267-275.
- Cai T., et al., Kernel Machine Approach to Testing the Significance of Multiple Genetic Markers for Risk Prediction. Biometrics. 2011. 67(3): 975-986.
- Ge T, et al. Massively Expedited Genome-Wide Heritability Analysis (MEGHA). PNAS. 2015. 112, 2479-2484.
- Xue G, et al. Functional Dissociations of Risk and Reward Processing in the Medial Prefrontal Cortex. Cerebral Cortex. 2009. 19, 1019-1027.
- Yang J, et al. GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics. (2011) 88, 76-82.
- Tzeng et al. (2009) Biometrics 65, 822.
- Visscher et al. (2014) Statistical power to detect genetic (co)variance of complex traits using SNP data. PLoS Genetics.
- GCTA Power, http://cnsgenomics.com/shiny/gctaPower/

Appendix

Adaptive Mantel Test

Computing the adaptive Mantel test can be done efficiently using either the SVD or a linear algebra trick, depending on the relative sizes of n and p.

SVD

- Computing the SVD $X = UDV^T$ can be completed in $O(np^2)$.
- When $rank(X) = r \leq n$, the Mantel statistic can be then be computed in $O(n^2)$:

$$T = \sum_{i=1}^{r} \eta_i z_i^2$$

- Using B permutations gives a total complexity of $O(np^2 + Bn^2)$.

Adaptive Mantel Test

Linear Algebra Trick

When $p \gg n$, it is better to instead use the following reformulation for K:

$$K_\lambda = X(X^T X + \lambda I_p)^{1} = (XX^T + \lambda I_n)^{-1} XX^T.$$

Calculating K_λ with this alternative form can be done in $O(n^2 p)$, giving a total computational cost of $O\left(n^2(p + B)\right)$.

- The computation for the adaptive test scales this cost linear relative the number of tuning parameters included.
- The computations can be easily parallelized.

EEG Pre-processing

- **EEG pre-processing:**

  1. Downsample from 1024 Hz to 128 Hz
  2. Remove bad channels
  3. Band-pass filter from 1 Hz to 45 Hz
  4. Interpolate/re-reference bad channels
  5. ICA to remove eyeblinks and motion artifacts
  6. Remove remaining bad trials. Exclude subjects if > 5% of trials removed.

- **Calculate coherence** for all subjects and all channels using the FFT, and compute mean coherence by frequency band.

Most Significant Channel Pairs for h^2

Estimating Heritability with GCTA

- While GCTA is a sensible and feasible approach to estimating heritability, it may prove impractical due to sample size limitations.

Power calculations for GCTA model.

Sample Size	Heritability `$h^2$`	Power	$SE(\hat h^2)$
1000	0.2	0.097	0.316
2000	0.2	0.24	0.158
3000	0.2	0.475	0.105
1000	0.5	0.353	0.316
2000	0.5	0.885	0.158
3000	0.5	0.997	0.105
400	1	---	0.79

http://cnsgenomics.com/shiny/gctaPower/

Multivariate Mantel

Hooper's trace correlation (Hooper 1959) is defined as

$$r_T^2 = \frac{1}{q}\mathrm{tr}\left((Y^TY)^{-1}Y^TX(X^TX)^{-1}X^TY)\right)$$

The relationship between Hooper's trace correlation and $r(H_F, K_F)$ follows from $\mathrm{tr}(H_F) = q, \mathrm{tr}(K_F) = p$:

$$r_T^2 = \sqrt{\frac{p}{q}}r(H_F, K_F).$$

Multivariate Mantel

Multivariate Heritability

Following Ge et al. (2016), a multivariate version of h^2 can be defined

$$h^2 = \frac{\mathrm{tr}(\Sigma_g)}{\mathrm{tr}(\Sigma_g) + \mathrm{tr}(\Sigma_\varepsilon)}.$$

A method of moments estimator for h^2 is

$$\hat{h}^2_{MOM} = \frac{tr(K_R)tr(\hat{\Sigma}_b)}{tr(H_R)} = \frac{tr(K_R)}{tr(H_R)} \frac{tr(H_R K_R) - tr(H_R)tr(K_R)/n}{tr(K_R^2) - tr^2(K_R)/n}.$$

When both X and Y are column-standardized and full column rank it can be shown that

$$\hat{h}^2_{MOM} \in \left[\frac{1}{\sqrt{q}} r(H_R, K_R), \sqrt{p}\, r(H_R, K_R)\right]$$

Multivariate Mantel

Interpretation of Tuning Parameters

In general, a large tuning parameter **reduces the adjustment for the correlation between features**, and so approaches the use of **Euclidean distance** in that modality as the parameter increases.

Limiting Relationships

$$\lim_{\lambda_x \to 0, \lambda_y \to 0} r(H_{\lambda_y}, K_{\lambda_x}) = r(H_F, K_F)$$

$$\lim_{\lambda_x \to \infty, \lambda_y \to \infty} r(H_{\lambda_y}, K_{\lambda_x}) = r(H_R, K_R)$$

Penalized Likelihood

$$-\ell \propto c_1 log|\Sigma_e| + tr[(Y - XB)\Sigma_e^{-1}(Y - XB)^T] + \lambda_x tr[B\Sigma_e^{-1}B^T] + \lambda_y tr[\Sigma_e^{-1}] + c_2$$

AdaMant Package

```
set.seed(1234)
X <- matrix(rnorm(500), nrow = 50, ncol = 10)
Y <- X %*% rep(c(0, 0.6), 5) + rnorm(50, 0, 2)

adamant(X, Y, lambdas_X = c(0, 1, 10, Inf),
        n_perms = 2000, P_val_only = TRUE)
```

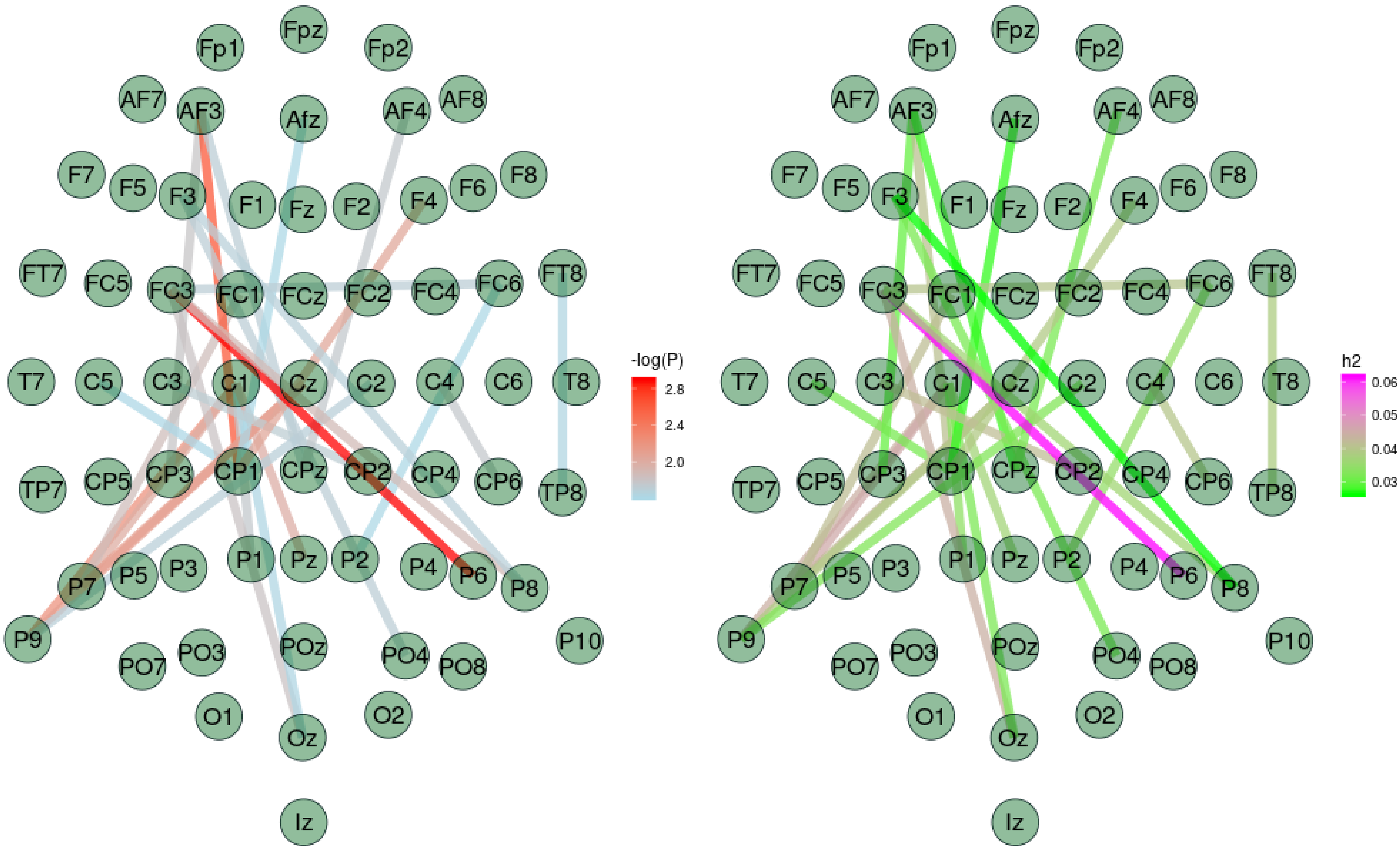
```
## -------------------------------
## Adaptive Mantel Output
## -------------------------------
## P_val            = 0.0055
## n                = 50
## p                = 10
## rank(X^TX)       = 10
## kappa            = 1.982
## Best Lambda      = 10
## n_perms          = 2000
## time             = 0.12 secs
## -------------------------------

## [1] 0.0055
```

SNPs Related to Alzheimer's Disease

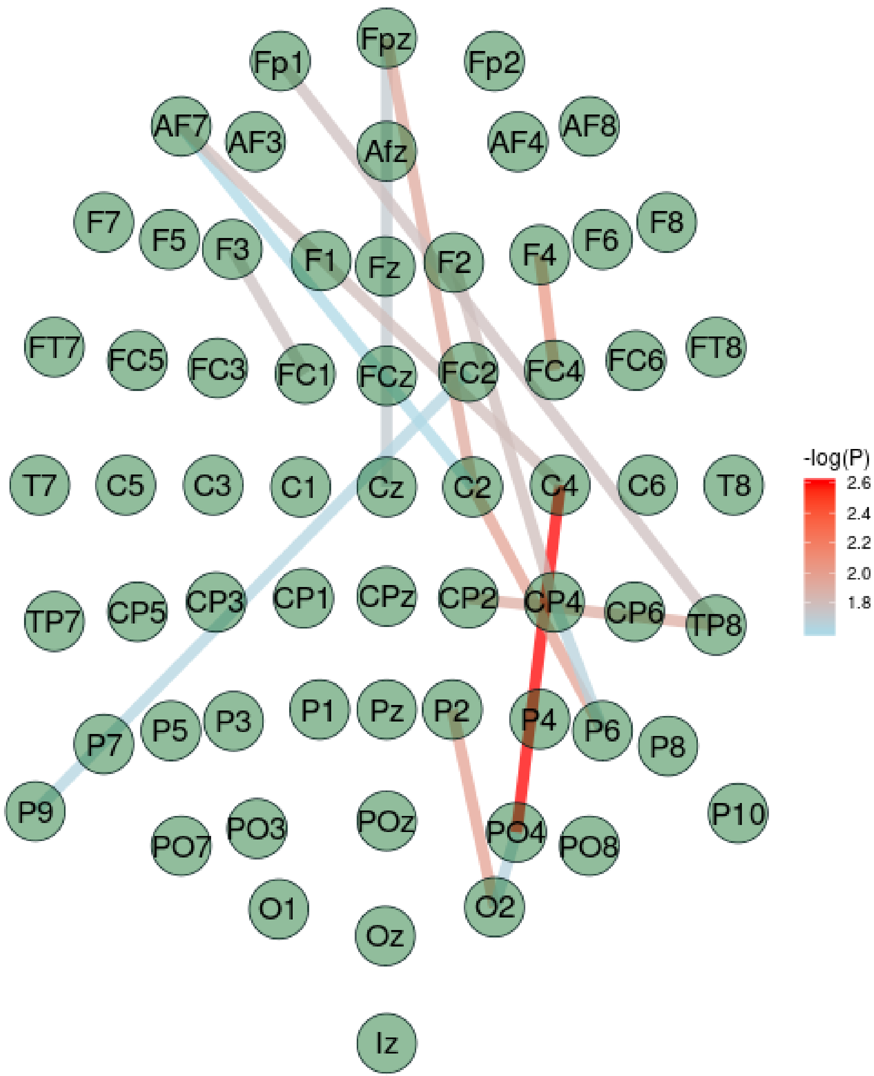
Info on SNPs from SNPedia:

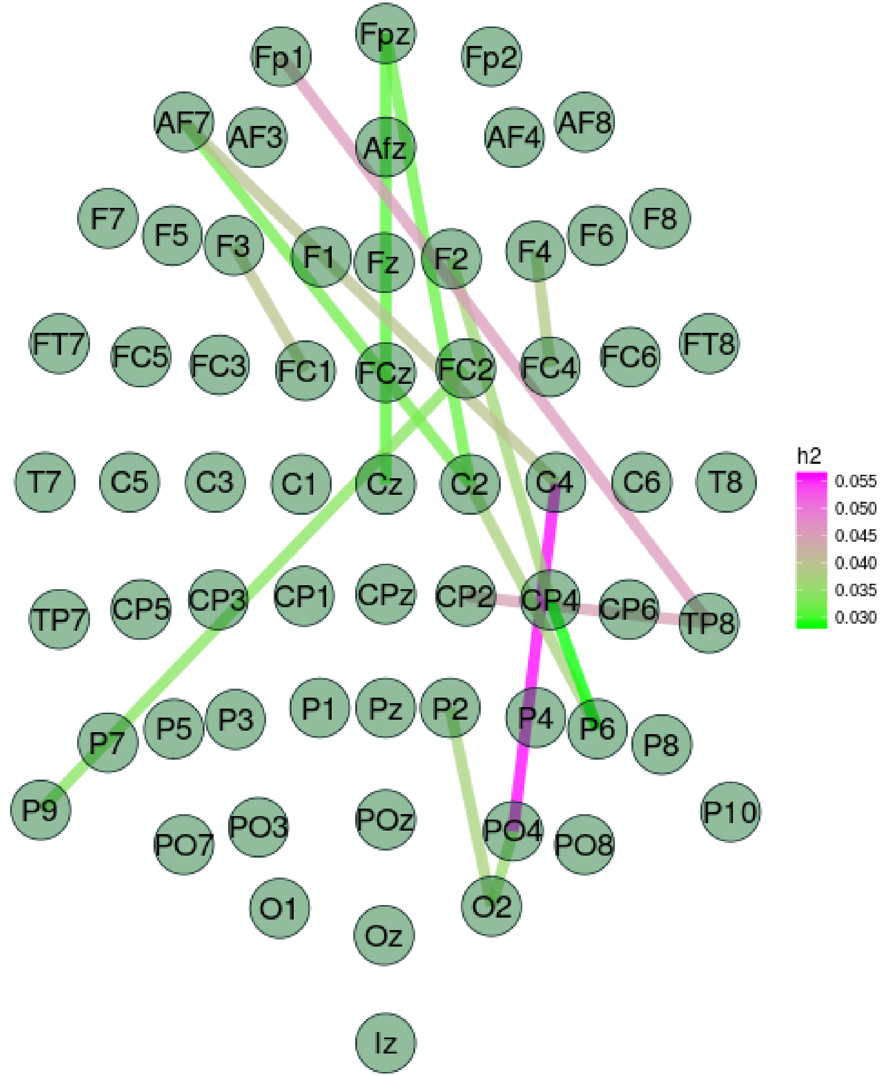
- *rs2227564*: A functional polymorphism within plasminogen activator urokinase (PLAU) which some studies have shown to be associated with Alzheimer's disease. *(Riemenschneider et al., 2006)*
- *rs3851179*: A slight protective effect of the (A) allele of this SNP was found in *(Carasquillo et al., 2010)*
- *rs3818361*: A SNP associated with the complement component (3b/4b) receptor 1 CR1 gene. *(Carasquillo et al., 2010)*
- *rs9886784*: An intergenic SNP on chromosome 9, is reported to influence the risk for Alzheimer's disease; the odds ratio is 3.23 (CI: 1.79 - 5.84). *(Li et al., 2007)*

h2 Theta Coherence

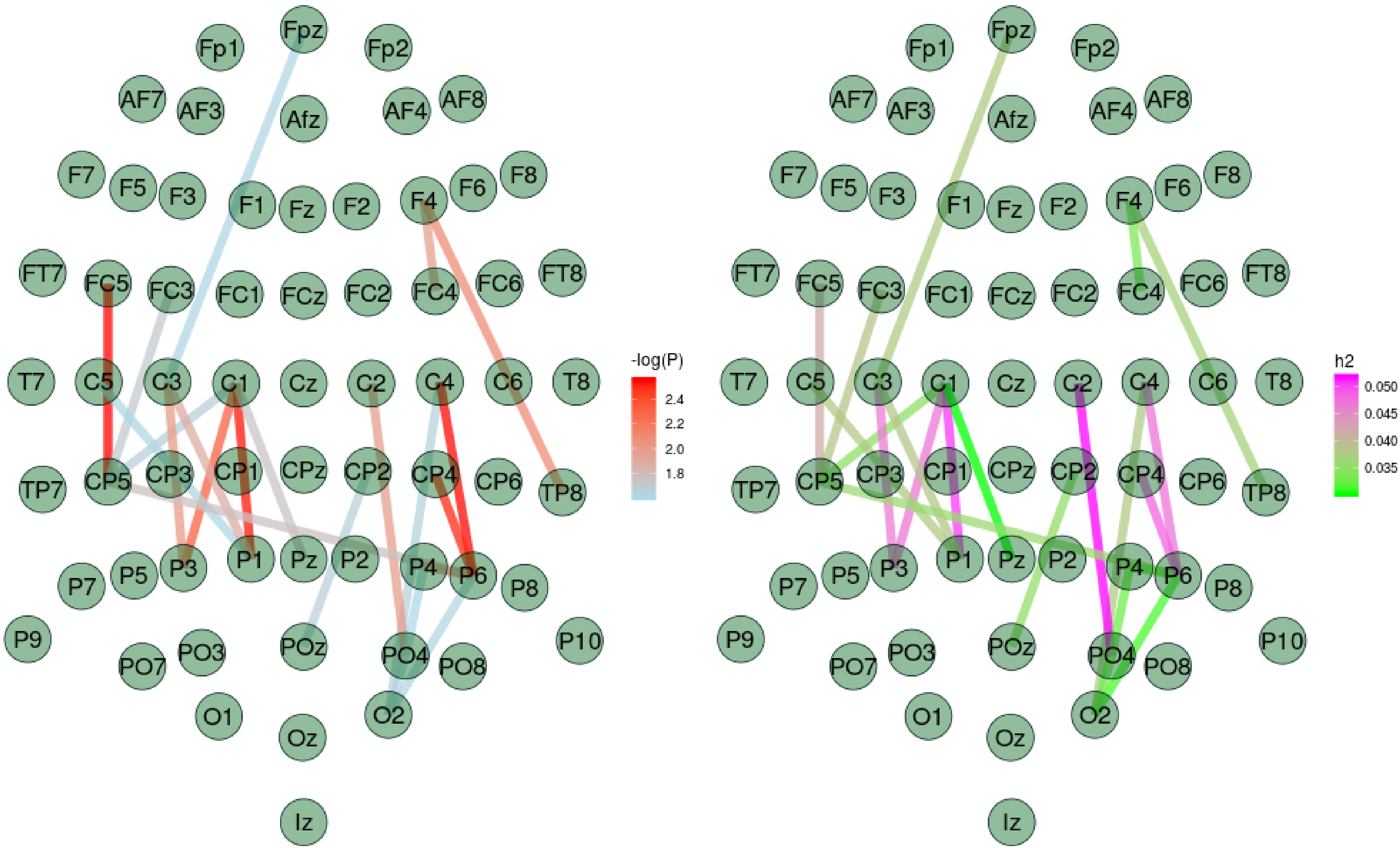


h2 Alpha Coherence





h2 Beta Coherence

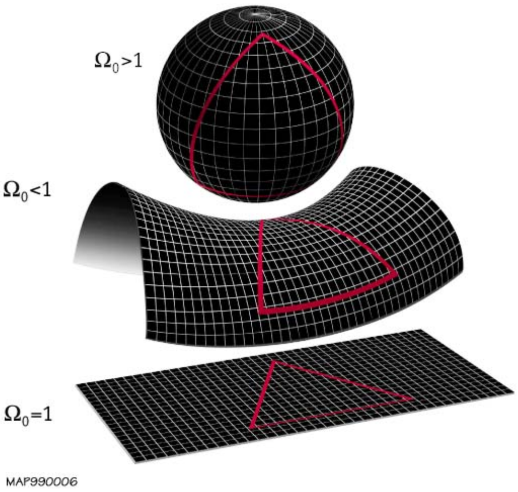


Manifold Regression

Manifold Regression

Riemannian Manifold

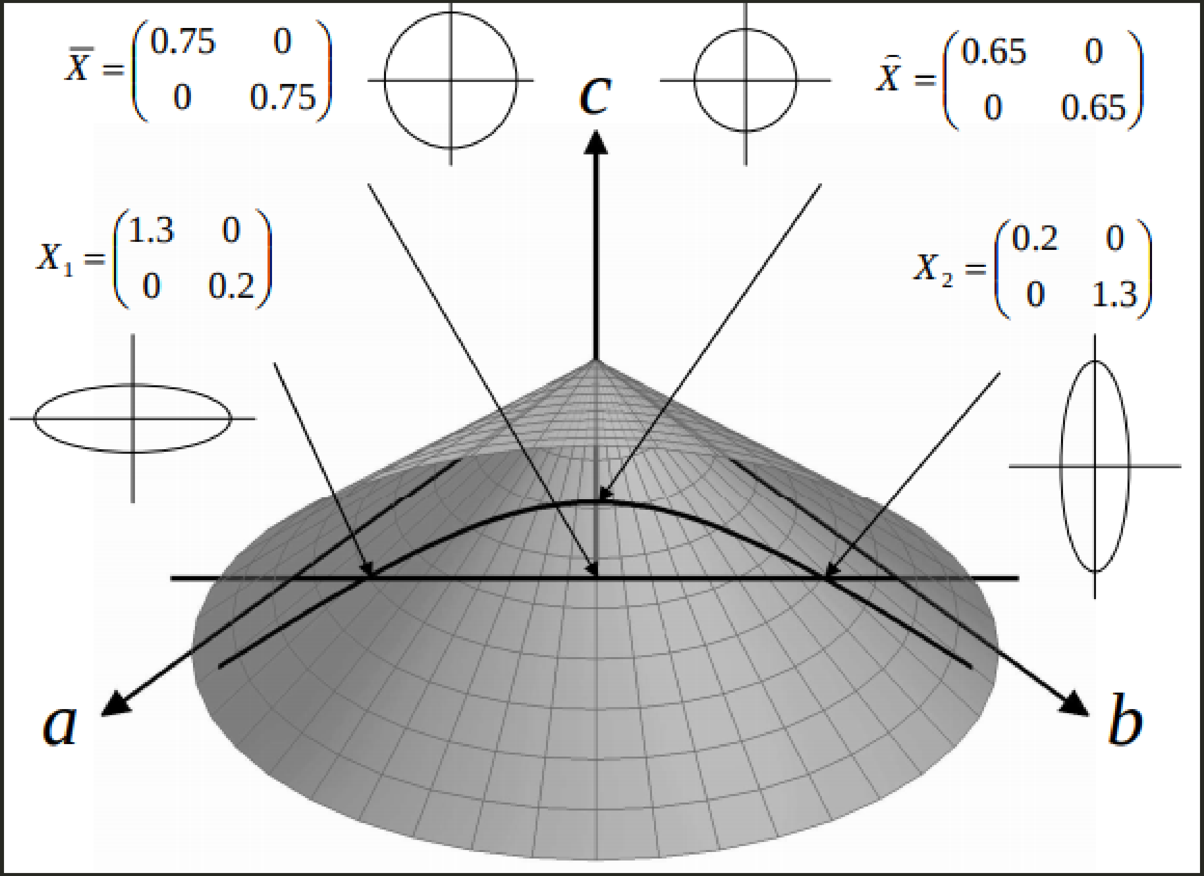
A smooth surface such that each point has a locally Euclidean neighborhood.



Some Manifolds of Interest

1. $\mathrm{Sym}^+(q)$, $q \times q$ **positive definite matrices**
2. $O(q)$ and $SO(q)$, the groups of **orthogonal** and **special orthogonal** $q \times q$ matrices
3. **Stiefel Manifold**, $q \times k$ matrices of orthonormal vectors.

Positive Definite Cone



Manifold Regression

Specification of the Geodesic Model

- Suppose a $q \times q$ PD matrix $S_i \in \mathrm{Sym}^+(q)$ and a $k \times 1$ vector of covariates are observed for each subject $i = 1, \cdots, n$.
- Let $\beta \in \mathbb{R}^p$ be a $p \times 1$ vector of regression coefficients
- $\Sigma(\cdot, \cdot) : \mathbb{R}^{\mathrm{k}} \times \mathbb{R}^{\mathrm{p}} \to \mathrm{Sym}^+(q)$.
- We are interested in modeling the **conditional mean** of S_i given x_i, denoted

$$\Sigma_i(\beta) = \Sigma(x_i, \beta) \in \mathrm{Sym}^+(q).$$

Manifold Regression

Specification of the Geodesic Model

- Further let:
    - $D \in \mathrm{Sym}^+(q)$ be the intercept matrix $D = \Sigma(0, \beta)$, with $D = BB^T$ for some $B \in GL(q)$
    - $Y_D(x_i, \beta) = Y_{D,i} \in \mathrm{Sym}(q)$ be a "directional" matrix
    - $C_i(\beta)$ be a Cholesky square root $\Sigma(x_i, \beta) = C_i(\beta)C_i(\beta)^T$.
- The **geodesic model** is given by

$$\Sigma(x_i, \beta) = B \exp(B^{-1} Y_{D,i}(\beta) B^{-T}) B^T = C_i(\beta)C_i(\beta)^T.$$

Manifold Regression

- The **geodesic model** is given by

$$\Sigma(x_i, \beta) = B \exp(B^{-1} Y_{D,i}(\beta) B^{-T}) B^T = C_i(\beta) C_i(\beta)^T.$$

- The **residuals** are defined as

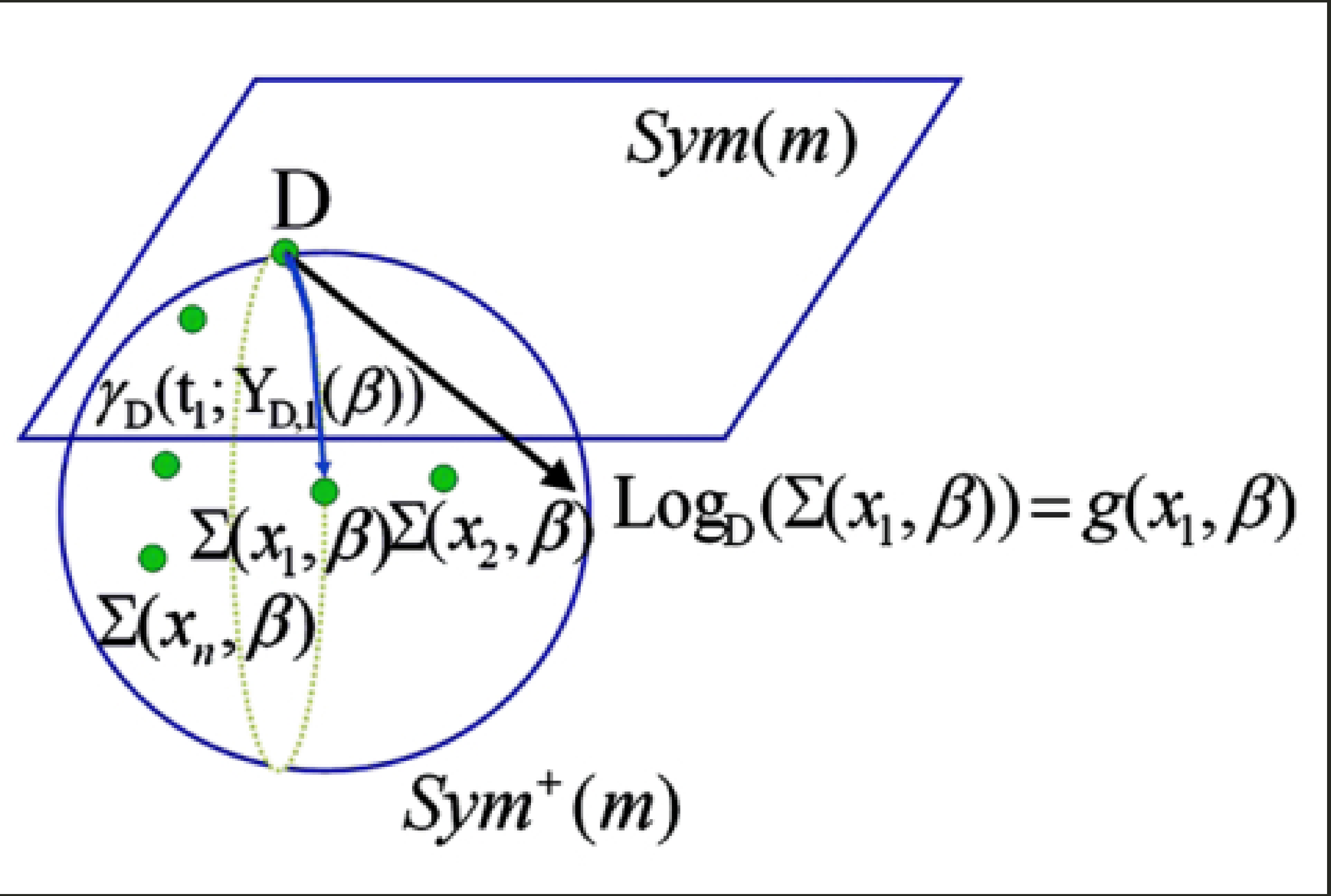
$$\mathcal{E}_i(\beta) = \log(C_i(\beta)^{-1} S_i C_i(\beta)^{-T}).$$

The intrinsic regression model is then specified by $\mathbb{E}[\mathcal{E}_i(\beta) | x_i] = 0$, where the intrinsic least squares estimate of β is defined as

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \operatorname{tr}\left(\mathcal{E}_i(\beta)^2\right).$$

Manifold Regression

Geodesic Model



Manifold Regression

Manifold Mixed Effects Model

- Kim et al. (2017) have proposed an extension of the linear mixed effects model to manifold-valued responses.
- Let $Y_{[ij]}, B, B_i \in \mathcal{M}, V \in T_B\mathcal{M}^p, U_i \in T_{h[ij]}\mathcal{M}^q, x_{[ij]} \in \mathbb{R}^p, z_{[ij]} \in \mathbb{R}^q$, and let $\Gamma_{B\to B_i}V$ be the parallel transport of V from B to B_i.
- A simplified form of the **manifold mixed effects model** can be stated as

$$Y_{[ij]} = \mathrm{Exp}(\mathrm{Exp}(B_i, \Gamma_{B\to B_i}(V)x_{[ij]}, \varepsilon_{[ij]}))$$

$$B_i = \mathrm{Exp}(B, U_i).$$

$$\Gamma_{B\to I}U_i \;\sim\; \mathcal{N}_{SYM}(0, \sigma_U^2),$$

where $\mathcal{N}_{SYM}$ is the normal distribution over symmetric positive definite matrices.

Manifold Regression

Heritability Estimation for Manifold-valued Phenotype

- Analagous to the definition of the variance components model for heritability analysis, we propose a random effects model for manifold-valued phenotypes
- For this model, we instead consider the random effects U_i to have a common aggregate genetic effect σ_U^2

$$\Gamma_{B \to I} U \sim \mathcal{N}_{SYM}(0, \sigma_U^2 G)$$