High-dimensional CCA

Lennie Wells

May 9, 2023

Introduction and outline

Canonical Correlation Analysis (CCA) is much like Principal Component Analysis (PCA).

Introduction and outline

Canonical Correlation Analysis (CCA) is much like Principal Component Analysis (PCA).

Similarly, the term may refer to:

- A mathematical object
- A class of algorithms
- Process for dimension reduction or interpretation

Introduction and outline

Canonical Correlation Analysis (CCA) is much like Principal Component Analysis (PCA).

Similarly, the term may refer to:

- A mathematical object
- A class of algorithms
- Process for dimension reduction or interpretation

Structure of talk

- Population CCA
- CCA as a tool for data analysis
- Our specific motivations and contributions

Population CCA

In words: Given random variables $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$ find linear combinations $u^T X, v^T Y$ with maximal correlation, successively, subject to orthogonality.

Population CCA

In words: Given random variables $X \in \mathbb{R}^p, Y \in \mathbb{R}^q$ find linear combinations $u^T X, v^T Y$ with maximal correlation, successively, subject to orthogonality.

'Function' from law of pair (X, Y) to an optimal pair of bases.

Population CCA

In words: Given random variables $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$ find linear combinations $u^T X, v^T Y$ with maximal correlation, successively, subject to orthogonality.

'Function' from law of pair (X, Y) to an optimal pair of bases.

Note: correlation is invariant to scaling, so normalise to have variance one.

Population CCA

In words: Given random variables $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$ find linear combinations $u^T X, v^T Y$ with maximal correlation, successively, subject to orthogonality.

'Function' from law of pair (X, Y) to an optimal pair of bases.

Note: correlation is invariant to scaling, so normalise to have variance one.

In maths: Inductively define u_k, v_k by

$$\begin{aligned}
\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \quad & \operatorname{Cov}\left(u^T X, v^T Y\right) \\
\text{subject to} \quad & \operatorname{Var}\left(u^T X\right) = \operatorname{Var}\left(v^T Y\right) = 1, \\
& \operatorname{Cov}\left(u^T X, u_j^T X\right) = \operatorname{Cov}\left(v^T Y, v_j^T Y\right) = 0 \quad \text{for } j = 1, \ldots, k-1.
\end{aligned}$$

Population CCA

In words: Given random variables $X \in \mathbb{R}^p$, $Y \in \mathbb{R}^q$ find linear combinations $u^T X, v^T Y$ with maximal correlation, successively, subject to orthogonality.

'Function' from law of pair (X, Y) to an optimal pair of bases.

Note: correlation is invariant to scaling, so normalise to have variance one.

In maths: Inductively define u_k, v_k by

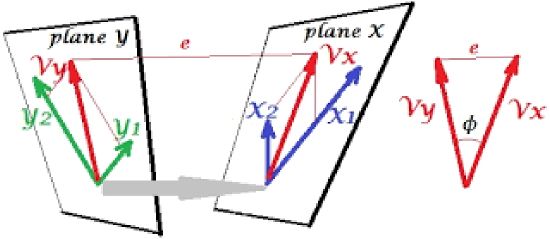
$$\begin{aligned} \underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \quad & \operatorname{Cov}\left(u^T X, v^T Y\right) \\ \text{subject to} \quad & \operatorname{Var}\left(u^T X\right) = \operatorname{Var}\left(v^T Y\right) = 1, \\ & \operatorname{Cov}\left(u^T X, u_j^T X\right) = \operatorname{Cov}\left(v^T Y, v_j^T Y\right) = 0 \quad \text{for } j = 1, \ldots, k-1. \end{aligned}$$

Notation:

- ρ_k the optimal value called k^{th} *canonical correlation*
- $u_k^T X, v_k^T Y$ called *canonical variates*
- u_k, v_k called *weights*

Visualisation of CCA

$$\underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \ \text{Cov}\left(u^T X, v^T Y\right)$$
$$\text{subject to} \ \text{Var}\left(u^T X\right) = \text{Var}\left(v^T Y\right) = 1,$$
$$\text{Cov}\left(u^T X, u_j^T X\right) = \text{Cov}\left(v^T Y, v_j^T Y\right) = 0 \quad \text{for } j = 1, \dots, k-1.$$



Matrix formulation

$$
\begin{aligned}
\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}}\ & \operatorname{Cov}\left(u^T X, v^T Y\right) \\
\text{subject to}\ & \operatorname{Var}\left(u^T X\right) = \operatorname{Var}\left(v^T Y\right) = 1, \\
& \operatorname{Cov}\left(u^T X, u_j^T X\right) = \operatorname{Cov}\left(v^T Y, v_j^T Y\right) = 0 \quad \text{for } j = 1, \ldots, k-1.
\end{aligned}
$$

Matrix formulation

$$\begin{aligned} \underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \quad & \text{Cov}\left(u^T X, v^T Y\right) \\ \text{subject to} \quad & \text{Var}\left(u^T X\right) = \text{Var}\left(v^T Y\right) = 1, \\ & \text{Cov}\left(u^T X, u_j^T X\right) = \text{Cov}\left(v^T Y, v_j^T Y\right) = 0 \quad \text{for } j = 1, \ldots, k-1. \end{aligned}$$

Weights are not the only interesting vectors - let's inspect matrix form - but skim links to rest of matrix analysis for now...

Matrix formulation

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \ \text{Cov}\left(u^T X, v^T Y\right)$$
$$\text{subject to} \ \text{Var}\left(u^T X\right) = \text{Var}\left(v^T Y\right) = 1,$$
$$\text{Cov}\left(u^T X, u_j^T X\right) = \text{Cov}\left(v^T Y, v_j^T Y\right) = 0 \quad \text{for } j = 1, \ldots, k-1.$$

Weights are not the only interesting vectors - let's inspect matrix form - but skim links to rest of matrix analysis for now...

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \ u^T \Sigma_{xy} v$$
$$\text{subject to} \ u^T \Sigma_{xx} u = 1, \ v^T \Sigma_{yy} v = 1,$$
$$u^T \Sigma_{xx} u_j = v^T \Sigma_{yy} v_j = 0 \text{ for } 1 \leq j \leq k-1.$$

Matrix formulation

$$\begin{aligned} \underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \ & \text{Cov}\left(u^T X, v^T Y\right) \\ \text{subject to} \ & \text{Var}\left(u^T X\right) = \text{Var}\left(v^T Y\right) = 1, \\ & \text{Cov}\left(u^T X, u_j^T X\right) = \text{Cov}\left(v^T Y, v_j^T Y\right) = 0 \quad \text{for } j = 1, \dots, k-1. \end{aligned}$$

Weights are not the only interesting vectors - let's inspect matrix form - but skim links to rest of matrix analysis for now...

$$\begin{aligned} \underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \ & u^T \Sigma_{xy} v \\ \text{subject to} \ & u^T \Sigma_{xx} u = 1, \ v^T \Sigma_{yy} v = 1, \\ & u^T \Sigma_{xx} u_j = v^T \Sigma_{yy} v_j = 0 \text{ for } 1 \leq j \leq k-1. \end{aligned}$$

Notation: Define *canonical loadings*

$$\Sigma_{xx} u_k = \text{Cov}(X, u_k^T X), \quad \Sigma_{yy} v_k = \text{Cov}(Y, v_k^T Y)$$

Reconstruction

Orthonormality constraints $u_k^T \Sigma_{xx} u_j = v_k^T \Sigma_{yy} v_j = \delta_{jk}$ for $1 \leq j \leq k - 1$.

So $(u_k), (\Sigma_{xx} u_k)$ and $(v_k), (\Sigma_{yy} v_k)$ are each pairs of *dual bases*.

Reconstruction

Orthonormality constraints $u_k^T \Sigma_{xx} u_j = v_k^T \Sigma_{yy} v_j = \delta_{jk}$ for $1 \leq j \leq k-1$.

So $(u_k), (\Sigma_{xx} u_k)$ and $(v_k), (\Sigma_{yy} v_k)$ are each pairs of *dual bases*.

Hence

$$X = \sum_{k=1}^{p} \Sigma_{xx} u_k \langle u_k, X \rangle, \qquad Y = \sum_{k=1}^{q} \Sigma_{yy} v_k \langle v_k, Y \rangle$$

Reconstruction

Orthonormality constraints $u_k^T \Sigma_{xx} u_j = v_k^T \Sigma_{yy} v_j = \delta_{jk}$ for $1 \leq j \leq k-1$.

So $(u_k), (\Sigma_{xx} u_k)$ and $(v_k), (\Sigma_{yy} v_k)$ are each pairs of *dual bases*.

Hence

$$X = \sum_{k=1}^{p} \Sigma_{xx} u_k \langle u_k, X \rangle, \qquad Y = \sum_{k=1}^{q} \Sigma_{yy} v_k \langle v_k, Y \rangle$$

Probabilistic CCA: Consider the model

$$\begin{aligned} Z &\sim \mathcal{N}(0, I_d), & \min(p,q) \geq d \geq 1 \\ X|Z &\sim \mathcal{N}(W_1 Z + \mu_1, \Psi_1), & W_1 \in \mathbb{R}^{p \times d}, \Psi_1 \succcurlyeq 0 \\ Y|Z &\sim \mathcal{N}(W_2 Z + \mu_2, \Psi_2), & W_2 \in \mathbb{R}^{p \times d}, \Psi_2 \succcurlyeq 0 \end{aligned}$$

Then MLEs for W_1, W_2 are essentially matrices of canonical loadings

Practical Uses

- **High-level:** two sets of data and want to understand interactions
- **Different perspectives:**
  - stats: estimation
  - bio-informatics: algorithm for data matrices, part of pipeline...
- **Uses:** Dimension reduction, visualisation / interpretation, multi-view / self-supervised learning
- **Motivating Example:** Multi-OMICS for human microbiome measuring $p = 200$ metabolites, and $q = 800$ enzymes for $n = 500$ individuals.

Classical Estimator / Sample CCA algorithm

Just stick in sample covariances

$$\begin{aligned} &\underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \widehat{\text{Cov}}(u^\top \mathbf{X}, v^\top \mathbf{Y}) \\ &\text{subject to } \widehat{\text{Var}}(u^\top \mathbf{X}) \leq 1,\ \widehat{\text{Var}}(v^\top \mathbf{Y}) \leq 1, \\ &\qquad \widehat{\text{Cov}}(u^\top \mathbf{X}, u_j^\top \mathbf{X}) = \widehat{\text{Cov}}(v^\top \mathbf{Y}, v_j^\top \mathbf{Y}) = 0 \text{ for } 1 \leq j \leq k-1. \end{aligned}$$

Classical Estimator / Sample CCA algorithm

Just stick in sample covariances

$$
\begin{aligned}
\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \ & \widehat{\text{Cov}}(u^\top \mathbf{X}, v^\top \mathbf{Y}) \\
\text{subject to } & \widehat{\text{Var}}(u^\top \mathbf{X}) \leq 1, \ \widehat{\text{Var}}(v^\top \mathbf{Y}) \leq 1, \\
& \widehat{\text{Cov}}(u^\top \mathbf{X}, u_j^\top \mathbf{X}) = \widehat{\text{Cov}}(v^\top \mathbf{Y}, v_j^\top \mathbf{Y}) = 0 \text{ for } 1 \leq j \leq k-1.
\end{aligned}
$$

Classical Theory: asymptotic distributions, significance tests...

Classical Estimator / Sample CCA algorithm

Just stick in sample covariances

$$\begin{aligned} &\underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \widehat{\text{Cov}}(u^\top \mathbf{X}, v^\top \mathbf{Y}) \\ &\text{subject to } \widehat{\text{Var}}(u^\top \mathbf{X}) \leq 1,\ \widehat{\text{Var}}(v^\top \mathbf{Y}) \leq 1, \\ &\qquad \widehat{\text{Cov}}(u^\top \mathbf{X}, u_j^\top \mathbf{X}) = \widehat{\text{Cov}}(v^\top \mathbf{Y}, v_j^\top \mathbf{Y}) = 0 \text{ for } 1 \leq j \leq k-1. \end{aligned}$$

Classical Theory: asymptotic distributions, significance tests...

Problem: Arbitrary correlations of 1 in high dimensions ($q \geq n$).

Classical Estimator / Sample CCA algorithm

Add regularisation

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \widehat{\text{Cov}}(u^\top \mathbf{X}, v^\top \mathbf{Y}) - \tau_u \|u\|_1 - \tau_v \|v\|_1$$

$$\text{subject to } \widehat{\text{Var}}(u^\top \mathbf{X}) \leq 1, \ \widehat{\text{Var}}(v^\top \mathbf{Y}) \leq 1,$$

$$\widehat{\text{Cov}}(u^\top \mathbf{X}, u_j^\top \mathbf{X}) = \widehat{\text{Cov}}(v^\top \mathbf{Y}, v_j^\top \mathbf{Y}) = 0 \text{ for } 1 \leq j \leq k-1.$$

Classical Theory: asymptotic distributions, significance tests...

Problem: Arbitrary correlations of 1 in high dimensions ($q \geq n$).

Setting the stage

- **Main perspective:** Exploratory Data Analysis (EDA) with an eye on physical interpretation. Loosely frequentist approach.
- **Regularised CCA algorithms:**
  - ridge CCA: l2 penalty
  - sparse CCA: l1 penalty
- **Practical considerations:**
  - model / tuning parameter selection
  - number of pairs to consider
  - interpretation
- **Unanswered question:** Are these regularised CCA methods appropriate for real data? Might other structural assumptions be more natural.

Our contributions

- **New algorithm:** using **graphical lasso**; motivated by graphical models and conditional independence.
- **Practical advice:** for model comparison and interpretation motivated by the fundamental geometry of population CCA.

Our contributions

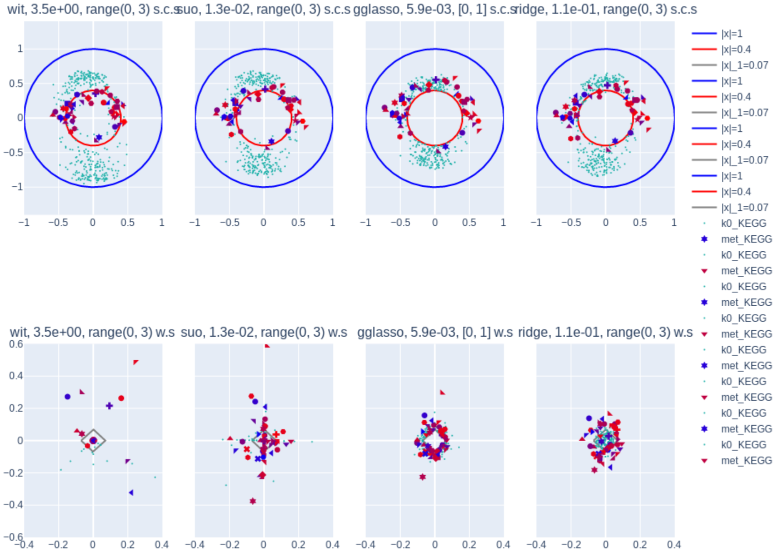
- **New algorithm:** using **graphical lasso**; motivated by graphical models and conditional independence.
- **Practical advice:** for model comparison and interpretation motivated by the fundamental geometry of population CCA.
- **Main conclusions:**
  - CCA is fundamentally a subspace problem
  - Variates and loadings are easier to estimate than weights
  - Powerful visualisation via biplots
  - Model selection is subtle
  - Graphical lasso approach works well!

Our contributions

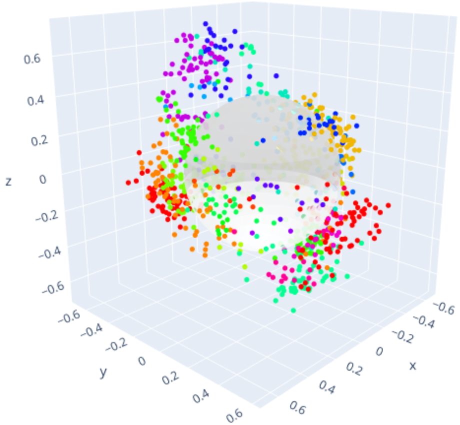
- **New algorithm:** using **graphical lasso**; motivated by graphical models and conditional independence.
- **Practical advice:** for model comparison and interpretation motivated by the fundamental geometry of population CCA.
- **Main conclusions:**
  - CCA is fundamentally a subspace problem
  - Variates and loadings are easier to estimate than weights
  - Powerful visualisation via biplots
  - Model selection is subtle
  - Graphical lasso approach works well!

These may seem natural, but are far from the accepted wisdom, and should be a welcome contribution to the field.

Questions?


wit, 3.5e+00, range(0, 3) s.c.s
suo, 1.3e-02, range(0, 3) s.c.s
gglasso, 5.9e-03, [0, 1] s.c.s
ridge, 1.1e-01, range(0, 3) s.c.s
wit, 3.5e+00, range(0, 3) w.s
suo, 1.3e-02, range(0, 3) w.s
gglasso, 5.9e-03, [0, 1] w.s
ridge, 1.1e-01, range(0, 3) w.s
|x|=1
|x|=0.4
|x|_1=0.07
k0_KEGG
met_KEGG

Pretty picture



Links with matrix analysis

$$\begin{aligned} &\underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}}\ u^T \Sigma_{xy} v \\ &\qquad \text{subject to } u^T \Sigma_{xx} u = 1,\ v^T \Sigma_{yy} v = 1, \\ &\qquad\qquad u^T \Sigma_{xx} u_j = v^T \Sigma_{yy} v_j = 0 \text{ for } 1 \le j \le k-1. \end{aligned}$$

- **Singular Value Decomposition**
  $(\rho_k, \Sigma_x^{1/2} u_k, \Sigma_y^{1/2} v_k)$ give SVD of $M := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2}$

Links with matrix analysis

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \; u^T \Sigma_{xy} v$$
$$\text{subject to } u^T \Sigma_{xx} u = 1, \; v^T \Sigma_{yy} v = 1,$$
$$u^T \Sigma_{xx} u_j = v^T \Sigma_{yy} v_j = 0 \text{ for } 1 \leq j \leq k-1.$$

- **Singular Value Decomposition**
  $(\rho_k, \Sigma_x^{1/2} u_k, \Sigma_y^{1/2} v_k)$ give SVD of $M := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2}$
- **Generalised Eigenvalue Problem:** $Aw = \lambda Bw$

$$A = \begin{pmatrix} 0 & \Sigma_{XY} \\ \Sigma_{YX} & 0 \end{pmatrix}, \quad B = \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{pmatrix}, \quad w = \begin{pmatrix} u \\ v \end{pmatrix}, \quad d = p + q.$$

Links with matrix analysis

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \; u^T \Sigma_{xy} v$$

$$\text{subject to } u^T \Sigma_{xx} u = 1, \; v^T \Sigma_{yy} v = 1,$$

$$u^T \Sigma_{xx} u_j = v^T \Sigma_{yy} v_j = 0 \text{ for } 1 \leq j \leq k-1.$$

- **Singular Value Decomposition**
  $(\rho_k, \Sigma_x^{1/2} u_k, \Sigma_y^{1/2} v_k)$ give SVD of $M := \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2}$
- **Generalised Eigenvalue Problem:** $Aw = \lambda Bw$

$$A = \begin{pmatrix} 0 & \Sigma_{XY} \\ \Sigma_{YX} & 0 \end{pmatrix}, \quad B = \begin{pmatrix} \Sigma_{XX} & 0 \\ 0 & \Sigma_{YY} \end{pmatrix}, \quad w = \begin{pmatrix} u \\ v \end{pmatrix}, \quad d = p + q.$$

- **Canonical Angles**
  $\mathcal{X} := \text{span}(\{X_j : j = 1, \dots, p\})$, $\mathcal{Y} := \text{span}(\{Y_j : j = 1, \dots, q\})$ then CCA is

$$\underset{W_k \in \mathcal{X}, Z_k \in \mathcal{Y}}{\text{maximize}} \; \langle W_k, Z_k \rangle$$

$$\text{subject to } \|W_k\|_2 \leq 1, \; \|Z_k\|_2 \leq 1,$$

$$\langle W_k, W_j \rangle = \langle Z_k, Z_j \rangle = 0 \text{ for } 1 \leq j \leq k-1.$$

Probabilistic CCA

Consider the model:

$$\begin{aligned} Z &\sim \mathcal{N}(0, I_d), & \min(p,q) \geq d \geq 1 \\ X|Z &\sim \mathcal{N}(W_1 Z + \mu_1, \Psi_1), & W_1 \in \mathbb{R}^{p\times d}, \Psi_1 \succcurlyeq 0 \\ Y|Z &\sim \mathcal{N}(W_2 Z + \mu_2, \Psi_2), & W_2 \in \mathbb{R}^{p\times d}, \Psi_2 \succcurlyeq 0 \end{aligned}$$

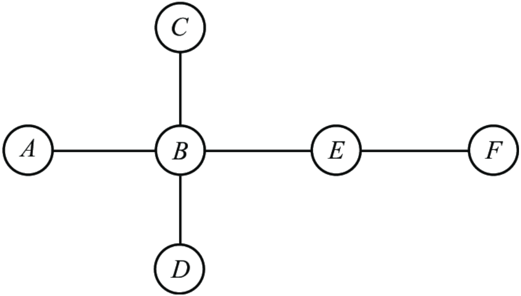
Then the MLEs of the parameters $W_1, W_2, \Psi_1, \Psi_2, \mu_1, \mu_2$ are

$$\begin{aligned} \hat{W}_1 &= \hat{\Sigma}_{xx} U_d M_1 \\ \hat{W}_2 &= \hat{\Sigma}_{yy} V_d M_2 \\ \hat{\Psi}_1 &= \hat{\Sigma}_{xx} - \hat{W}_1 \hat{W}_1^T \\ \hat{\Psi}_2 &= \hat{\Sigma}_{yy} - \hat{W}_2 \hat{W}_2^T \end{aligned}$$

and $\hat{\mu_1} = \bar{X}$, $\hat{\mu_2} = \bar{Y}$ where $M_1, M_2 \in \mathbb{R}^{d\times d}$ are arbitrary matrices with $M_1 M_2^T = R$, $\|M_1\| \leq 1, \|M_2\| \leq 1$

Graphical Models

- Setup: $X = (X^1, \dots, X^p)$ a random vector; $G = (V, E)$ graph with $V = \{1, \dots, p\}$
- Key idea: graph structure constrains distribution of X
- Gaussian graphical models: $X^i \perp X^j | (X^k)_{k \neq i,j}$ whenever $(i, j) \notin E$



Graphical Lasso

- Aim: Solve structure estimation problem
- Setup: samples $x_1, \dots, x_n$ from some zero-mean Gaussian with precision matrix $\in \mathbb{R}^{p \times p}$
- Write: $\mathbf{S} = \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T \in \mathbb{R}^{p \times p}$
- Log-likelihood:

$$l(\Omega; \mathbf{X}) = \log\det \; - \mathrm{trace}(\mathbf{S})$$

- Penalised Objective:

$$\hat{} \in \mathrm{argmax}_{\succeq \mathbf{0}} \{\log\det \; - \mathrm{trace}(\mathbf{S}) - \alpha \rho_1()\}$$

where: $\alpha \in (0, \infty)$ is a penalty parameter, $\rho_1() = \sum_{i \neq j} |\Omega_{ij}|$

Our Approach

Observations:

- $\Sigma = \Omega^{-1}$ is differentiable function of Ω.
- M is a differentiable function of Σ
- So $M = f(\Omega)$ where f is differentiable

Algorithm: Estimate Ω with graphical lasso, plug-in $\hat{M} = f(\hat{\Omega})$, then apply SVD

Our Approach

Observations:

- $\Sigma = \Omega^{-1}$ is differentiable function of Ω.
- M is a differentiable function of Σ
- So $M = f(\Omega)$ where f is differentiable

Algorithm: Estimate Ω with graphical lasso, plug-in $\hat{M} = f(\hat{\Omega})$, then apply SVD

Theorem: can combine results in the literature to see convergence rate

$$\sin \Theta(\hat{B}, B) \lesssim s_0(p) \left(\frac{\log p}{n} \right)^{1/2}$$

Existing Methods: PMD [WTH09]

Penalised Matrix Decomposition

A natural but difficult objective:

$$\begin{aligned}
&\underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \ \frac{1}{n} u^T \mathbf{X}^T \mathbf{Y} v \\
&\text{subject to } \frac{1}{n}\|\mathbf{X}u\|_2^2 \leqslant 1, \ \ \frac{1}{n}\|\mathbf{Y}v\|_2^2 \leqslant 1, \ \ \|u\|_1 \leqslant c_1, \ \ \|v\|_1 \leqslant c_2
\end{aligned}$$

Make constraints tractable:

$$\begin{aligned}
&\underset{u\in\mathbb{R}^p, v\in\mathbb{R}^q}{\text{maximize}} \ \frac{1}{n} u^T \mathbf{X}^T \mathbf{Y} v \\
&\text{subject to } \|u\|_2^2 \leqslant 1, \ \ \|v\|_2^2 \leqslant 1, \ \ \|u\|_1 \leqslant c_1, \ \ \|v\|_1 \leqslant c_2.
\end{aligned}$$

Justification: Identity approximation valid

Algorithm: alternate between solving for u, v by soft-thresholding

Existing Methods: PMD [WTH09]

Penalised Matrix Decomposition

A natural but difficult objective:

$$\begin{aligned} &\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \; \frac{1}{n} u^T \mathbf{X}^T \mathbf{Y} v \\ &\text{subject to } \frac{1}{n}\|\mathbf{X}u\|_2^2 \leqslant 1, \;\; \frac{1}{n}\|\mathbf{Y}v\|_2^2 \leqslant 1, \;\; \|u\|_1 \leqslant c_1, \;\; \|v\|_1 \leqslant c_2 \end{aligned}$$

Make constraints tractable:

$$\begin{aligned} &\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \; \frac{1}{n} u^T \mathbf{X}^T \mathbf{Y} v \\ &\text{subject to } \|u\|_2^2 \leqslant 1, \;\; \|v\|_2^2 \leqslant 1, \;\; \|u\|_1 \leqslant c_1, \;\; \|v\|_1 \leqslant c_2. \end{aligned}$$

Justification: Identity approximation valid

Algorithm: alternate between solving for u, v by soft-thresholding

Later pairs: Apply with $\frac{1}{n}\mathbf{X}^T\mathbf{Y} - \sum_{j=1}^{k-1} d_j u_j v_j^T$

Existing Methods: AMA [SMN+17]

Alternating Minimisation Algorithm

Use Lagrange multiplier penalty rather than explicit constraints:

$$\underset{u,v}{\text{minimize}} \; -\frac{1}{n}u^T\mathbf{X}^T\mathbf{Y}v + \tau_1\|u\|_1 + \tau_2\|v\|_1 \\ + \mathbb{1}\left\{u : \frac{1}{n}\|\mathbf{X}u\|_2^2 \leq 1\right\} + \mathbb{1}\left\{v : \frac{1}{n}\|\mathbf{Y}v\|_2^2 \leq 1\right\}$$

Algorithm: alternate between solving for u, v.
For v fixed get:

$$\underset{u\in\mathbb{R}^p, z\in\mathbb{R}^n}{\text{minimize}} \; \underbrace{-u^T\mathbf{X}^T\mathbf{Y}v + \tau_1\|u\|_1}_{f(u)} + \underbrace{\mathbb{1}\left\{\|z\|_2 \leq 1\right\}}_{g(z)}$$

$$\text{subject to } \mathbf{X}u = z$$

Tool: Linearized ADMM

Existing Methods: AMA [SMN+17]

Alternating Minimisation Algorithm

Use Lagrange multiplier penalty rather than explicit constraints:

$$\underset{u,v}{\text{minimize}} \; -\frac{1}{n}u^T\mathbf{X}^T\mathbf{Y}v + \tau_1\|u\|_1 + \tau_2\|v\|_1 + \mathbb{1}\left\{u : \frac{1}{n}\|\mathbf{X}u\|_2^2 \leq 1\right\} + \mathbb{1}\left\{v : \frac{1}{n}\|\mathbf{Y}v\|_2^2 \leq 1\right\}$$

Algorithm: alternate between solving for u, v.
For v fixed get:

$$\underset{u\in\mathbb{R}^p, z\in\mathbb{R}^n}{\text{minimize}} \; \underbrace{-u^T\mathbf{X}^T\mathbf{Y}v + \tau_1\|u\|_1}_{f(u)} + \underbrace{\mathbb{1}\left\{\|z\|_2 \leq 1\right\}}_{g(z)}$$

$$\text{subject to } \mathbf{X}u = z$$

Tool: Linearized ADMM

Later pairs: Add constraint $U^T\mathbf{X}^T\mathbf{X}u = 0$; $V^T\mathbf{Y}^T\mathbf{Y}v = 0$

Xiaotong Suo, Victor Minden, Bradley Nelson, Robert Tibshirani, and Michael Saunders.
Sparse canonical correlation analysis.
arXiv:1705.10865 [stat], June 2017.
arXiv: 1705.10865.

Daniela M. Witten, Robert Tibshirani, and Trevor Hastie.
A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis.
Biostatistics, 10(3):515–534, July 2009.