Adaptive Mantel Test for High-dimensional Inference

Dustin Pluta

26 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 Motivation

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

Data Description

- **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**.

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

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

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** $\mathcal{K}_X : \mathrm{R}^p \times \mathrm{R}^p \to \mathrm{R}$ and $\mathcal{K}_Y : \mathrm{R} \times \mathrm{R} \to \mathrm{R}$, 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 pairwise distance 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 $u, v \in \mathrm{R}^{\mathrm{p}}$, 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

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

Kernel Mantel Test

Similarity with Weighted Inner Products

For two vectors $u, v \in \mathrm{R}^p$, 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

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

Question

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

How should the weight matrix be chosen?

Weight Matrices

Euclidean Inner Product

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

$$K_E = XX^T,$$

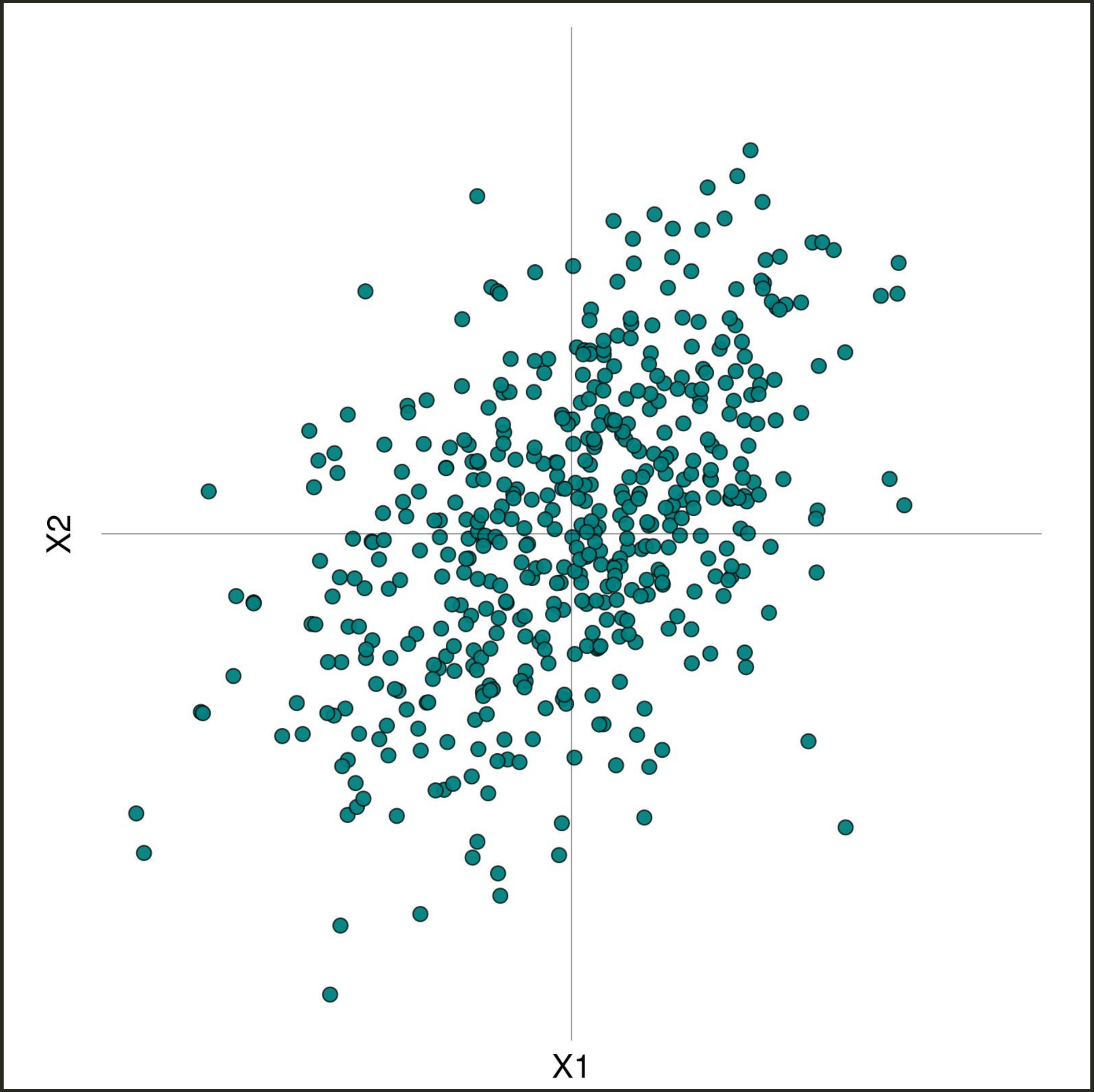
which is the Gram matrix for the standard Euclidean inner product.

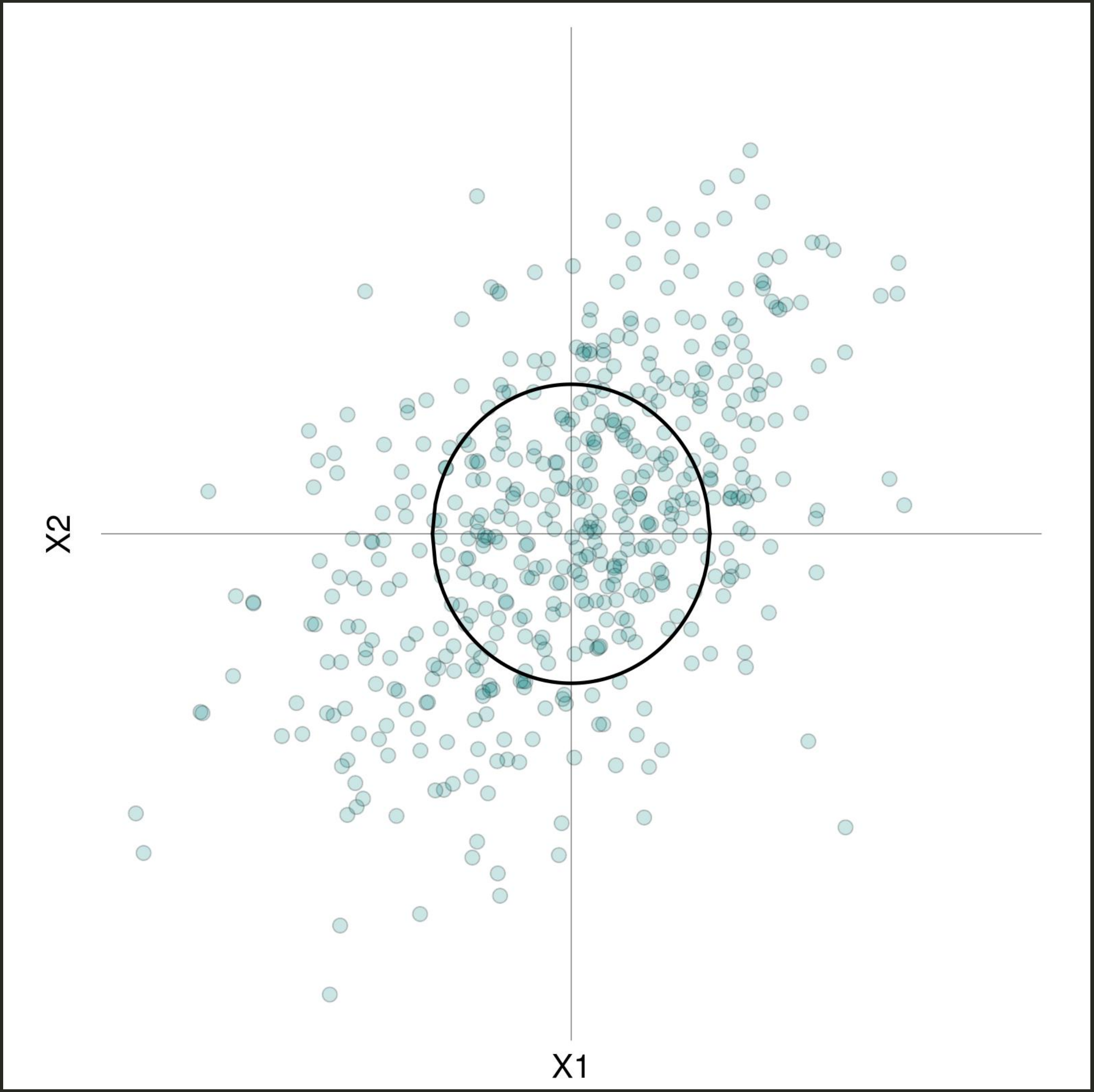
Mahalanobis Similarity

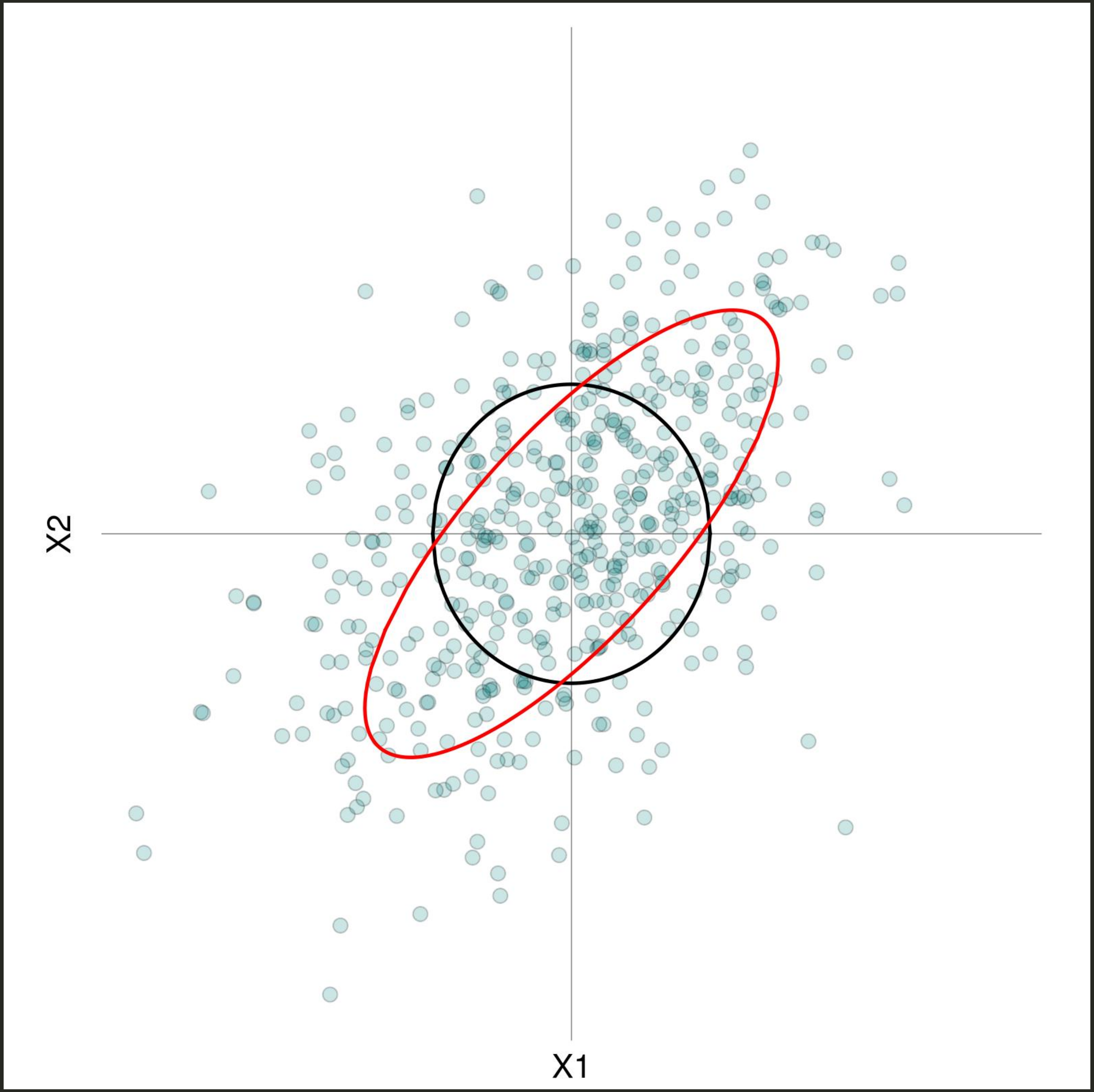
- Choosing $\mathcal{W} = (X^T X)^{-1}$ gives

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

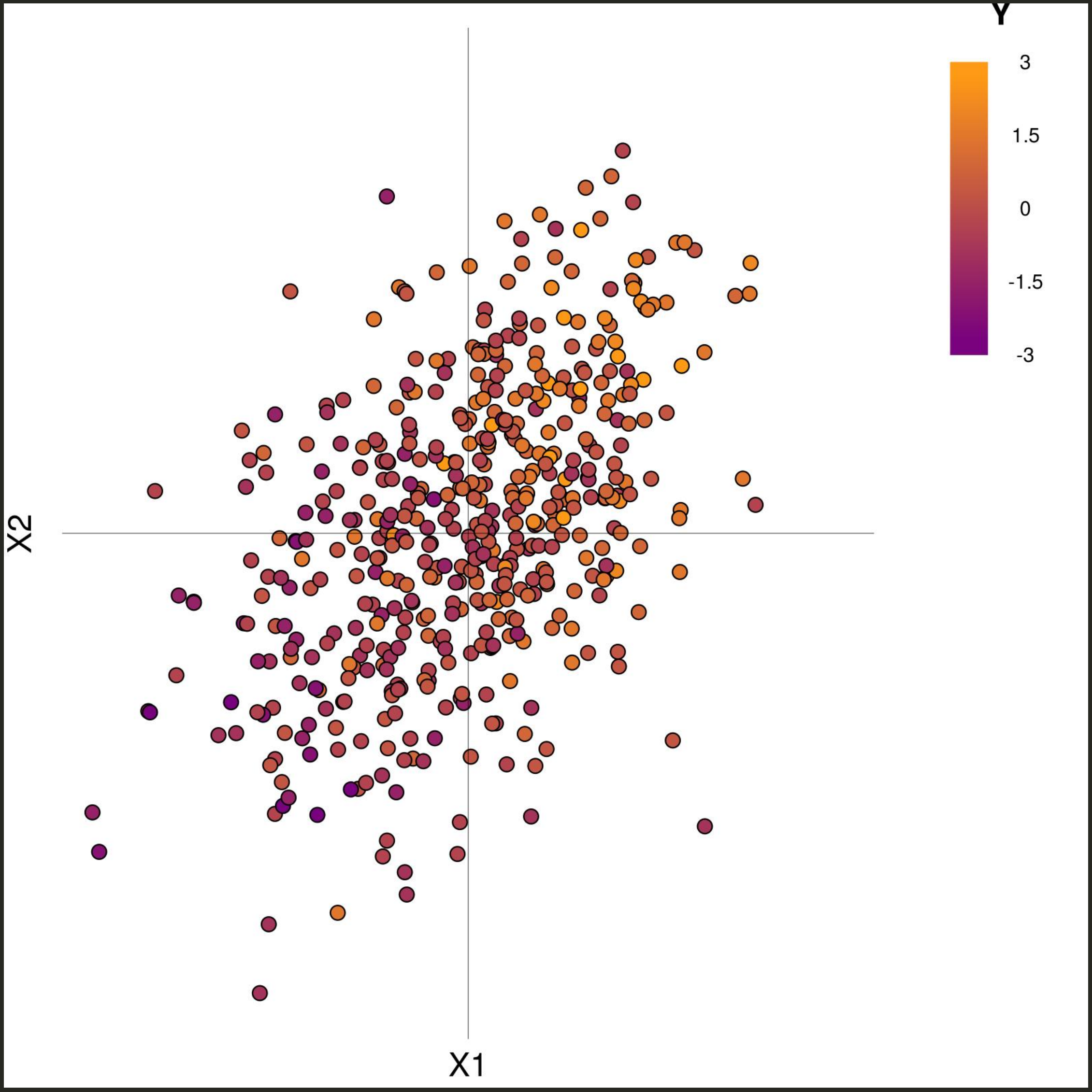
which is a similarity matrix related to the Mahalanobis distance, and is the projection matrix for the column space of X.

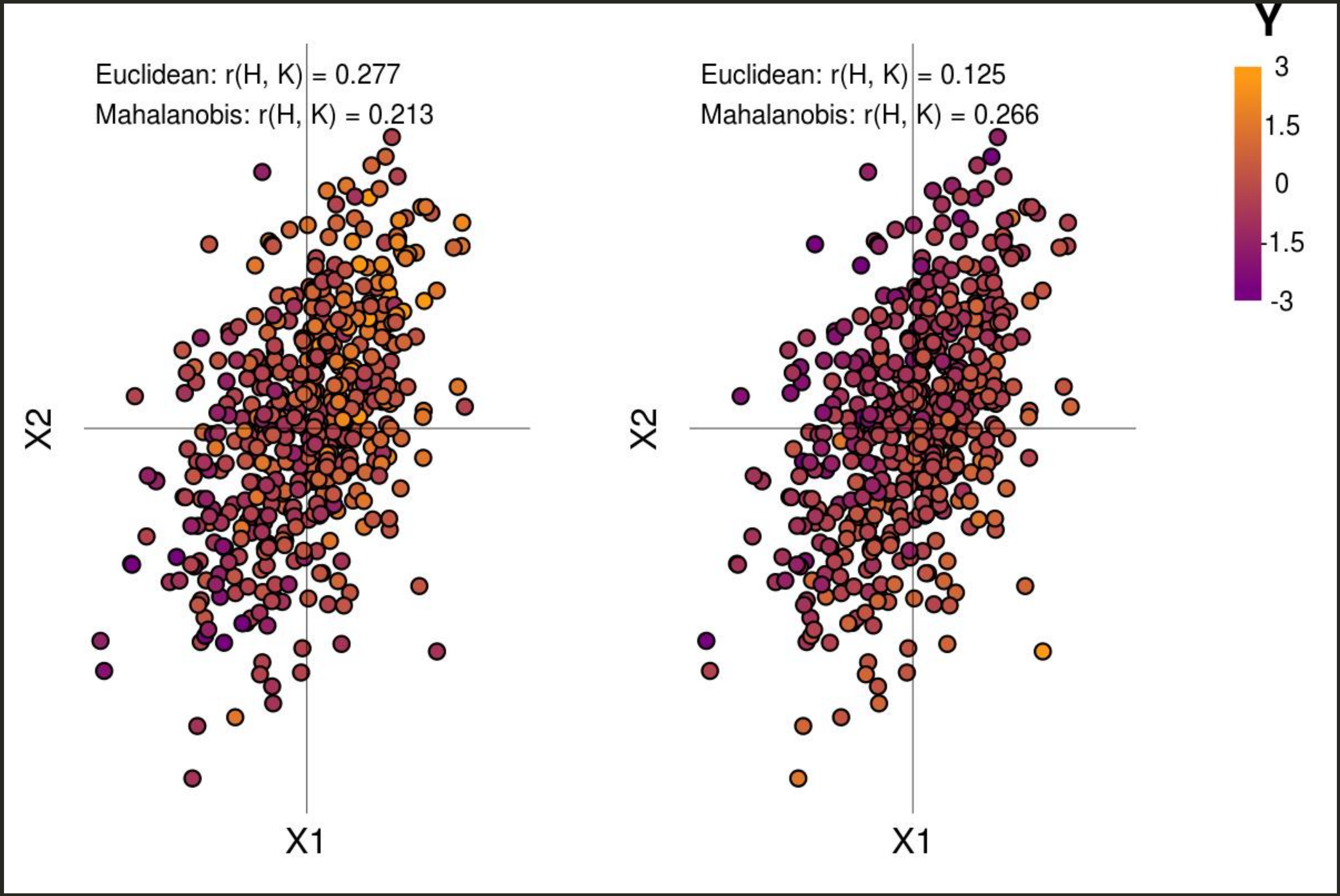

X2
X1


X2
X1


X2
X1

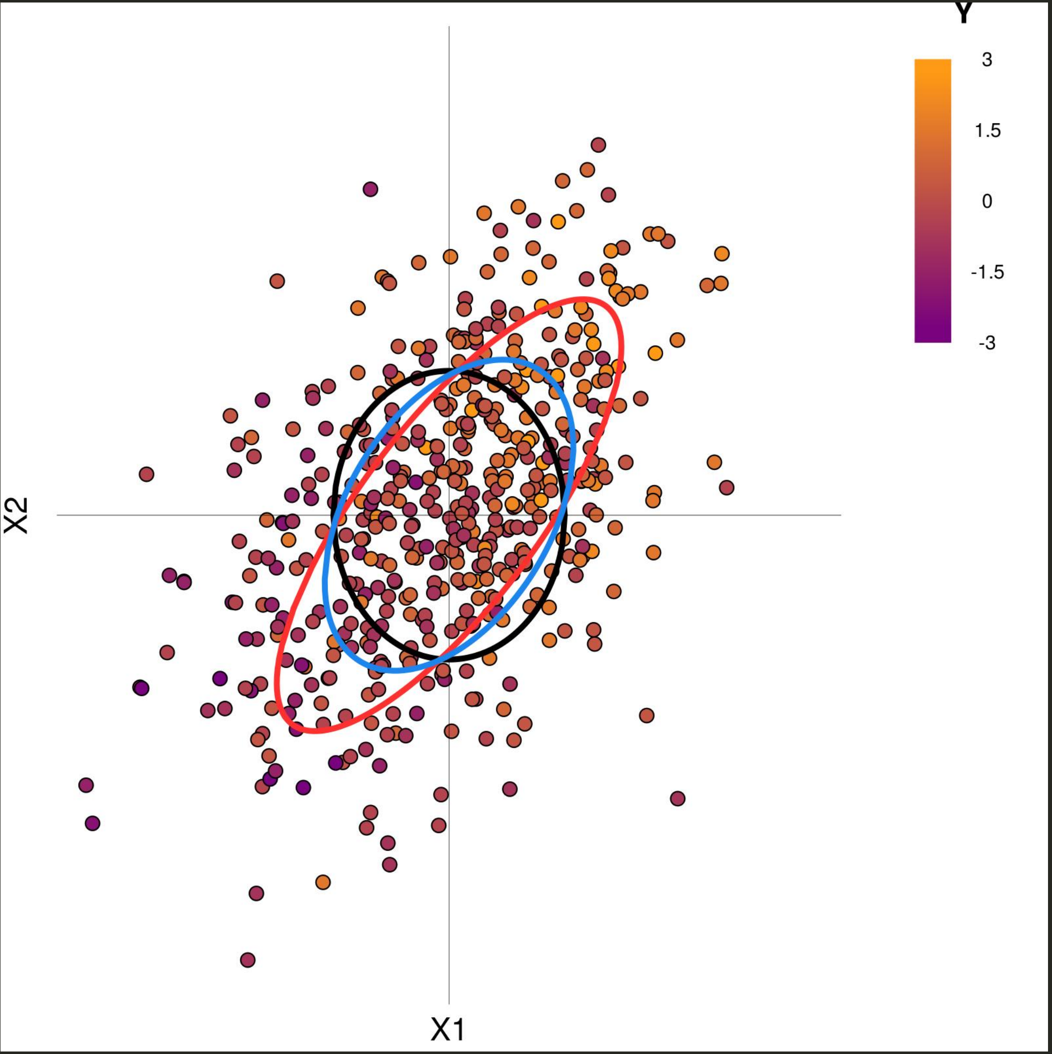
Data generated from variance components model with $\sigma_b^2 = 1$





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

Kernel Mantel Test

Assume $\mathrm{rank}(X) = r$ with singular value decomposition $X = U_{n\times r} D_{r\times r} V^T_{p\times r}$, 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 $$r(H, K_\lambda) = \frac{\sum_{j=1}^r \frac{\eta_j}{\lambda+\eta_j} z_j^2}{\sqrt{\sum_{j=1}^r \left(\frac{\eta_j}{\eta_j+\lambda}\right)^2}\sum_{i=1}^n y_i^2}.$$

This implies limit relationships between the three measures of correlation:

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

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	$Y \sim N(X\beta, \sigma_\varepsilon^2 I_N), \qquad \lVert\beta\rVert_2^2 < c(\lambda)$

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	$S = Z^T D(D^T D + \lambda I_p)^{-1} D^T Z$	$T_\lambda = \mathrm{tr}(K_\lambda H)$	$c_3 \sum_{j=1}^{r} \frac{\eta_j}{\lambda + \eta_j} \chi_1^2$

Linear Model Score Tests

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 $(\lambda + \eta_j)^{-1}$.

Adaptive Mantel Test

Problem

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, and approximate the reference distribution using permutations.

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**

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

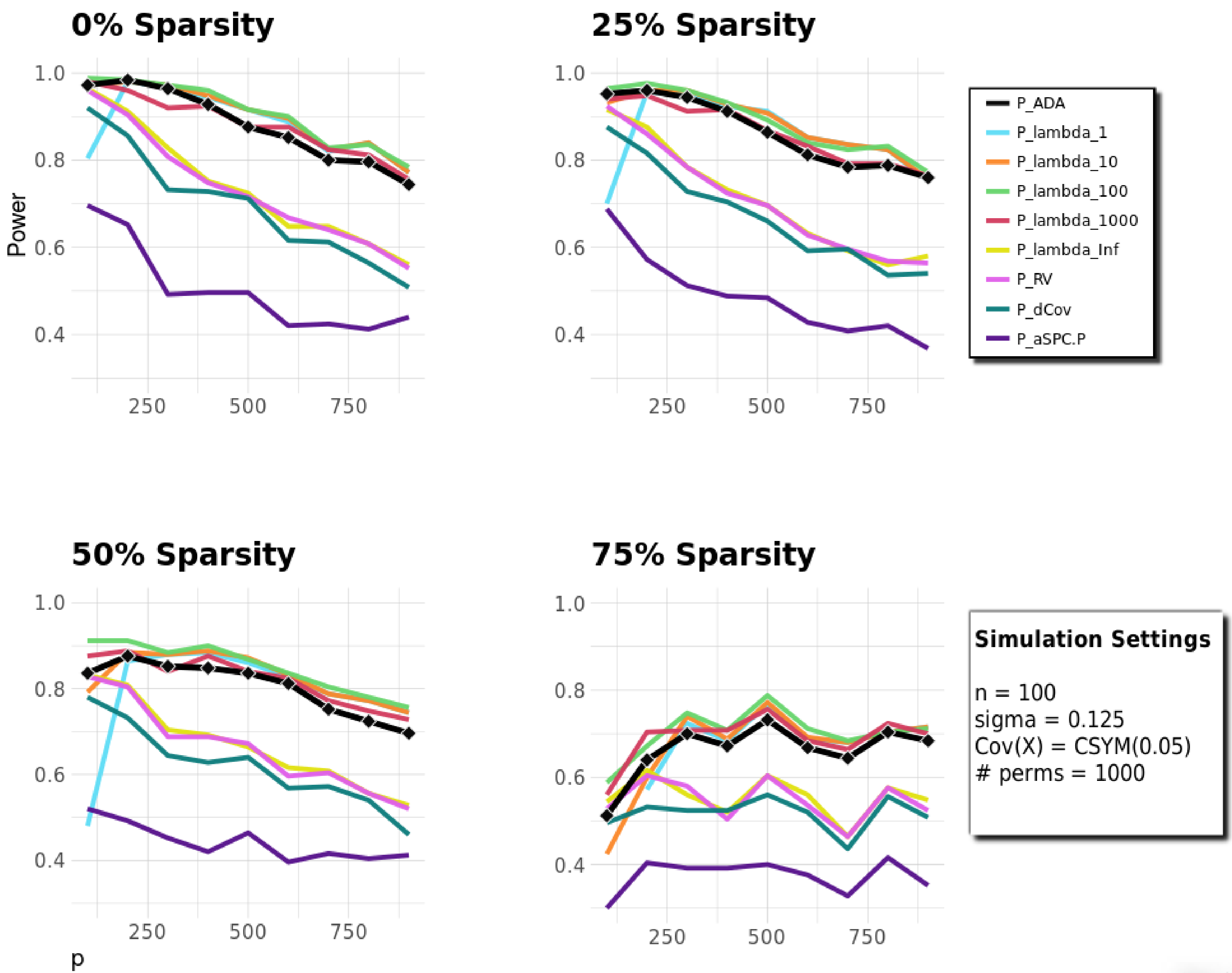
Generate B permutations of $H^{(0)}$, labeled $H^{(b)}, b = 1, \ldots, B$

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

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

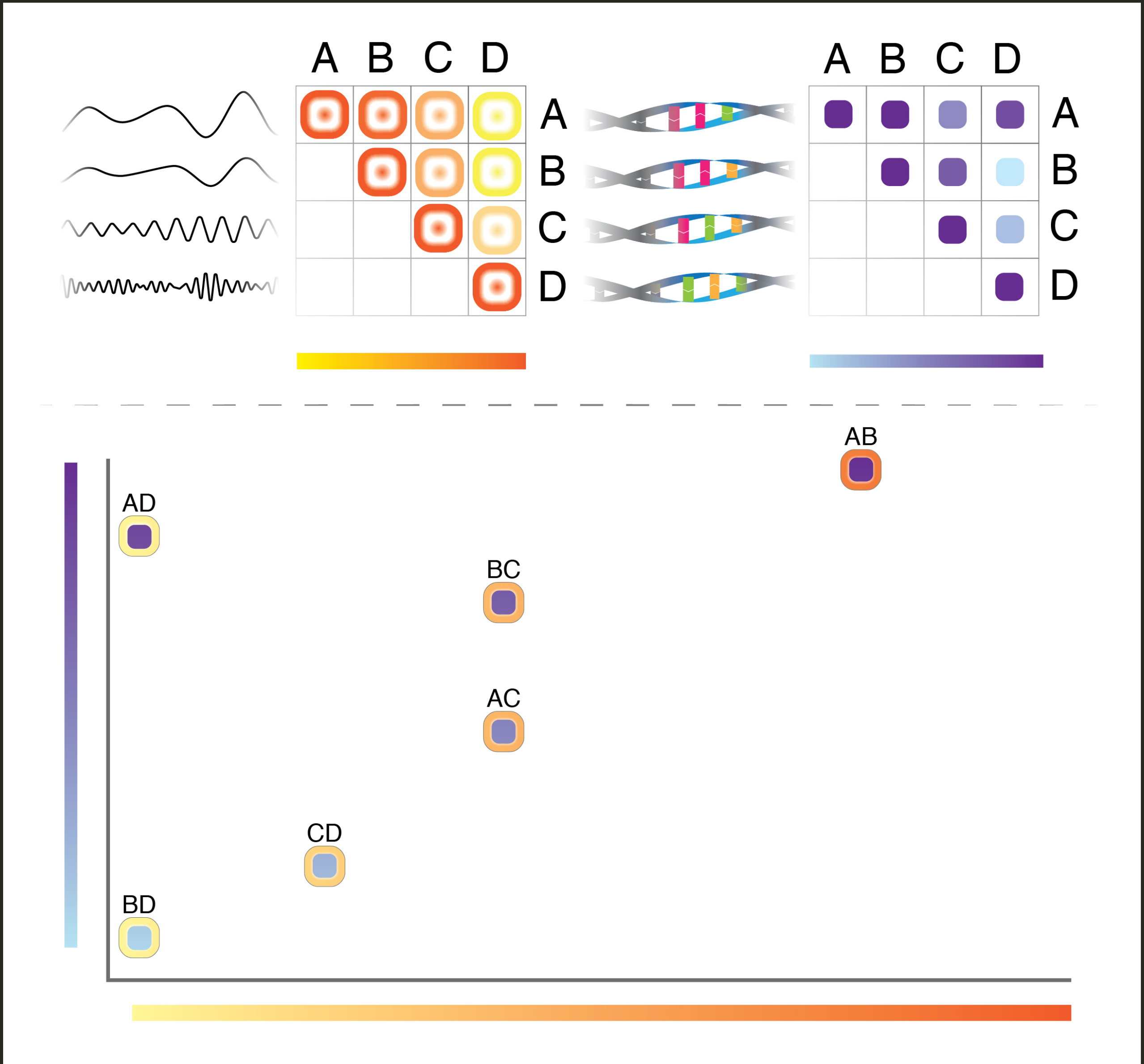
$P^{(b)} \leftarrow \min_j P_j^{(b)}, \quad \forall b = 0, \ldots, B$

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

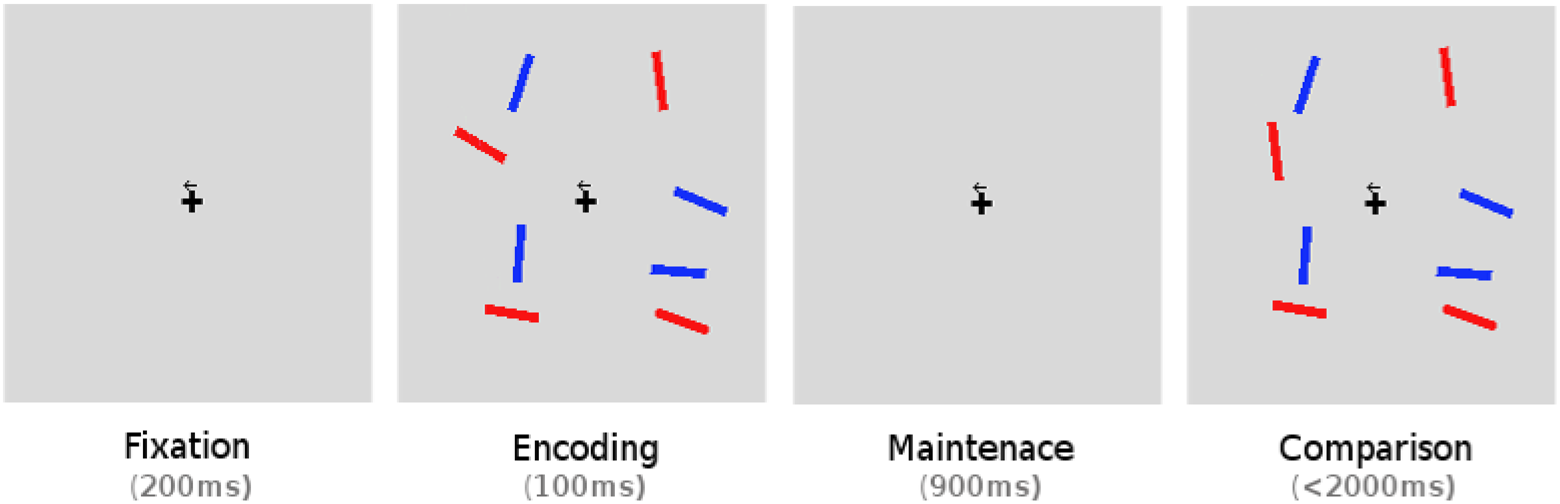

0% Sparsity
25% Sparsity
50% Sparsity
75% Sparsity
Power
p
1.0
0.8
0.6
0.4
250
500
750
P_ADA
P_lambda_1
P_lambda_10
P_lambda_100
P_lambda_1000
P_lambda_Inf
P_RV
P_dCov
P_aSPC.P
Simulation Settings
n = 100
sigma = 0.125
Cov(X) = CSYM(0.05)
# perms = 1000

Application Example

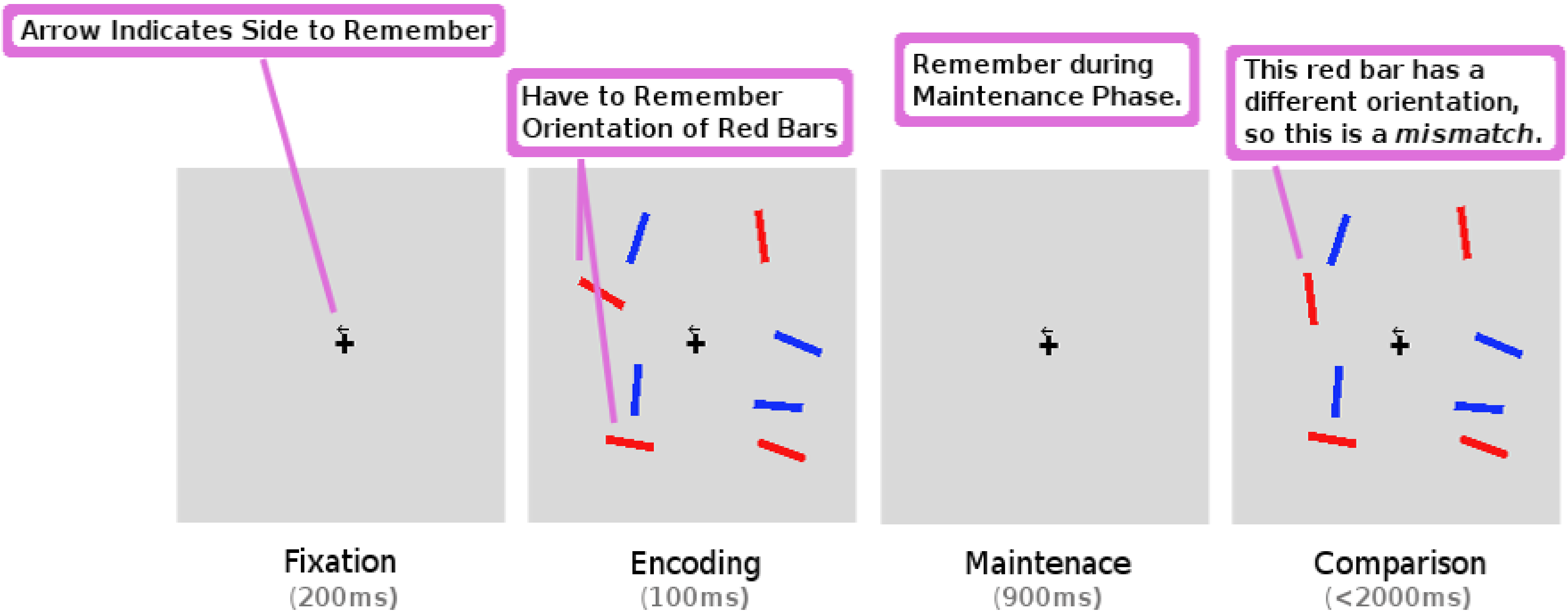
- 350 Subjects from the BNU data set
- ~10 minute 64 channel EEG recording during VWM task
  - Preprocessed according to standard pipeline
  - Coherence measures for each channel pair was calculated by the FFT, and grouped into five frequency bands (in Hz): $\delta\,(1-4)$, $\theta\,(4-8)$, $\alpha\,(8-16)$, $\beta\,(16-32)$, $\gamma\,(32+)$
- 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



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

Links

- **Adaptive Mantel Test Paper:** arxiv.org/pdf/1712.07270.pdf
- **Slides available:** 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.