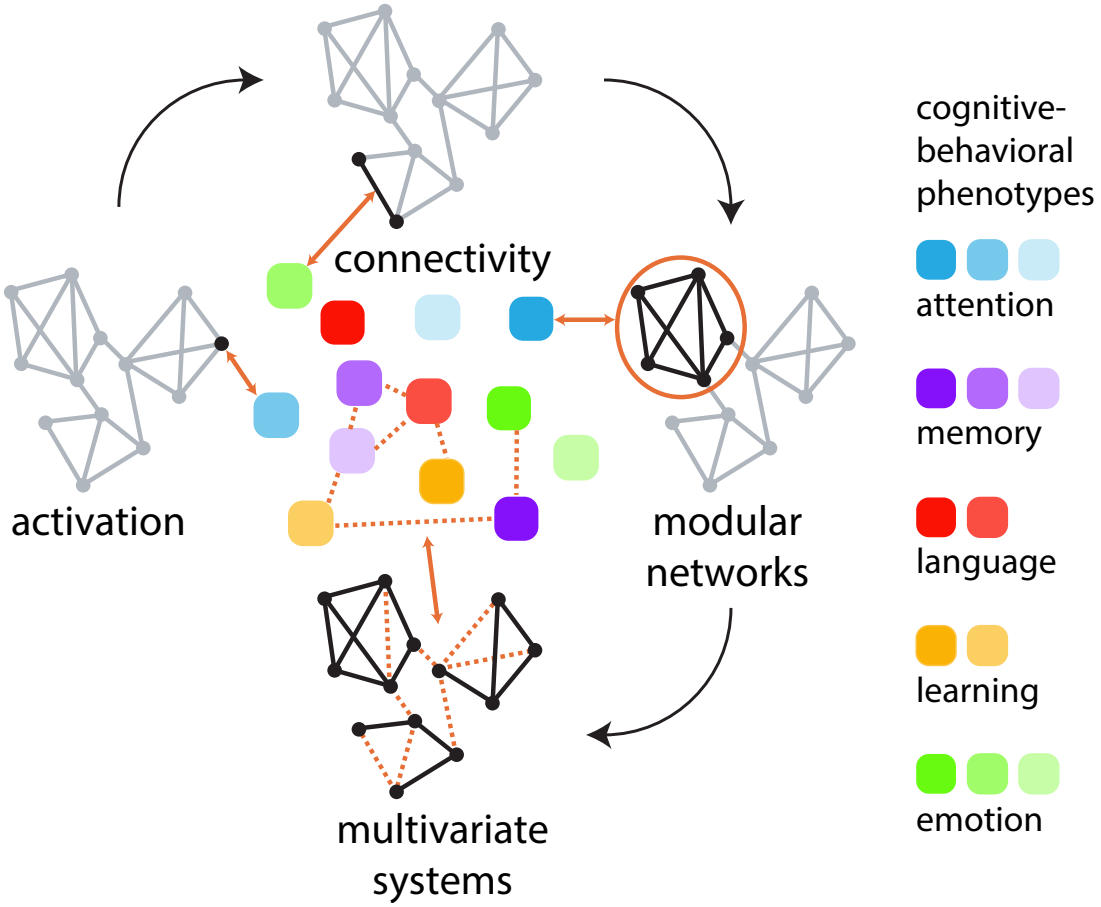
Multivariate statistical analyses for MEG data

Bratislav Mišić

Montréal Neurological Institute

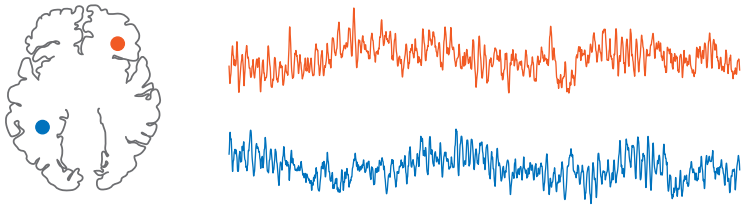
November 7th 2016

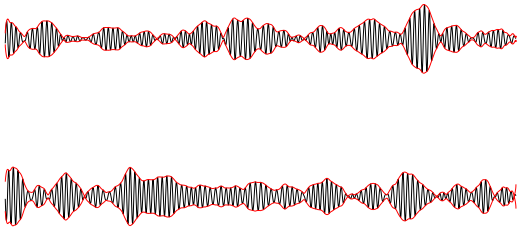
Towards multivariate analysis



Mišić & Sporns (2016) *Curr Opin Neurobiol*

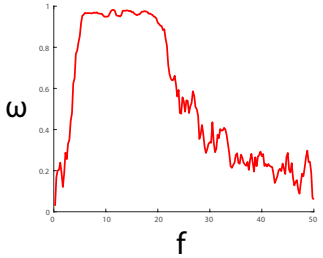
Estimating functional connectivity





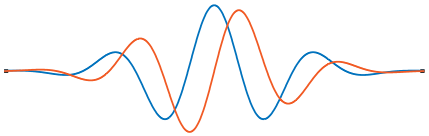
amplitude coupling

Brookes et al (2012) NeuroImage
Hipp et al (2012) Nat Neurosci



spectral coupling

Nolte et al (2004) J Clin Neurophysiol
Baccala & Sameshima (2001) Biol Cybern
Kaminski & Blinowska (1991) Biol Cybern



phase coupling

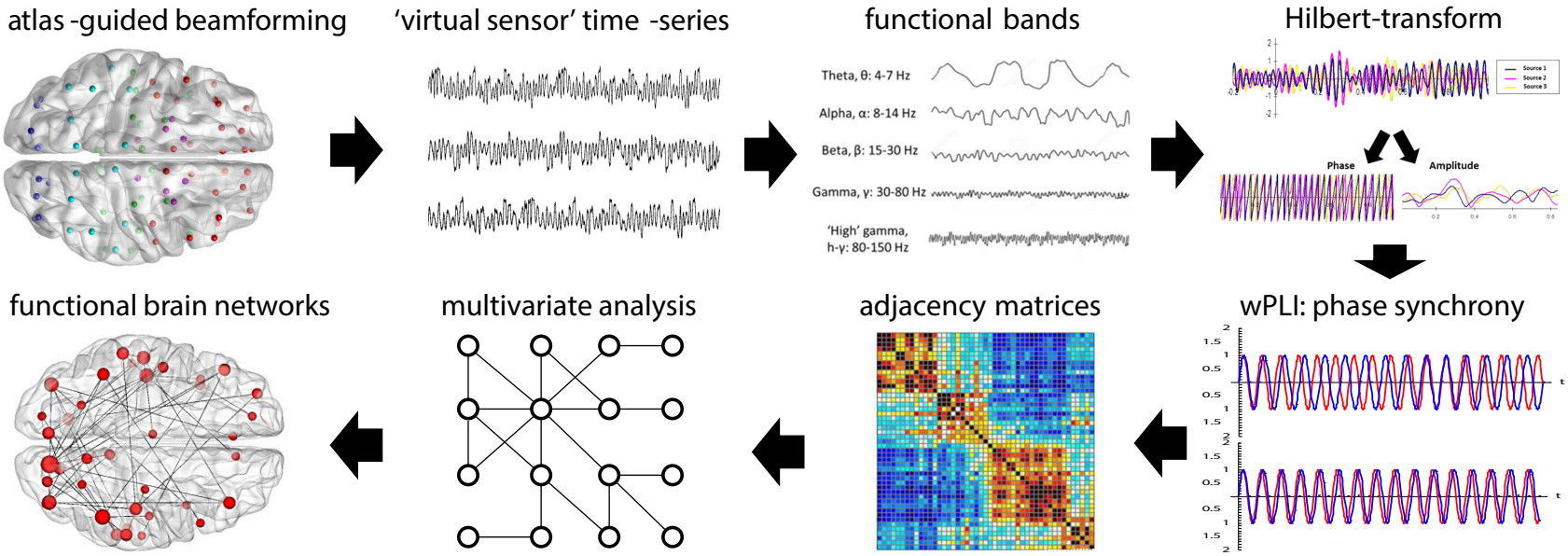
Lachaux et al (1999) Hum Brain Mapp
Stam et al (2007) Hum Brain Mapp
Vinck et al (2011) NeuroImage

Niso et al. (2013) *Neuroinformatics*

Bastos & Schoffelen (2016) *Front Syst Neurosci*

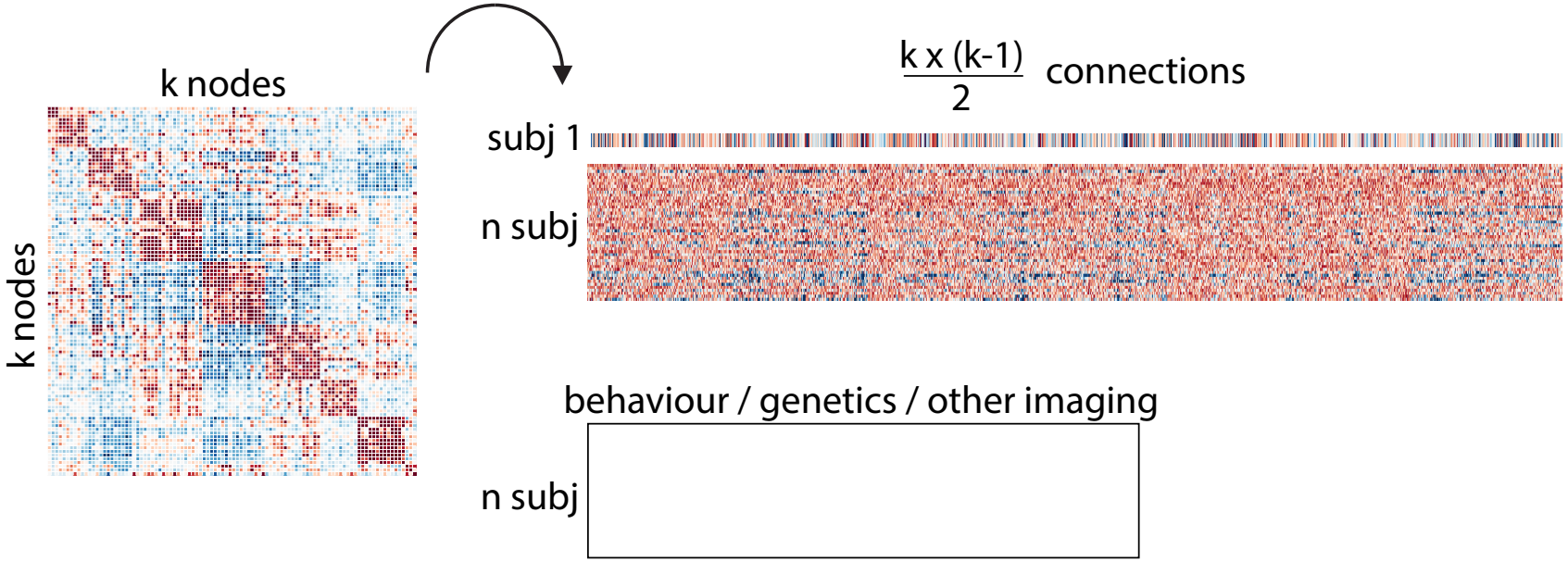
Colcough et al. (2016) *NeuroImage*

Constructing functional networks



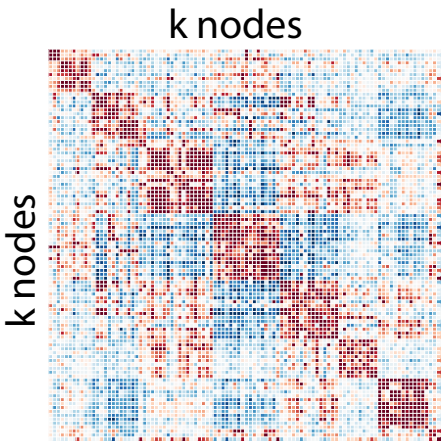
Dunkley et al. (2015) *NeuroImage: Clinical*

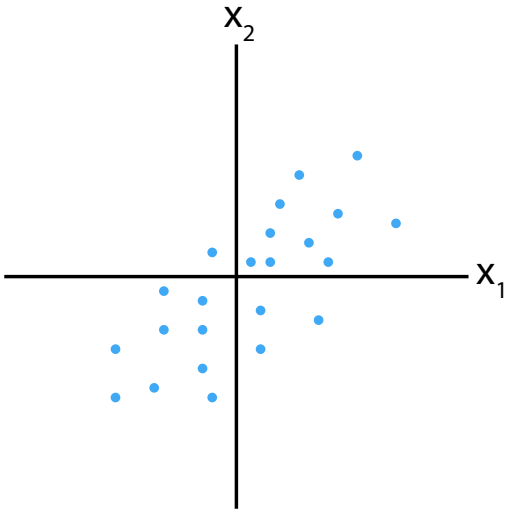
Why multivariate statistics?



1. how to operationalize network property?
2. how to deal with with more variables than observations?
3. how to relate multiple data sets to one another?

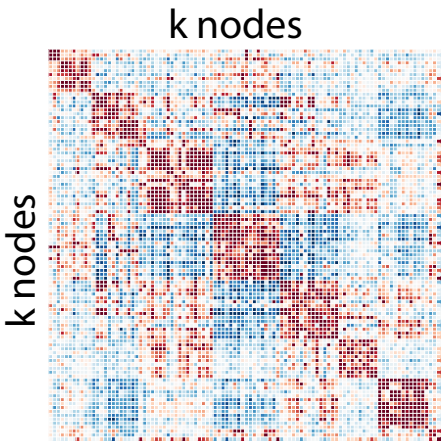
Principal component analysis (PCA)

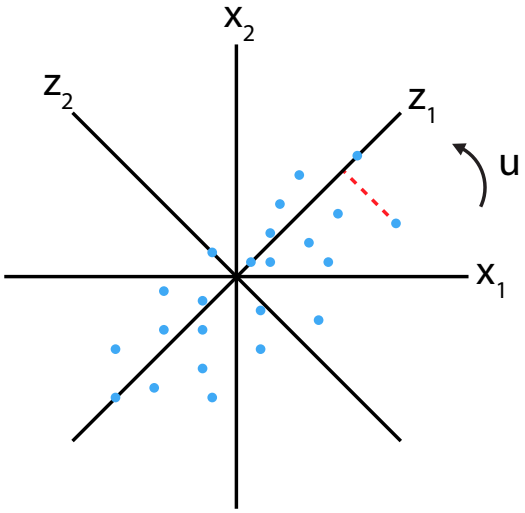




Hotelling (1933) *J Educ Psychol*

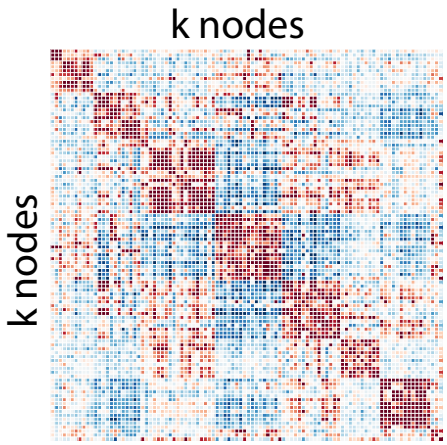
Principal component analysis (PCA)

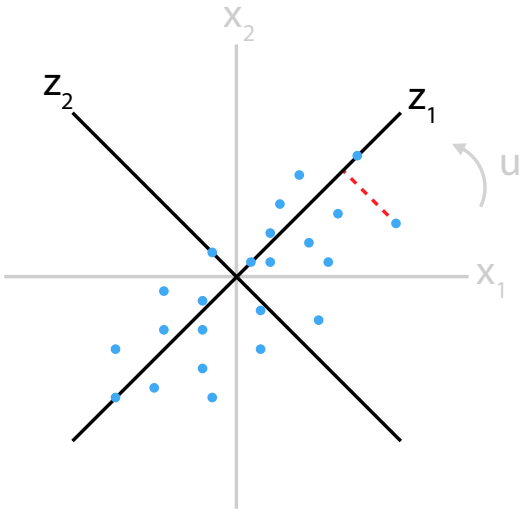




Hotelling (1933) *J Educ Psychol*

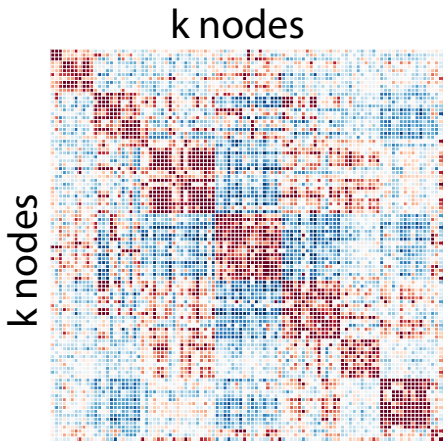
Principal component analysis (PCA)

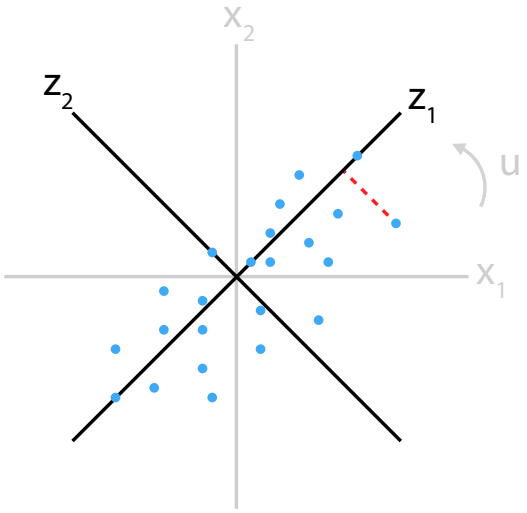


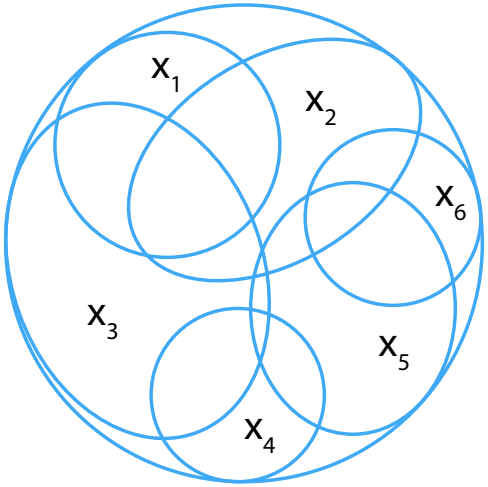


Hotelling (1933) *J Educ Psychol*

Principal component analysis (PCA)

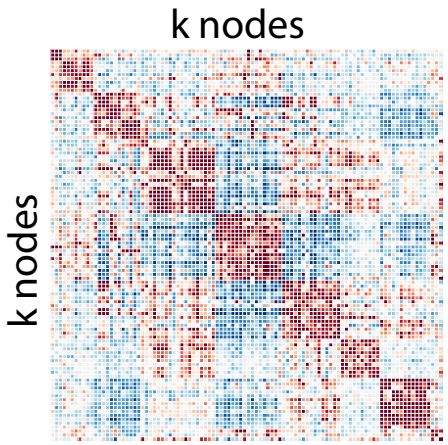


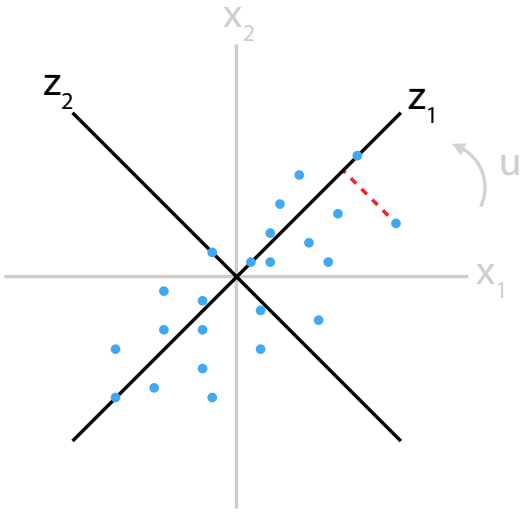


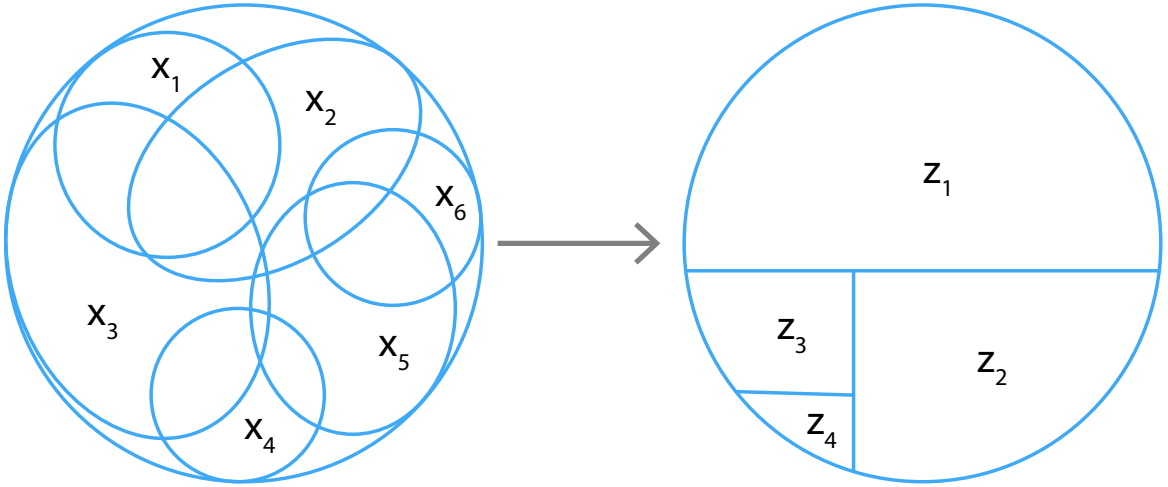


Hotelling (1933) *J Educ Psychol*

Principal component analysis (PCA)







Hotelling (1933) *J Educ Psychol*

Maximizing variance

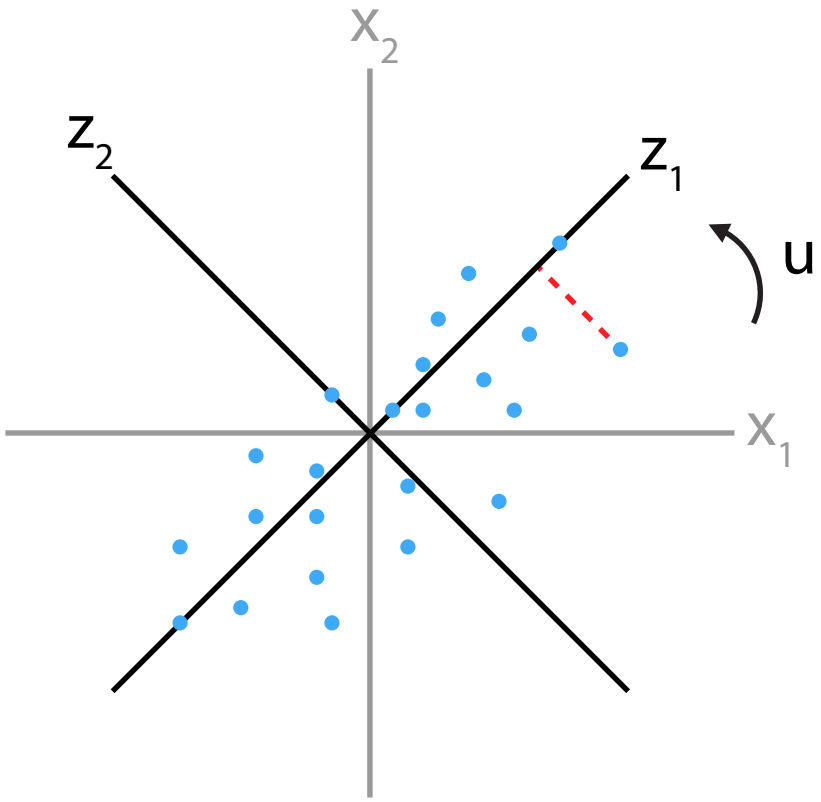
find new variable $\mathbf{z} = \mathbf{X}\mathbf{u}$

choose $\mathbf{u}$ to maximize $var(\mathbf{z})$

under the constraint $\mathbf{u}'\mathbf{u} = 1$

$var(\mathbf{z}) = \frac{1}{n-1}\mathbf{u}'\mathbf{X}'\mathbf{X}\mathbf{u} = \mathbf{u}'\mathbf{R}\mathbf{u}$

since $\mathbf{R} = \frac{1}{n-1}\mathbf{X}'\mathbf{X}$



Maximizing variance

$L = \mathbf{u}'\mathbf{R}\mathbf{u} - \lambda(\mathbf{u}'\mathbf{u} - 1)$

$\frac{\partial L}{\partial \mathbf{u}} = 2\mathbf{R}\mathbf{u} - 2\mathbf{u}\lambda = 0$

$\mathbf{R}\mathbf{u} = \mathbf{u}\lambda$

$(\mathbf{R} - \lambda\mathbf{I})\mathbf{u} = 0$

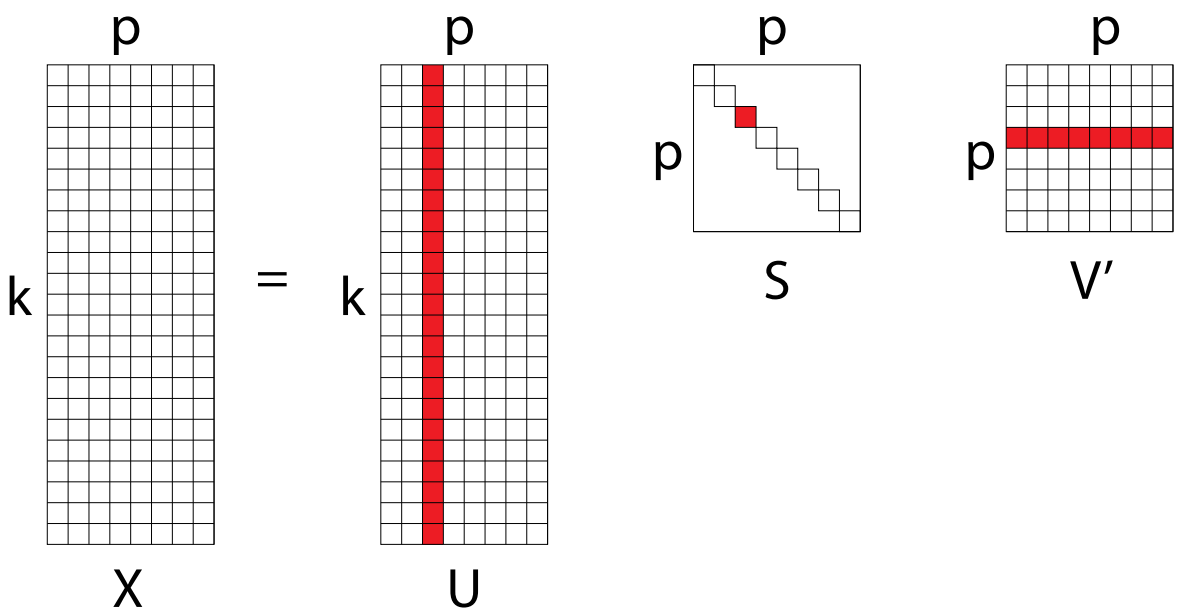
eigenvalue λ (variance) & eigenvector $\mathbf{u}$ (weights)

$var(\mathbf{z}) = \mathbf{u}'\mathbf{R}\mathbf{u} = \mathbf{u}'\mathbf{u}\lambda = \lambda$

Singular value decomposition

Spectral decomposition:
$\mathrm{EIG}(\mathbf{X}'\mathbf{X}) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$
$\mathrm{EIG}(\mathbf{X}\mathbf{X}') = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$

Singular value decomposition:
$\mathrm{SVD}(\mathbf{X}) = \mathbf{U}\mathbf{S}\mathbf{V}'$

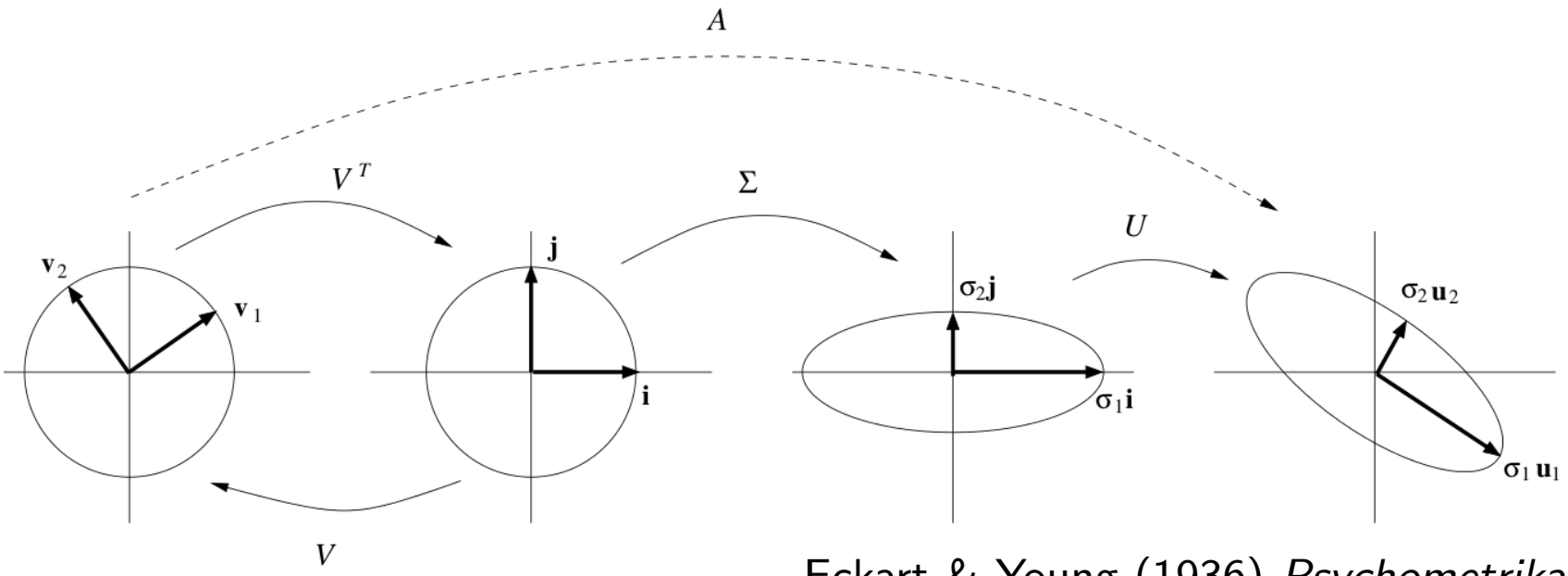


Eckart & Young (1936) *Psychometrika*

Singular value decomposition

Spectral decomposition:
$\text{EIG}(\mathbf{X}'\mathbf{X}) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$
$\text{EIG}(\mathbf{X}\mathbf{X}') = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$

Singular value decomposition:
$\text{SVD}(\mathbf{X}) = \mathbf{U}\mathbf{S}\mathbf{V}'$



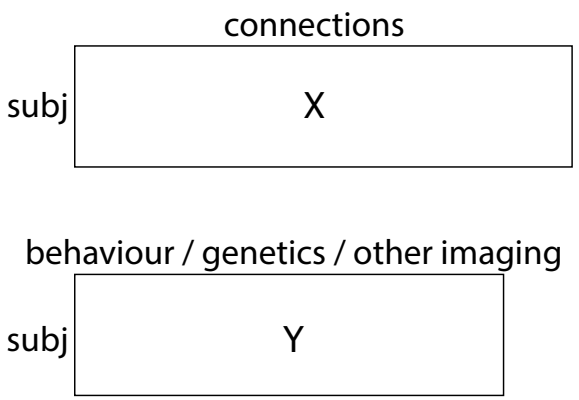
Eckart & Young (1936) *Psychometrika*

A family of techniques

PCA: SVD($\mathbf{X}$)

PLS: SVD($\mathbf{X}'\mathbf{Y}$)

CCA: SVD($(\mathbf{X}'\mathbf{X}')^{-1/2}(\mathbf{X}'\mathbf{Y})(\mathbf{Y}'\mathbf{Y})^{-1/2}$)



Worsley et al. (1997) *NeuroImage*

Tijl et al. (2005) *Handbook of Geometric Computing - Springer*

McIntosh & Mišić (2013) *Annu Rev Psychol*

Related techniques: ICA and FA

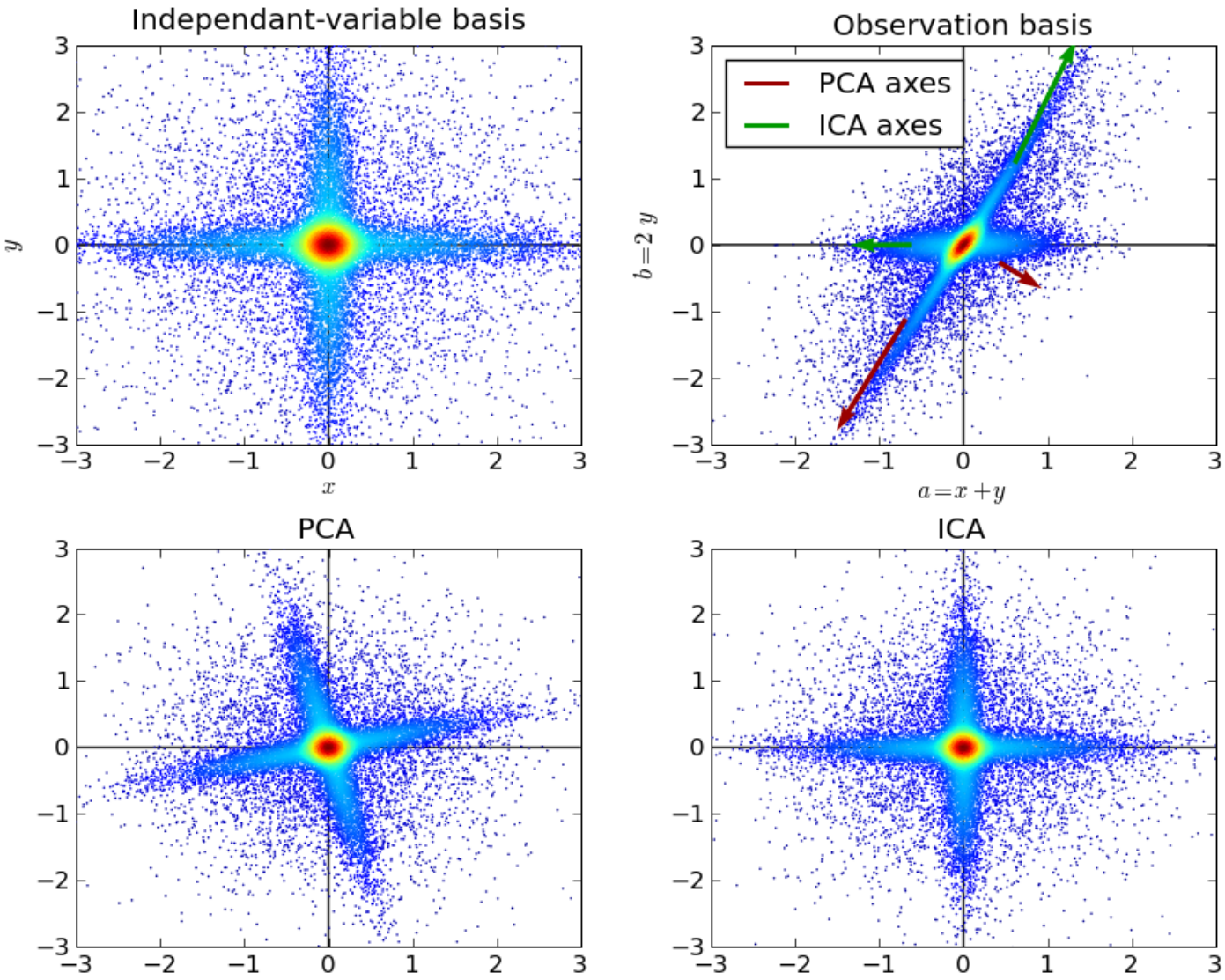
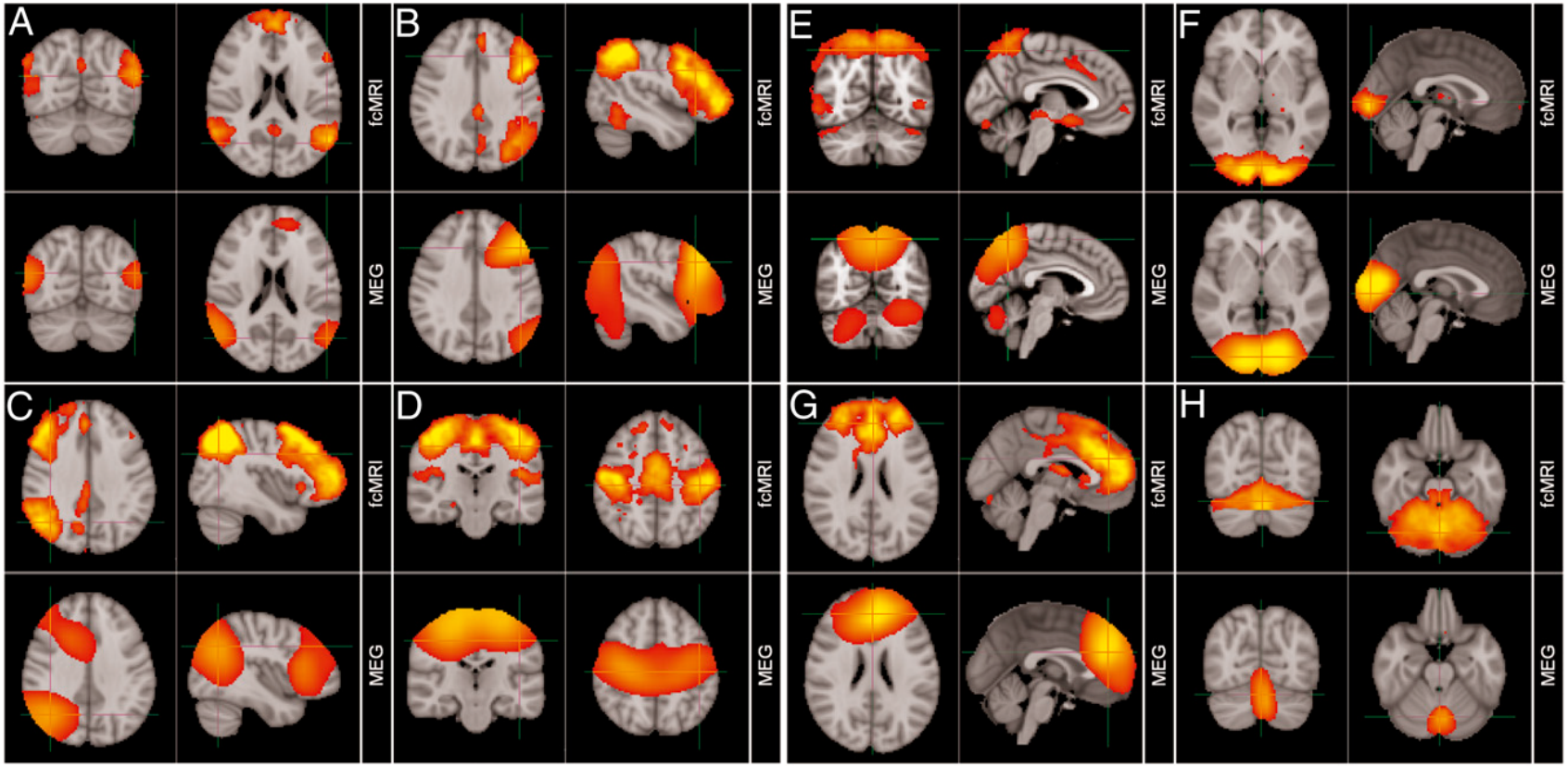


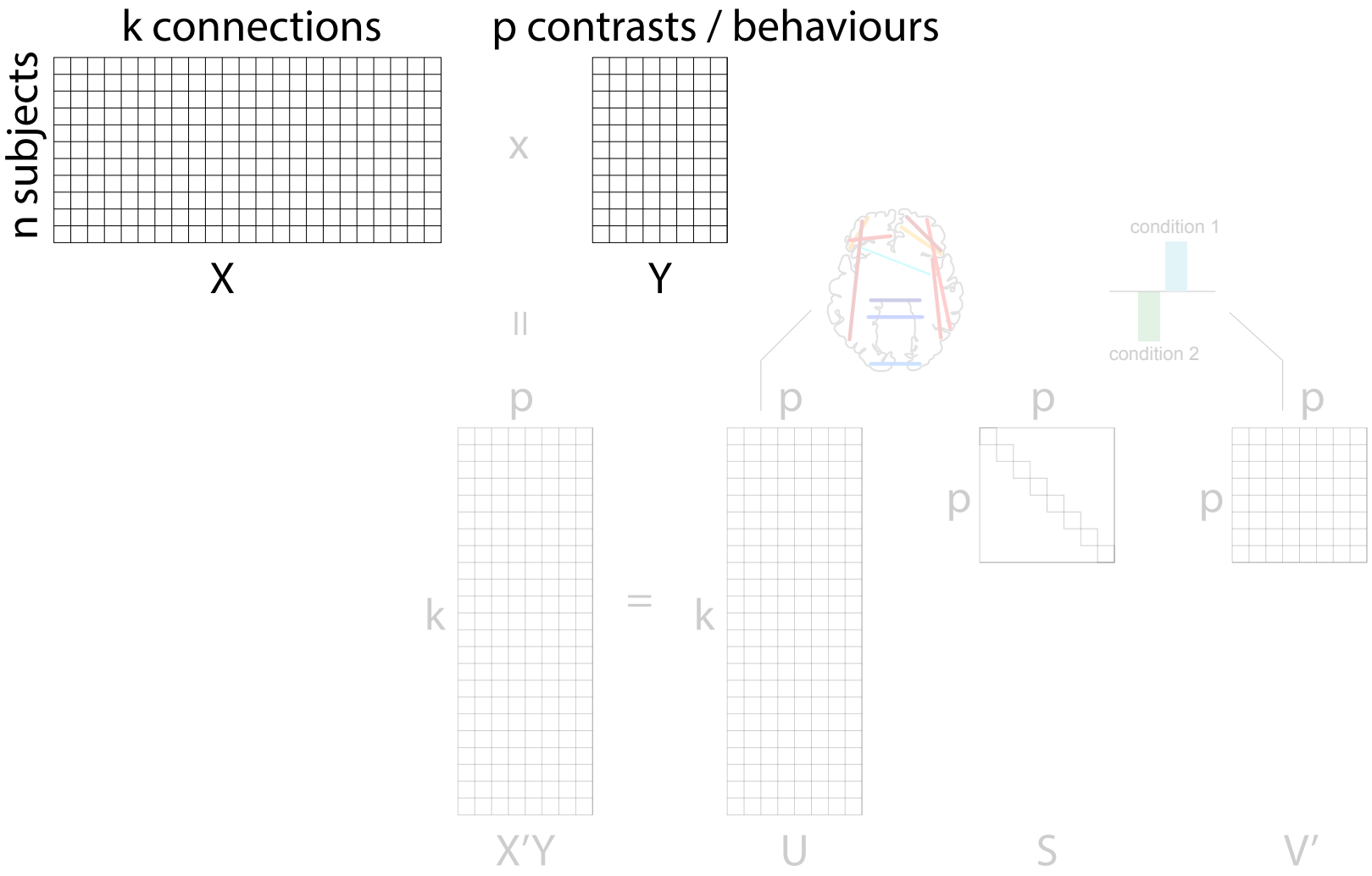
image: Gaël Varoqaux

Example: resting-state MEG networks



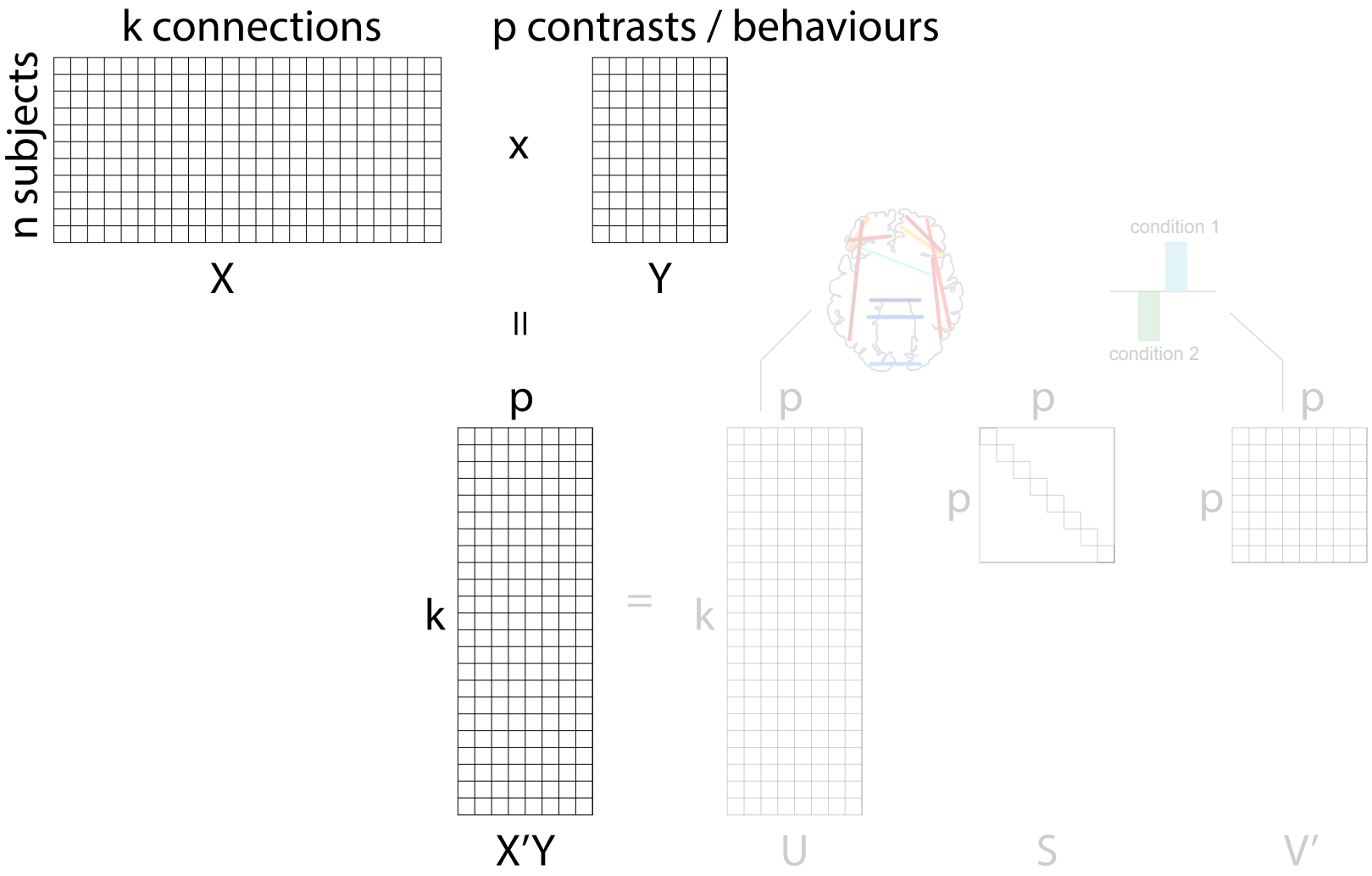
Brookes et al. (2011) *Proc Natl Acad Sci USA*

Partial least squares (PLS)



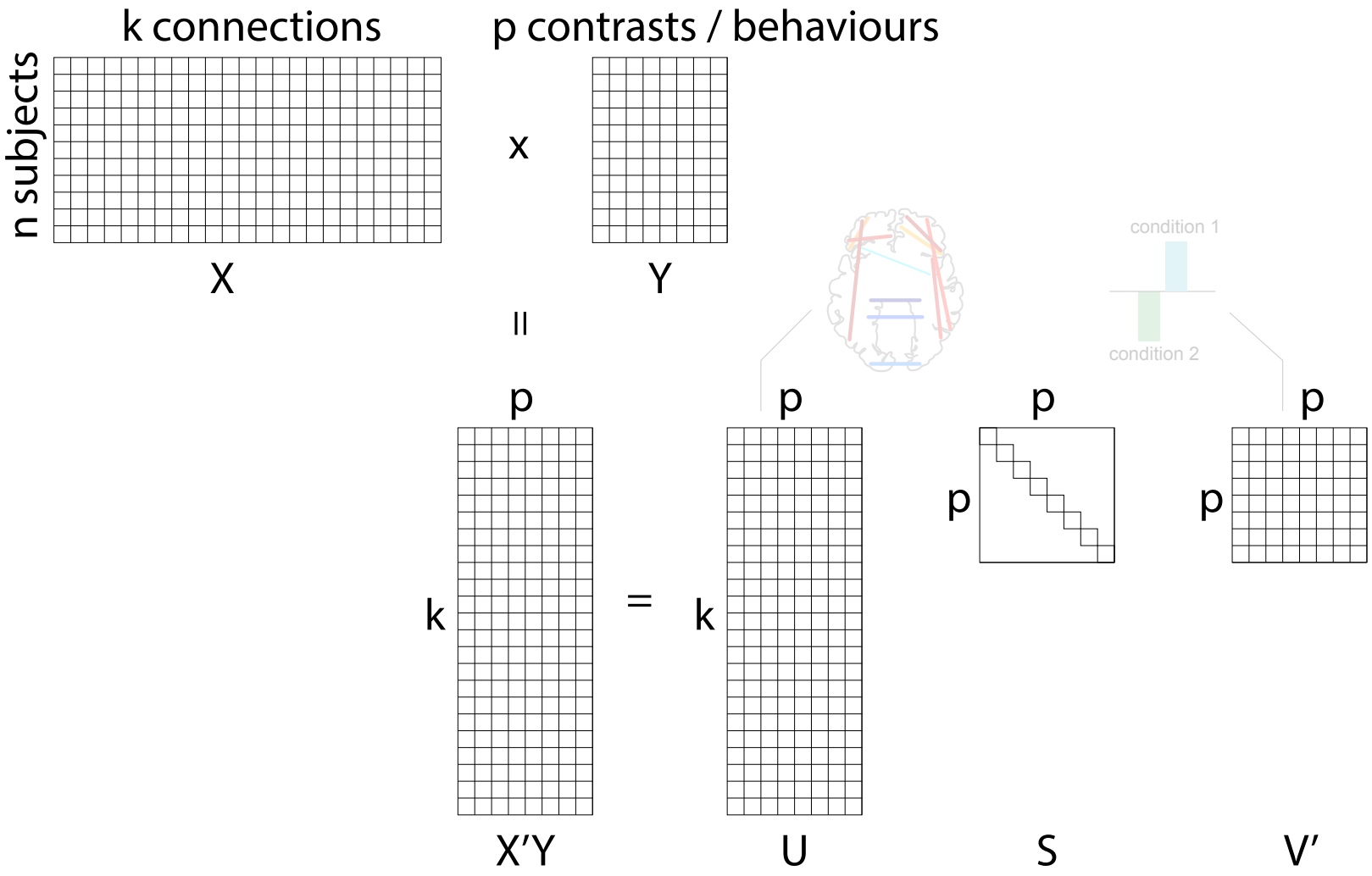
Wold et al. (1982, 1984) *Mathematics and statistics in chemistry*

Partial least squares (PLS)



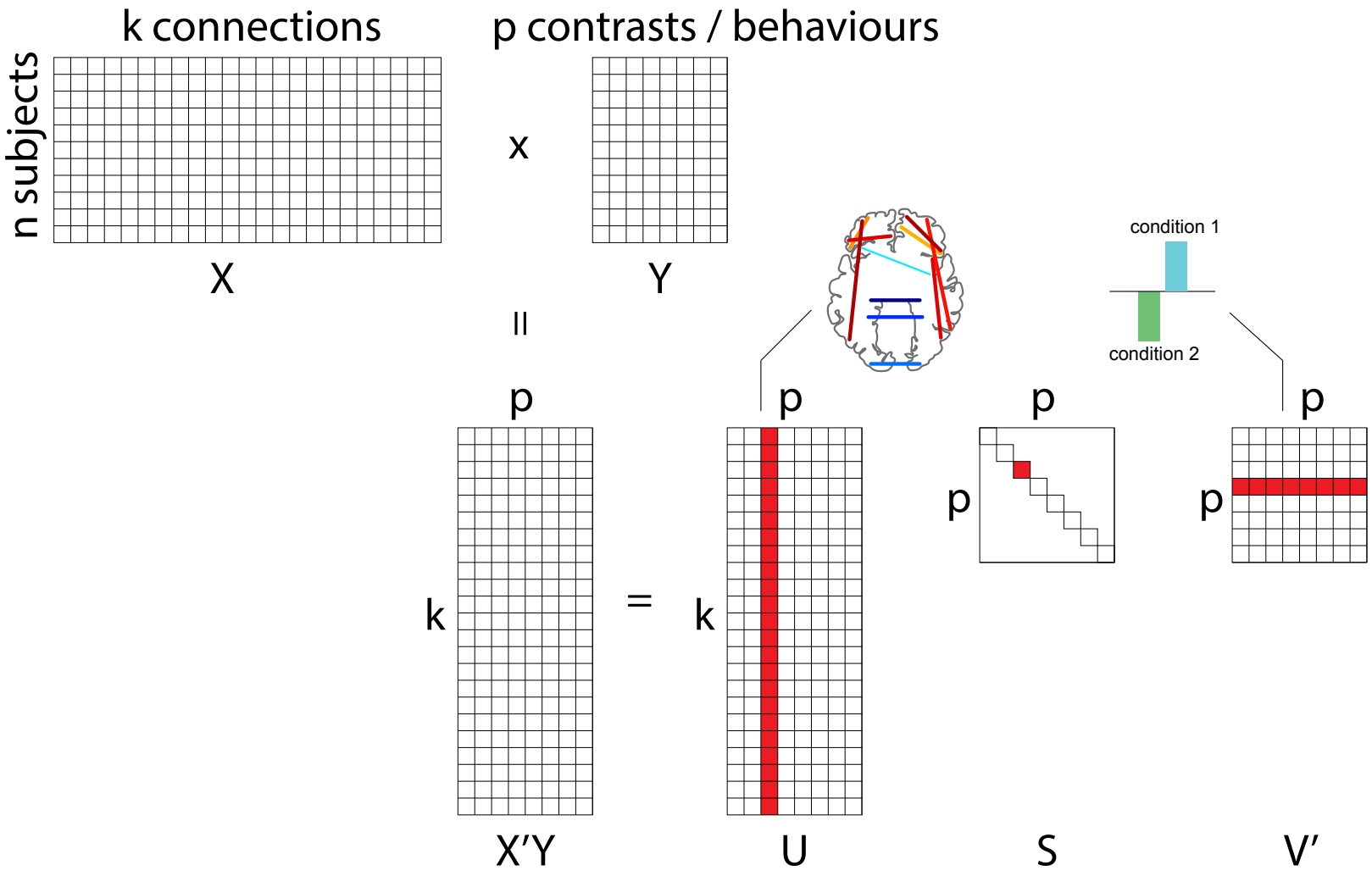
Wold et al. (1982, 1984) *Mathematics and statistics in chemistry*

Partial least squares (PLS)



Wold et al. (1982, 1984) *Mathematics and statistics in chemistry*

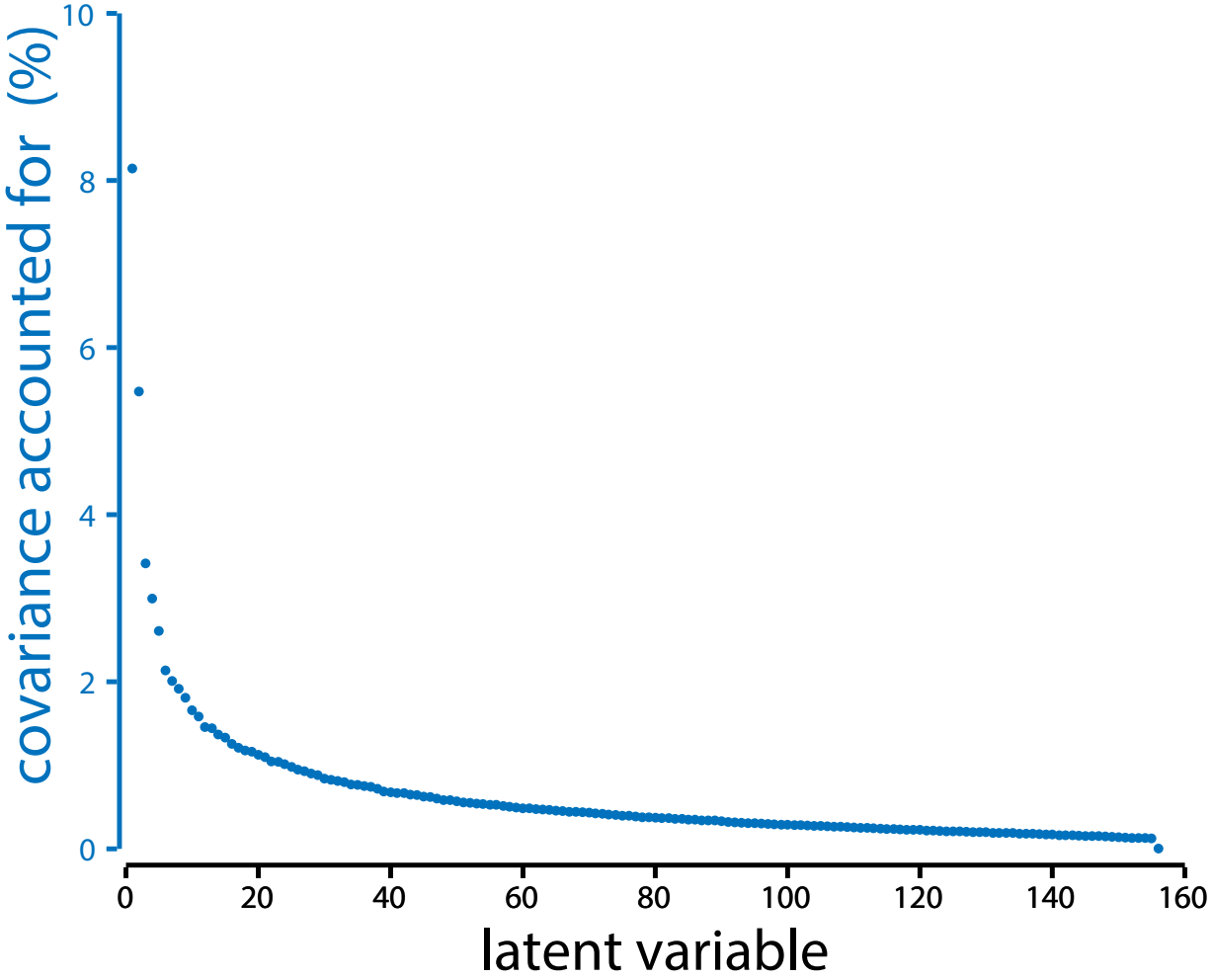
Partial least squares (PLS)



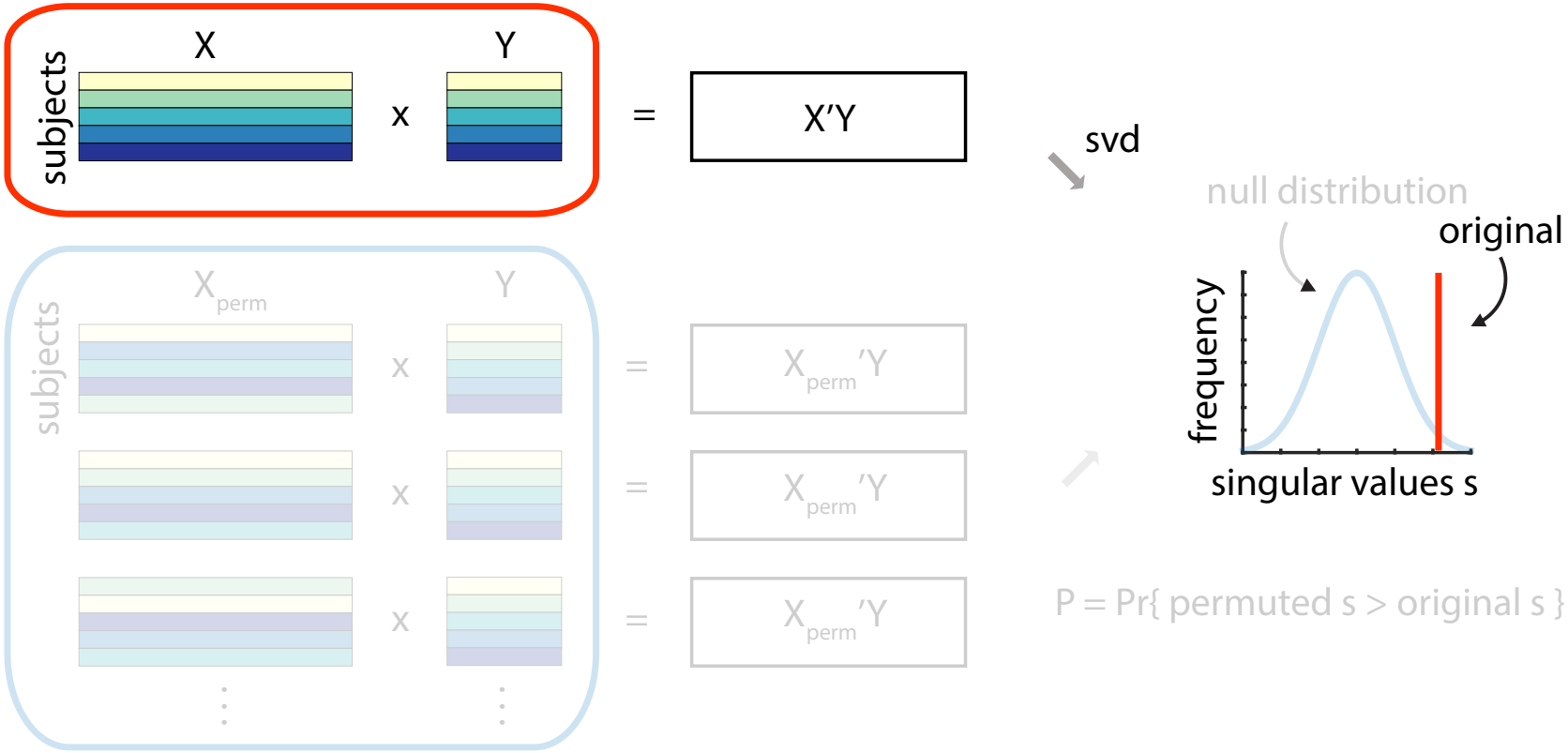
Wold et al. (1982, 1984) *Mathematics and statistics in chemistry*

How many components to retain?

$$\% \text{ covariance} = s_i^2 / \sum_j s_j^2$$



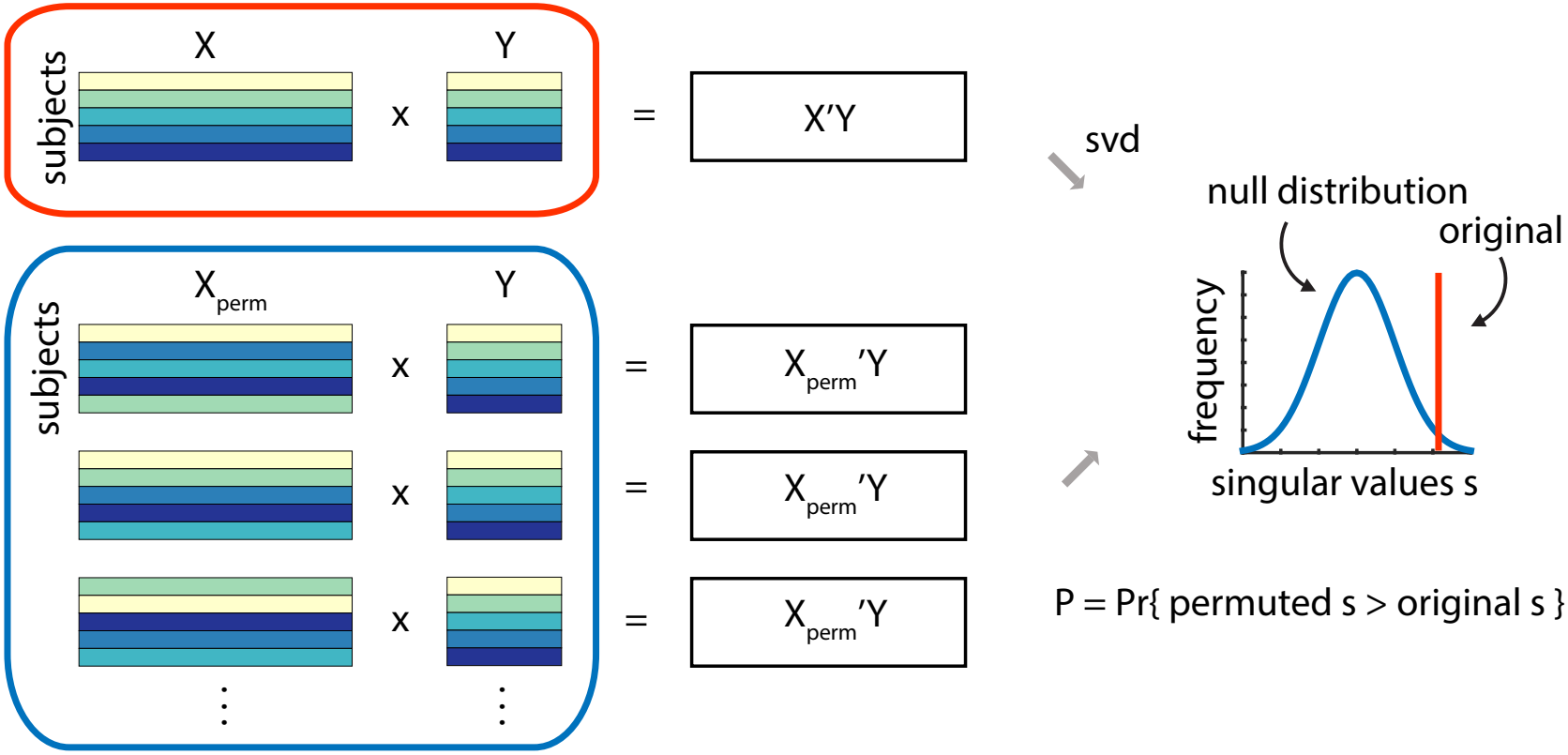
Statistical significance: permutation tests



Edgington (1965, 1969) *J Psychol*

McIntosh et al. (1996, 2004) *NeuroImage*

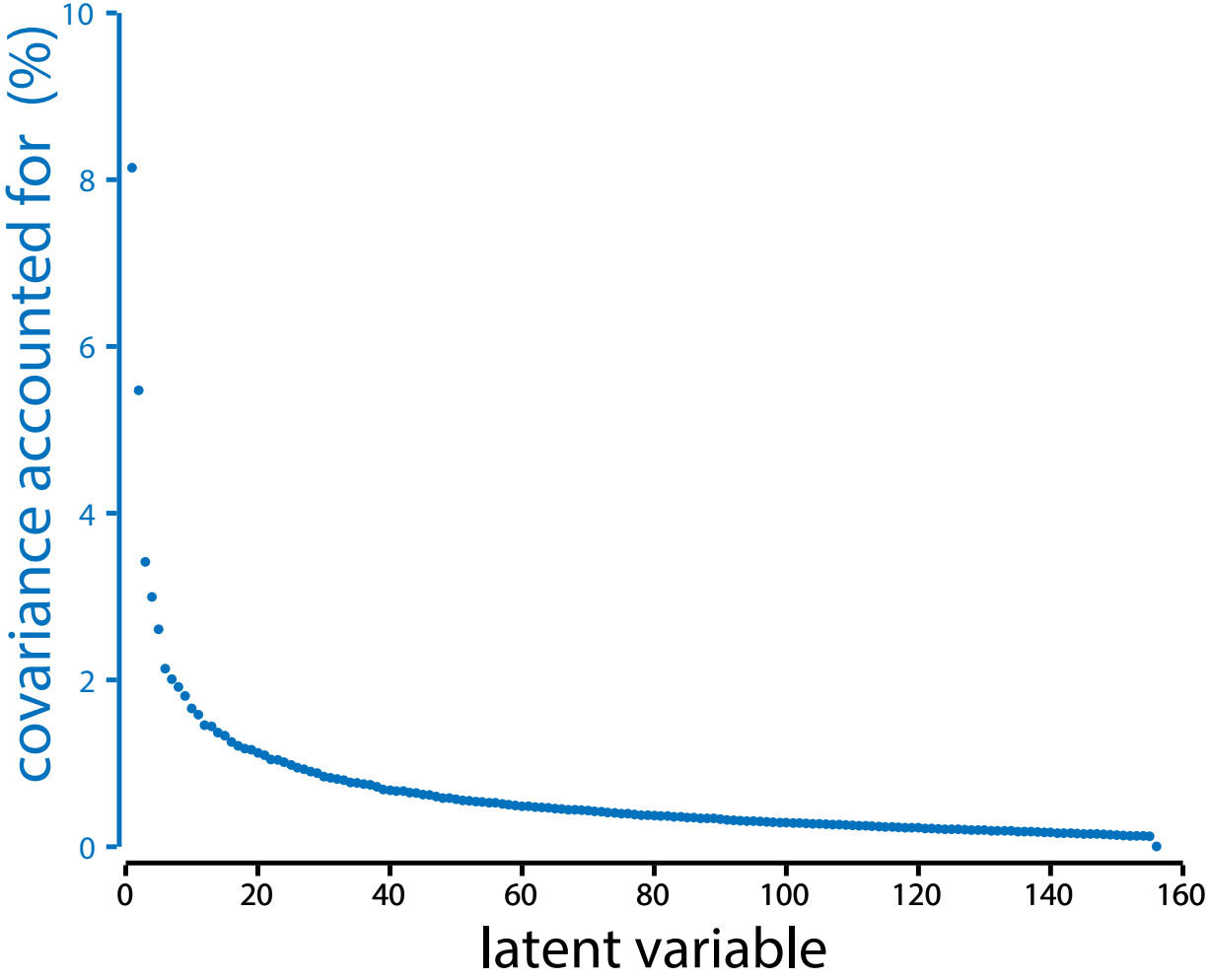
Statistical significance: permutation tests



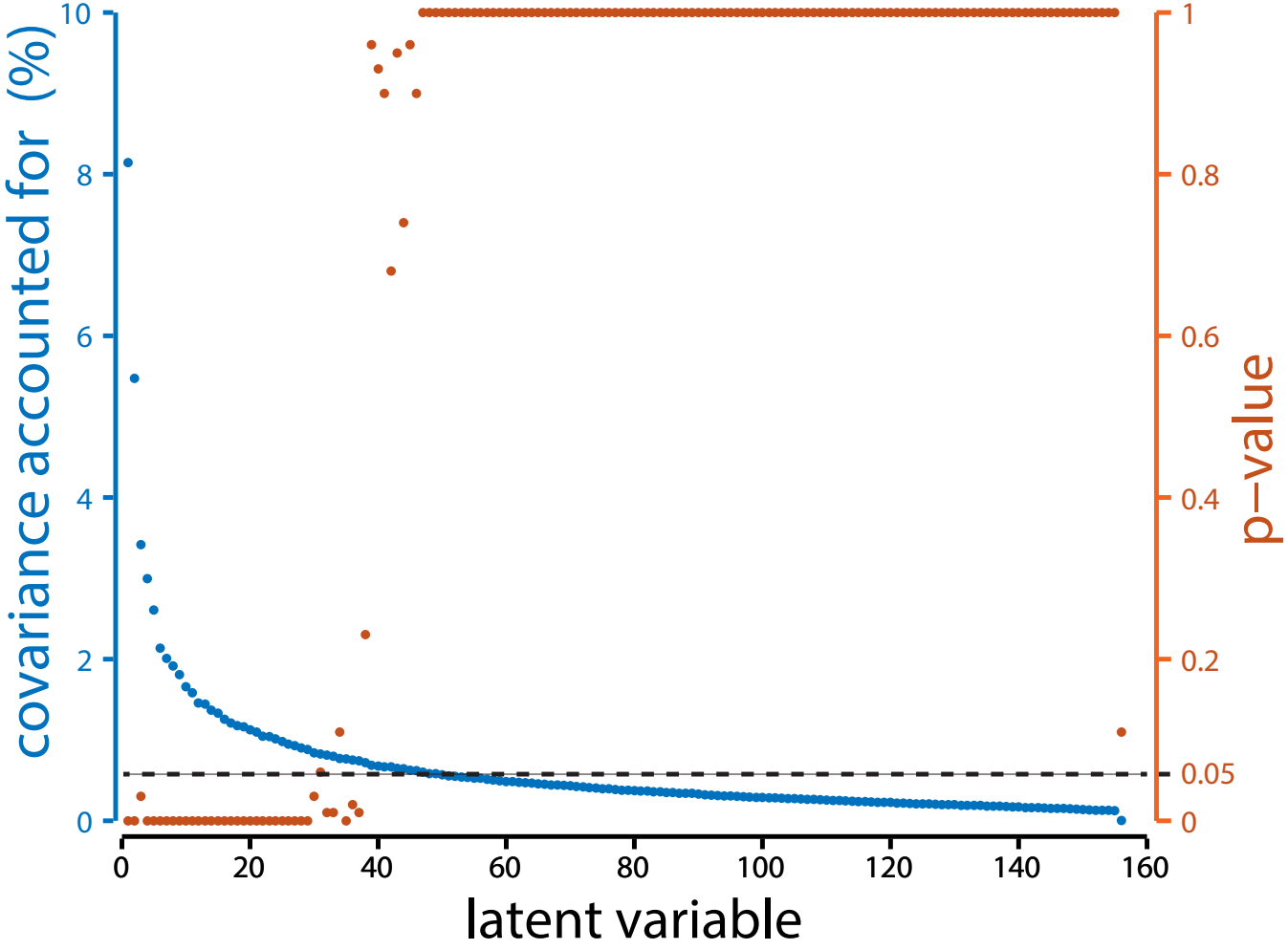
Edgington (1965, 1969) *J Psychol*

McIntosh et al. (1996, 2004) *NeuroImage*

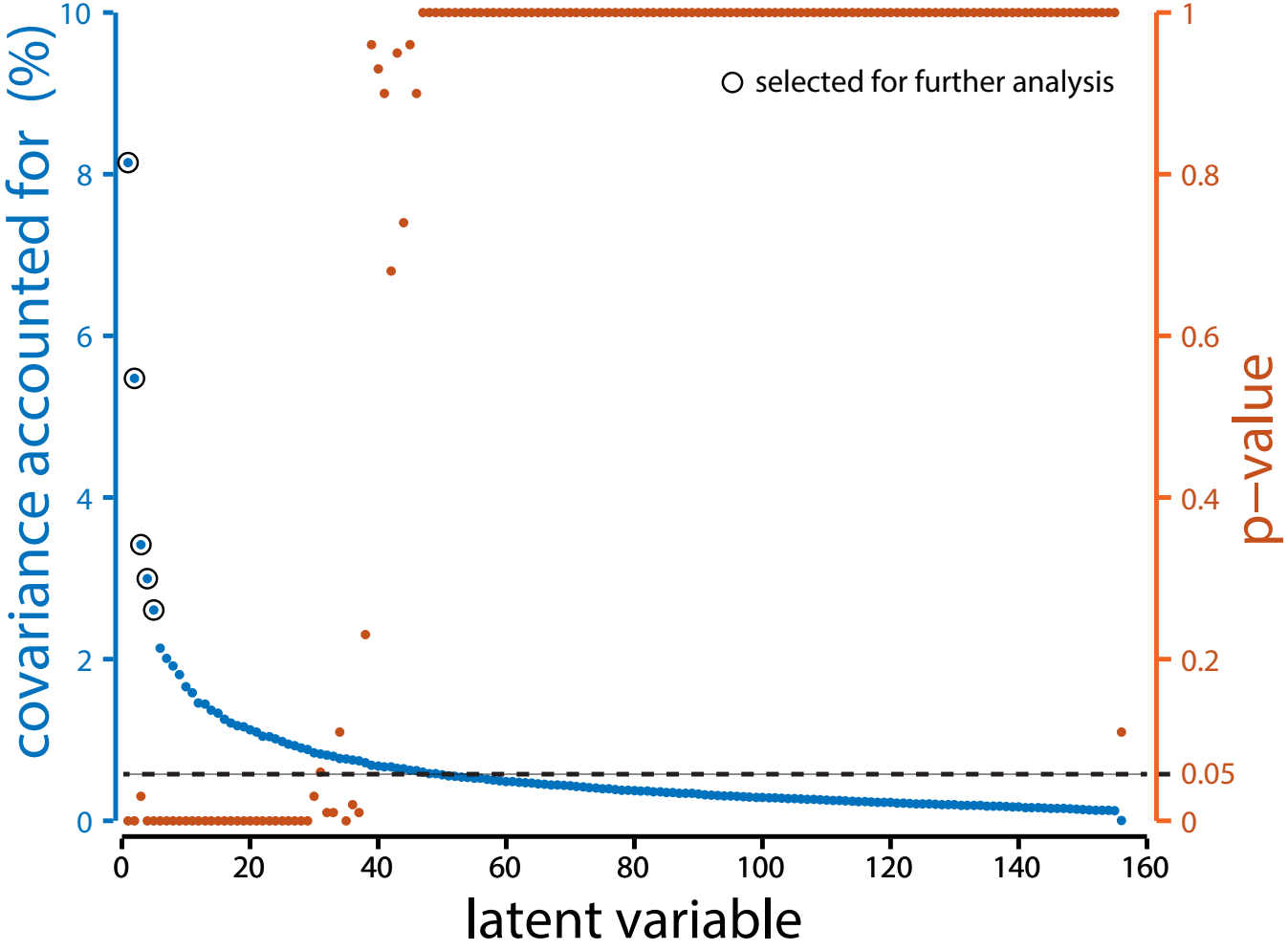
How many components to retain?



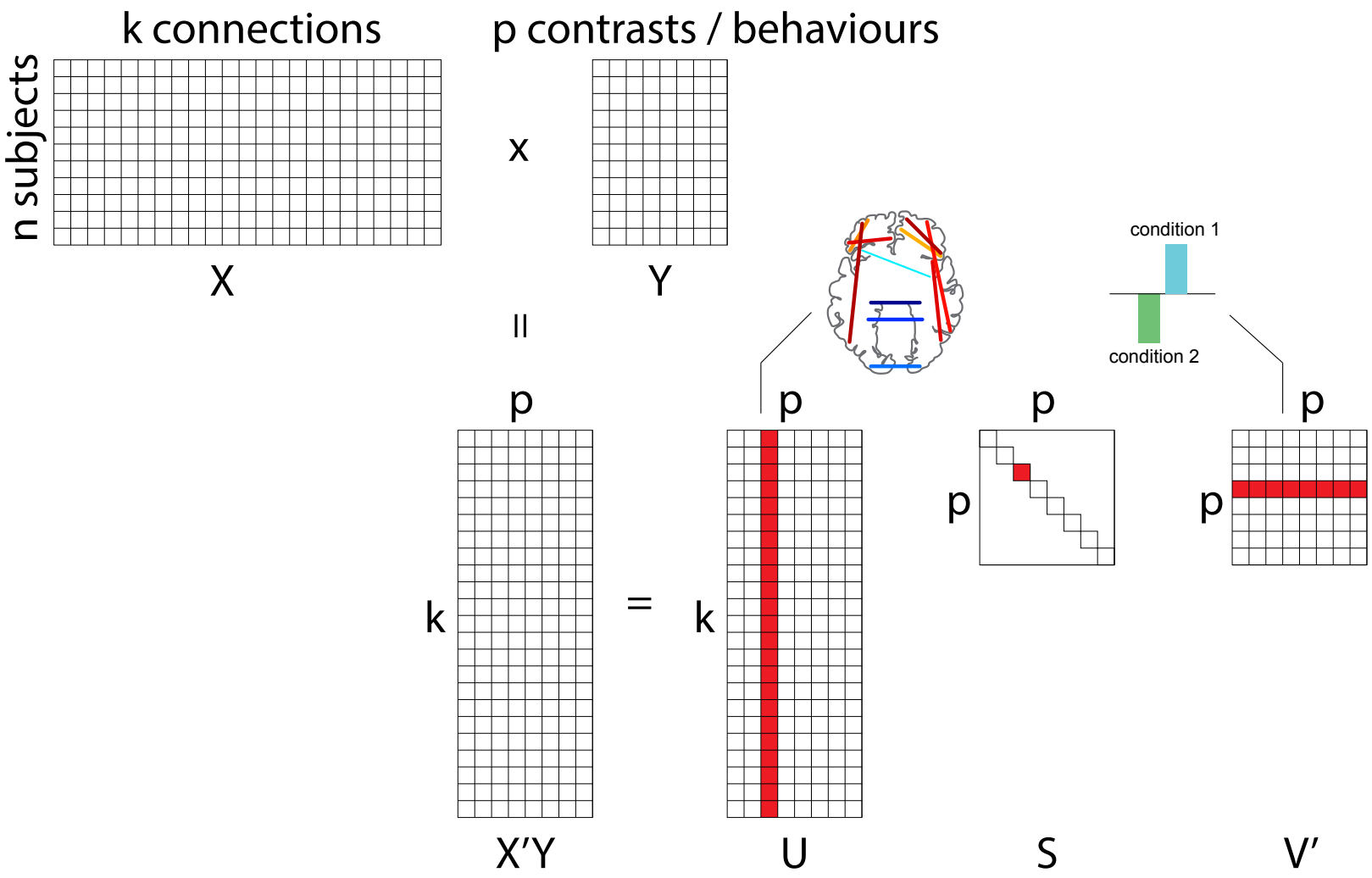
How many components to retain?



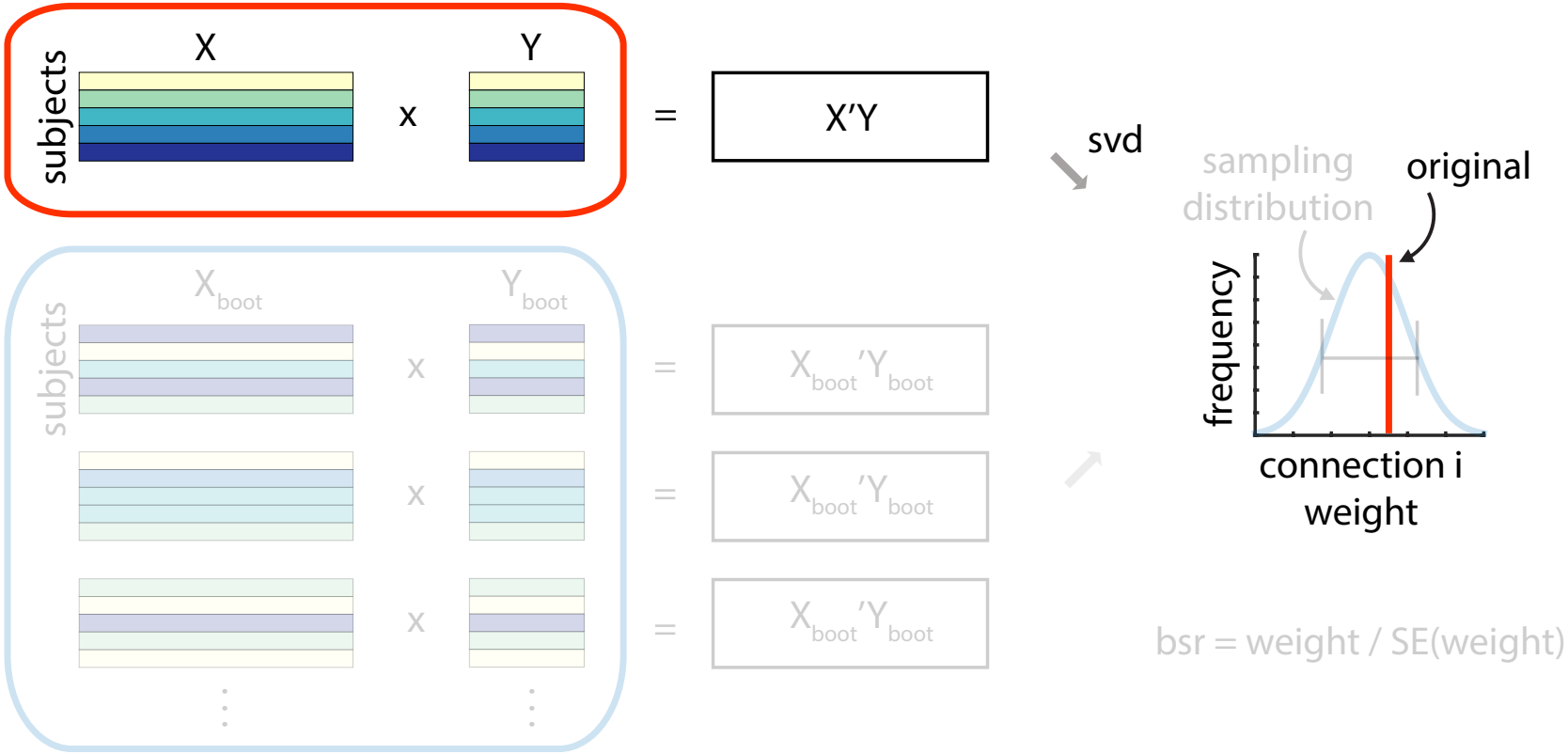
How many components to retain?



Which connections to focus on?



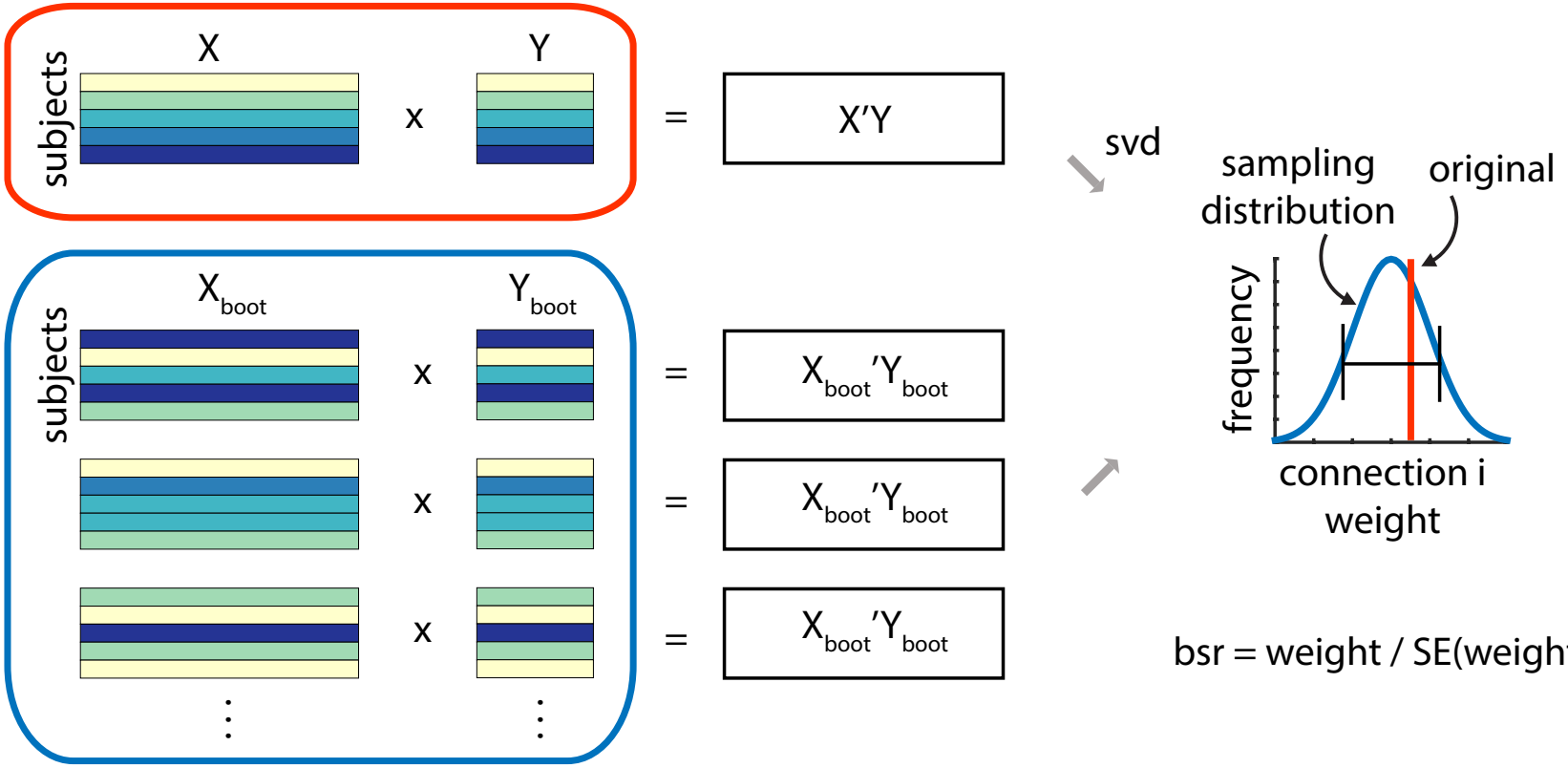
Reliability: bootstrapping



Efron & Tibshirani (1986) *Stat Sci*

McIntosh et al. (1996, 2004) *NeuroImage*

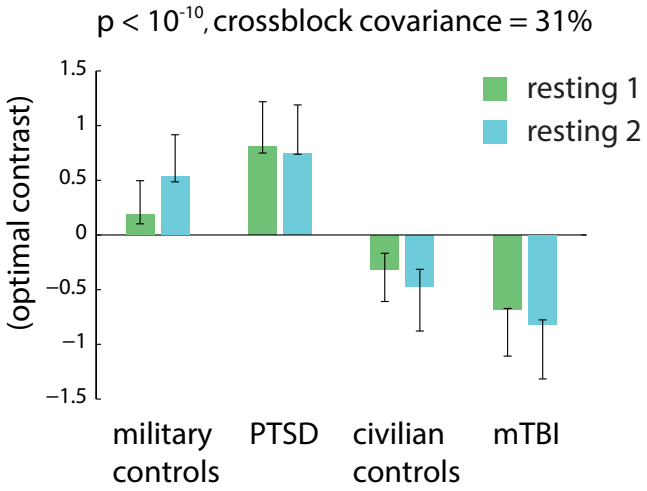
Reliability: bootstrapping

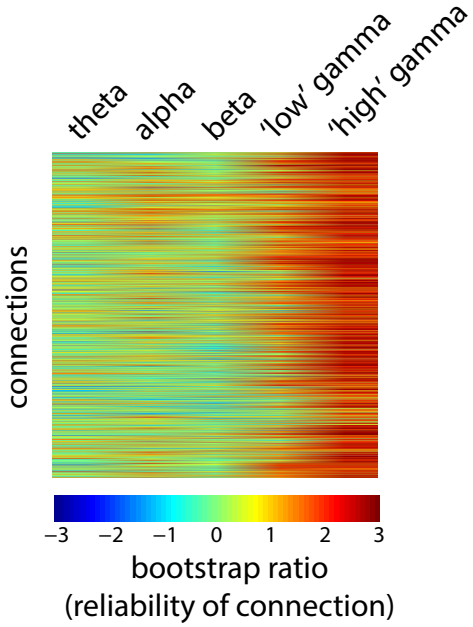


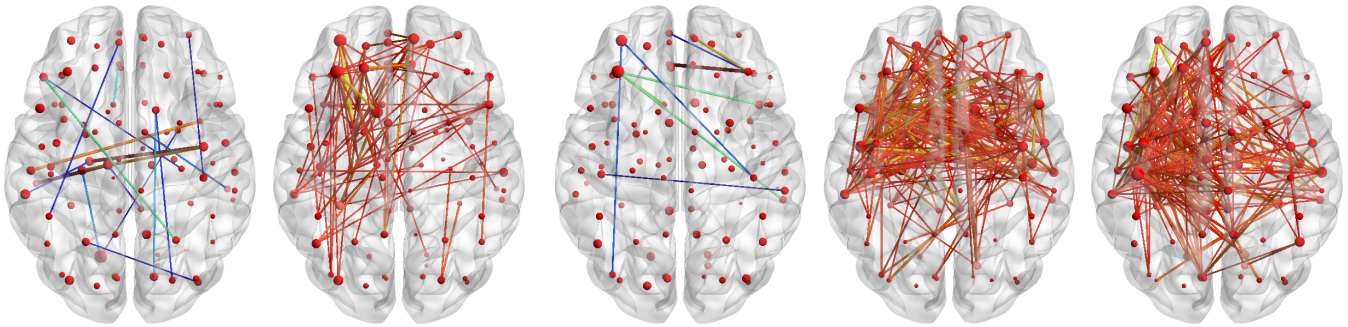
Efron & Tibshirani (1986) *Stat Sci*

McIntosh et al. (1996, 2004) *NeuroImage*

Example: connectivity differentiates PTSD from mTBI

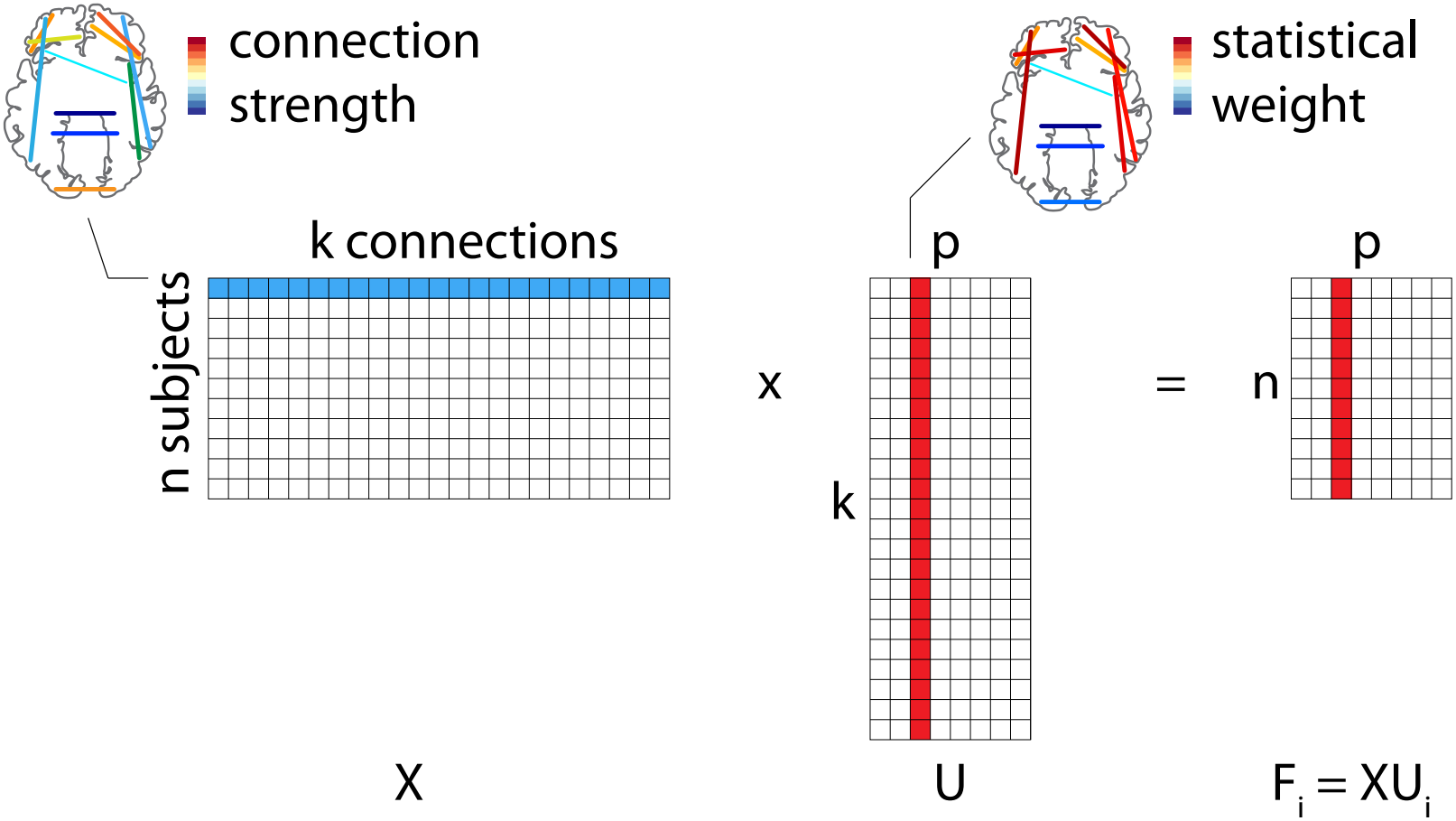




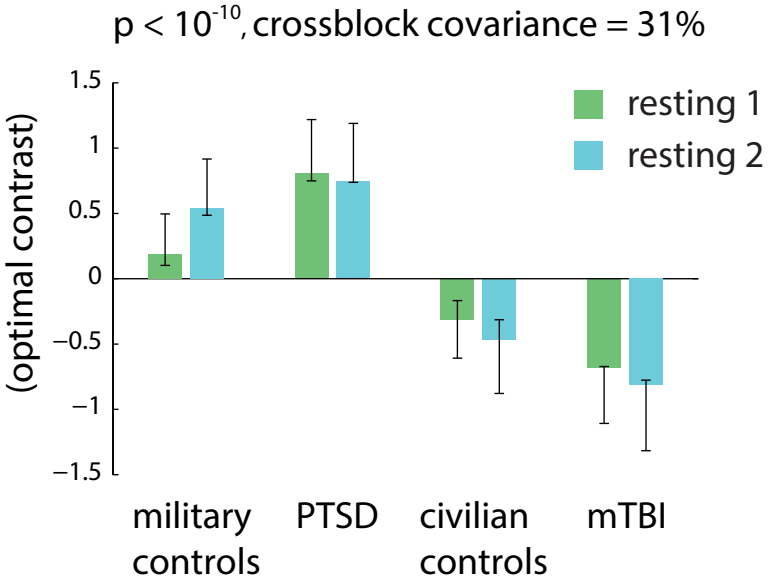


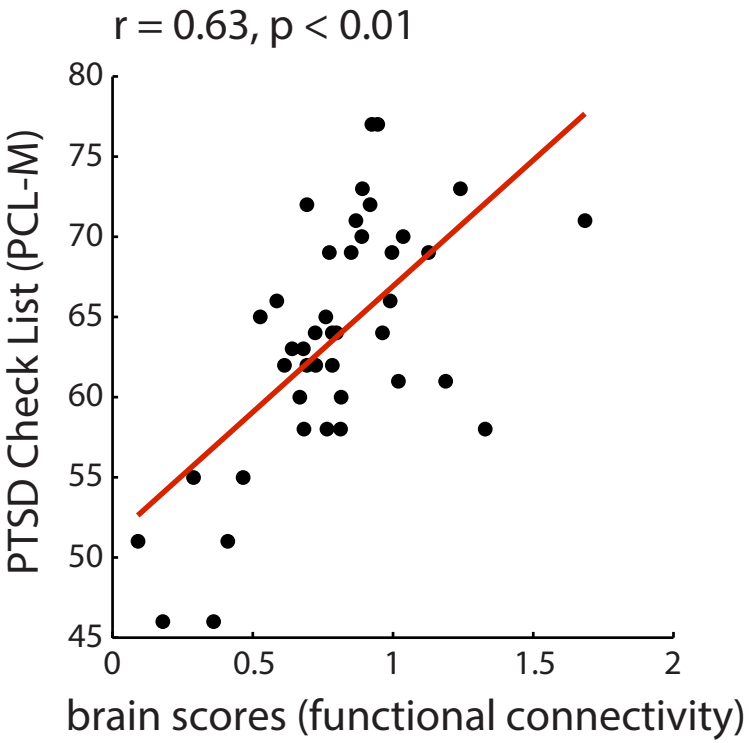
theta (4-7 Hz) | alpha (8-14 Hz) | beta (15-30 Hz) | 'low' gamma (30-80 Hz) | 'high' gamma (80-150 Hz)

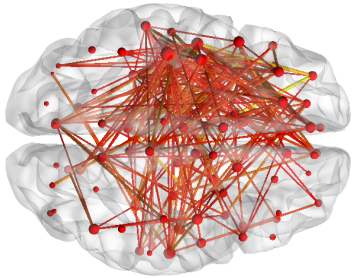
Individual participants



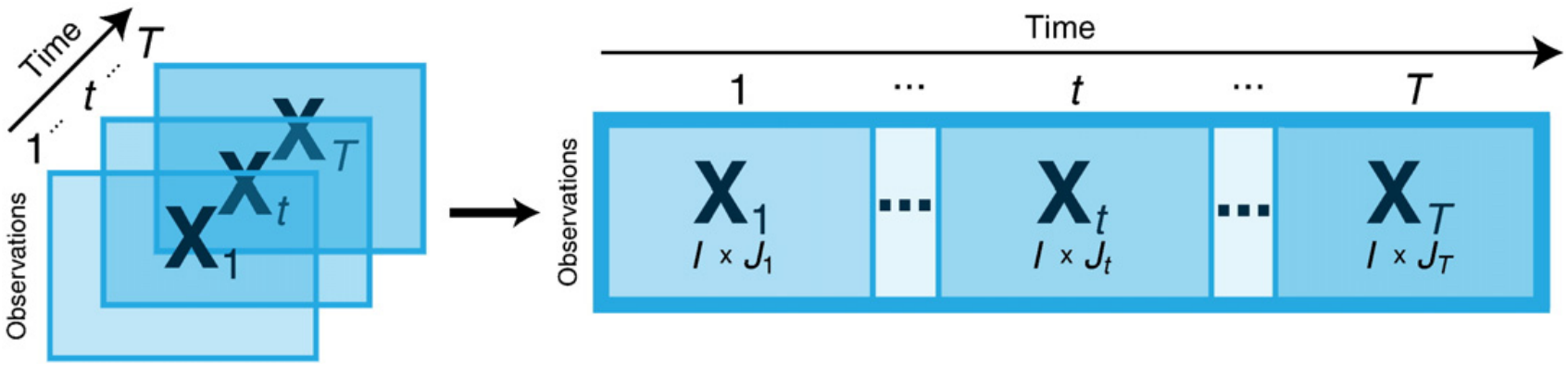
Example: connectivity and symptom severity in PTSD





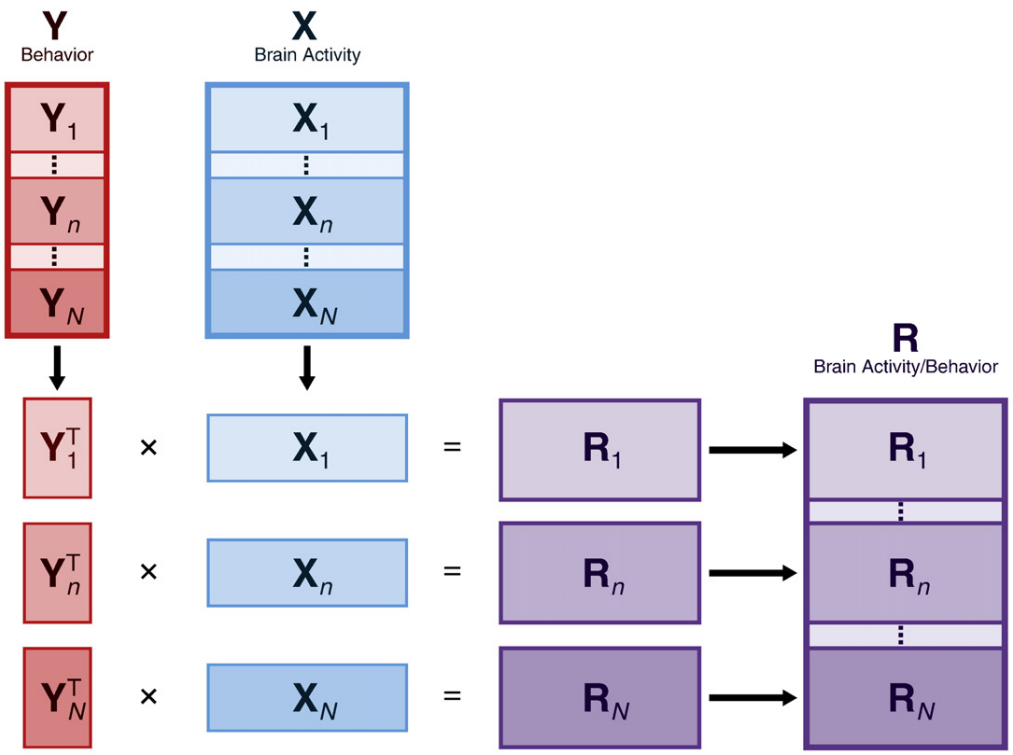


Adding other dimensions



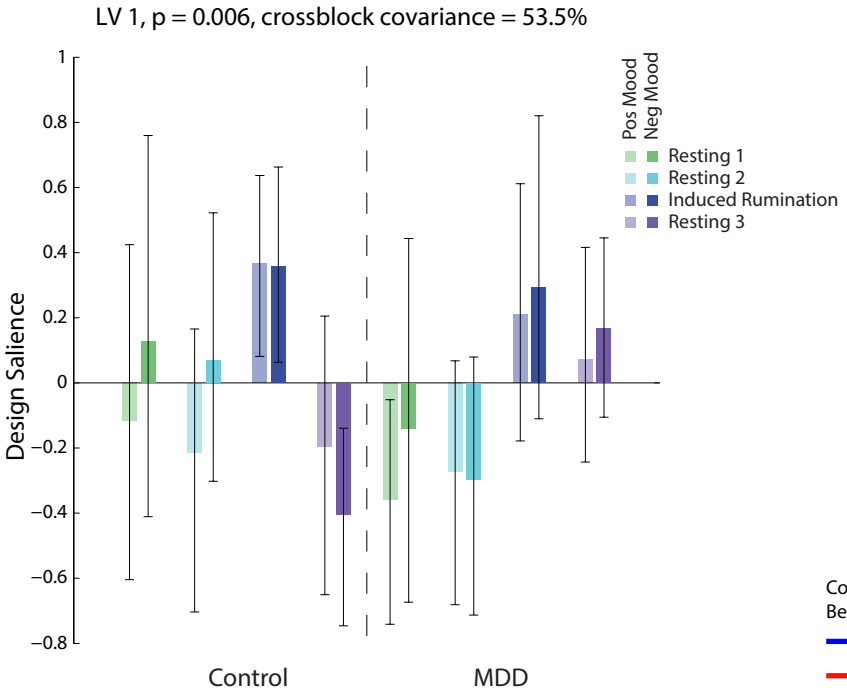
Krishnan et al. (2011) *NeuroImage*

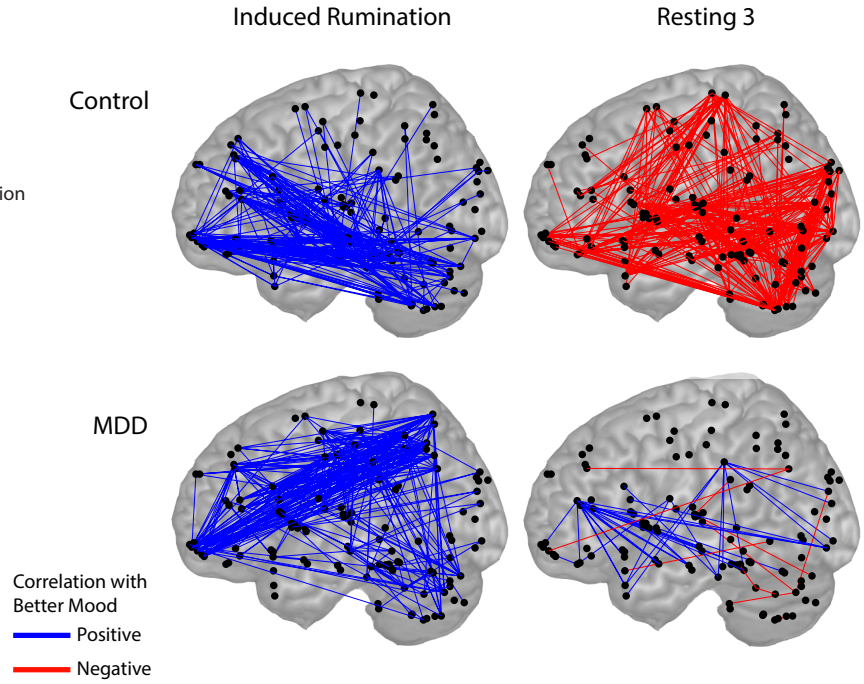
Adding other dimensions



Krishnan et al. (2011) *NeuroImage*

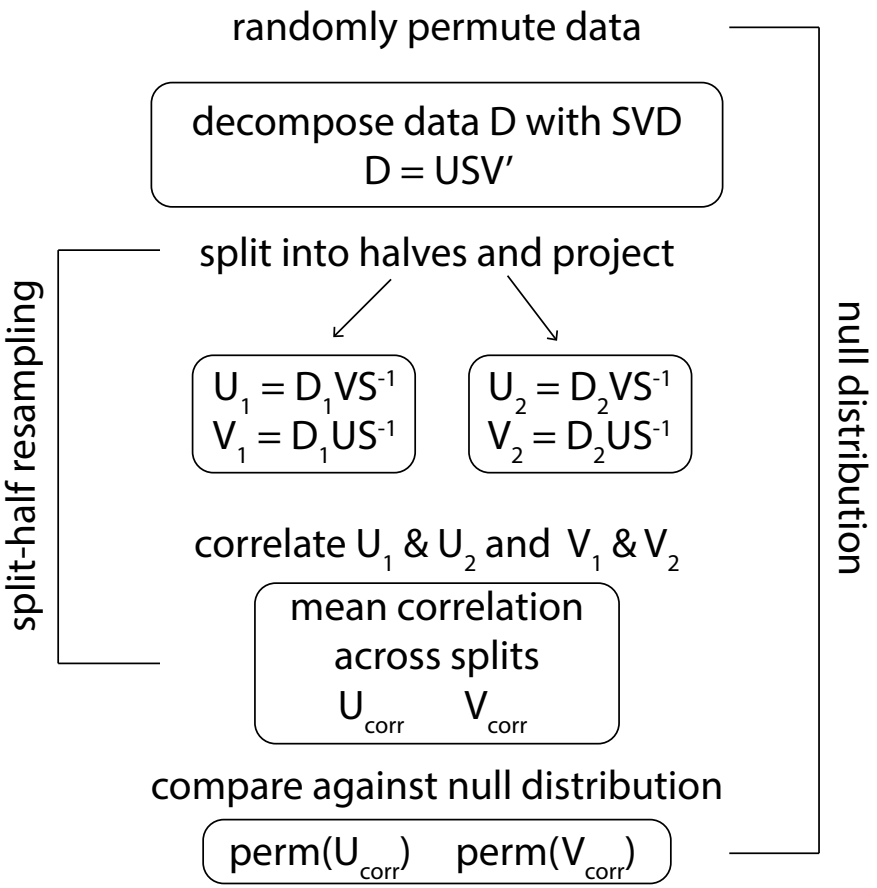
Example: connectivity and mood in major depression





Berman et al. (2014) *NeuroImage*

Cross-validation: split-half resampling

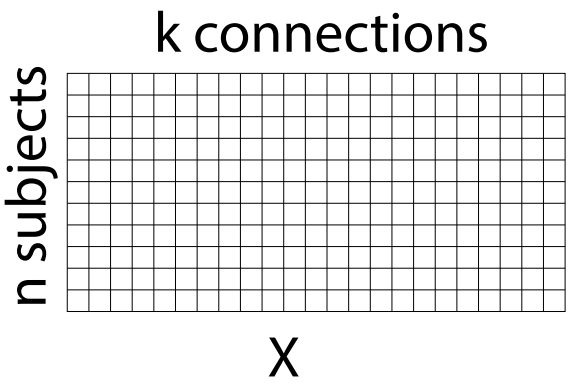


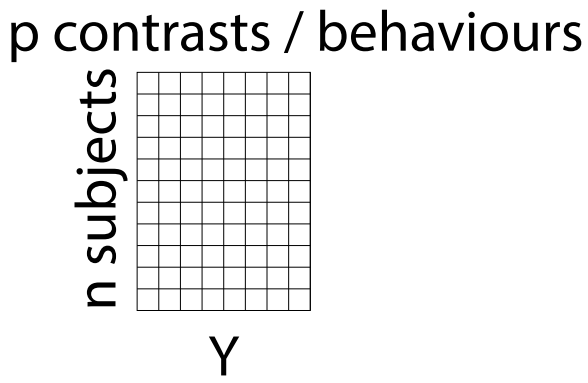
Strother et al. (2002) *NeuroImage*

Kovacevic et al. (2013) *New perspectives in Partial Least Squares* (Ed: Abdi)

Canonical correlation analysis (CCA)

$$\mathrm{SVD}((\mathbf{X}'\mathbf{X}')^{-1/2}(\mathbf{X}'\mathbf{Y})(\mathbf{Y}'\mathbf{Y})^{-1/2})$$

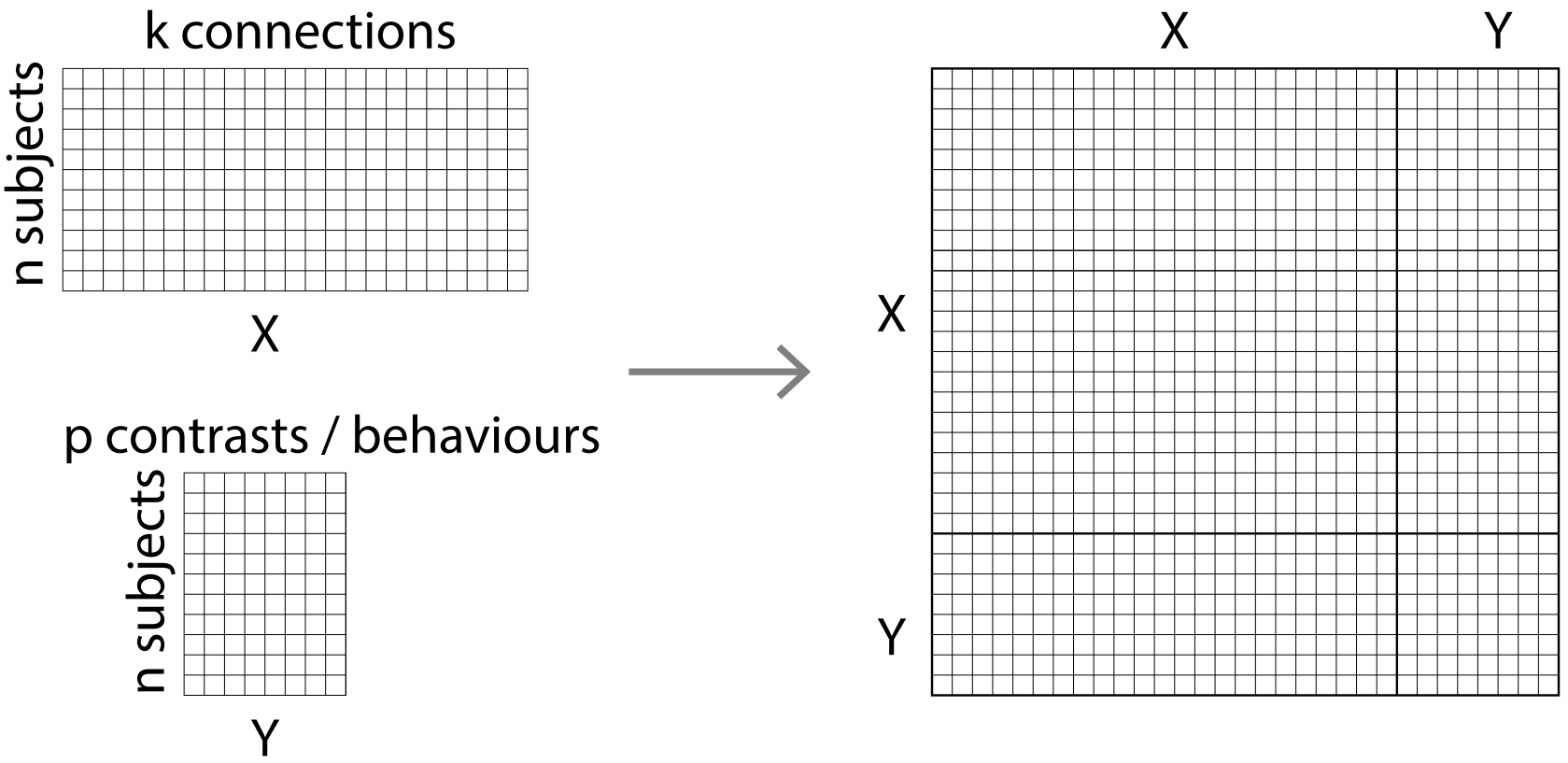




Hotelling (1936) *Biometrika*

Canonical correlation analysis (CCA)

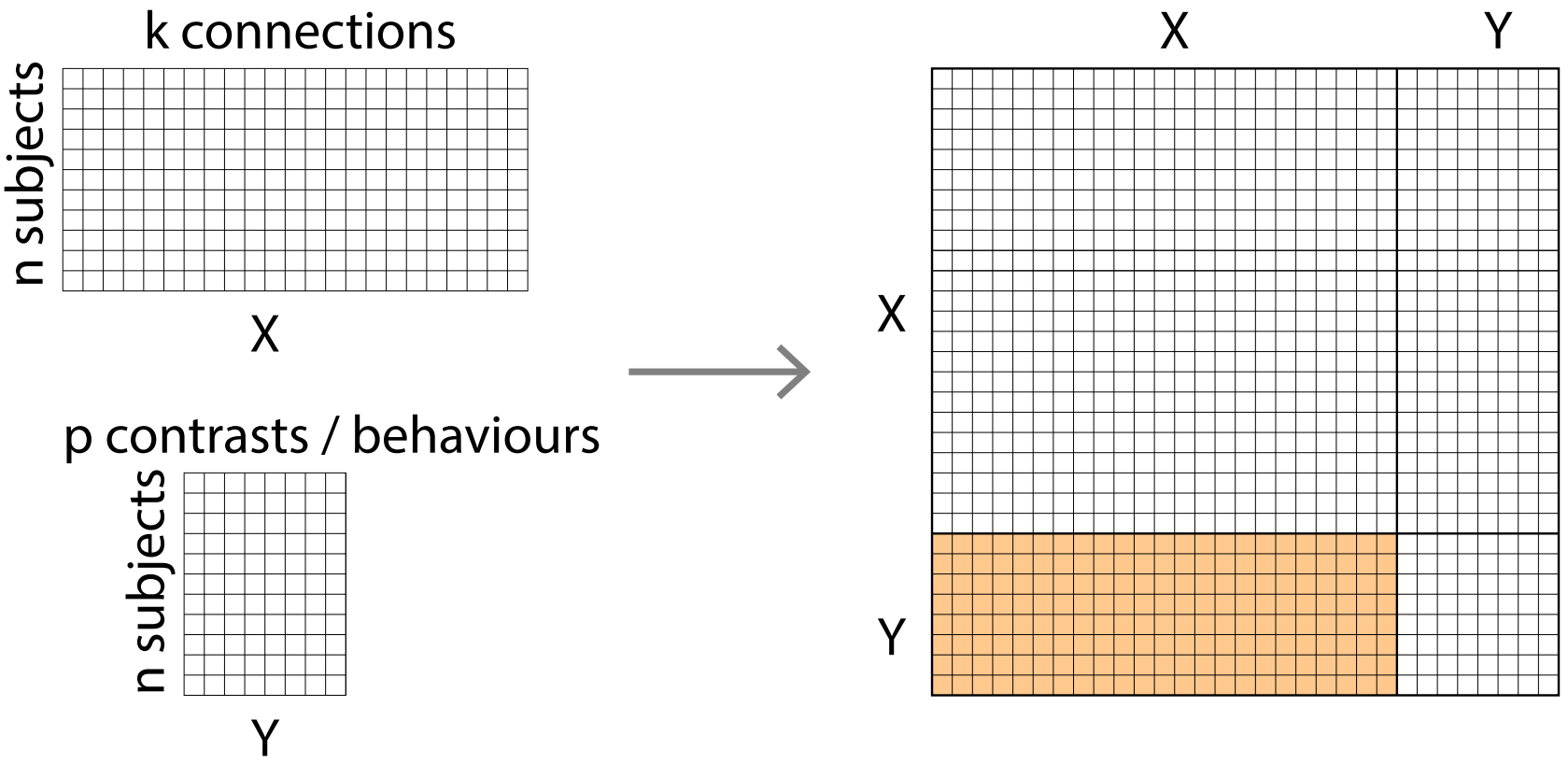
$$\mathrm{SVD}((\mathbf{X}'\mathbf{X}')^{-1/2}(\mathbf{X}'\mathbf{Y})(\mathbf{Y}'\mathbf{Y})^{-1/2})$$



Hotelling (1936) *Biometrika*

Canonical correlation analysis (CCA)

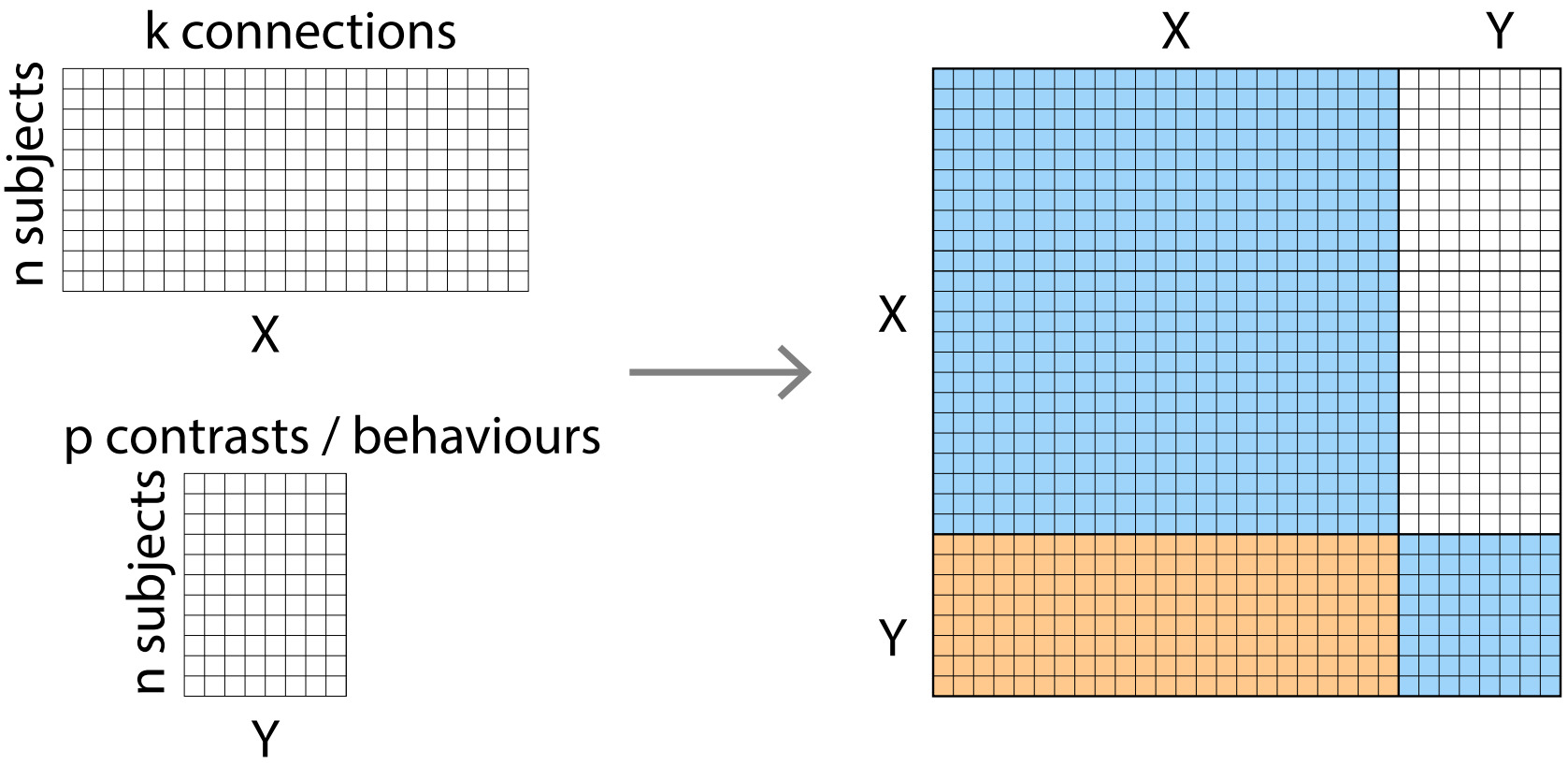
$$\mathrm{SVD}((\mathbf{X}'\mathbf{X}')^{-1/2}(\mathbf{X}'\mathbf{Y})(\mathbf{Y}'\mathbf{Y})^{-1/2})$$



Hotelling (1936) *Biometrika*

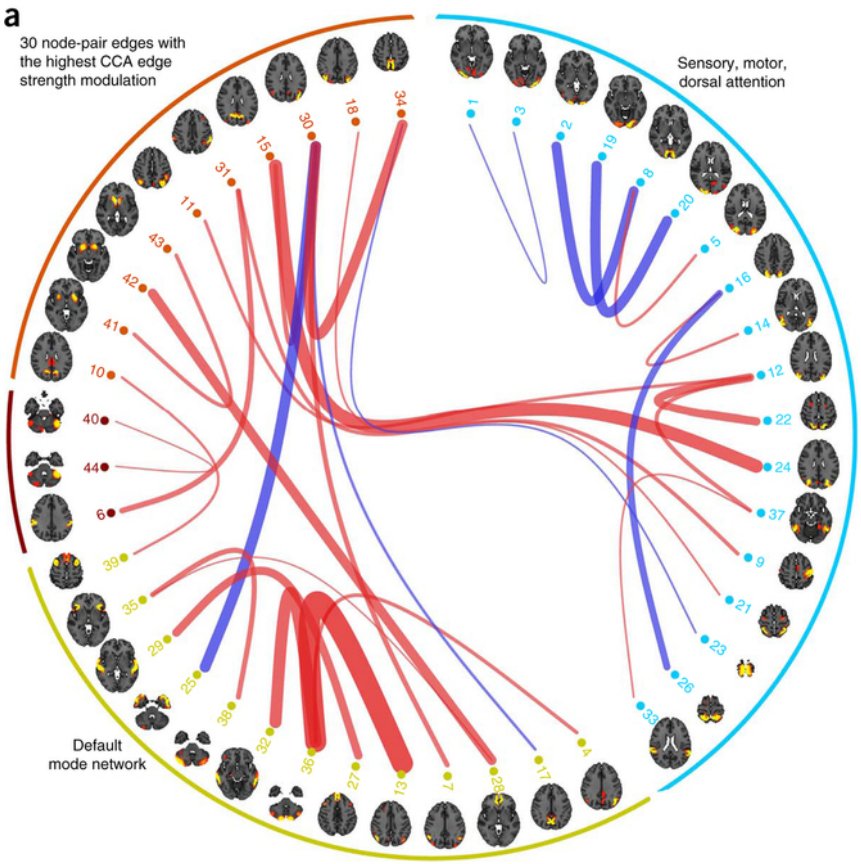
Canonical correlation analysis (CCA)

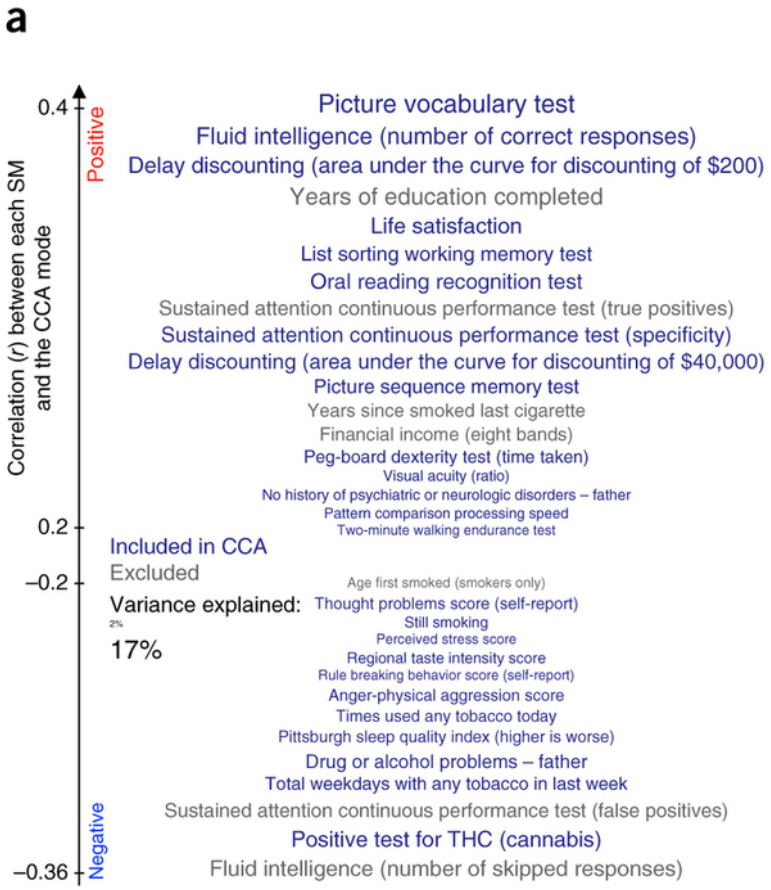
$$\mathrm{SVD}((\mathbf{X}'\mathbf{X}')^{-1/2}(\mathbf{X}'\mathbf{Y})(\mathbf{Y}'\mathbf{Y})^{-1/2})$$



Hotelling (1936) *Biometrika*

Example: relating connectivity and behaviour





Smith et al. (2015) *Nat Neurosci*

Linear discriminant analysis (LDA)

$$\mathrm{SVD}(\mathbf{W}^{-1}\mathbf{B})$$

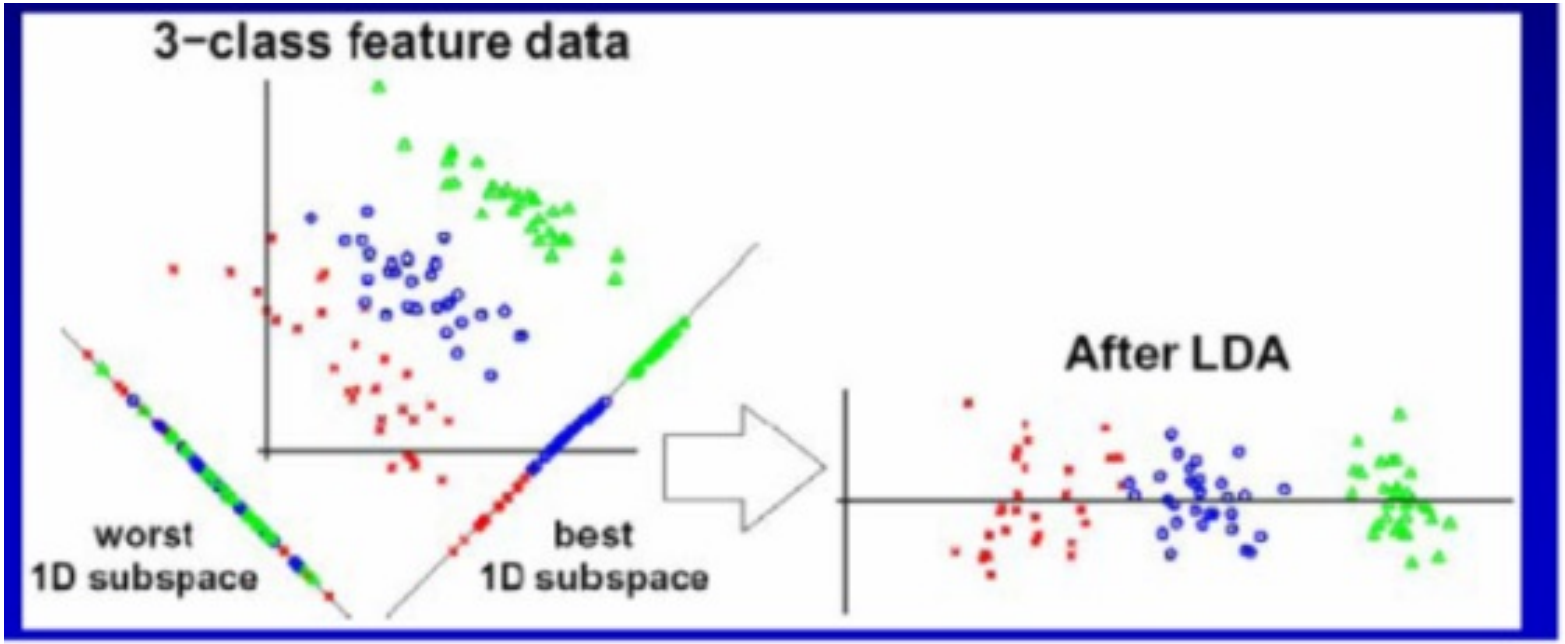


image: https://mlalgorithm.wordpress.com/

Friston et al. (1995) *NeuroImage*

Extensions

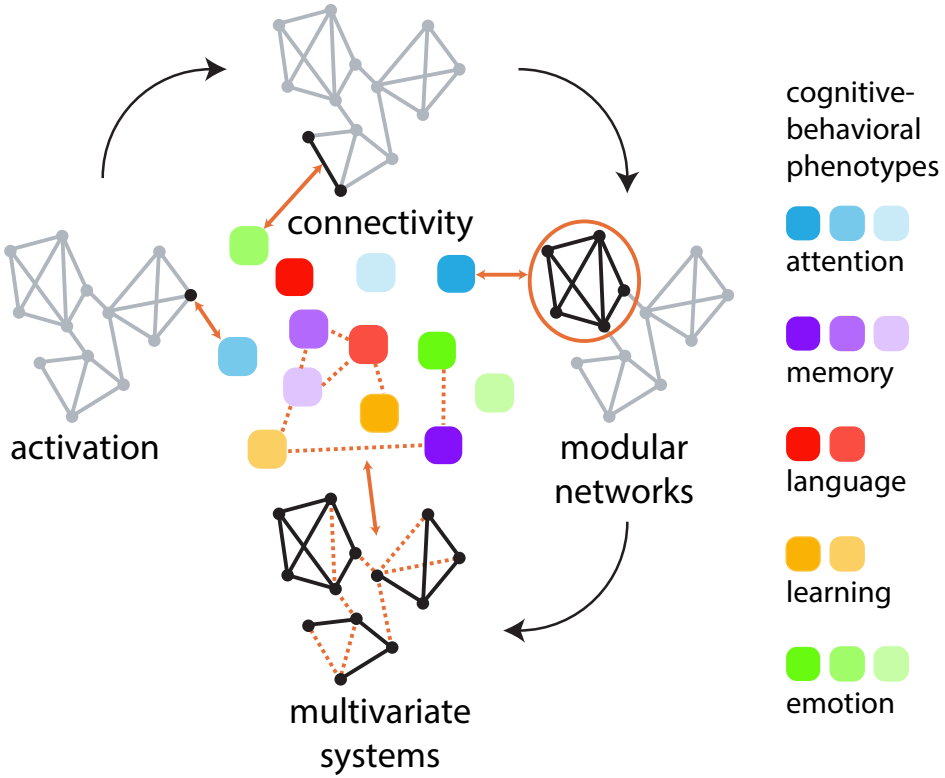
- sparse/regularized solutions, e.g.
  sCCA: Witten & Tibshirani (2009) *Stat Appl Genet Mol Biol*
  PLS-CA: Beaton et al. (2015) *Psychol Meth*
- extensions to 3+ data sets, e.g.
  PARAFAC: Bro (1997) *Chemometr Intell Lab*
  Multiway PLS: Wold et al. (1987) *J Chemometrics*
- nonlinear dependencies
- Bayesian implementations, e.g.
  IBFA: Virtanen et al. (2011) *ICML-II*
- prediction

Limitations and considerations

- overfitting
- linear
- unique partitioning of variance/covariance
- inference on individual variables

Summary

- multivariate models embody network property
- all techniques entail unique assumptions
- many linear multivariate techniques are related
- multivariate techniques are versatile

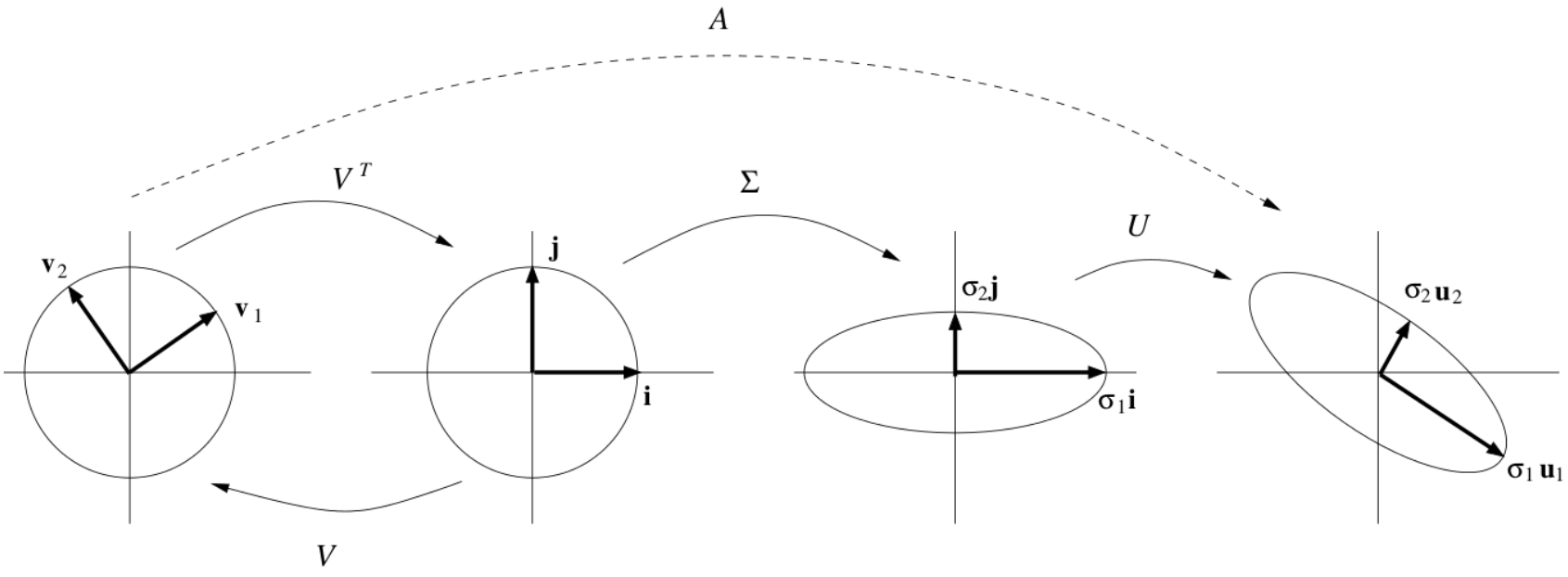


Resources

- Matlab toolbox for dimensionality reduction
  https://lvdmaaten.github.io/drtoolbox/
- PLS toolbox
  https://www.rotman-baycrest.on.ca/index.php?section=84
- sparse CCA
  http://statweb.stanford.edu/~tibs/Correlate/
- coming soon to Brainstorm

Matching randomized components

$$\mathbf{X}'\mathbf{Y} = \mathbf{U}\mathbf{S}\mathbf{V}' \longleftrightarrow \mathbf{X}'_{\text{boot}}\mathbf{Y} = \mathbf{U}_{\text{boot}}\mathbf{S}_{\text{boot}}\mathbf{V}'_{\text{boot}}$$



Milan & Whittaker (1995) *J Roy Stat Soc C*