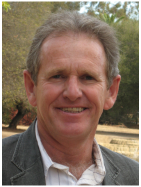
# Canonical Correlation Analysis in high dimensions with structured regularization

Elena Tuzhilina

Stanford University, Department of Statistics

*elenatuz@stanford.edu*



**Trevor Hastie**

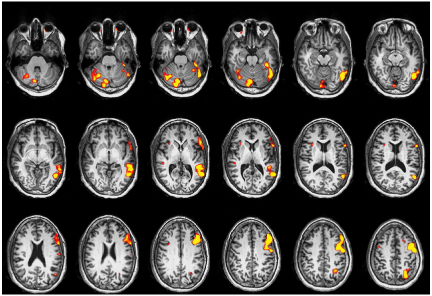
**Leonardo Tozzi**

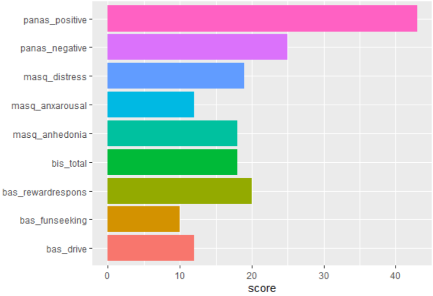
# Data

1. **brain activations**: magnetic resonance imaging obtained during a gambling task designed to probe the brain circuits underlying reward
2. **behavioral performance measures**: self-reports assessing various aspects of reward-related behaviors, depression symptoms and positive as well as negative affective states

# Data

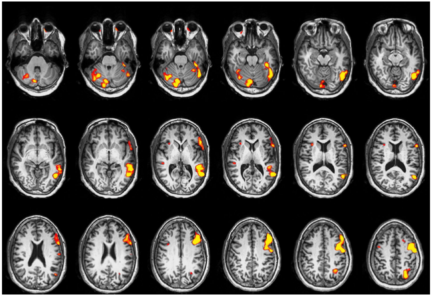
1. **brain activations**: magnetic resonance imaging obtained during a gambling task designed to probe the brain circuits underlying reward
2. **behavioral performance measures**: self-reports assessing various aspects of reward-related behaviors, depression symptoms and positive as well as negative affective states

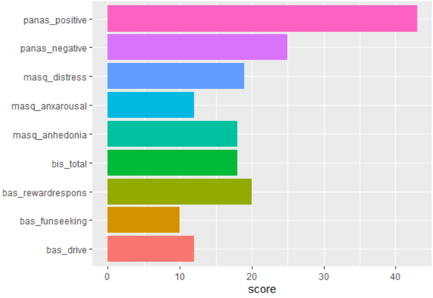




# Data

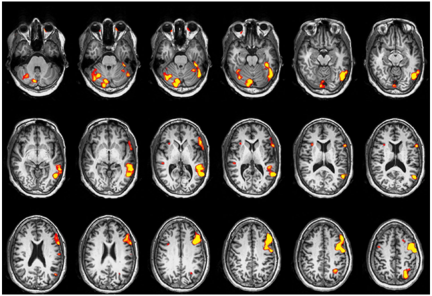
1. **brain activations**: magnetic resonance imaging obtained during a gambling task designed to probe the brain circuits underlying reward
2. **behavioral performance measures**: self-reports assessing various aspects of reward-related behaviors, depression symptoms and positive as well as negative affective states



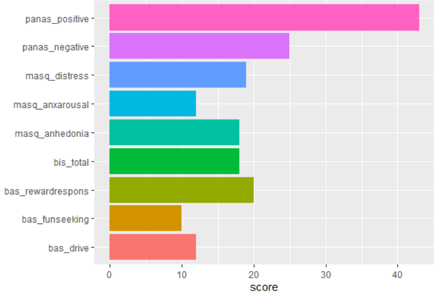



**Question:** is there any correlation between brain activity and behavioral measures of performance during the cognitive tasks?

# Notations



$X \in \mathbb{R}^{n \times p}$ – brain activations




$Y \in \mathbb{R}^{n \times q}$ – behavior test scores

**Dimensions**:

- $n = 153$ participants
- $p = 90,368$ greyordinates
- $q = 9$ scores

# Canonical Correlation Analysis

**Goal**: given two random vectors $x = (x_1, \ldots, x_p)$ and $y = (y_1, \ldots, y_q)$

$$\text{maximize } \operatorname{cor}(\alpha^T x, \beta^T y) \text{ w.r.t. } \alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^q$$

# Canonical Correlation Analysis

**Goal**: given two random vectors $x = (x_1, \dots, x_p)$ and $y = (y_1, \dots, y_q)$

$$\text{maximize } \operatorname{cor}(\alpha^T x, \beta^T y) \text{ w.r.t. } \alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^q$$

- canonical coefficients $\alpha$ and $\beta$
- canonical variates $u = \alpha^T x$ and $v = \beta^T y$
- canonical correlation $\rho(\alpha, \beta) = \operatorname{cor}(u, v)$

# Canonical Correlation Analysis

**Goal**: given two random vectors $x = (x_1, \dots, x_p)$ and $y = (y_1, \dots, y_q)$

$$\text{maximize } \operatorname{cor}(\alpha^T x, \beta^T y) \text{ w.r.t. } \alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^q$$

- canonical coefficients $\alpha$ and $\beta$
- canonical variates $u = \alpha^T x$ and $v = \beta^T y$
- canonical correlation $\rho(\alpha, \beta) = \operatorname{cor}(u, v)$

Correlation coefficient

$$\rho(\alpha, \beta) = \operatorname{cor}(\alpha^T x, \beta^T y) = \frac{\alpha^T \operatorname{cov}(x, y)\beta}{\sqrt{\alpha^T \operatorname{var}(x)\alpha}\ \sqrt{\beta^T \operatorname{var}(y)\beta}},$$

# Canonical Correlation Analysis

**Goal**: given two random vectors $x = (x_1, \ldots, x_p)$ and $y = (y_1, \ldots, y_q)$

$$\text{maximize } \operatorname{cor}(\alpha^T x, \beta^T y) \text{ w.r.t. } \alpha \in \mathbb{R}^p, \beta \in \mathbb{R}^q$$

- canonical coefficients $\alpha$ and $\beta$
- canonical variates $u = \alpha^T x$ and $v = \beta^T y$
- canonical correlation $\rho(\alpha, \beta) = \operatorname{cor}(u, v)$

Correlation coefficient

$$\rho(\alpha, \beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T \Sigma_{XX} \alpha}\,\sqrt{\beta^T \Sigma_{YY} \beta}}$$

# Correlation Coefficient

Correlation coefficient

$$\rho(\alpha, \beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T \Sigma_{XX} \alpha} \sqrt{\beta^T \Sigma_{YY} \beta}}$$

# Correlation Coefficient

Correlation coefficient

$$\rho(\alpha, \beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T \Sigma_{XX} \alpha} \sqrt{\beta^T \Sigma_{YY} \beta}}$$

**CCA optimization problem**:

$$\text{maximize } \alpha^T \Sigma_{XY} \beta \text{ w.r.t. } \alpha \in \mathbb{R}^p \text{ and } \beta \in \mathbb{R}^q$$
$$\text{s.t. } \alpha^T \Sigma_{XX} \alpha = 1 \text{ and } \beta^T \Sigma_{YY} \beta = 1$$

# Correlation Coefficient

Correlation coefficient

$$\rho(\alpha, \beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T \Sigma_{XX} \alpha} \sqrt{\beta^T \Sigma_{YY} \beta}}$$

**CCA optimization problem**:

$$\text{maximize } \widetilde{\alpha}^T \Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-\frac{1}{2}} \widetilde{\beta} \text{ w.r.t. } \widetilde{\alpha} \in \mathbb{R}^p \text{ and } \widetilde{\beta} \in \mathbb{R}^q$$
$$\text{s.t. } \|\widetilde{\alpha}\| = 1 \text{ and } \|\widetilde{\beta}\| = 1$$

# Correlation Coefficient

Correlation coefficient

$$\rho(\alpha, \beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T \Sigma_{XX} \alpha} \sqrt{\beta^T \Sigma_{YY} \beta}}$$

**CCA optimization problem**:

$$\begin{aligned} &\text{maximize } \widetilde{\alpha}^T \Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-\frac{1}{2}} \widetilde{\beta} \text{ w.r.t. } \widetilde{\alpha} \in \mathbb{R}^p \text{ and } \widetilde{\beta} \in \mathbb{R}^q \\ &\text{s.t. } \|\widetilde{\alpha}\| = 1 \text{ and } \|\widetilde{\beta}\| = 1 \end{aligned}$$

**Solution**: via Singular Value Decomposition of $\Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-\frac{1}{2}}$

# Correlation Coefficient

Correlation coefficient

$$\rho(\alpha, \beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T \Sigma_{XX} \alpha} \sqrt{\beta^T \Sigma_{YY} \beta}}$$

**CCA optimization problem**:

$$\begin{aligned} \text{maximize } & \widetilde{\alpha}^T \Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-\frac{1}{2}} \widetilde{\beta} \text{ w.r.t. } \widetilde{\alpha} \in \mathbb{R}^p \text{ and } \widetilde{\beta} \in \mathbb{R}^q \\ \text{s.t. } & \|\widetilde{\alpha}\| = 1 \text{ and } \|\widetilde{\beta}\| = 1 \end{aligned}$$

**Solution**: via Singular Value Decomposition of $\Sigma_{XX}^{-\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-\frac{1}{2}}$

**Problem**: does not work for $p > n$!

# Regularization

Modified correlation coefficient

$$\rho(\alpha,\beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T(\Sigma_{XX} + \lambda_1 I)\alpha}\ \sqrt{\beta^T \Sigma_{YY} \beta}}$$

**RCCA optimization problem**:

$$\begin{aligned} &\text{maximize } \alpha^T \Sigma_{XY} \beta \text{ w.r.t. } \alpha \in \mathbb{R}^p \text{ and } \beta \in \mathbb{R}^q \\ &\quad \text{s.t. } \alpha^T \Sigma_{XX} \alpha = 1,\ \beta^T \Sigma_{YY} \beta = 1 \text{ and } \|\alpha\| \leq t_1 \end{aligned}$$

**Solution**: via Singular Value Decomposition of $(\Sigma_{XX} + \lambda I)^{-\frac{1}{2}} \Sigma_{XY} \Sigma_{YY}^{-\frac{1}{2}}$

# CCA package

rcc {CCA} R Documentation

## Regularized Canonical Correlation Analysis

### Description

The function performs the Regularized extension of the Canonical Correlation Analysis to seek correlations between two data matrices when the number of columns (variables) exceeds the number of rows (observations)

### Usage

```
rcc(X, Y, lambda1, lambda2)
```

### Arguments

| | |
|---|---|
| `X` | numeric matrix (n * p), containing the X coordinates. |
| `Y` | numeric matrix (n * q), containing the Y coordinates. |
| `lambda1` | Regularization parameter for X |
| `lambda2` | Regularization parameter for Y |

### Details

When the number of columns is greater than the number of rows, the matrice X'X (and/or Y'Y) may be ill-conditioned. The regularization allows the inversion by adding a term on the diagonal.

### Value

A list containing the following components:

| | |
|---|---|
| `corr` | canonical correlations |
| `names` | a list containing the names to be used for individuals and variables for graphical outputs |
| `xcoef` | estimated coefficients for the 'X' variables as returned by `cancor()` |
| `ycoef` | estimated coefficients for the 'Y' variables as returned by `cancor()` |
| `scores` | a list returned by the internal function comput() containing individuals and variables coordinates on the canonical variates basis. |

### Author(s)

Sébastien Déjean, Ignacio González

# CCA package

```
library(CCA)
rcc(X = activation, Y = behavior, lambda1 = 10, lambda2 = 0)
```

```
Error: cannot allocate vector of size 62.1 Gb
Traceback:

1. rcc(X = activation, Y = behavior, lambda1 = 10, lambda2 = 0)
2. var(X, na.rm = TRUE, use = "pairwise")
```

# CCA package

```
library(CCA)
rcc(X = activation, Y = behavior, lambda1 = 10, lambda2 = 0)
```

```
Error: cannot allocate vector of size 62.1 Gb
Traceback:

1. rcc(X = activation, Y = behavior, lambda1 = 10, lambda2 = 0)
2. var(X, na.rm = TRUE, use = "pairwise")
```

```
"rcc" <-
function (X, Y, lambda1, lambda2)
{
    Xnames <- dimnames(X)[[2]]
    Ynames <- dimnames(Y)[[2]]
    ind.names <- dimnames(X)[[1]]
    Cxx <- var(X, na.rm = TRUE, use = "pairwise") + diag(lambda1,
        ncol(X))
    Cyy <- var(Y, na.rm = TRUE, use = "pairwise") + diag(lambda2,
        ncol(Y))
    Cxy <- cov(X, Y, use = "pairwise")
    res <- geigen(Cxy, Cxx, Cyy)
    names(res) <- c("cor", "xcoef", "ycoef")
    scores <- comput(X, Y, res)
    return(list(cor = res$cor, names = list(Xnames = Xnames,
        Ynames = Ynames, ind.names = ind.names), xcoef = res$xcoef,
        ycoef = res$ycoef, scores = scores))
}
```

# CCA package

```r
library(CCA)
rcc(X = activation, Y = behavior, lambda1 = 10, lambda2 = 0)
```

```
Error: cannot allocate vector of size 62.1 Gb
Traceback:

1. rcc(X = activation, Y = behavior, lambda1 = 10, lambda2 = 0)
2. var(X, na.rm = TRUE, use = "pairwise")
```

```r
"rcc" <-
function (X, Y, lambda1, lambda2)
{
    Xnames <- dimnames(X)[[2]]
    Ynames <- dimnames(Y)[[2]]
    ind.names <- dimnames(X)[[1]]
    Cxx <- var(X, na.rm = TRUE, use = "pairwise") + diag(lambda1,
        ncol(X))
    Cyy <- var(Y, na.rm = TRUE, use = "pairwise") + diag(lambda2,
        ncol(Y))
    Cxy <- cov(X, Y, use = "pairwise")
    res <- geigen(Cxy, Cxx, Cyy)
    names(res) <- c("cor", "xcoef", "ycoef")
    scores <- comput(X, Y, res)
    return(list(cor = res$cor, names = list(Xnames = Xnames,
        Ynames = Ynames, ind.names = ind.names), xcoef = res$xcoef,
        ycoef = res$ycoef, scores = scores))
}
```

$$C_{XX} = \boxed{p \times p}$$

$$C_{YY} = \boxed{q \times q}$$

$$C_{XY} = \boxed{p \times q}$$

**Problem**: $C_{XX}$, $C_{XY}$ are large for $p \gg n$

# Kernel trick

**Goal**: find a linear transformation such that RCCA for $(X, Y)$ is equivalent to RCCA for $(R, Y)$ and

$$V = \boxed{p \times n} \qquad\qquad R = XV = \boxed{\quad n \times p \quad}\ \boxed{p \times n} = \boxed{n \times n}$$

# Kernel trick

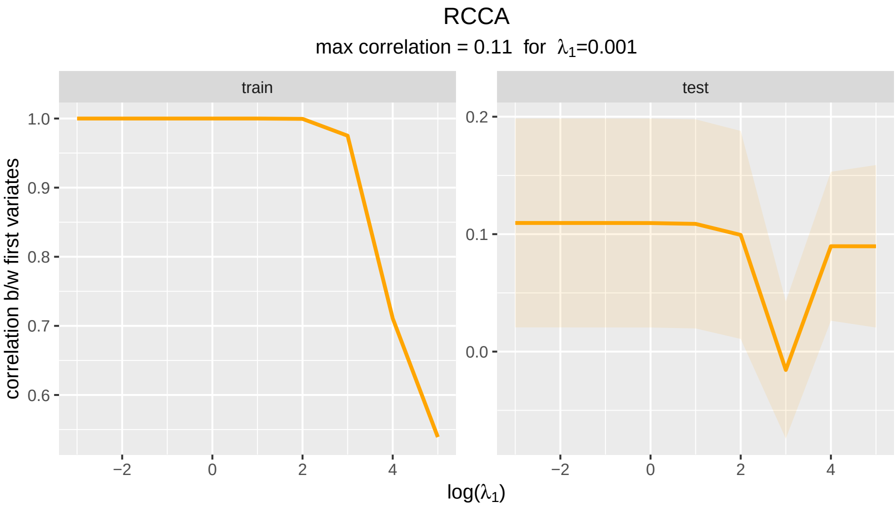
**Goal**: find a linear transformation such that RCCA for $(X, Y)$ is equivalent to RCCA for $(R, Y)$ and

$$V = \boxed{p \times n} \qquad R = XV = \boxed{n \times p}\ \boxed{p \times n} = \boxed{n \times n}$$

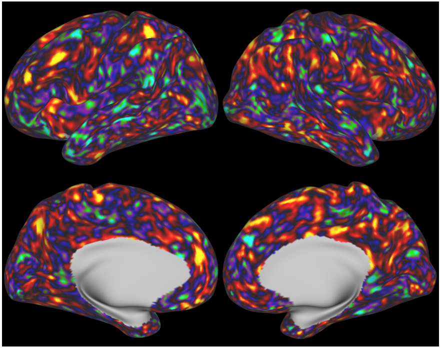
**Solution**:

1. $X = UDV^T = \boxed{n \times n}\ \boxed{n \times n}\ \boxed{n \times p}$
2. set $R = XV = UD$ and solve RCCA problem for $(R, Y) \Longrightarrow$ canonical coefficients $\alpha_R, \beta_R$
3. apply inverse transformation $\alpha_X = V\alpha_R$ and $\beta_X = \beta_R$
4. the variates stay the same $v_R = R\alpha_R = X\alpha_X = v_X$ and $u_R = u_X$

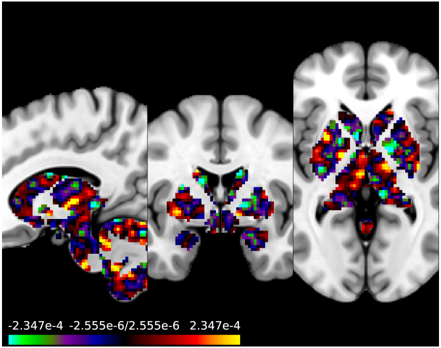
# Brain data: RCCA best model



**Visualization**: plot canonical coefficients $\alpha$ for the optimal RCCA model with $\lambda_1 = 0.001$



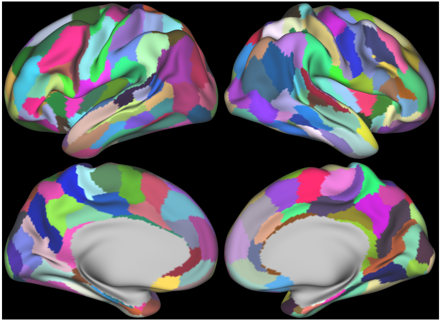
(a) Cortical coefficients.



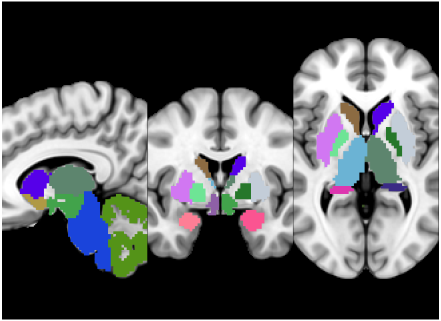

(b) Subcortical coefficients.

# Brain regions

**Motivation**: brain features come in groups (aka brain regions). How to take into account the group structure?



(a) Cortical parcellation (210 regions).



(b) Subcortical parcellation (19 regions).

# Grouped structure

**Notations**:

- $K = 229$ groups
- $p_k = \#$ features in group $k$
- $X_k$ – set of features in group $k$
- $\alpha_k$ – set of coefficients in group $k$

$$\alpha = (\underbrace{\alpha_1}_{p_1}, \ldots, \underbrace{\alpha_K}_{p_K}) \text{ and } X = (\underbrace{X_1}_{p_1}, \ldots, \underbrace{X_K}_{p_K})$$

# Grouped structure

**Notations**:

- $K = 229$ groups
- $p_k = \#$ features in group $k$
- $X_k$ – set of features in group $k$
- $\alpha_k$ – set of coefficients in group $k$

**Assumptions**:

1. group homogeneity
$$\alpha_k \approx \bar{\alpha}_k$$
2. sparsity on a group level
$$\bar{\alpha}_k \approx 0$$

$$\alpha = (\underbrace{\alpha_1}_{p_1}, \dots, \underbrace{\alpha_K}_{p_K}) \text{ and } X = (\underbrace{X_1}_{p_1}, \dots, \underbrace{X_K}_{p_K})$$

# Grouped structure

**Notations**:

- $K = 229$ groups
- $p_k = \#$ features in group $k$
- $X_k$ – set of features in group $k$
- $\alpha_k$ – set of coefficients in group $k$

**Assumptions**:

1. group homogeneity
$$\alpha_k \approx \bar{\alpha}_k$$
2. sparsity on a group level
$$\bar{\alpha}_k \approx 0$$

$$\alpha = (\underbrace{\alpha_1}_{p_1}, \ldots, \underbrace{\alpha_K}_{p_K}) \text{ and } X = (\underbrace{X_1}_{p_1}, \ldots, \underbrace{X_K}_{p_K})$$

**GRCCA optimization problem**:

$$\text{maximize } \alpha^T \Sigma_{XY} \beta \text{ w.r.t. } \alpha \in \mathbb{R}^p \text{ and } \beta \in \mathbb{R}^q$$

$$\text{s.t } \alpha^T \Sigma_{XX} \alpha = 1, \; \beta^T \Sigma_{YY} \beta = 1,$$

$$\sum_{k=1}^{K} \|\alpha_k - \bar{\alpha}_k\|^2 \leq t_1 \textit{ and } \sum_{k=1}^{K} p_k \bar{\alpha}_k^2 \leq s_1$$

# GRCCA

**GRCCA optimization problem**:

$$\text{maximize } \alpha^T \Sigma_{XY} \beta \text{ w.r.t. } \alpha \in \mathbb{R}^p \text{ and } \beta \in \mathbb{R}^q$$
$$\text{s.t } \alpha^T \Sigma_{XX} \alpha = 1,\ \beta^T \Sigma_{YY} \beta = 1,$$
$$\sum_{k=1}^{K} \|\alpha_k - \bar{\alpha}_k\|^2 \leq t_1 \textit{ and } \sum_{k=1}^{K} p_k \bar{\alpha}_k^2 \leq s_1$$

Modified correlation coefficient

$$\rho(\alpha, \beta) = \frac{\alpha^T \Sigma_{XY} \beta}{\sqrt{\alpha^T (\Sigma_{XX} + \lambda_1 (I - C) + \mu_1 C) \alpha} \sqrt{\beta^T \Sigma_{YY} \beta}}$$

$$C = \begin{bmatrix} \frac{11^T}{p_1} & 0 & \dots & 0 \\ 0 & \frac{11^T}{p_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{11^T}{p_K} \end{bmatrix}$$

# GRCCA vs. RCCA

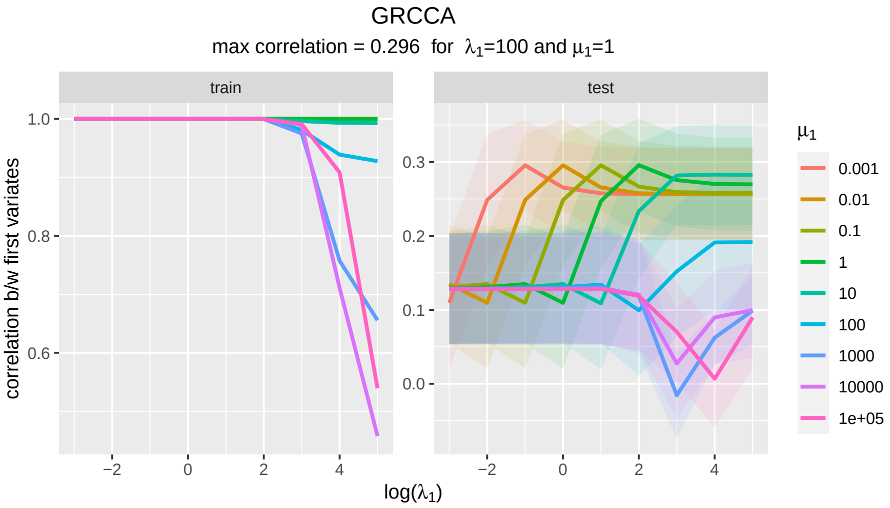
**Lemma**

GRCCA for $(X, Y)$ is equivalent to RCCA for $(\widetilde{X}, Y)$ where

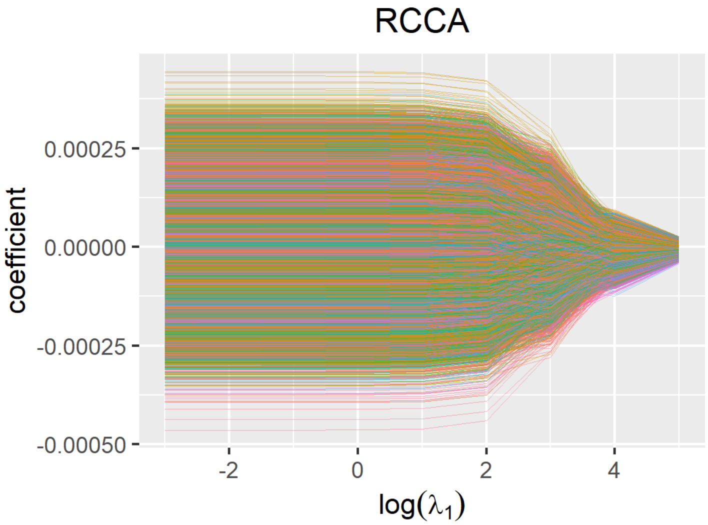
$$\widetilde{X} = \left(\mathrm{X}_1 - \bar{\mathrm{X}}_1, \sqrt{\frac{p_1\lambda_1}{\mu_1}}\bar{\mathrm{X}}_1, \ldots, \mathrm{X}_K - \bar{\mathrm{X}}_K, \sqrt{\frac{p_K\lambda_1}{\mu_1}}\bar{\mathrm{X}}_K\right)$$

Can use Kernel trick!

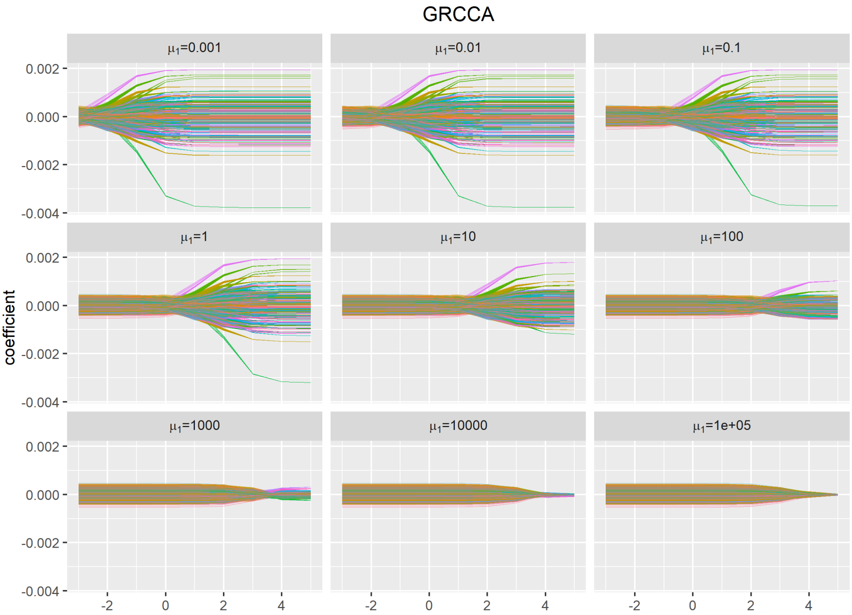
# Brain data: GRCCA best model



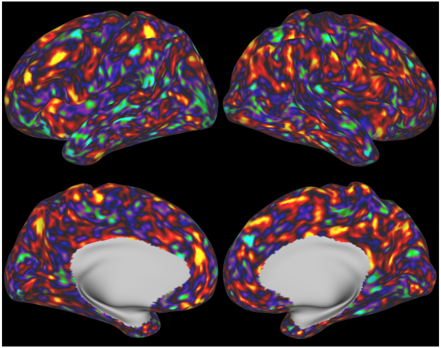
# Brain data: coefficient paths



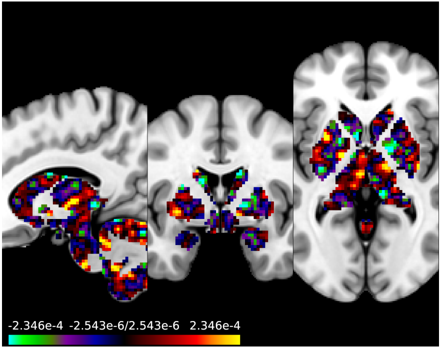
# Brain data: coefficient paths



**Visualization**: plot canonical coefficients $\alpha$ for GRCCA model with $\lambda_1 = 1$ and $\mu_1 = 1$



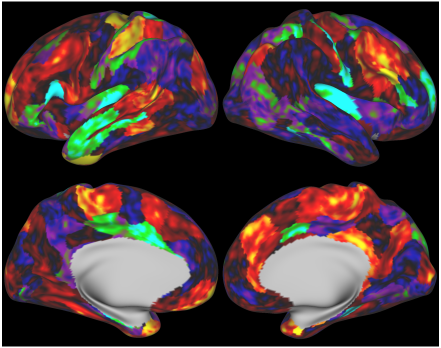
(a) Cortical coefficients.



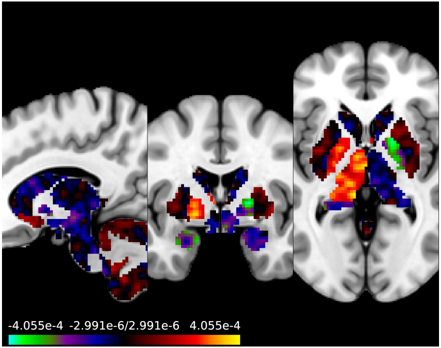

(b) Subcortical coefficients.

# Brain data: improved interpretability

**Visualization**: plot canonical coefficients $\alpha$ for GRCCA model with $\lambda_1 = 10$ and $\mu_1 = 1$



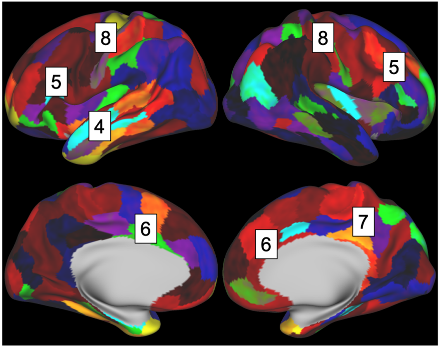
(a) Cortical coefficients.



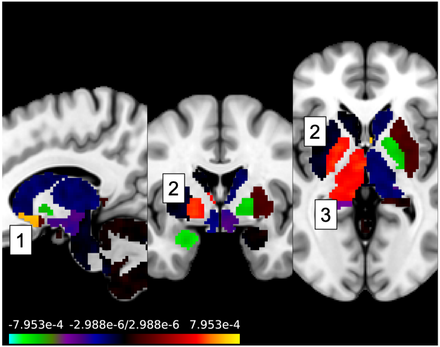

(b) Subcortical coefficients.

# Brain data: improved interpretability

**Visualization**: plot canonical coefficients $\alpha$ for GRCCA model with optimal $\lambda_1 = 100$ and $\mu_1 = 1$




(a) Cortical coefficients.




(b) Subcortical coefficients.

**Annotation of brain regions**: [1] nucleus accumbens, [2] putamen, [3] thalamus, [4] temporal lobe, [5] dorsolateral prefrontal cortex, [6] dorsomedial prefrontal cortex, [7] posterior cingulate cortex, [8] precentral cortex.

# References


D. M. Barch, G. C. Burgess, M. P. Harms, S. E. Petersen, B. L. Schlaggar, M. Corbetta, M. F. Glasser, S. Curtiss, S. Dixit, C. Feldt, D. Nolan, E. Bryant, T. Hartley, O. Footer, J. M. Bjork, R. Poldrack, S. Smith, H. Johansen-Berg, A. Z. Snyder, and D. C. Van Essen, *Function in the human connectome: Task-fMRI and individual differences in behavior*, NeuroImage, 80 (2013), pp. 169–189.

D. C. V. Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, and W.-M. H. Consortium, *The WU-Minn Human Connectome Project: an overview*, Neuroimage, 80 (2013), pp. 62–79.

M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson, S. M. Smith, and D. C. V. Essen, *A multi-modal parcellation of human cerebral cortex*, Nature, 536 (2016), pp. 171–178.

I. Gonzalez, D. S, P. Martin, and A. Baccini, *CCA: An R Package to Extend Canonical Correlation Analysis*, Journal of Statistical Software, 23(12) (2008), pp. 1–14.

H. Hotelling, *Relations between two sets of variables*, Biometrika, 28 (1936), pp. 321—-377.

W. Härdle and L. Simar, *Applied Multivariate Statistical Analysis: Second edition*, Springer, 2007.

S. Leurgans, R. Moyeed, and B. Silverman, *Canonical Correlation Analysis when the Data are Curves*, Journal of the Royal Statistical Society, 55(3) (1993), pp. 725—-740.

K. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, New York, NY, 1979.

H. Vinod, *Canonical Ridge and Econometrics of Joint Production*, Journal of Econometrics, 4(2) (1976), pp. 147—-166.

S. Weintraub, S. S. Dikmen, R. K. Heaton, D. S. Tulsky, P. D. Zelazo, P. J. Bauer, N. E. Carlozzi, J. Slotkin, D. Blitz, K. Wallner-Allen, N. A. Fox, J. L. Beaumont, D. Mungas, C. J. Nowinski, J. Richler, J. A. Deocampo, J. E. Anderson, J. J. Manly, B. Borosh, R. Havlik, K. Conway, E. Edwards, L. Freund, J. W. King, C. Moy, E. Witt, and R. C. Gershon, *Cognition assessment using the NIH Toolbox*, Neurology, 80 (2013), pp. S54–64.

Thank you for your attention!