# Statistical learning: Advanced

Advanced

Alain Celisse

PCA PLS

CCA

### Alain Celisse

SAMM

Paris 1-Panthéon Sorbonne University

alain.celisse@univ-paris1.fr

Lecture 1: Components-based methods

\_

Master 2 TIDE - Paris 1 - Fall 2021

PCA

### Successive topics of the coming lectures:

- 1. Component-based analysis: PCA, PLS, CCA,...
- Beyond quantitative data: Categorical data Correpondence Analysis (CA), MCA, FAMD
- Clustering from dissimilarity measures: Similarity, K-means and kernelized K-means, Hierarchical clustering
- 4. Change-point detection
  Parametric models and non-parametric ones
- 5. Multiple testing and FDR control

# Outline of the lecture

Statistical learning: Advanced

Alain Celisse

PCA

PLS CCA

► PCA: Principal Component Analysis

► PLS: Partial Least-Squares

► CCA: Canonical Correlation Analysis

Statistical learning: Advanced

Alain Celisse

#### PCA

Unsupervised PCA algorithm

PLS

CCA

# Principal Component Analysis (PCA)

#### Notation

- n individuals
- ► Each individual i described by d (quantitative) variables  $X_i = (X_i^1, \dots, X_i^d)^\top \in \mathbb{R}^d$ : features for individual i
- **X**:  $n \times d$  "full" design matrix
- $X_{.,j} \in \mathbb{R}^n$ : jth column vector of X (n measurements for variable  $X^j$ )
- $lackbrack X_{i\cdot} \in \mathbb{R}^d$ : *i*th row vector of  $\boldsymbol{X}$

CCA

#### Notation

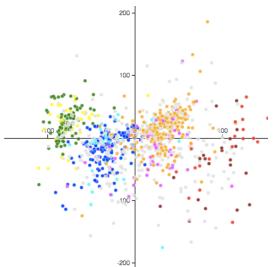
- n individuals
- ► Each individual i described by d (quantitative) variables  $X_i = (X_i^1, \dots, X_i^d)^\top \in \mathbb{R}^d$ : features for individual i
- **X**:  $n \times d$  "full" design matrix
- ▶  $X_{.,j} \in \mathbb{R}^n$ : jth column vector of X (n measurements for variable  $X^j$ )
- $X_i \in \mathbb{R}^d$ : ith row vector of X

There is no label Y in the PCA story!

# Assumption for having PCA work well

### Key assumption

Variability carries information about heterogeneity



Statistical learning: Advanced

Alain Celisse

PCA

Unsupervised

PCA algorithm

PLS CCA

# PCA algorithm

$$\widehat{\operatorname{Var}}(X^{\top}\theta) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top}\theta = \frac{1}{n} (\boldsymbol{X}\theta)^{\top} (\boldsymbol{X}\theta) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle$$

# PCA algorithm

- ► Step 1:
  - Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle\}$$

Statistical learning: Advanced

Alain Celisse

PCA

Unsupervised

PCA algorithm

- $\widehat{\operatorname{Var}}(\boldsymbol{X}^{\top}\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top}\boldsymbol{\theta} = \frac{1}{n} \left( \boldsymbol{X} \boldsymbol{\theta} \right)^{\top} \left( \boldsymbol{X} \boldsymbol{\theta} \right) = \left\langle \boldsymbol{X} \boldsymbol{\theta}, \boldsymbol{X} \boldsymbol{\theta} \right\rangle$
- PCA algorithm
  - ► Step 1:
    - Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle \right\}$$

 $m{h}^1 = \widetilde{m{X}}_{\cdot 1} = m{X} heta^1 \in \mathbb{R}^n$ : first principal component

 $\widehat{\operatorname{Var}}(X^{\top}\theta) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top}\theta = \frac{1}{n} (\boldsymbol{X}\theta)^{\top} (\boldsymbol{X}\theta) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle$ 

# PCA algorithm

- ► Step 1:
  - Find

$$heta^1 \in Arg\max_{\| heta\|=1}\left\{\left\langle oldsymbol{X} heta,oldsymbol{X} heta
ight
angle
ight\}$$

- $h^1 = \widetilde{X}_{\cdot 1} = X\theta^1 \in \mathbb{R}^n$ : first principal component
- ▶ Step t  $(2 \le t \le rank(X) = r)$ :
  - Find

$$\begin{array}{ll} \boldsymbol{\theta}^t \in \textit{Arg} & \max \\ \|\boldsymbol{\theta}\| = 1 \\ \left\langle \boldsymbol{X}\boldsymbol{\theta}, \boldsymbol{h}^j \right\rangle = 0 \\ j = 1, \dots, t-1 \end{array}$$

 $\widehat{\operatorname{Var}}(X^{\top}\theta) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top}\theta = \frac{1}{n} (\boldsymbol{X}\theta)^{\top} (\boldsymbol{X}\theta) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle$ 

# PCA algorithm

- ► Step 1:
  - Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \boldsymbol{X} \theta, \boldsymbol{X} \theta \rangle \right\}$$

- $h^1 = X_{\cdot 1} = X\theta^1 \in \mathbb{R}^n$ : first principal component
- ▶ Step t  $(2 \le t \le rank(X) = r)$ :
  - Find

$$egin{aligned} heta^t \in \mathit{Arg} & \max \ \| heta\| = 1 \ \left\langle oldsymbol{X} heta, oldsymbol{h}^j 
ight
angle = 0 \ j = 1, \ldots, t-1 \end{aligned}$$

 $lackbox{h}^t = \widetilde{m{X}}_{\cdot t} = m{X} \theta^t \in \mathbb{R}^n$ : tth principal component

# PCA algorithm

Alain Celisse

PCA

Unsupervised

PCA algorithm

PLS CCA

 $\widehat{\operatorname{Var}}(X^{\top}\theta) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top}\theta = \frac{1}{n} (\boldsymbol{X}\theta)^{\top} (\boldsymbol{X}\theta) = \langle \boldsymbol{X}\theta, \boldsymbol{X}\theta \rangle$ 

# PCA algorithm ► Step 1:

► Step 1: ► Find

$$heta^1 \in Arg\max_{\| heta\|=1}\left\{\left\langle oldsymbol{X} heta,oldsymbol{X} heta
ight
angle
ight\}$$

- $h^1 = \tilde{\mathbf{X}}_{\cdot 1} = \mathbf{X}\theta^1 \in \mathbb{R}^n$ : first principal component
- ▶ Step t  $(2 \le t \le \operatorname{rank}(X) = r)$ :
  - Find

$$\begin{array}{ll} \boldsymbol{\theta}^t \in \textit{Arg} & \max \\ \|\boldsymbol{\theta}\| = 1 \\ \left\langle \boldsymbol{X}\boldsymbol{\theta}, h^j \right\rangle = 0 \\ j = 1, \dots, t-1 \end{array} \left. \left\{ \left\langle \boldsymbol{X}\boldsymbol{\theta}, \boldsymbol{X}\boldsymbol{\theta} \right\rangle \right\} \right. \\ \end{array}$$

- $lackbrack h^t = oldsymbol{X}_{\cdot t} = oldsymbol{X} heta^t \in \mathbb{R}^n$ : tth principal component
- Output:

$$m{X} pprox \widetilde{m{X}} = \left[ h^1; h^2; \dots; h^r \right] = m{X} \left[ \theta^1; \dots; \theta^r \right]$$

# Step 1

Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \mathbf{X}\theta, \mathbf{X}\theta \rangle\}$$

 $h^1 = X_{\cdot,1} = X\theta^1 \in \mathbb{R}^n$ : first principal component

### Interpretation

- $\langle \mathbf{X}\theta, \mathbf{X}\theta \rangle = \|\mathbf{X}\theta\|^2 = \theta^{\top} \mathbf{X}^{\top} \mathbf{X}\theta = n\widehat{\operatorname{Var}}(\mathbf{X}^{\top}\theta)$
- $\triangleright$   $h^1$  is the first (one-dimensional) approximation to X for describing each individual i by  $h_i^1 \in \mathbb{R}$  instead of  $X_{i.} \in \mathbb{R}^d$

Statistical learning: Advanced

Alain Celisse

PCA

Unsupervised

PLS

CCA

Step t  $(2 \le t \le r = \operatorname{rank}(X))$ 

► Find

 $\begin{array}{ll} \boldsymbol{\theta^t} \in \textit{Arg} & \max \\ \|\boldsymbol{\theta}\| = 1 \\ \left\langle \mathbf{X}\boldsymbol{\theta}, \mathbf{h^j} \right\rangle = 0 \\ j = 1, \dots, t-1 \end{array}$ 

 $lackbox{h}^t = \widetilde{m{X}}_{t} = m{X}\theta^t \in \mathbb{R}^n$ : tth principal component

# PCA: Focus on the other steps

Advanced

Alain Celisse

PCA Unsupervised

PCA algorithm

PLS CCA

Step t  $(2 \le t \le r = \operatorname{rank}(X))$ 

Find

 $egin{aligned} heta^{f t} \in \mathit{Arg} & \max \ \| heta\| = 1 \ \left<m{X} heta, h^{j}
ight> = 0 \ j = 1, \ldots, t-1 \end{aligned}$ 

 $m{h}^t = m{X}_{\cdot t} = m{X} heta^t \in \mathbb{R}^n$ : tth principal component

### Interpretation

- lacktriangle Any new component must be ot previous ones
- $\blacktriangleright$   $(h^1,\ldots,h^r)$ : orthogonal family
- Each individual *i* described by:

$$egin{aligned} ig(h_i^1,h_i^2ig) &\in \mathbb{R}^2 ext{ instead of } oldsymbol{X}_{i\cdot} \in \mathbb{R}^d \ \end{aligned}$$
 (with two components)

Statistical learning: Advanced

Alain Celisse

#### PCA

#### PL

Supervised

PLS algorithm
PLS: Have a new look

CCA

# Partial Least-Squares (PLS 1)

#### Notation

- n individuals
- Each individual *i* described by *d* (quantitative) variables  $X^1, \ldots, X^d$
- $lacksquare X_i = (X_i^1, \dots, X_i^d)^{\top} \in \mathbb{R}^d$ : features for individual i
- ightharpoonup X:  $n \times d$  "full" design matrix
- ▶  $X_{.,j} \in \mathbb{R}^n$ : jth column vector of X (n measurements for variable  $X^j$ )
- ▶  $X_{i.} \in \mathbb{R}^d$ : *i*th row vector of X
- Label:

Each individual i described by  $Y_i$ 

 $\mathbf{Y} = (Y_1, \dots, Y_n)^{\top} \in \mathbb{R}^n$ 

#### Notation

- n individuals
- Each individual i described by d (quantitative) variables  $X^1, \ldots, X^d$
- $lacksquare X_i = (X_i^1, \dots, X_i^d)^{\top} \in \mathbb{R}^d$ : features for individual i
- $\triangleright$  **X**:  $n \times d$  "full" design matrix
- ▶  $X_{.,j} \in \mathbb{R}^n$ : jth column vector of X (n measurements for variable  $X^j$ )
- ▶  $X_{i.} \in \mathbb{R}^d$ : *i*th row vector of X
- Label:

Each individual i described by  $Y_i$ 

 $\mathbf{Y} = (Y_1, \dots, Y_n)^{\top} \in \mathbb{R}^n$ 

Unlike PCA, the PLS framework is a supervised one!

# Comparison PCA/PLS

Statistical learning: Advanced

Alain Celisse

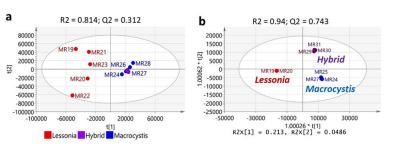


PLS

PLS algorithm PLS: Have a new look

CCA

Supervised



▶ Picture (a): If one class exhibits more variability than the others, PCA focuses on this class for defining axes → blue and purple points gathered within one cluster

# Comparison PCA/PLS

Statistical learning: Advanced

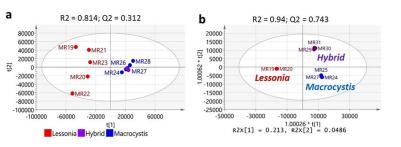
Alain Celisse



Supervised

PLS

PLS algorithm PLS: Have a new look



- Picture (a): If one class exhibits more variability than the others, PCA focuses on this class for defining axes
   → blue and purple points gathered within one cluster
- ▶ Picture (b): With PLS, the axes split labeled classes into well separated clusters

# PLS algorithm: Step 1

$$\operatorname{Cov}(\widehat{X^{\top}\theta}, Y) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \theta \cdot Y_{i} = \frac{1}{n} \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle$$

(columns of **X** centered and reduced) (vector **Y** is centered)

Statistical learning: Advanced

Alain Celisse

PCA

PLS Supervised

PLS algorithm
PLS: Have a new look

PLS algorithm

PLS: Have a new look

 $\operatorname{Cov}(\widehat{X^{\top}\theta}, Y) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \theta \cdot Y_{i} = \frac{1}{n} \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle$ 

(columns of **X** centered and reduced) (vector **Y** is centered)

## PLS-1 algorithm: Step 1

► Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \mathbf{X}\theta, \mathbf{Y} \rangle \right\}$$

 $lackbrack h^1 = \widetilde{m{X}}_{\cdot 1} = m{X} \theta^1 \in \mathbb{R}^n$ : first PLS component

Supervised PLS algorithm

PLS: Have a new look

CCA

 $\operatorname{Cov}(\widehat{X^{\top}\theta}, Y) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \theta \cdot Y_{i} = \frac{1}{n} \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle$ 

(columns of **X** centered and reduced)
(vector **Y** is centered)

### PLS-1 algorithm: Step 1

► Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \left\langle \mathbf{X}\theta, \mathbf{Y} \right\rangle \right\}$$

 $lackbrack h^1 = \widetilde{m{X}}_{\cdot 1} = m{X} \theta^1 \in \mathbb{R}^n$ : first PLS component

### Remarks:

 $\blacktriangleright$   $h^1$  maximizes the covariance with Y (explicability)

 $\operatorname{Cov}(\widehat{X^{\top}\theta}, Y) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_{i}^{\top} \theta \cdot Y_{i} = \frac{1}{n} \langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle$ 

(columns of **X** centered and reduced) (vector **Y** is centered)

## PLS-1 algorithm: Step 1

► Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \mathbf{X}\theta, \mathbf{Y} \rangle \right\}$$

▶  $h^1 = \widetilde{\mathbf{X}}_{\cdot 1} = \mathbf{X}\theta^1 \in \mathbb{R}^n$ : first PLS component

### Remarks:

- $ightharpoonup h^1$  maximizes the covariance with ightharpoonup (explicability)
- ► Focuses on covariance (≠ correlation!)
- ► Not scale invariant
  - $\longrightarrow$  centering and reducing columns of  ${m X}$

Supervised

PLS algorithm

PLS: Have a new look

CCA

PLS-1 algorithm: Step t  $(2 \le t \le r = rank(X))$ 

► Find

$$\begin{array}{ll} \boldsymbol{\theta}^t \in \textit{Arg} & \max \\ \|\boldsymbol{\theta}\| = 1 \\ \left\langle \mathbf{X}\boldsymbol{\theta}, \mathbf{h^j} \right\rangle = 0 \\ j = 1, \dots, t-1 \end{array} \left. \left\{ \left\langle \mathbf{X}\boldsymbol{\theta}, \mathbf{Y} \right\rangle \right\} \right. \\ \end{array}$$

 $lackbox{h}^t = \widetilde{m{X}}_{\cdot t} = m{X} \theta^t \in \mathbb{R}^n$ : tth PLS component

CCA

PLS-1 algorithm: Step t  $(2 \le t \le r = \text{rank}(X))$ 

► Find

$$egin{aligned} eta^t \in \mathit{Arg} & \max \ & \{\langle oldsymbol{X} eta, oldsymbol{Y} 
angle \} \ & \langle oldsymbol{X} eta, oldsymbol{h^j} 
angle = 0 \ & j = 1, \dots, t-1 \end{aligned}$$

 $lackbox{h}^t = \widetilde{m{X}}_{\cdot t} = m{X} \theta^t \in \mathbb{R}^n$ : tth PLS component

### Interpretation

- lacktriangle Any new PLS component must be ot previous ones

CCA

- PLS-1 algorithm: Step t  $(2 \le t \le r = \text{rank}(X))$ 
  - Find

$$egin{aligned} heta^t \in \mathit{Arg} & \max \ \| heta\| = 1 \ & \langle oldsymbol{X} heta, oldsymbol{h^j} 
angle = 0 \ & i = 1, \dots, t-1 \end{aligned}$$

 $lackbox{h}^t = \widetilde{m{X}}_{\cdot t} = m{X} \theta^t \in \mathbb{R}^n$ : tth PLS component

### Interpretation

- ▶ Any new PLS component must be ⊥ previous ones
- $\blacktriangleright$   $(h^1,\ldots,h^r)$ : orthogonal family

- PLS-1 algorithm: Step t  $(2 \le t \le r = rank(X))$ 
  - ► Find

$$\begin{array}{ll} \boldsymbol{\theta}^t \in \textit{Arg} & \max \\ \|\boldsymbol{\theta}\| = 1 \\ \left\langle \mathbf{X}\boldsymbol{\theta}, \mathbf{h'} \right\rangle = \mathbf{0} \\ j = 1, \dots, t-1 \end{array} \left. \left\{ \left\langle \mathbf{X}\boldsymbol{\theta}, \mathbf{Y} \right\rangle \right\} \right. \\ \end{array}$$

- lacksquare  $h^t = \widetilde{m{X}}_{\cdot t} = m{X} \theta^t \in \mathbb{R}^n$ : tth PLS component
- Interpretation

  - ▶ Any new PLS component must be ⊥ previous ones
  - $\blacktriangleright$   $(h^1,\ldots,h^r)$ : orthogonal family
  - **Each** individual *i* described by:

$$ig(h_i^1,h_i^2ig)\in\mathbb{R}^2$$
 instead of  $m{X}_{i\cdot}\in\mathbb{R}^d$  (with two components)

► Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \mathbf{X}\theta, \mathbf{Y} \rangle\}$$

ightharpoonup With  $h^1 = \widetilde{X}_{.1} = X \theta^1 \in \mathbb{R}^n$ 

### Calculating the solution

$$\begin{aligned} & \theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \left\langle \mathbf{X}\theta, \mathbf{Y} \right\rangle \right\} \\ \Leftrightarrow & \theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \left\langle \theta, \mathbf{X}^\top \mathbf{Y} \right\rangle \right\} \\ \Leftrightarrow & \theta^1 = \frac{\mathbf{X}^\top \mathbf{Y}}{\|\mathbf{X}^\top \mathbf{Y}\|} \end{aligned}$$

Statistical learning: Advanced

Alain Celisse

PLS: Have a new look

PCA

PLS

Supervised PLS algorithm

Statistical learning: Advanced

Alain Celisse

PCA

PLS Supervised

PLS algorithm

PLS: Have a new look

CCA

Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \{\langle \mathbf{X}\theta, \mathbf{Y} \rangle\}$$

lacksquare With  $h^1 = \widetilde{\pmb{X}}_{\cdot 1} = \pmb{X} heta^1 \in \mathbb{R}^n$ 

### Calculating the solution

$$\begin{array}{l} \theta^1 \in \textit{Arg} \max_{\|\theta\|=1} \left\{ \left\langle \textbf{\textit{X}} \theta, \textbf{\textit{Y}} \right\rangle \right\} \\ \Leftrightarrow \quad \theta^1 \in \textit{Arg} \max_{\|\theta\|=1} \left\{ \left\langle \theta, \textbf{\textit{X}}^\top \textbf{\textit{Y}} \right\rangle \right\} \\ \Leftrightarrow \quad \theta^1 = \frac{\textbf{\textit{X}}^\top \textbf{\textit{Y}}}{\|\textbf{\textit{X}}^\top \textbf{\textit{Y}}\|} \\ \longrightarrow \text{The first PLS component is } h^1 = \textbf{\textit{X}} \theta^1 = \frac{\textbf{\textit{X}} \textbf{\textit{X}}^\top \textbf{\textit{Y}}}{\|\textbf{\textit{X}}^\top \textbf{\textit{Y}}\|} \end{array}$$

learning: Advanced Alain Celisse

Alain Celisse

PCA PLS

Supervised
PLS algorithm
PLS: Have a new look

CCA

Find

$$\theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \langle \mathbf{X}\theta, \mathbf{Y} \rangle \right\}$$

ightharpoonup With  $h^1 = \widetilde{m{X}}_{\cdot 1} = m{X} heta^1 \in \mathbb{R}^n$ 

### Calculating the solution

$$\begin{array}{l} \theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \left\langle \textbf{\textit{X}}\theta, \textbf{\textit{Y}} \right\rangle \right\} \\ \Leftrightarrow \quad \theta^1 \in Arg \max_{\|\theta\|=1} \left\{ \left\langle \theta, \textbf{\textit{X}}^\top \textbf{\textit{Y}} \right\rangle \right\} \\ \Leftrightarrow \quad \theta^1 = \frac{\textbf{\textit{X}}^\top \textbf{\textit{Y}}}{\|\textbf{\textit{X}}^\top \textbf{\textit{Y}}\|} \\ \longrightarrow \text{The first PLS component is } h^1 = \textbf{\textit{X}}\theta^1 = \frac{\textbf{\textit{X}}\textbf{\textit{X}}^\top \textbf{\textit{Y}}}{\|\textbf{\textit{X}}^\top \textbf{\textit{Y}}\|} \end{array}$$

Interpretation of  $h^1$ 

 $h^1 = \widehat{m{Y}}^1$ : first explanation of  $m{Y}$  by "the model"  $m{X} heta$  (regression)

# Revisiting PLS algorithm: Step 2 (t = 2)

Orthogonality constraint

$$(V_1 = \operatorname{Vect}(\mathbf{X}^{\top} h^1))$$

$$ig\langle m{X} heta, h^1 ig
angle = 0 \quad \Leftrightarrow \quad \Big\langle heta, m{X}^{ op} h^1 \Big
angle = 0$$

$$\Leftrightarrow \quad \theta \in \mathrm{Vect} \left( m{X}^{ op} h^1 \right)^{\perp} = V_1^{\perp}$$

Statistical learning: Advanced

Alain Celisse

PCA

PLS

Supervised PLS algorithm

PLS: Have a new look

# Revisiting PLS algorithm: Step 2 (t = 2)

Orthogonality constraint

$$(V_1 = \operatorname{Vect}(\boldsymbol{X}^{\top} h^1))$$

$$\left\langle \boldsymbol{X}\boldsymbol{\theta},h^{1}\right\rangle =0 \quad\Leftrightarrow\quad \left\langle \boldsymbol{\theta},\boldsymbol{X}^{\top}h^{1}\right\rangle =0$$
 
$$\Leftrightarrow\quad \boldsymbol{\theta}\in\operatorname{Vect}\left(\boldsymbol{X}^{\top}h^{1}\right)^{\perp}=V_{1}^{\perp}$$

#### Constrained maximization

Under this orthogonality constraint, solving

$$\max_{\|\boldsymbol{\theta}\|=1} \left\{ \left\langle \boldsymbol{X} \boldsymbol{\theta}, \boldsymbol{Y} \right\rangle \right\} \quad \Leftrightarrow \quad \max_{\|\boldsymbol{\theta}\|=1} \left\{ \left\langle \boldsymbol{\theta}, \boldsymbol{X}^{\top} \boldsymbol{Y} - \Pi_{V_1} (\boldsymbol{X}^{\top} \boldsymbol{Y}) \right\rangle \right\}$$

where  $\Pi_{V_1}(\cdot)$  denotes the orthogonal projection onto  $V_1$  Reminder:

$$u - \Pi_{V_1}(u) \in V_1^{\perp}, \quad \forall u \in \mathbb{R}^d$$

Statistical learning: Advanced

Alain Celisse

PCA

PLS

Supervised
PLS algorithm
PLS: Have a new look

Solution at t=2

Statistical learning: Advanced

Alain Celisse

PCA

PLS

Supervised
PLS algorithm
PLS: Have a new look

$$\theta^{2} = Arg \quad \max_{\|\boldsymbol{\theta}\| = 1} \left\{ \left\langle \boldsymbol{\theta}, \boldsymbol{X}^{\top} \boldsymbol{Y} \right\rangle \right\}$$

$$\langle \boldsymbol{\theta}, \boldsymbol{X}^{\top} \boldsymbol{h}^{1} \rangle = 0$$

$$\Leftrightarrow \quad \theta^{2} = Arg \quad \max_{\|\boldsymbol{\theta}\| = 1} \left\{ \left\langle \boldsymbol{\theta}, \boldsymbol{X}^{\top} \boldsymbol{Y} - \Pi_{V_{1}} (\boldsymbol{X}^{\top} \boldsymbol{Y}) \right\rangle \right\}$$

$$\langle \boldsymbol{\theta}, \boldsymbol{X}^{\top} \boldsymbol{h}^{1} \rangle = 0$$

$$\Leftrightarrow \quad \theta^{2} = \frac{\boldsymbol{X}^{\top} \boldsymbol{Y} - \Pi_{V_{1}} (\boldsymbol{X}^{\top} \boldsymbol{Y})}{\|\boldsymbol{X}^{\top} \boldsymbol{Y} - \Pi_{V_{1}} (\boldsymbol{X}^{\top} \boldsymbol{Y})\|}$$

Solution at 
$$t=2$$

$$\theta^2 = Arg \quad \max_{ \|\boldsymbol{\theta}\| = 1 } \quad \left\{ \left\langle \boldsymbol{\theta}, \boldsymbol{X}^\top \boldsymbol{Y} \right\rangle \right\}$$
$$\left\langle \boldsymbol{\theta}, \boldsymbol{X}^\top \boldsymbol{h}^1 \right\rangle = 0$$

$$\Leftrightarrow \quad \theta^2 = Arg \quad \max_{ \|\theta\| = 1 } \left\{ \left\langle \theta, \boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1} (\boldsymbol{X}^\top \boldsymbol{Y}) \right\rangle \right\}$$
$$\left\langle \theta, \boldsymbol{X}^\top h^1 \right\rangle = 0$$

$$\Leftrightarrow \quad \theta^2 = \frac{\mathbf{X}^{\top} \mathbf{Y} - \Pi_{V_1} (\mathbf{X}^{\top} \mathbf{Y})}{\|\mathbf{X}^{\top} \mathbf{Y} - \Pi_{V_1} (\mathbf{X}^{\top} \mathbf{Y})\|}$$

### Reminder:

$$\Pi_{V_1}(\boldsymbol{X}^{\top}\boldsymbol{Y}) = \left\langle \boldsymbol{X}^{\top}\boldsymbol{Y}, \frac{\boldsymbol{X}^{\top}h^1}{\left\|\boldsymbol{X}^{\top}h^1\right\|} \right\rangle \frac{\boldsymbol{X}^{\top}h^1}{\left\|\boldsymbol{X}^{\top}h^1\right\|}$$

CCA

#### Pseudo-code

Input: X, Y

▶ Initialization t = 0:

$$R^0 = \boldsymbol{X}^{ op} \boldsymbol{Y}$$

▶ Step t = 1:

$$m{ heta}^1 = rac{R^{m{0}}}{\|R^{m{0}}\|}$$
 and  $h^1 = m{X} heta^1$ 

$$P^1 = R^0 - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})$$

#### Pseudo-code

Input: X, Y

▶ Initialization t = 0:

$$R^0 = \boldsymbol{X}^{\top} \boldsymbol{Y}$$

▶ Step t = 1:

$$m{ heta}^1 = rac{R^{m{0}}}{\|R^{m{0}}\|}$$
 and  $h^1 = m{X} heta^1$ 

$$P^1 = R^0 - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})$$

For t = 2 to rank (X) = r:

$$m{ heta}^t = rac{R^{t-1}}{\|R^{t-1}\|}$$
 and  $h^t = m{X} \theta^t$ 

$$P^t = R^{t-1} - \Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y})$$

### Pseudo-code

Input: X, Y

lnitialization t=0:

$$R^0 = \boldsymbol{X}^{ op} \boldsymbol{Y}$$

ightharpoonup Step t=1:

$$m{ heta}^1=rac{R^{m{0}}}{\parallel R^{m{0}}\parallel}$$
 and  $h^1=m{X} heta^1$ 

$$P^1 = R^0 - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})$$

For 
$$t = 2$$
 to rank  $(X) = r$ :

$$lackbox{mlack}{ heta} heta^t = rac{R^{t-1}}{\|R^{t-1}\|} ext{ and } h^t = m{X} heta^t$$

$$\triangleright R^t = R^{t-1} - \Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y})$$

**Output:** 
$$(h^1, h^2, \dots, h^r)$$
 such that

$$\widetilde{\mathbf{X}} = [h^1; h^2; \dots; h^r] = \mathbf{X} \cdot [\theta^1; \theta^2; \dots; \theta^r]$$

 $(r = \operatorname{rank}(\boldsymbol{X}))$ 

PLS

PLS algorithm PLS: Have a new look

Statistical learning: Advanced

Alain Celisse

**PCA** 

PLS

Supervised PLS algorithm

PLS: Have a new lo

CCA

### Covariance and not correlation

Covariance is not scale invariante

Statistical learning: Advanced

Alain Celisse

PCA

PLS

Supervised PLS algorithm

PLS: Have a new look

CCA

- Covariance is not scale invariante
- ► The **X**<sub>i</sub>s have been normalized (centering and reducing)

Statistical learning: Advanced

Alain Celisse

PCA

PLS

Supervised PLS algorithm

PLS: Have a new look

CCA

- Covariance is not scale invariante
- ► The X.<sub>j</sub>s have been normalized (centering and reducing)
- ▶ By contrast, the  $h^{j}$ s (PLS components):

Statistical learning: Advanced

Alain Celisse

PCA

PLS

Supervised PLS algorithm

PLS: Have a new look

CCA

- Covariance is not scale invariante
- ▶ The  $X_{.j}$ s have been normalized (centering and reducing)
- ▶ By contrast, the h<sup>j</sup>s (PLS components):
  - are an orthogonal basis

Statistical learning: Advanced

Alain Celisse

PCA

PLS Supervised

PLS algorithm
PLS: Have a new look

CCA

- Covariance is not scale invariante
- ightharpoonup The  $X_{.j}$ s have been normalized (centering and reducing)
- ▶ By contrast, the h<sup>i</sup>s (PLS components):
  - ► are an orthogonal basis
  - are not unitary vectors! (unlike the X.js)

Statistical learning: Advanced

Alain Celisse

PCA

PLS Supervised

PLS algorithm
PLS: Have a new look

CCA

- Covariance is not scale invariante
- ▶ The  $X_{.j}$ s have been normalized (centering and reducing)
- ▶ By contrast, the h<sup>j</sup>s (PLS components):
  - ► are an orthogonal basis
  - ightharpoonup are not unitary vectors! (unlike the  $X_{.j}$ s)
- ► Can induce some bias in the vizualization

Statistical learning: Advanced

Alain Celisse

**PCA** 

PLS

CCA

Supervised CCA-1 algorithm

CCA: Have a new look Extension to two design matrices

Canonical Correlation Analysis (CCA)

#### Notation

- n individuals
- Each individual i described by d (quantitative) variables  $X^1, \ldots, X^d$
- $X_i = (X_i^1, \dots, X_i^d)^\top \in \mathbb{R}^d$ : features for individual i
- ightharpoonup X:  $n \times d$  "full" design matrix
- ▶  $X_{.,j} \in \mathbb{R}^n$ : jth column vector of X (n measurements for variable  $X^j$ )
- $igwedge X_{i\cdot} \in \mathbb{R}^d$ : *i*th row vector of X
- Label:

Each individual i described by  $Y_i$ 

 $\mathbf{Y} = (Y_1, \dots, Y_n)^{\top} \in \mathbb{R}^n$ 

### Notation

- n individuals
- Each individual i described by d (quantitative) variables  $X^1, \ldots, X^d$
- $lacksquare X_i = (X_i^1, \dots, X_i^d)^{\top} \in \mathbb{R}^d$ : features for individual i
- **X**:  $n \times d$  "full" design matrix
- ▶  $X_{.,j} \in \mathbb{R}^n$ : jth column vector of X (n measurements for variable  $X^j$ )
- ▶  $X_{i.} \in \mathbb{R}^d$ : *i*th row vector of X
- ► Label:

Each individual i described by  $Y_i$ 

 $\mathbf{Y} = (Y_1, \ldots, Y_n)^{\top} \in \mathbb{R}^n$ 

Unlike PCA, the CCA framework is a supervised one!

## **CCA** algorithm



Alain Celisse

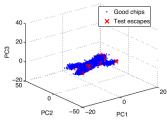


PLS

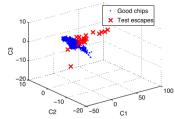
CCA Supervised

#### CCA-1 algorithm

CCA: Have a new look Extension to two design matrices



3-D Principal Component Space



3-D Canonical Space

# CCA-1 algorithm: Step 1

$$\operatorname{Corr}(\widehat{\boldsymbol{X}\theta}, \boldsymbol{Y}) = \frac{\langle \boldsymbol{X}\theta, \boldsymbol{Y} \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{Y}\|}$$

(columns of **X** centered and reduced) (vector **Y** is centered)

### CCA-1 algorithm: Step 1

► Find

$$heta^1 \in Arg\max_{ heta \in \mathbb{R}^d} \left\{ rac{\langle oldsymbol{X} heta, oldsymbol{Y}
angle}{\|oldsymbol{X} heta\|\cdot\|oldsymbol{Y}\|}
ight\}$$

▶  $h^1 = \widetilde{\mathbf{X}}_{\cdot 1} = \mathbf{X}\theta^1 \in \mathbb{R}^n$ : first CCA component

### Remark:

- Focuses on maximizing correlation!
- $ightharpoonup h^1$  maximizes the correlation with  $m{Y}$  (explicability)

Statistical learning: Advanced

Alain Celisse

PCA

PLS

CCA Supervised

CCA-1 algorith

CCA: Have a new look Extension to two design matrices

CCA Supervised

----

CCA: Have a new look

Extension to two design
matrices

CCA-1 algorithm: Step t  $(2 \le t \le r = rank(X))$ 

► Find

$$heta^t \in Arg \quad \max_{egin{array}{c} igl(m{X} heta, h^jigr) = 0 \ j = 1, \dots, t-1 \ \end{array}} \left\{ rac{igl(m{X} heta, m{Y}igr)}{\|m{X} heta\|\cdot\|m{Y}\|} 
ight\}$$

 $h^t = \widetilde{X}_{t} = X\theta^t \in \mathbb{R}^n$ : tth CCA component

CCA Supervised

CCA-1 algorit

Extension to two design matrices

- CCA-1 algorithm: Step t  $(2 \le t \le r = rank(X))$ 
  - ▶ Find

$$\theta^t \in Arg \quad \max_{\begin{subarray}{c} \left\langle m{X} heta, m{h}^j 
ight
angle = 0 \\ j = 1, \dots, t-1 \end{subarray}} \quad \left\{ \frac{\left\langle m{X} heta, m{Y} 
ight
angle}{\|m{X} heta\| \cdot \|m{Y}\|} \right\}$$

- $lackbox{h}^t = \widetilde{m{X}}_{\cdot t} = m{X} \theta^t \in \mathbb{R}^n$ : tth CCA component
- Interpretation

  - lacktriangle Any new component must be ot previous ones
  - $\blacktriangleright$   $(h^1, \ldots, h^r)$ : orthonormal family
  - Each individual *i* described by:

$$(h_i^1, h_i^2) \in \mathbb{R}^2$$
 instead of  $oldsymbol{X}_{i\cdot} \in \mathbb{R}^d$  (with two components)

# Revisiting CCA algorithm: Step 1

► Find

$$heta^1 \in \mathit{Arg}\max_{ heta \in \mathbb{R}^d} \left\{ rac{\langle oldsymbol{X} heta, oldsymbol{Y}
angle}{\|oldsymbol{X} heta\|\cdot\|oldsymbol{Y}\|}
ight\}$$

$$h^1 = \widetilde{\boldsymbol{X}}_{.1} = \boldsymbol{X}\theta^1 \in \mathbb{R}^n$$

Calculating the solution

Statistical learning: Advanced

Alain Celisse

PCA

PLS

CCA

matrices

Supervised CCA-1 algorithm

CCA: Have a new look

Extension to two design

# Revisiting CCA algorithm: Step 1

Find

$$heta^1 \in \mathit{Arg}\max_{ heta \in \mathbb{R}^d} \left\{ rac{\langle oldsymbol{X} heta, oldsymbol{Y}
angle}{\|oldsymbol{X} heta\|\cdot\|oldsymbol{Y}\|}
ight\}$$

 $h^1 = \widetilde{X}_{\cdot,1} = X\theta^1 \in \mathbb{R}^n$ 

### Calculating the solution

$$\theta^{1} \in Arg \max_{\theta \in \mathbb{R}^{d}} \left\{ \frac{\langle \mathbf{X}\theta, \mathbf{Y} \rangle}{\|\mathbf{X}\theta\| \cdot \|\mathbf{Y}\|} \right\}$$

$$\Leftrightarrow \quad \theta^{1} \in Arg \max_{\|\mathbf{X}\theta\|=1} \left\{ \langle \mathbf{X}\theta, \mathbf{Y} \rangle \right\}$$

$$\Leftrightarrow \quad \theta^{1} \in Arg \max_{\|\mathbf{X}\theta\|=1} \left\{ \left\langle \theta, \mathbf{X}^{\top} \mathbf{Y} \right\rangle \right\}$$

$$\Leftrightarrow \quad \theta^{1} = \frac{\mathbf{X}^{\top} \mathbf{Y}}{\|\mathbf{X}\mathbf{X}^{\top} \mathbf{Y}\|}$$

 $\longrightarrow$  The first CCA component is  $h^1 = X\theta^1 = \frac{XX^\top Y}{\|XX^\top Y\|}$ 

Statistical learning: Advanced

Alain Celisse

PCA

PLS

CCA

matrices

Supervised CCA-1 algorithm

CCA: Have a new look

CCA: Have a new look Extension to two design

## Revisiting CCA algorithm: Step 2 (t = 2)

Orthogonality constraint

$$(V_1 = \operatorname{Vect}(\boldsymbol{X}^{\top} h^1))$$

$$\begin{split} \left\langle \boldsymbol{X}\boldsymbol{\theta},\boldsymbol{h}^{1}\right\rangle &= 0 \quad \Leftrightarrow \quad \left\langle \boldsymbol{\theta},\boldsymbol{X}^{\top}\boldsymbol{h}^{1}\right\rangle &= 0 \\ &\Leftrightarrow \quad \boldsymbol{\theta} \in \operatorname{Vect}\left(\boldsymbol{X}^{\top}\boldsymbol{h}^{1}\right)^{\perp} &= \boldsymbol{V}_{1}^{\perp} \end{split}$$

### Constrained maximization

Under this orthogonality constraint, solving

$$\max_{\|\boldsymbol{X}\boldsymbol{\theta}\|=1} \left\{ \left\langle \boldsymbol{X}\boldsymbol{\theta}, \boldsymbol{Y} \right\rangle \right\} \quad \Leftrightarrow \quad \max_{\|\boldsymbol{X}\boldsymbol{\theta}\|=1} \left\{ \left\langle \boldsymbol{\theta}, \boldsymbol{X}^{\top} \boldsymbol{Y} - \Pi_{V_1} (\boldsymbol{X}^{\top} \boldsymbol{Y}) \right\rangle \right\}$$

where  $\Pi_{V_1}(\cdot)$  denotes the orthogonal projection onto  $V_1$  **Reminder**:

$$u - \Pi_{V_1}(u) \in V_1^{\perp}, \quad \forall u \in \mathbb{R}^d$$

Statistical learning: Advanced

Alain Celisse

PCA

PLS

CCA

Supervised
CCA-1 algorith

matrices

CCA: Have a new look

Extension to two design

CCA

matrices

Supervised

CCA: Have a new look
Extension to two design

Solution at t=2

$$\begin{array}{ll} \theta^2 = \textit{Arg} & \max \\ \| \boldsymbol{X} \boldsymbol{\theta} \| = 1 \\ \left< \boldsymbol{\theta}, \boldsymbol{X}^\top \boldsymbol{h}^1 \right> = 0 \end{array} \left. \left\{ \left< \boldsymbol{\theta}, \boldsymbol{X}^\top \boldsymbol{Y} \right> \right\} \end{array} \right.$$

$$\Leftrightarrow \quad \theta^2 = Arg \quad \max_{ \left\| \boldsymbol{X} \boldsymbol{\theta} \right\| = 1 \\ \left\langle \boldsymbol{\theta}, \boldsymbol{X}^\top \boldsymbol{h}^1 \right\rangle = 0 } \quad \left\{ \left\langle \boldsymbol{\theta}, \boldsymbol{X}^\top \boldsymbol{Y} - \Pi_{V_1} (\boldsymbol{X}^\top \boldsymbol{Y}) \right\rangle \right\}$$

$$\Leftrightarrow \quad \theta^2 = \frac{\boldsymbol{X}^{\top} \boldsymbol{Y} - \Pi_{V_1} (\boldsymbol{X}^{\top} \boldsymbol{Y})}{\|\boldsymbol{X} (\boldsymbol{X}^{\top} \boldsymbol{Y} - \Pi_{V_1} (\boldsymbol{X}^{\top} \boldsymbol{Y}))\|}$$

Reminder:

$$\Pi_{V_1}(\boldsymbol{X}^{\top}\boldsymbol{Y}) = \left\langle \boldsymbol{X}^{\top}\boldsymbol{Y}, \frac{\boldsymbol{X}^{\top}h^1}{\left\|\boldsymbol{X}^{\top}h^1\right\|} \right\rangle \frac{\boldsymbol{X}^{\top}h^1}{\left\|\boldsymbol{X}^{\top}h^1\right\|}$$

### Pseudo-code

Input: X, Y

▶ Initialization t = 0:

$$R^0 = \boldsymbol{X}^{ op} \boldsymbol{Y}$$

▶ Step t = 1:

$$lackbox{lack}{ heta^1} = rac{R^{f 0}}{\|m{X}R^{f 0}\|} ext{ and } h^1 = m{X} heta^1$$

$$P^1 = R^0 - \Pi_{V_1}(\boldsymbol{X}^\top \boldsymbol{Y})$$

For 
$$t = 2$$
 to rank  $(X) = r$ :

$$m{\theta}^t = rac{R^{t-1}}{\|m{X}R^{t-1}\|}$$
 and  $h^t = m{X}\theta^t$ 

$$\blacktriangleright \ \, \Pi_{V_t}(\boldsymbol{X}^\top\boldsymbol{Y}) = \left\langle \boldsymbol{X}^\top\boldsymbol{Y}, \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|} \right\rangle \frac{\boldsymbol{X}^\top h^t}{\|\boldsymbol{X}^\top h^t\|}$$

$$\triangleright R^t = R^{t-1} - \Pi_{V_t}(\boldsymbol{X}^\top \boldsymbol{Y})$$

**Output:** 
$$(h^1, h^2, \dots, h^r)$$
 such that

$$\widetilde{\mathbf{X}} = [h^1; h^2; \dots; h^r] = \mathbf{X} \cdot [\theta^1; \theta^2; \dots; \theta^r]$$

 $(r = \operatorname{rank}(\boldsymbol{X}))$ 

Extension to two design matrices

Statistical learning: Advanced

Alain Celisse

PCA

PLS

CCA

Supervised CCA-1 algorithm

CCA: Have a new look

Extension to two design

## Unsupervised situation

#### Statistical learning: Advanced

#### Alain Celisse

#### PCA

PLS

#### CCA

Supervised

CCA-1 algorithm

CCA: Have a new look

Extension to two design

matrices

### Notation

- n individuals
- First design matrix X:
  - lacksquare  $X_i = (X_i^1, \dots, X_i^d)^{\top} \in \mathbb{R}^d$ : features for individual i
  - $\triangleright$  **X**:  $n \times d$  "full" design matrix
  - ▶  $X_{.,j} \in \mathbb{R}^n$ : jth column vector of X
  - $lackbrack X_{i\cdot} \in \mathbb{R}^d$ : *i*th row vector of  $oldsymbol{X}$
- ► Second design matrix **W**:
  - $W_i = (W_i^1, \dots, W_i^d)^{\top} \in \mathbb{R}^p$ : features for individual i
  - **W**:  $n \times p$  "full" design matrix
  - $lackbox{W}_{\cdot,j} \in \mathbb{R}^n$ : jth column vector of  $oldsymbol{W}$
  - ▶  $W_{i.} \in \mathbb{R}^{p}$ : *i*th row vector of W

## Examples for $\boldsymbol{W}$ and $\boldsymbol{X}$

- Results from two different experiments (Two psychologic tests, bio. exper. at two time-stamps)
- ► Wine critic ratings and wine biological features

CCA

Supervised

CCA-1 algorithm

CCA: Have a new look

Extension to two design

 $\operatorname{Corr}(\widehat{\boldsymbol{X}\theta}, \boldsymbol{W}\gamma) = \frac{\langle \boldsymbol{X}\theta, \boldsymbol{W}\gamma \rangle}{\|\boldsymbol{X}\theta\| \cdot \|\boldsymbol{W}\gamma\|}$ 

(columns of  $\boldsymbol{X}, \boldsymbol{W}$  centered and reduced)

CCA algorithm: Step 1

► Find

$$\theta^1, \gamma^1 \in \textit{Arg} \max_{\theta \in \mathbb{R}^d, \gamma \in \mathbb{R}^p} \left\{ \frac{\langle \boldsymbol{X} \theta, \boldsymbol{W} \gamma \rangle}{\|\boldsymbol{X} \theta\| \cdot \|\boldsymbol{W} \gamma\|} \; \right\}$$

- ▶  $h^1 = \widetilde{\boldsymbol{X}}_{\cdot 1} = \boldsymbol{X} \theta^1 \in \mathbb{R}^n$ : first CCA component for  $\boldsymbol{X}$
- $ightharpoonup g^1 = \widetilde{W}_{\cdot 1} = W \gamma^1 \in \mathbb{R}^n$ : first CCA component for W

CCA algorithm: Step t  $(2 \le t \le \min(\operatorname{rank}(\boldsymbol{X}), \operatorname{rank}(\boldsymbol{W})))$ 

► Find

$$egin{aligned} heta^t, \gamma^t \in Arg & \max & \left\{ egin{aligned} & \left\{ \left\{ egin{aligned} & \left\{ egin{aligned} & \left\{ \left\{ egin{aligned} & \left\{ \left\{ egin{aligned} & \left\{ \left\{ \left\{ \left\{ egin{aligned} & \left\{ \left\{ \left\{ \left\{ \left\{ \left\{ \left\{ \left\{ \left\{$$

 $ightharpoonup r = \min(\operatorname{rank}(\boldsymbol{X}), \operatorname{rank}(\boldsymbol{W}))$ 

### Interpretation

- $(h^1, \ldots, h^r)$  and  $(g^1, \ldots, g^r)$ : orthonormal families since the above problem is equivalent to

$$\begin{aligned} \max & \left\{ \left\langle \boldsymbol{X}\boldsymbol{\theta},\,\boldsymbol{W}\boldsymbol{\gamma}\right\rangle \right\} \\ \left\langle \boldsymbol{X}\boldsymbol{\theta},\,\boldsymbol{h}^{j}\right\rangle &=0 \\ \left\langle \boldsymbol{W}\boldsymbol{\gamma},\,\boldsymbol{g}^{j}\right\rangle &=0 \\ j&=1,\ldots,t-1 \end{aligned}$$

### Summary

- Impelmentations similar to those of PLS and CCA-1 do exist for efficiently solving the CCA optimization problem
- CCA is used in a lot of applications for combining different types of information ("tables") used for describing "individuals"
- Extensions of CCA do exist for multiple tables (see Generalized CCA...)