Probabilisic Joint and Individual Variation Explained (ProJIVE) for Data Integration

Benjamin B. Risk

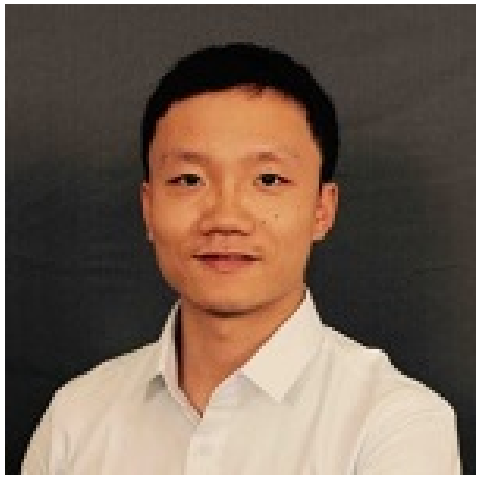
Department of Biostatistics & Bioinformatics
Rollins School of Public Health
Emory University

benjamin.risk@emory.edu

github.com/thebrisklab

Joint work

Raphiel Murden (Biostatistics, Emory), Ganzhong Gavin Tian (Biostatistics, Emory), Deqiang Qiu (Department of Radiology, Emory University School of Medicine; Department of Biomedical Engineering, Georgia Tech) and Benjamin Risk.


Data fusion / data integration

- Goal: find information shared by multiple datasets collected on the same participants.
- Subject scores can be used to summarize a subject's phenotype from multiple data sources (dimension reduction).
- Subject scores can be investigated as possible biomarkers in behavior and neurological disorders [Sui et al., 2011].
- Find shared information from neuroimaging and behavioral/clinical data.
- Application: brain morphometry and cognitive test batteries from the Alzheimer's Disease Neuroimaging Initiative.

Previous approach: JIVE

Joint & **I**ndividual **V**ariation **E**xplained
[Lock et al., 2013, Feng et al., 2018]: shared information in subject score subspaces $\in \mathbb{R}^n$.

Consider $\mathbf{X}_k \in \mathbb{R}^{p_k \times n}$, where p_k is the number of features/variables in the k^{th} data set for $k = 1, ..., K$.

$$\mathbf{X}_k = \mathbf{J}_k + \mathbf{A}_k + \mathbf{E}_k, \text{ for } k = 1, \ldots, K.$$

- Signal $\mathbf{B}_k = \mathbf{J}_k + \mathbf{A}_k$ is low rank
- $\text{Row}(\mathbf{J}_k) = \text{Row}(\mathbf{J}_{k'})$ for all $k, k' \in \{1, \ldots, K\}$
- $\text{Row}(\mathbf{J}_k) \perp \text{Row}(\mathbf{A}_k)$
- $\cap_{k=1}^{K} \text{Row}(\mathbf{A}_k) = \mathbf{0}$.
- $\mathbf{E}_k$ is isotropic (singular values are approximately equal).

Limitations of current approaches

- Defining joint and individual structure in terms of subspaces can be challenging to understand.
- A probabilistic model may improve interpretation.
- AJIVE estimates individual subspaces after estimation of the joint subspace.
- Maximum likelihood framework for simultaneously estimating joint and individual subspaces may improve accuracy.

Motivating method: Probabilistic PCA

Probabilistic PCA [Tipping and Bishop, 1999]: let $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, \ldots, n$:

$$\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{W}\mathbf{z}_i + \boldsymbol{e}_i$$
$$\mathbf{z}_i \overset{iid}{\sim} N(0, \mathbf{I})$$
$$\boldsymbol{e}_i \overset{iid}{\sim} N(0, \sigma^2\mathbf{I})$$

- Low rank signal.
- Factor analysis with isotropic noise.
- The MLE corresponds to the classic PCA solution.
- Although solution can be derived from Gaussian assumptions, it applies much more generally.

Probabilistic JIVE

We propose Probabilistic JIVE (ProJIVE).

Let $\mathbf{x}_{ik} \in \mathbb{R}^{p_k}$, $k = 1, \ldots, K$, $i = 1, \ldots, n$.

$\mathbf{W}_{Jk} \in \mathbb{R}^{p_k \times r_J}$, $\mathbf{W}_{Ik} \in \mathbb{R}^{p_k \times r_{Ik}}$

$$\mathbf{x}_{ik} = \mathbf{W}_{Jk}\mathbf{z}_i + \mathbf{W}_{Ik}\mathbf{b}_{ik} + \boldsymbol{\epsilon}_{ik},$$

$$(\mathbf{z}_i^\top, \mathbf{b}_{i1}^\top, \ldots, \mathbf{b}_{iK}^\top)^\top \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}),$$

$$\boldsymbol{\epsilon}_{ik} \overset{iid}{\sim} N(\mathbf{0}, \sigma_k^2 \mathbf{I}),\ \mathrm{Cov}\left(\boldsymbol{\epsilon}_{ik}, \boldsymbol{\epsilon}_{ik'}\right) = \mathbf{0},$$

$$\mathrm{Cov}\left[(\mathbf{z}_i^\top, \mathbf{b}_{i1}^\top, \ldots, \mathbf{b}_{iK}^\top)^\top, \boldsymbol{\epsilon}_{ik}\right] = \mathbf{0},$$

$$\mathrm{rank}(\mathbf{W}_{Jk}) = r_J,\ \mathrm{rank}(\mathbf{W}_{Jk} + \mathbf{W}_{Ik}) < p_k, \quad k, k' = 1, \ldots, K.$$

Inter-battery factor analysis

It turns out a similar model was proposed in [Tucker, 1958]: inter-battery factor analysis.

An MLE for the joint signal ignoring individual components was derived in [Browne, 1979].

More recently, Bayesian Canonical Correlation analysis with variational inference was proposed [Klami et al., 2013] and the related group factor analysis [Klami et al., 2015]. Use shrinkage to approximate block-wise sparsity.

For the case of $K = 2$:

$$\text{Cov}\begin{pmatrix}\mathbf{x}_{i1}\\ \mathbf{x}_{i2}\end{pmatrix} = \begin{pmatrix}\mathbf{W}_{J1}\mathbf{W}_{J1}^\top + \mathbf{W}_{I1}\mathbf{W}_{I1}^\top + \sigma_1^2\mathbf{I} & \mathbf{W}_{J1}\mathbf{W}_{J2}^\top \\ \mathbf{W}_{J2}\mathbf{W}_{J1}^\top & \mathbf{W}_{J2}\mathbf{W}_{J2}^\top + \mathbf{W}_{I2}\mathbf{W}_{I2}^\top + \sigma_2^2\mathbf{I}\end{pmatrix}$$
$$= \mathbf{C}.$$

For $\boldsymbol{\mu}_k = \mathbf{0}$, the log-likelihood of the data is

$$\ell = -\frac{n}{2}\left\{(p_1 + p_2)\log(2\pi) + \log(|\mathbf{C}|) + tr(\mathbf{C}^{-1}\mathbf{S})\right\}$$

where $\mathbf{S} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$.

ProJIVE Identifiability

Theorem

Suppose $K = 2$ and let f_Φ define the multivariate normal density with parameters $\{\mathbf{W}_{J1}, \mathbf{W}_{J2}, \mathbf{W}_{I1}, \mathbf{W}_{I2}, \sigma_1^2, \sigma_2^2\}$ in the ProJIVE model. Let f_{Φ^*} denote the MVN with parameters $\{\mathbf{W}^*_{J1}, \mathbf{W}^*_{J2}, \mathbf{W}^*_{I1}, \mathbf{W}^*_{I2}, \sigma_1^{*2}, \sigma_2^{*2}\}$. Then $f_\Phi = f_{\Phi^*}$ if and only if

1. *(identifiability of error variance) $\sigma_1^{*2} = \sigma_1^2,\ \sigma_2^{*2} = \sigma_2^2$,*
2. *(identifiability of joint loadings up to certain linear transformations) There exists an $r_J \times r_J$ non-singular matrix $\mathbf{T}_1$ such that $\mathbf{W}^*_{J1} = \mathbf{W}_{J1}\mathbf{T}_1$ and $\mathbf{W}^*_{J2} = \mathbf{W}_{J2}(\mathbf{T}_1^{-1})^\top$ and*

$$\mathbf{W}_{I1}\mathbf{W}_{I1}^\top + \mathbf{W}_{J1}(\mathbf{I} - \mathbf{T}_1\mathbf{T}_1^\top)\mathbf{W}_{J1}^\top \succeq 0,$$
$$\mathbf{W}_{I2}\mathbf{W}_{I2}^\top + \mathbf{W}_{J2}(\mathbf{I} - (\mathbf{T}_1^{-1})^\top\mathbf{T}_1^{-1})\mathbf{W}_{J2}^\top \succeq 0,$$

3. *(individual components are not identifiable) $\mathbf{W}^*_{I1}$ is defined such that $\mathbf{W}^*_{I1}(\mathbf{W}^*_{I1})^\top = \mathbf{W}_{I1}\mathbf{W}_{I1}^\top + \mathbf{W}_{J1}(\mathbf{I} - \mathbf{T}_1\mathbf{T}_1^\top)\mathbf{W}_{J1}^\top$, and $\mathbf{W}^*_{I2}$ is defined such that $\mathbf{W}^*_{I2}(\mathbf{W}^*_{I2})^\top = \mathbf{W}_{I2}\mathbf{W}_{I2}^\top + \mathbf{W}_{J2}(\mathbf{I} - (\mathbf{T}_1^{-1})^\top\mathbf{T}_1^{-1})\mathbf{W}_{J2}^\top$.*

ProJIVE Identifiability, part II

Lemma

Suppose $K = 2$ *and assume* $Col(\mathbf{W}_{Jk}) \cap Col(\mathbf{W}_{Ik}) = \{\mathbf{0}\}$, $k = 1, 2$, *and assume* $rank(\mathbf{W}_{Ik}) = r_{Ik}$. *Suppose there exists a* $\mathbf{T}_1$ *as defined in Theorem 1 such that* $f_\Phi = f_{\Phi^*}$. *Then a)* $\mathbf{T}_1 \in \mathcal{O}$, *and as a consequence, the joint loadings are identifiable up to orthogonal transformations; and b) the individual loadings are identifiable up to orthogonal transformations.*

ProJIVE Identifiability, part III

Theorem

Suppose $K > 2$ *and consider parameter sets* $\{\mathbf{W}_{Jk}, \mathbf{W}_{Ik}, \sigma_k^2, k = 1, \ldots, K\}$ *and* $\{\mathbf{W}^*_{Jk}, \mathbf{W}^*_{Ik}, \sigma_k^{*2}, k = 1, \ldots, K\}$ *in the ProJIVE model. Then* $f_\Phi = f_{\Phi^*}$ *if and only if for* $k = 1, \ldots, K$

1. *(identifiability of error variance)* $\sigma_k^{*2} = \sigma_k^2$,
2. *(identifiability of joint components up to orthogonal transformations)* $\mathbf{W}^*_{Jk} = \mathbf{W}_{Jk}\mathbf{O}_J$ *for* $\mathbf{O}_J \in \mathcal{O}$,
3. *(identifiability of individual components up to orthogonal transformations)* $\mathbf{W}^*_{Ik} = \mathbf{W}_{Ik}\mathbf{O}_k$ *for* $\mathbf{O}_k \in \mathcal{O}$.

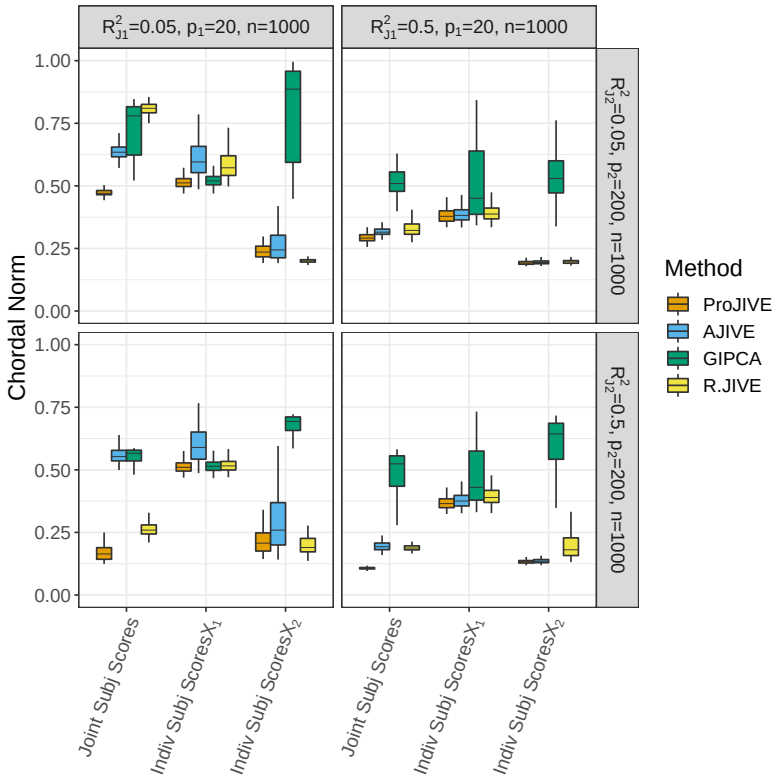
Estimation

- We derived an EM algorithm to fit this model.
- Closed form for M-step.
- Computational costs: more than AJIVE but usually less than r.jive and generalized integrative PCA [Zhu et al., 2020].
- Selecting number of components: for signal rank, use screeplots on separate PCA, then for joint scores, use a permutation test detecting significant correlation between PC scores.
- LRTs or BIC possible but may be computationally costly.

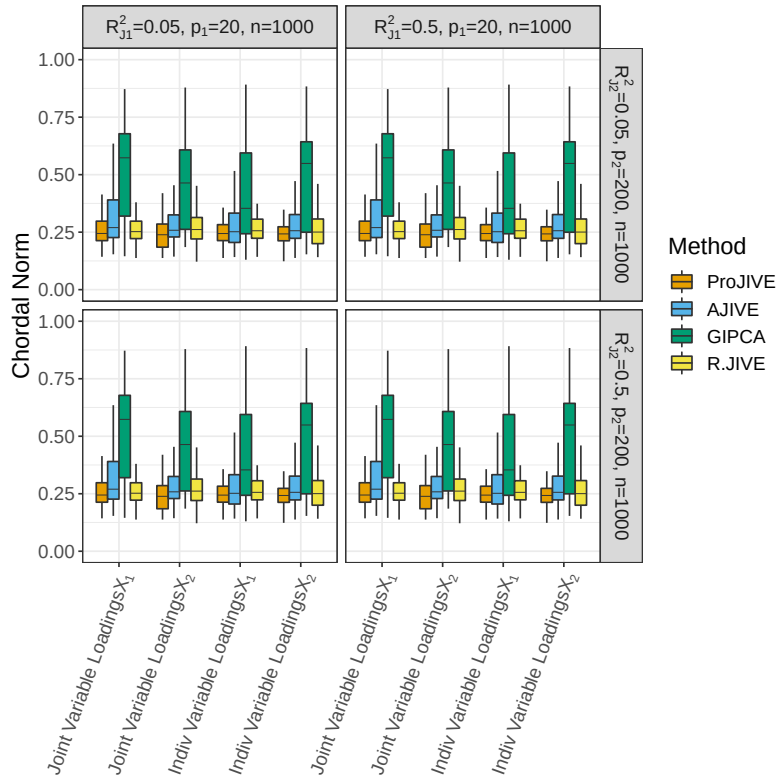
Simulations

1. p_1=20 and $p_2 = 200$, $r_J = 3$, $r_{I1} = r_{I2} = 2$.
2. Joint Variation Explained in $\mathbf{X}_1$: (a) $R^2_{J1} = 0.05$ and (b) $R^2_{J1} = 0.5$.
3. Joint Variation Explained in $\mathbf{X}_2$: (a) $R^2_{J2} = 0.05$ and (b) $R^2_{J2} = 0.5$.
4. Data generating distributions: (a) Gaussian scores and loadings and (b) mixture of Gaussian joint scores ($\pi_1 = 0.2$, $\mu_1 = -4$, unit variance; $\pi_2 = 0.50$, $\mu_2 = 0$; and $\pi_3 = 0.30$, $\mu_3 = 4$) and Rademacher loadings (joint and individual).
5. $n = 1000$.
6. $R^2_{I1} = R^2_{I2} = 0.25$.

Simulations: Gaussian assumptions met



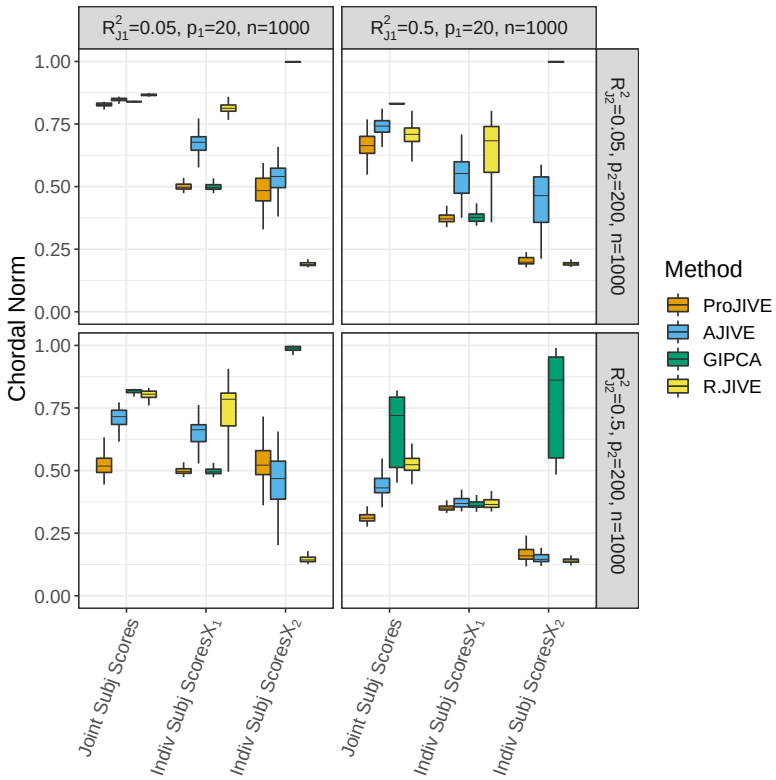
(a)



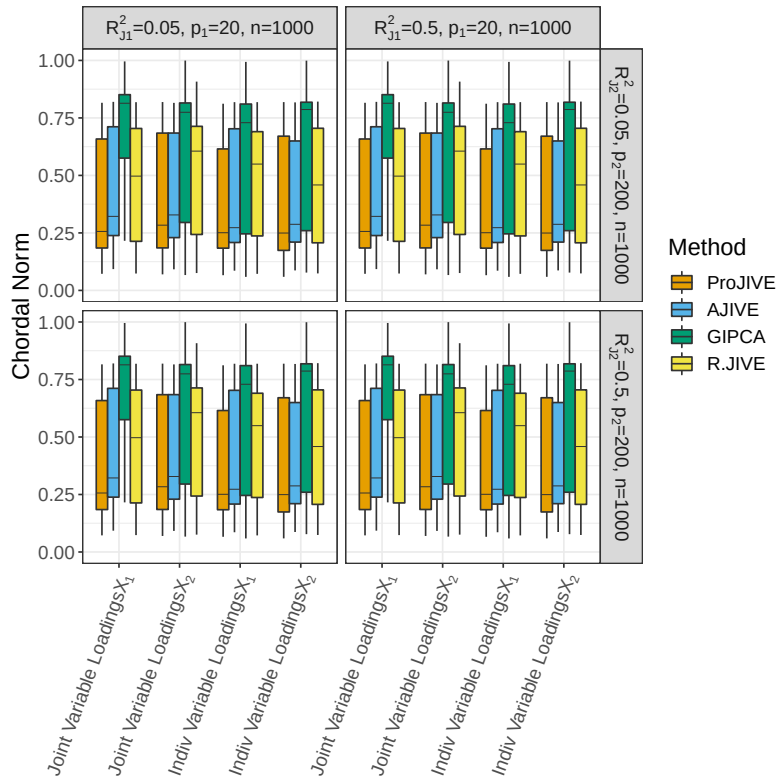
(b)

Figure: Gaussian scores and loadings (generating mechanism = model). a) subject scores for $p_1 = 20$ and $p_2 = 200$; b) variable loadings. ProJIVE: Our method. AJIVE: [Feng et al., 2018]. R.JIVE: [Lock et al., 2013]. GIPCA: [Zhu et al., 2020].

Simulations: robustness



(a)



(b)

Figure: MOG scores and rademacher loadings (generating mechanism $\neq$ model). a) subject scores for $p_1 = 20$ and $p_2 = 200$; b) variable loadings.

Brain morphometry and cognition in ADNI

- Preprocessed data from the Alzheimer's Disease Prediction of Longitudinal Evolution (TADPOLE) Challenge (Alzheimer's Disease Neuroimaging Initiative data) in 2018.
- Single time point with n=587 adults with both cognition and T1 imaging ($K = 2$).
- Dataset 1: standardized cognition variables (CDR-SB, ADAS, MMSE, RAVLT, MOCA, ECOG). $p_1 = 22$.
- Dataset 2: cortical thickness, surface area, and cortical volume for 34 ROIs per hemisphere [Desikan et al., 2006], along with cortical volumes for other regions/structures. $p_2 = 245$.
- For each feature, regressed out age and sex, then standardized residuals.
- We also conducted analysis with $K = 5$ treating each morphometry as a separate dataset: results were similar.

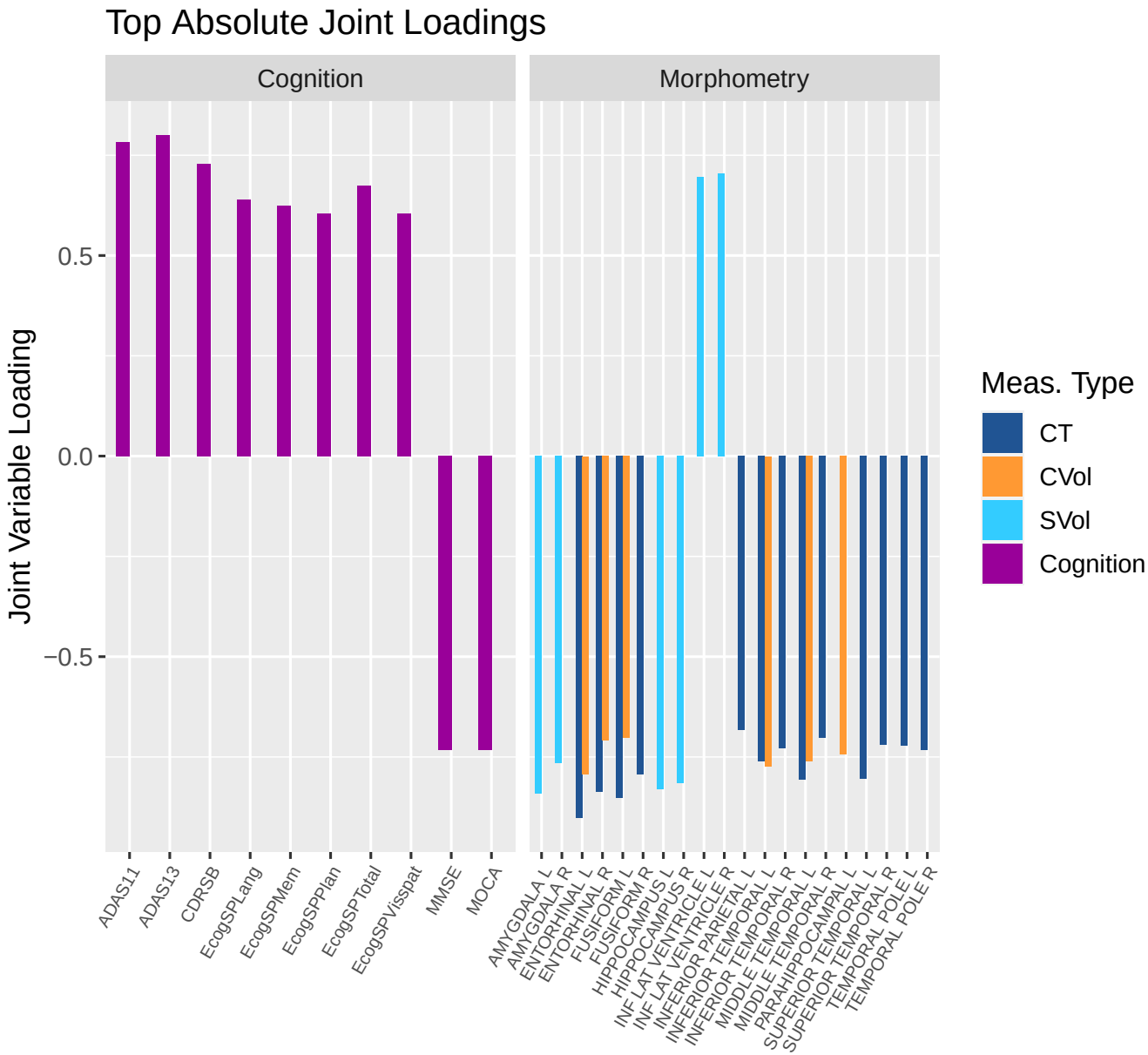
Table: Summary statistics for selected covariates of participants in ADNI-GO and ADNI2.

	AD (N=88)	MCI (N=340)	CN (N=159)	Total (N=587)
	Mean (S.D.) or N (%)			
Age				
	74.0 (7.92)	71.5 (7.57)	72.8 (5.85)	72.2 (7.25)
Gender				
Female	28 (31.8%)	150 (44.1%)	84 (52.8%)	262 (44.6%)
ApoE4				
0	21 (23.9%)	178 (52.4%)	111 (69.8%)	310 (52.8%)
1	45 (51.1%)	126 (37.1%)	46 (28.9%)	217 (37.0%)
2	22 (25.0%)	36 (10.6%)	2 (1.3%)	60 (10.2%)

Joint and individual variance

- Screeplots: cognition $r_1 = 5$; morphometry $r_2 = 10$.
- Joint rank from permutation test: $r_J = 1$.
- Proportion of variance explained by the joint signal: 0.21 for cognition; 0.10 for morphometry.
- Individual proportions: 0.50 for cognition; 0.36 for morphometry.

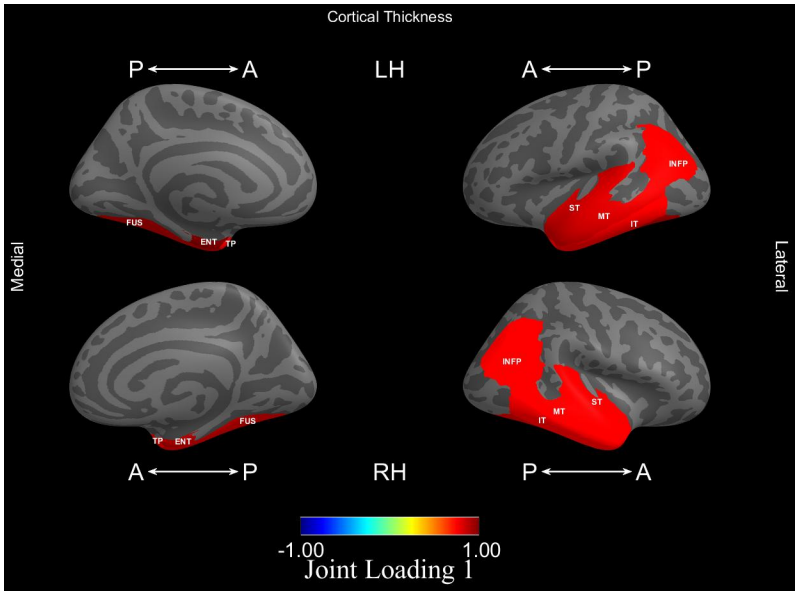
Joint Loadings



(a)

Joint loadings: morphometry

(b)



(c)

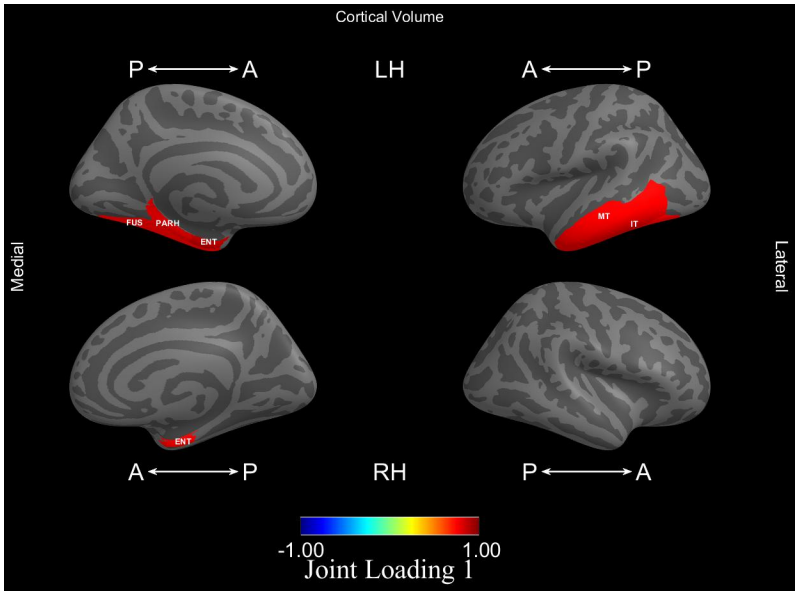
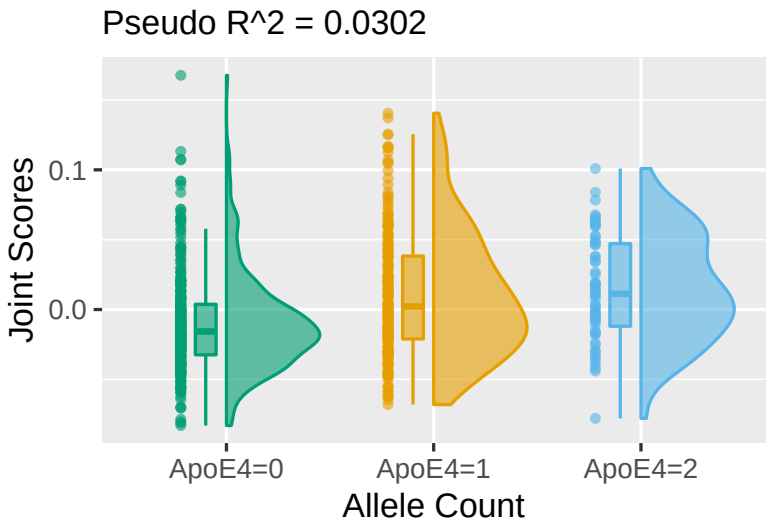


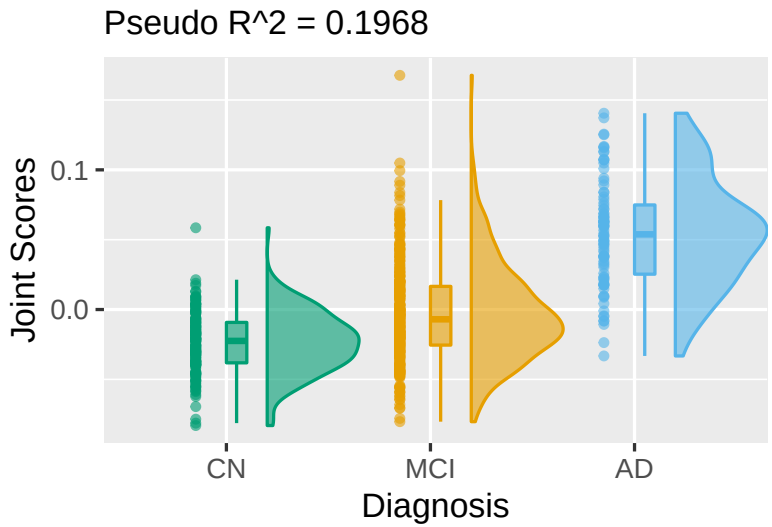
Figure: (b) 90th percentile of absolute brain loadings which occur in measures of cortical thickness. (c) 90th percentile of absolute brain loadings which occur in measures of cortical volume.

Joint scores and external variables

To gain insight into joint scores, relate to other measures:



(a)



(b)

Figure: Joint subject scores estimated via ProJIVE show separation by (a) the count of ApoE4 allele counts and (b) diagnosis via raincloud plots.

Joint scores and external variables

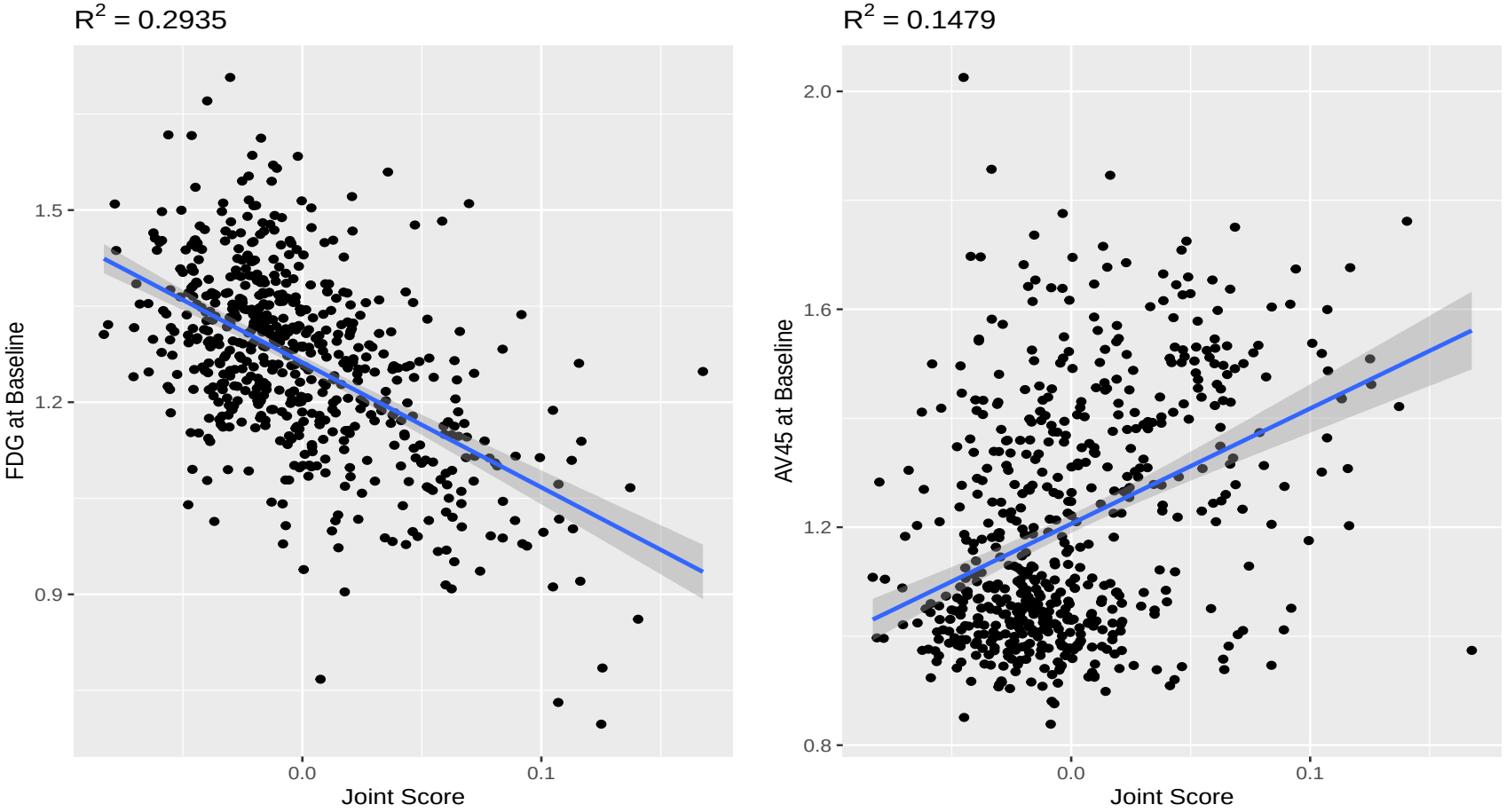


Figure: Can we use cheap, non-invasive MRI instead of expensive PET? PET uses radioactive tracers. FDG PET: measure of brain metabolism and indicative of neurodegeneration. AV45 PET: measures amyloid beta. Used to diagnose AD

Conclusions

- We propose a probablistic model for JIVE, called ProJIVE.
- Intuitive latent variable formulation.
- Extends probabilistic PCA to multiple datsets.
- EM algorithm to simultaneously estimate joint and individual subspaces.
- In simulations, improves accuracy of both joint and individual scores over existing JIVE methods in many settings.
- Joint loadings extract brain regions associated with cognition and dementia.
- Joint scores are related to biomarkers of AD and dementia, including APOE4, FDG PET, and AV45 PET.

Acknowledgments

- Thank you!
- Data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`
- This study was supported by R21AG066970 (B.R.), R21AG064405 (D.Q.), R01AG072603 (D.Q.), and an ADRC pilot grant (D.Q. and B.R.) from parent grant P30AG066511.

References I

Browne, M. W. (1979).

The maximum-likelihood solution in inter-battery factor analysis.
British Journal of Mathematical and Statistical Psychology, 32(1):75–86.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., and others (2006).

An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.
Neuroimage, 31(3):968–980.

Feng, Q., Jiang, M., Hannig, J., and Marron, J. S. (2018).

Angle-based joint and individual variation explained.
Journal of Multivariate Analysis, 166:241–265.

Klami, A., Virtanen, S., and Kaski, S. (2013).

Bayesian Canonical Correlation Analysis.
Journal of Machine Learning Research, 14:965–1003.

Klami, A., Virtanen, S., Leppaaho, E., and Kaski, S. (2015).

Group Factor Analysis.
IEEE Transactions on Neural Networks and Learning Systems, 26(9):2136–2147.

Lock, E. F., Hoadley, K. A., Marron, J. S., and Nobel, A. B. (2013).

Joint and individual variation explained (JIVE) for integrated analysis of multiple data types.
The annals of applied statistics, 7(1):523.

Sui, J., Pearlson, G., Caprihan, A., Adali, T., Kiehl, K. A., Liu, J., Yamamoto, J., and Calhoun, V. D. (2011).

Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model.
NeuroImage, 57(3).

References II

Tipping, M. E. and Bishop, C. M. (1999).

Probabilistic principal component analysis.

Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61(3):611–622.

Tucker, L. R. (1958).

An inter-battery method of factor analysis.

Psychometrika 1958 23:2, 23(2):111–136.

Zhu, H., Li, G., and Lock, E. F. (2020).

Generalized integrative principal component analysis for multi-type data with block-wise missing structure.

Biostatistics, 21(2):302–318.