Canonical Correlation Analysis as Reduced Rank Regression in High Dimensions

Elena Tuzhilina

Department of Statistical Sciences, University of Toronto

elena.tuzhilina@utoronto.ca

08 June 2024

Acknowledgment

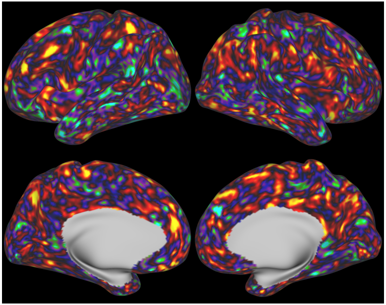
Claire Donnat

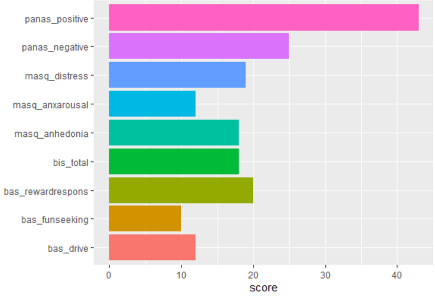
Elena Tuzhilina

The paper is available at ArXiv, the code can be accessed at GitHub.

Motivating example

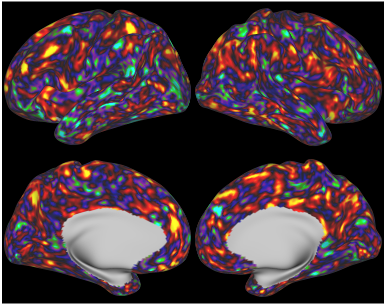
1. **brain activations**: magnetic resonance imaging obtained during a gambling task designed to probe the brain circuits underlying reward
2. **behavioral performance measures**: self-reports assessing various aspects of reward-related behaviors, depression symptoms and positive as well as negative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



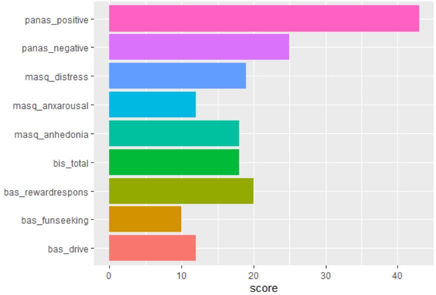


Question: is there any correlation between brain activity and behavioral measures of performance during the cognitive tasks?

Motivating example



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

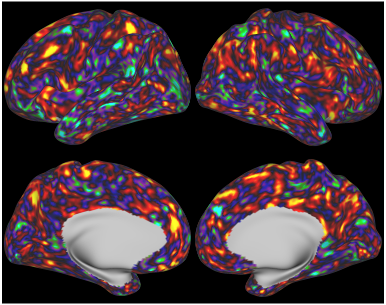


$Y \in \mathbb{R}^{n \times q}$ – behavior test scores

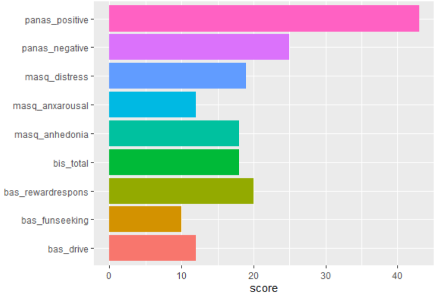
Data dimensions:

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

Motivating example



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



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

Data dimensions:

$n \ll p$ and $q \ll n$

Motivating example

Study	Objective	n	p	q
Wang et al. [2023]	X: features extracted from EEG data Y: Emotion Data	$n = 15$	$p = 310$	$q = 3$
Looden et al. [2022]	X: Connectivity features extracted from fMRI Y: Behavioural Measurements	$n = 299$	$p = 30, 3081$	$q = 3$
Luo et al. [2023]	X: Connectivity features extracted from fMRI Y: Clinical and Behavioural Measurements	$n = 122$	$p = 25, 651$	$q \in [30, 50]$
Tozzi et sl. [2021]	X: fMRI activations of each voxel in the brain Y: Questionnaire Data	$n = 269$	$p = 90, 000$	$q = 89$
Pigeau et al. [2022]	X: vertex-level cortical thickness measurements Y: Behavioural Data	$n = 25, 043$	$p = 38, 561$	$q = 52$

Table: Recent applications of high-dimensional CCA.

Canonical Correlation Analysis (CCA)

Given two random vectors $x = (x_1, \ldots, x_p)$ and $y = (y_1, \ldots, y_q)$

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \quad \lambda(u, v) = \text{cor}(u^T x, v^T y) = \frac{u^T \Sigma_{XY} v}{\sqrt{u^T \Sigma_X u}\,\sqrt{v^T \Sigma_Y v}}$$

Canonical Correlation Analysis (CCA)

Given two random vectors $x = (x_1, \ldots, x_p)$ and $y = (y_1, \ldots, y_q)$

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \quad \lambda(u, v) = \text{cor}(u^T x, v^T y) = \frac{u^T \Sigma_{XY} v}{\sqrt{u^T \Sigma_X u}\,\sqrt{v^T \Sigma_Y v}}$$

Constrained formulation: the first pair of canonical directions (u_1, v_1) is defined as

$$\underset{u \in \mathbb{R}^p,\ v \in \mathbb{R}^q}{\text{argmax}} \quad u^T \Sigma_{XY} v \ \text{ subject to } \ u^T \Sigma_{XX} u = v^T \Sigma_{YY} v = 1$$

Canonical Correlation Analysis (CCA)

Given two random vectors $x = (x_1, \dots, x_p)$ and $y = (y_1, \dots, y_q)$

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \quad \lambda(u, v) = \text{cor}(u^T x, v^T y) = \frac{u^T \Sigma_{XY} v}{\sqrt{u^T \Sigma_X u} \sqrt{v^T \Sigma_Y v}}$$

Constrained formulation: the i^{th} pair of canonical directions (u_i, v_i) is defined as

$$\underset{u \in \mathbb{R}^p,\ v \in \mathbb{R}^q}{\text{argmax}} \quad u^T \Sigma_{XY} v \quad \text{subject to} \quad u^T \Sigma_X u = v^T \Sigma_Y v = 1$$

$$u^\top \Sigma_X u_j = v^\top \Sigma_Y v_j = 0 \quad \text{for} \quad j < i$$

Canonical Correlation Analysis (CCA)

Given two random vectors $x = (x_1, \ldots, x_p)$ and $y = (y_1, \ldots, y_q)$

$$\underset{u \in \mathbb{R}^p, v \in \mathbb{R}^q}{\text{maximize}} \quad \lambda(u, v) = \text{cor}(u^T x, v^T y) = \frac{u^T \Sigma_{XY} v}{\sqrt{u^T \Sigma_X u}\,\sqrt{v^T \Sigma_Y v}}$$

Constrained formulation: the i^{th} pair of canonical directions (u_i, v_i) is defined as

$$\underset{u \in \mathbb{R}^p,\ v \in \mathbb{R}^q}{\text{argmax}} \quad u^T \Sigma_{XY} v \;\; \text{subject to} \;\; u^T \Sigma_X u = v^T \Sigma_Y v = 1$$

$$u^\top \Sigma_X u_j = v^\top \Sigma_Y v_j = 0 \;\; \text{for} \;\; j < i$$

The pairs $(u_i^\top x, v_i^\top y)$ are called the canonical variates, and the corresponding correlation values $\lambda_i = u_i^\top \Sigma_{XY} v_i$ are called the canonical correlations.

Canonical Correlation Analysis (CCA)

Find the first r directions $U = [u_1|\dots|u_r]$ and $V = [v_1|\dots|v_r]$

$$\underset{U\in\mathbb{R}^{p\times r},\ V\in\mathbb{R}^{q\times r}}{\text{maximize}} \quad \mathrm{Tr}(U^T\Sigma_{XY}V) \ \text{ subject to } \ U^T\Sigma_X U = V^T\Sigma_Y V = I_r$$

Canonical Correlation Analysis (CCA)

Transformed canonical directions $U_0 = \Sigma_X^{\frac{1}{2}} U$ and $V_0 = \Sigma_Y^{\frac{1}{2}} V$

$$\underset{U_0 \in \mathbb{R}^{p\times r},\ V_0 \in \mathbb{R}^{q\times r}}{\text{maximize}} \operatorname{Tr}\left(U_0^T \left(\Sigma_X^{-\frac{1}{2}} \Sigma_{XY} \Sigma_Y^{-\frac{1}{2}} \right) V_0 \right) \text{ subject to } U_0^\top U_0 = V_0^T V_0 = I_r$$

Solution: take singular value decomposition (SVD) of

$$\Sigma_X^{-\frac{1}{2}} \Sigma_{XY} \Sigma_Y^{-\frac{1}{2}} = U_0 \Lambda V_0^\top$$

and apply inverse transformation

$$U = \Sigma_X^{-\frac{1}{2}} U_0 \text{ and } V = \Sigma_Y^{-\frac{1}{2}} V_0$$

Canonical Correlation Analysis (CCA)

When we are given data $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$

Solution: take singular value decomposition (SVD) of

$$\widehat{\Sigma}_X^{-\frac{1}{2}} \widehat{\Sigma}_{XY} \widehat{\Sigma}_Y^{-\frac{1}{2}} = U_0 \Lambda V_0^\top$$

and apply inverse transformation

$$U = \widehat{\Sigma}_X^{-\frac{1}{2}} U_0 \text{ and } V = \widehat{\Sigma}_Y^{-\frac{1}{2}} V_0$$

This does not work for $p, q > n$.

Sparse CCA

[$r = 1$**]** Witten et al. [2009] and Parkhomenko et al. [2009] propose adding the ℓ_1 constraints.

$$\underset{u_1 \in \mathbb{R}^p,\ v_1 \in \mathbb{R}^q}{\text{maximize}} \ u_1^T \Sigma_{XY} v_1 \ \text{ subject to } \ u_1^T \Sigma_X u_1 = v_1^T \Sigma_Y v_1 = 1$$

$$\|u_1\|_1 \leq t_u \text{ and } \|v_1\|_1 \leq t_v$$

Sparse CCA

[$r = 1$**]** Witten et al. [2009] and Parkhomenko et al. [2009] propose adding the ℓ_1 constraints.

[$r > 1$**]** Wilms et al. [2015] propose to find (u_i, v_i) in a sequential progressively deflating X and Y.

$$\underset{u_1 \in \mathbb{R}^p,\ v_1 \in \mathbb{R}^q}{\text{minimize}} \|Yv_1 - Xu_1\|_2^2 + \rho_u \|u_1\|_1 + \rho_v \|v_1\|_1$$

$$\underset{u_2 \in \mathbb{R}^{p \times r},\ v_2 \in \mathbb{R}^{q \times r}}{\text{minimize}} \|Y^* v_2 - X^* u_2\|_2^2 + \rho_u \|u_2\|_1 + \rho_v \|v_2\|_1$$

$$\ldots$$

Sparse CCA

[$r = 1$**]** Witten et al. [2009] and Parkhomenko et al. [2009] propose adding the ℓ_1 constraints.

[$r > 1$**]** Wilms et al. [2015] propose to find (u_i, v_i) in a sequential progressively deflating X and Y.

[$r > 1$**]** Gao et al. [2015, 2017, 2023] propose to bound the number of non-zero rows in U and V.

$$\Sigma_{XY} = \Sigma_X U \Lambda V^\top \Sigma_Y \quad \text{subject to} \quad \begin{array}{l} U^\top \Sigma_X U = V^\top \Sigma_Y V = I_r \\ \color{red}{supp(U) \leq s_u \text{ and } supp(V) \leq s_v} \end{array}$$

Sparse CCA

[$r = 1$**]** Witten et al. [2009] and Parkhomenko et al. [2009] propose adding the ℓ_1 constraints.

[$r > 1$**]** Wilms et al. [2015] propose to find (u_i, v_i) in a sequential progressively deflating X and Y.

Fast but tend to yield poor estimates of the canonical directions.

[$r > 1$**]** Gao et al. [2015, 2017, 2023] propose to bound the number of non-zero rows in U and V.

Produce an accurate estimates but impractical on large datasets.

Sparse CCA

[$r = 1$**]** Witten et al. [2009] and Parkhomenko et al. [2009] propose adding the ℓ_1 constraints.

[$r > 1$**]** Wilms et al. [2015] propose to find (u_i, v_i) in a sequential progressively deflating X and Y.

Fast but tend to yield poor estimates of the canonical directions.

[$r > 1$**]** Gao et al. [2015, 2017, 2023] propose to bound the number of non-zero rows in U and V.

Produce an accurate estimates but impractical on large datasets.

Both assume symmetry in U and V.

ALS formulation:

$$\underset{U\in\mathbb{R}^{p\times r},\ V\in\mathbb{R}^{q\times r}}{\text{minimize}} \|YV - XU\|_F^2 \quad \text{subject to} \quad U^\top \Sigma_X U = V^\top \Sigma_Y V = I_r.$$

CCA via reduced rank regression

ALS formulation:

$$\underset{U\in\mathbb{R}^{p\times r},\ V\in\mathbb{R}^{q\times r}}{\text{minimize}} \|YV - XU\|_F^2 \quad \text{subject to} \quad U^\top \Sigma_X U = V^\top \Sigma_Y V = I_r.$$

RRR formulation: $Y_0 = Y\Sigma_Y^{-\frac{1}{2}}$ and $V_0 = V\Sigma_Y^{-\frac{1}{2}}$

$$\underset{U\in\mathbb{R}^{p\times r},\ V\in\mathbb{R}^{q\times r}}{\text{minimize}} \|Y_0 - XUV_0^T\|_F^2 \quad \text{subject to} \quad U^\top \Sigma_X U = V_0^\top V_0 = I_r.$$

CCA via reduced rank regression

ALS formulation:

$$\underset{U\in\mathbb{R}^{p\times r},\ V\in\mathbb{R}^{q\times r}}{\text{minimize}} \|YV - XU\|_F^2 \quad \text{subject to} \quad U^\top \Sigma_X U = V^\top \Sigma_Y V = I_r.$$

RRR formulation: $Y_0 = Y\Sigma_Y^{-\frac{1}{2}}$ and $V_0 = V\Sigma_Y^{-\frac{1}{2}}$, $B = UV_0^\top$

$$\underset{B\in\mathbb{R}^{p\times q}}{\text{minimize}} \|Y_0 - XB\|_F^2 \quad \text{subject to} \quad \text{rank}(B) = r$$

$$B^\top \Sigma_X B \;\text{ and }\; \Sigma_X^{\frac{1}{2}} B B^\top \Sigma_X^{\frac{1}{2}} \text{ are projection matrices}$$

CC A via reduced rank regression

Algorithm CCAR[3]

Input: $X \in \mathbb{R}^{n\times p}$, $Y \in \mathbb{R}^{n\times q}$ with $q, p \leq n$.

- Normalize $Y_0 = Y\Sigma_Y^{-\frac{1}{2}}$.
- Compute the OLS solution $B = \mathrm{argmin}_{B\in\mathbb{R}^{p\times q}} \|Y_0 - XB\|_F^2$.
- Compute the singular value decomposition $\Sigma_X^{\frac{1}{2}} B = U_0 \Lambda V_0^\top$.

Output: CCA directions $V = \Sigma_Y^{-\frac{1}{2}} V_0$ and $U = BV_0\Lambda^{-1}$

Algorithm CCAR[3] in high-dimensions

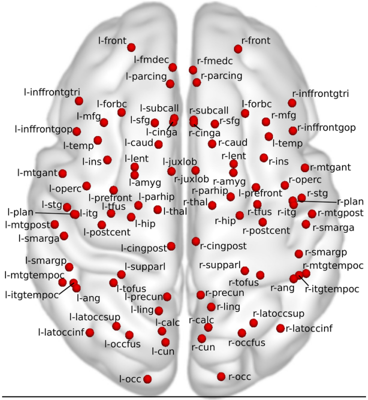
Input: $X \in \mathbb{R}^{n\times p}$, $Y \in \mathbb{R}^{n\times q}$ with $q, p \leq n$.

- Normalize $Y_0 = Y\Sigma_Y^{-\frac{1}{2}}$.
- Compute the OLS solution $B = \text{argmin}_{B\in\mathbb{R}^{p\times q}} \|Y_0 - XB\|_F^2 + \rho Pen(B)$
- Compute the singular value decomposition $\Sigma_X^{\frac{1}{2}} B = U_0 \Lambda V_0^\top$.

Output: CCA directions $V = \Sigma_Y^{-\frac{1}{2}} V_0$ and $U = BV_0\Lambda^{-1}$

$$B = \underset{B \in \mathbb{R}^{p \times q}}{\operatorname{argmin}} \|Y_0 - XB\|_F^2 + \rho Pen(B)$$

Sparse CCAR^3 $Pen(B) = \|B\|_{21}$

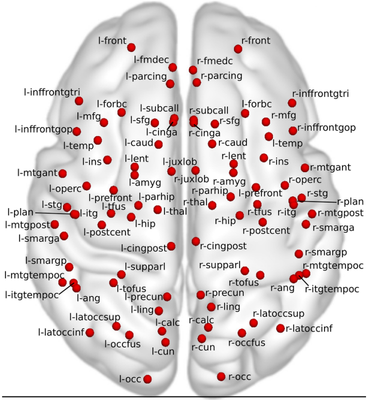


[1]

[1] Figure source: https://pubmed.ncbi.nlm.nih.gov/25188284/

$$B = \underset{B \in \mathbb{R}^{p \times q}}{\operatorname{argmin}} \| Y_0 - XB \|_F^2 + \rho Pen(B)$$

Group-sparse CCAR^3 $Pen(B) = \sum_{g \in G} \sqrt{T_g} \| B_g \|_{21}$



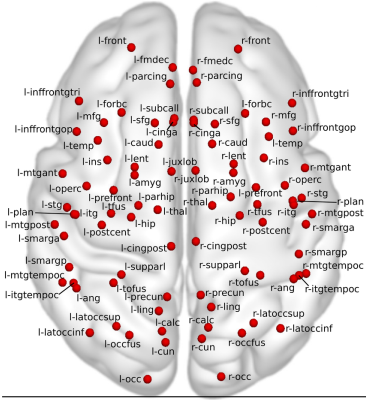
1

[1] Figure source: https://pubmed.ncbi.nlm.nih.gov/25188284/

Sparse CCAR[3]

$$B = \underset{B \in \mathbb{R}^{p \times q}}{\operatorname{argmin}} \| Y_0 - XB \|_F^2 + \rho Pen(B)$$

Graph-sparse CCAR[3] $Pen(B) = \|\Gamma B\|_{21}$



[1]

[1] Figure source: https://pubmed.ncbi.nlm.nih.gov/25188284/

Simulation: sparse CCAR^3

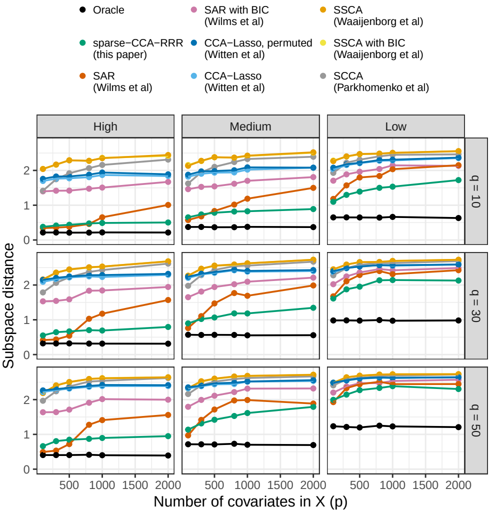


Figure: Simulation for sparse model with $Supp(B) = 10$ and $n = 500$.

Simulation: sparse CCAR[3]

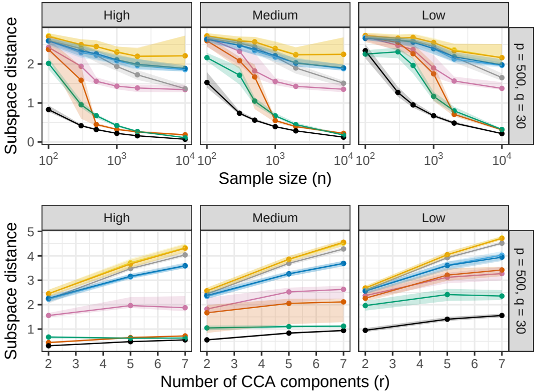


Figure: Simulation for sparse model with $Supp(B) = 10$, $q = 30$ and $p = 500$.

Simulation: group CCAR[3]

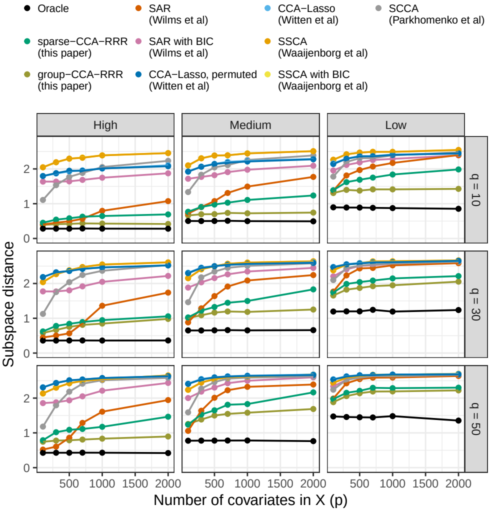


Figure: Simulation for group model with 5 nonzero groups and $n = 500$.

Simulation: group CCAR[3]

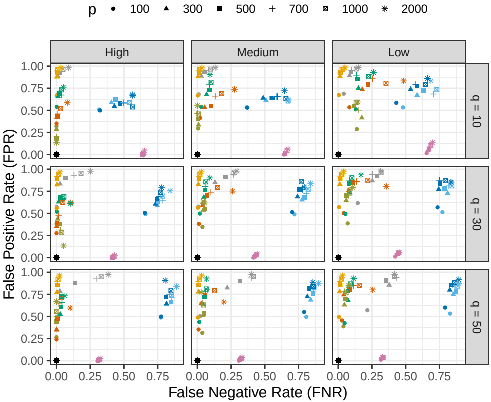


Figure: Simulation for group model with 5 nonzero groups and $n = 500$.

Simulation: graph CCAR[3]

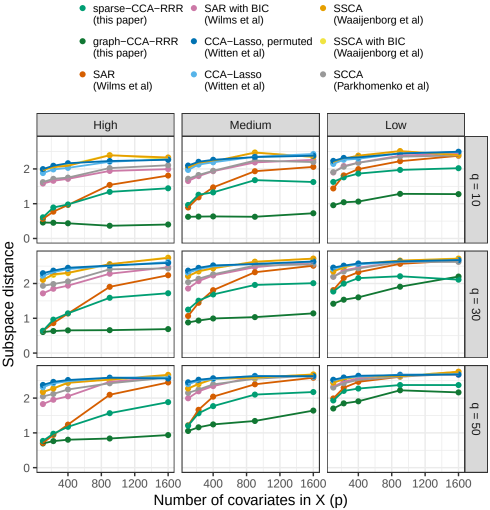


Figure: Simulation for sparse model with $Supp(\Gamma B) = 5$ and $n = 500$.

Nutrimouse dataset

Contains genomic and nutritional data from a Martin et al. [2007] study:

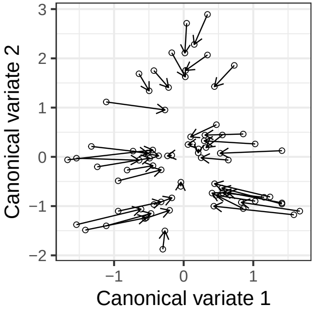
- $n = 40$ mice, each mouse was assigned to one of five diets
- X: expression profiles of $p = 120$ pre-selected genes in the liver
- Y: measurements of $q = 21$ hepatic fatty acid concentrations observed under various dietary treatments

Nutrimouse dataset

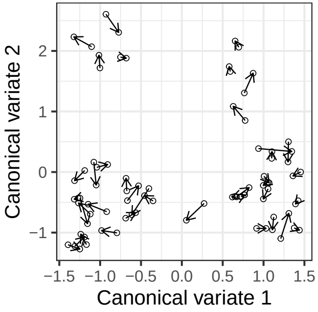
	Validation Subspace Distance	Validation Correlation	Diet Clustering Accuracy (LDA / MG)
sparse CCAR^3	**2.48**	**0.621**	**0.9** / **0.6**
SAR (Wilms et al.)	3.69	0.544	0.8 / 0.5
CCA-Lasso (Witten et al.)	5.92	0.464	0.6 / 0.525
SSCA (Waaijenborg et al.)	9.87	0.0172	0.8 / 0.475
SCCA (Parkhomenko et al.)	16.5	0.288	0.5 / 0.35

Table: Results of the 8-fold cross-validation on the Nutrimouse dataset.

Nutrimouse dataset



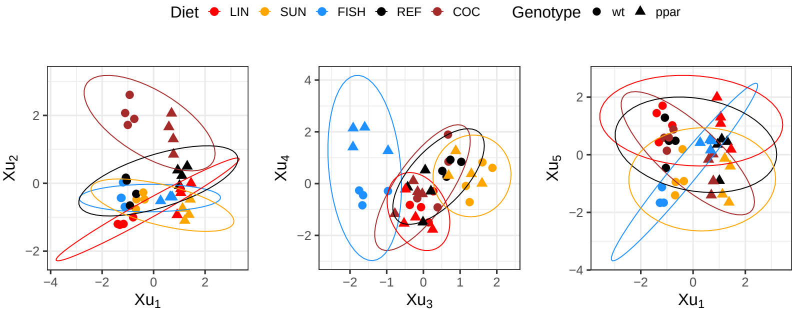
(a) SAR by Wilms et al. [2015]



(b) our method (sparse CCAR^3)

Figure: Arrow plots produced the Nutrimouse dataset. Each arrow in the plot connects the point (Xu_1, Xu_2) with its counterpart in Y-space, (Yv_1, Yv_2).

Nutrimouse dataset



(a) our method (sparse CCAR^3)

Figure: Scatter plots for canonical variates produced for the Nutrimouse dataset. The plots include 95% confidence ellipses to emphasize diet clustering.

Election dataset

Contains results of the US presidential elections of 2008, 2012, 2016 and 2020[2]..

- $n = 18$ entries representing the main election candidates
- X: the election scores in the $p = 49$ contiguous US states
- Y: the answers on $q = 10$ questions related to the candidates' program; the responses are reported as -1 if they disagree, 1 if they agree, and 0 otherwise.

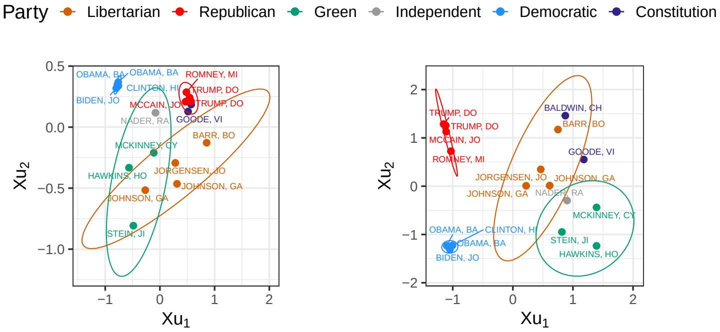
[2] The complete dataset can be found at https://www.fec.gov/data/

Election dataset

	Validation Subspace Distance	Clustering Accuracy
sparse R^3-CCA (this paper)	0.571	0.823
graph R^3-CCA (this paper)	**0.551**	**0.941**
SAR (Wilms et al.)	0.575	0.706
CCA-Lasso (Witten et al.)	1.92	0.941
SSCA (Waaijenborg et al.)	3.90	0.882

Table: Results on the leave-one-out analysis on the Election dataset.

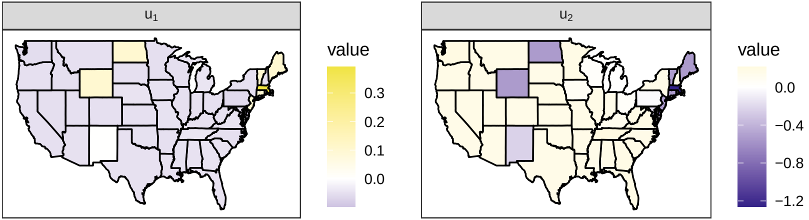
Election dataset



(a) SAR by Wilms et al. [2015] (b) our method (graph CCAR[3])

Figure: Scatter plots for 1^{st} vs 2^{nd} canonical variates computed for the Election dataset.

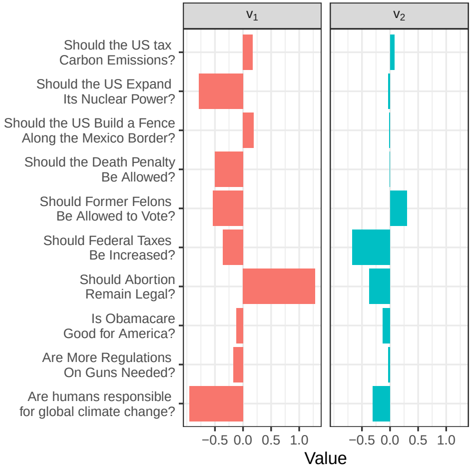
Election dataset



(a) our method (graph CCAR^3)

Figure: Canonical directions corresponding to votes of the Election dataset.

Election dataset



(a) our method (graph CCAR^3)

Figure: Canonical directions corresponding to candidate's opinion.

Neuroscience dataset

The gambling task dataset from Tozzi et al. [2021]:

- $n = 153$ individuals performing gambling task
- X: fMRI brain activations from $p = 229$ distinct brain regions of interest (ROI) in response to monetary gains versus losses
- Y: self-reported measures on $q = 9$ variables, i.e. drive, fun seeking, reward responsiveness, and overall behavioral inhibition, distress, anhedonia, and anxious arousal, positive and negative emotional states
- $|G| = 30$ groups of brain regions

Neuroscience dataset

	Validation Subspace Distance	Validation Correlation
sparse CCAR^3 (this paper)	1.74 ± 0.56	0.07 ± 0.2
group CCAR^3 (this paper)	$\mathbf{1.58 \pm 0.51}$	$\mathbf{0.12 \pm 0.19}$
graph CCAR^3 (this paper)	1.88 ± 0.66	$\mathbf{0.12 \pm 0.15}$
SAR (Wilms et al.)	1.70 ± 0.45	$\mathbf{0.12 \pm 0.22}$
CCA-Lasso (Witten et al.)	1.89 ± 0.5	-0.01 ± 0.19
SSCA (Waaijenborg et al.)	1.93 ± 0.48	-0.1 ± 0.18
SCCA (Parkhomenko et al.)	1.81 ± 0.53	0.03 ± 0.2

Table: Results of the 20-fold cross-validation on the Neuroscience dataset.

Neuroscience dataset

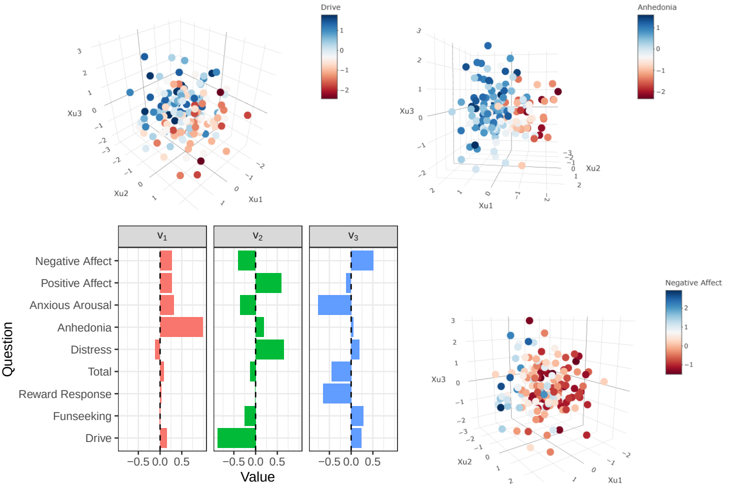


Figure: Results on the Neuroscience dataset using group CCAR^3. Scatter plots: canonical variates $(X\widehat{u}_i)_{i=1}^3$. Bar plots: canonical directions representing the questionnaire.

Neuroscience dataset

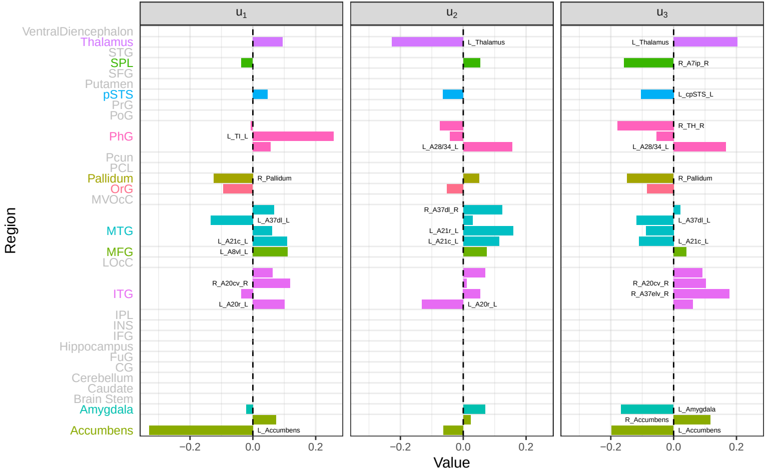


Figure: Results on the Neuroscience dataset using group CCAR^3. Top 20 brain activation directions $(\widehat{u}_i)_{i=1}^3$ grouped along the y-axis and colored according to the brain gyri.

Thank you for your attention!

References

C. Gao, Z. Ma, Z. Ren, and H. H. Zhou, *Minimax estimation in sparse canonical correlation analysis*, The Annals of Statistics, 43 (2015), pp. 2168–2197.

C. Gao, Z. Ma, and H. H. Zhou, *Sparse CCA: Adaptive estimation and computational barriers*, The Annals of Statistics, 45 (2017), pp. 2074–2101.

S. Gao and Z. Ma, *Sparse gca and thresholded gradient descent*, Journal of Machine Learning Research, 24 (2023), pp. 1–61.

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

P. G. Martin, H. Guillou, F. Lasserre, S. Déjean, A. Lan, J.-M. Pascussi, M. SanCristobal, P. Legrand, P. Besse, and T. Pineau, *Novel aspects of ppar α-mediated regulation of lipid and xenobiotic metabolism revealed through a nutrigenomic study*, Hepatology, 45 (2007), pp. 767–777.

E. Parkhomenko, D. Tritchler, and J. Beyene, *Sparse canonical correlation analysis with application to genomic data integration*, Statistical applications in genetics and molecular biology, 8 (2009).

L. Tozzi, E. Tuzhilina, M. F. Glasser, T. J. Hastie, and L. M. Williams, *Relating whole-brain functional connectivity to self-reported negative emotion in a large sample of young adults using group regularized canonical correlation analysis*, Neuroimage, 237 (2021), p. 118137.

S. Waaijenborg and A. H. Zwinderman, *Sparse canonical correlation analysis for identifying, connecting and completing gene-expression networks*, BMC Bioinformatics, 10 (2009).

I. Wilms and C. Croux, *Robust sparse canonical correlation analysis*, BMC Systems Biology, 10 (2016).

D. M. Witten and R. J. Tibshirani, *Extensions of sparse canonical correlation analysis with applications to genomic data*, Statistical applications in genetics and molecular biology, 8 (2009).