Canonical Correlation Analysis (CCA) and Extensions

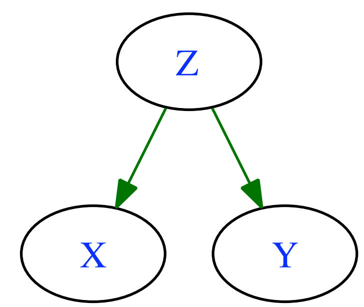
Max-Planck-Institute
for Biological Cybernetics

Outline

- Canonical Correlation Analysis (CCA)
  - Standard CCA Solution using Covariance Matrices
  - Example: Unsupervised categorisation of car types
- Kernel Canonical Correlation Analysis (kCCA)
  - For high-dimensional data and non-linear dependencies
  - Example: Cross-Language Semantic Content Extraction
- Temporal Kernel CCA (tkCCA)
  - For data with non-instantaneous couplings
  - Example: Multi-modal neuronal signals (invasive vs. non-invasive)
  - tkCCA estimates convolution linking non-invasive to invasive signals

Canonical Correlation Analysis (CCA)

Latent variable Z is measured in multivariate variables X and Y



$$X \in \mathbb{R}^M, Y \in \mathbb{R}^N$$

Which dimensions of X and Y reflect Z best?

CCA: Those dimensions that maximise the correlation between X and Y.

Given two (or more) multivariate variables

$$X \in \mathbb{R}^M, Y \in \mathbb{R}^N$$

CCA finds projections

$$w_x \in \mathbb{R}^M, w_y \in \mathbb{R}^N$$

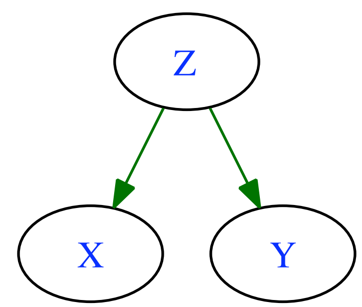
that maximise the covariance between the variables

$$\underset{w_x, w_y}{\operatorname{argmax}} \left(w_x^\top X Y^\top w_y \right) \quad \text{s.t.} \quad \begin{array}{c} w_x^\top X X^\top w_x = 1 \\ w_y^\top Y Y^\top w_y = 1 \end{array}$$

[Hotelling, Biometrika, 1936]

Canonical Correlation Analysis (CCA)

- Latent Variable **Z: Car Types**
- Measurements
  - **X**: Displacement, Horsepower, Weight
  - **Y**: Acceleration, Miles/Gallon



X =

Dimensions

Car models

Displacement	Horsepower	Weight
307	130	3504
350	165	3693
318	150	3436
	...	

Y =

Acceleration	Miles/Gallon
12.0000	18.0000
11.5000	15.0000
11.0000	18.0000
...	

Assuming centered data

$$\sum_i x_i = \sum_i y_i = 0$$

We can compute empirical
cross-covariance matrices
and auto-covariance matrices

$$C_{xy} = \frac{1}{N} X Y^\top$$

$$C_{xx} = \frac{1}{N} X X^\top$$

CCA objective

$$\underset{w_x, w_y}{\operatorname{argmax}} \left(w_x^\top X Y^\top w_y \right) \quad \text{s.t.} \quad \begin{array}{c} w_x^\top X X^\top w_x = 1 \\ w_y^\top Y Y^\top w_y = 1 \end{array}$$

Lagrangian

$$\mathcal{L} = w_x^\top C_{xy} w_y - \frac{1}{2}\alpha(w_x^\top C_{xx} w_x - 1) - \frac{1}{2}\beta(w_y^\top C_{yy} w_y - 1)$$

Partial Derivatives

$$\frac{\partial \mathcal{L}}{\partial w_x^\top} = C_{xy} w_y - \alpha C_{xx} w_x \qquad \frac{\partial \mathcal{L}}{\partial w_y^\top} = C_{yx} w_x - \beta C_{yy} w_y$$

We set the partial derivatives to 0 and multiply with $w_x^\top,\ w_y^\top$

$$w_x^\top C_{xy} w_y = \alpha w_x^\top C_{xx} w_x$$

$$w_x^\top C_{xy} w_y = \beta w_y^\top C_{yy} w_y$$

Thus from the auto-covariance constraints

$$1 = w_x^\top C_{xx} w_x = w_y^\top C_{yy} w_y$$

follows

$$\alpha = \beta$$

Given $\alpha = \beta$

the partial derivatives become

$$C_{xy} w_y = \alpha \; C_{xx} w_x$$
$$C_{yx} w_x = \alpha \; C_{yy} w_y$$

We can now reformulate these equations in block matrix form

$$\begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \alpha \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix}$$

which is just a *generalised eigenvalue equation*

Canonical Correlation Analysis (CCA)

Latent Variable

Z: Car Types

Measurements

X: Displacement, Horsepower, Weight

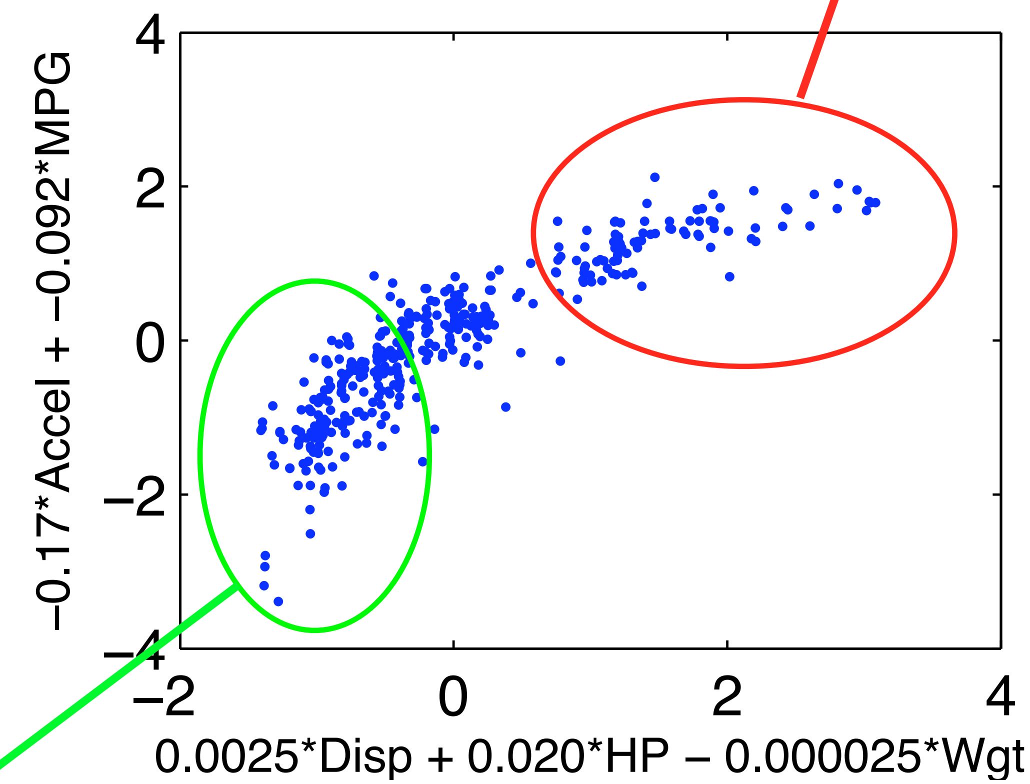
Y: Acceleration, Miles/Gallon

$$w_x = \begin{bmatrix} 0.0025 \\ 0.0202 \\ -0.000025 \end{bmatrix}$$

$$w_y = \begin{bmatrix} -0.17 \\ -0.092 \end{bmatrix}$$

Sport Cars
(big engine, strong, thirsty)

Yw_y

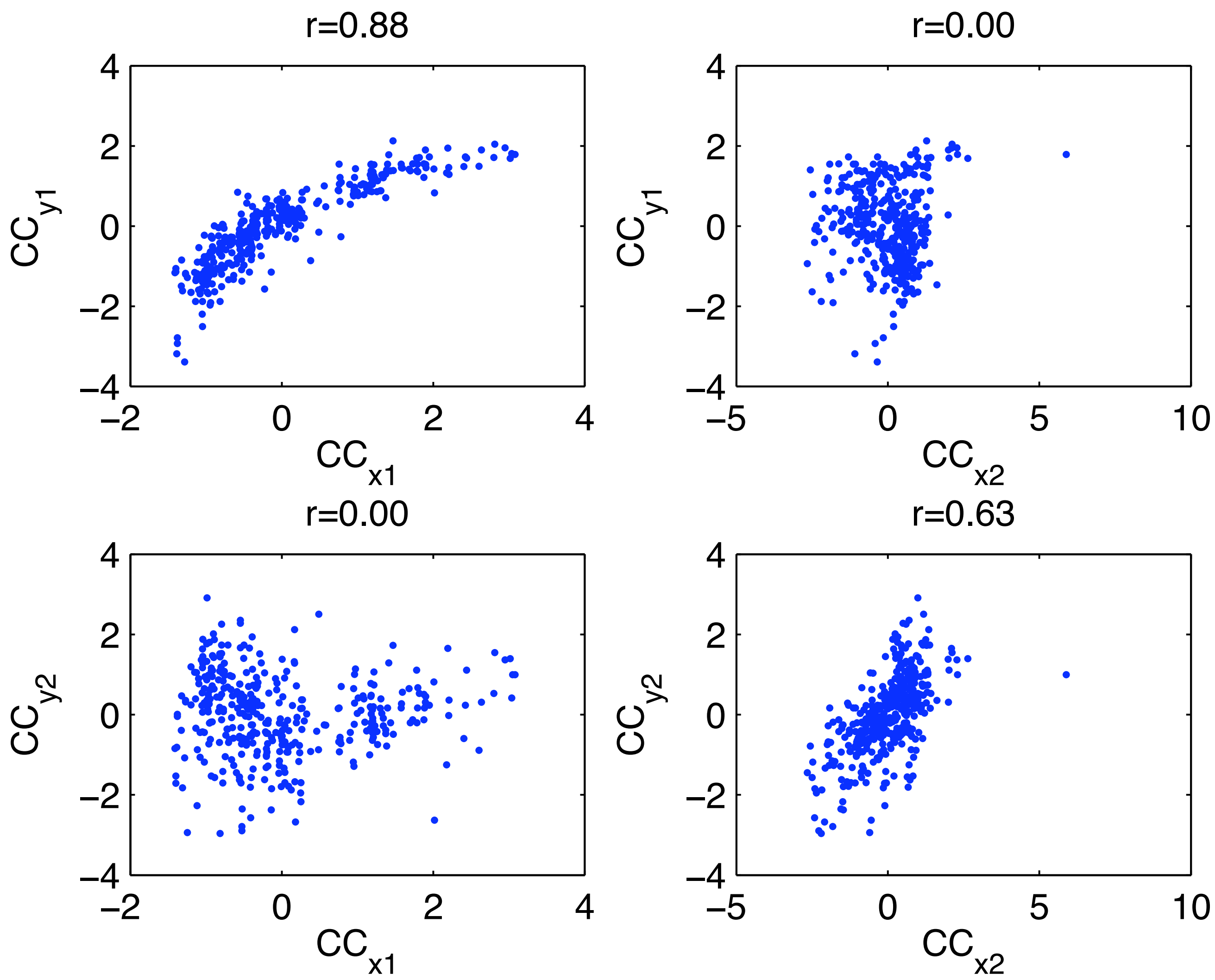


Xw_x

Commercial Cars
(small engine, low consumption)

Canonical Correlation Analysis (CCA)

After CCA



A short history of CCA

- Extensions of CCA
  - more than two variables [Kettenring 1971]
  - Kernel CCA (kCCA) [Akaho 2001]
    - finds non-linear dependencies
    - applicable to high-dimensional data
- Recently CCA became popular in
  - Machine Learning
    - Objective function for kernel ICA [Bach 2002]
    - Mutual information estimation [Gretton 2005]
  - Neuroscience
    - Receptive fields without spike triggering [Macke 2008]
    - Analysis of fMRI and multivariate stimuli [Hardoon 2007]
    - Analysis of multi-modal recordings [Bießmann 2009]

- Sometimes covariance matrices are too big to compute
  - Example: Bag-of-Words feature space (potentially infinite dimensional)
- CCA does not capture non-linear dependencies
- Solution:
  - Kernel Canonical Correlation Analysis (kCCA)
  - Operates on kernels of the data (not covariance matrices)

Bag-of-Words Feature Representation

U.K.

Germany

Time [Days] →

Words ↓

Extracted Word Counts

Volcano	0	4	3	
ash	0	3	2	...
flights	1	5	2	
⋮				

Vulkan	0	3	2	
Asche	0	1	2	...
Flüge	2	4	3	
⋮				

Kernel Canonical Correlation Analysis (kCCA)

Intuition behind the Kernel Trick:

Any solution found by CCA has to lie
in the subspace spanned by the data points

A sufficient representation of this subspace can be obtained by
the inner products of all data points (linear kernels)

$$K_x = X^\top X$$

$$K_y = Y^\top Y$$

No need to compute big covariance matrices!

The solution of CCA in kernel space is obtained by solving the generalised eigenvalue problem

$$\begin{bmatrix} 0 & K_x K_y \\ K_y K_x & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} K_x^2 & 0 \\ 0 & K_y^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

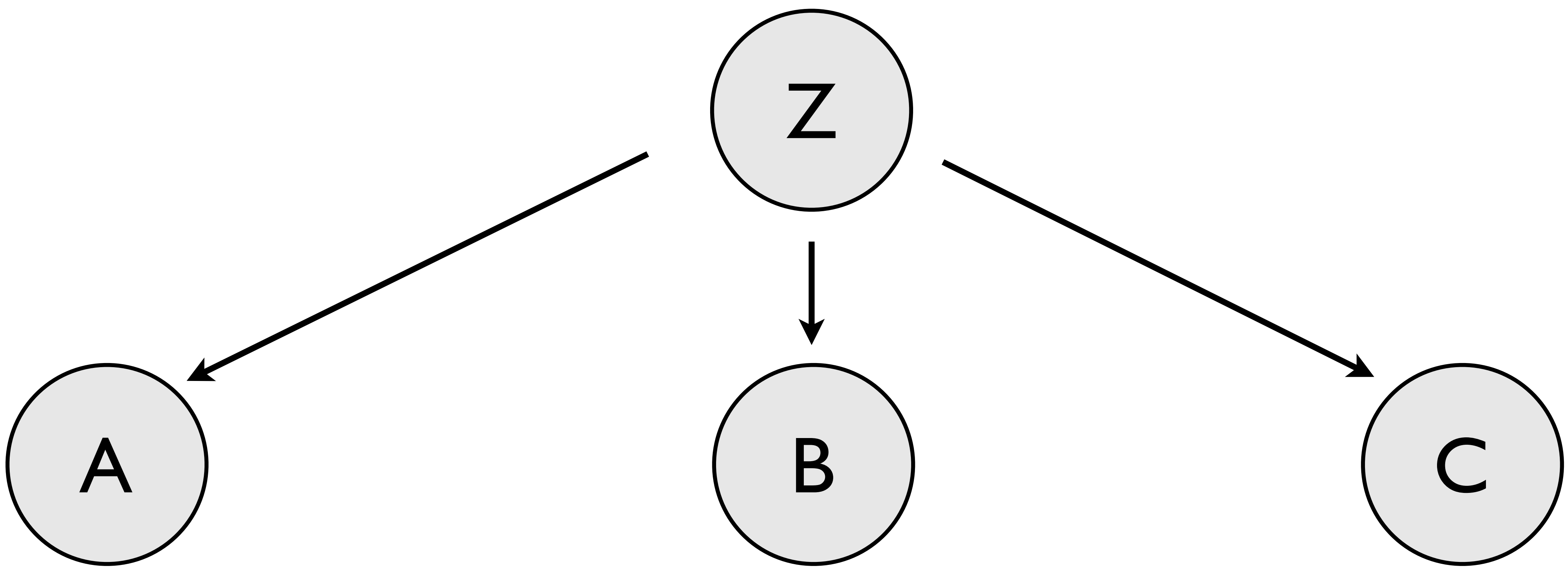
The solutions in the input space can be recovered by

$$w_x = X\alpha_x$$

$$w_y = Y\alpha_y$$

Kernel Canonical Correlation Analysis (kCCA)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra



Preamble
In the name of God Almighty!
We, the Swiss People and Cantons,

Präambel
Im Namen Gottes des Allmächtigen!
Das Schweizervolk und die Kantone,

Preambolo
In nome di Dio Onnipotente,
Il Popolo svizzero e i Cantoni,

Example: Suisse Constitution
Common Semantic Content Extraction
in multi-lingual text corpora

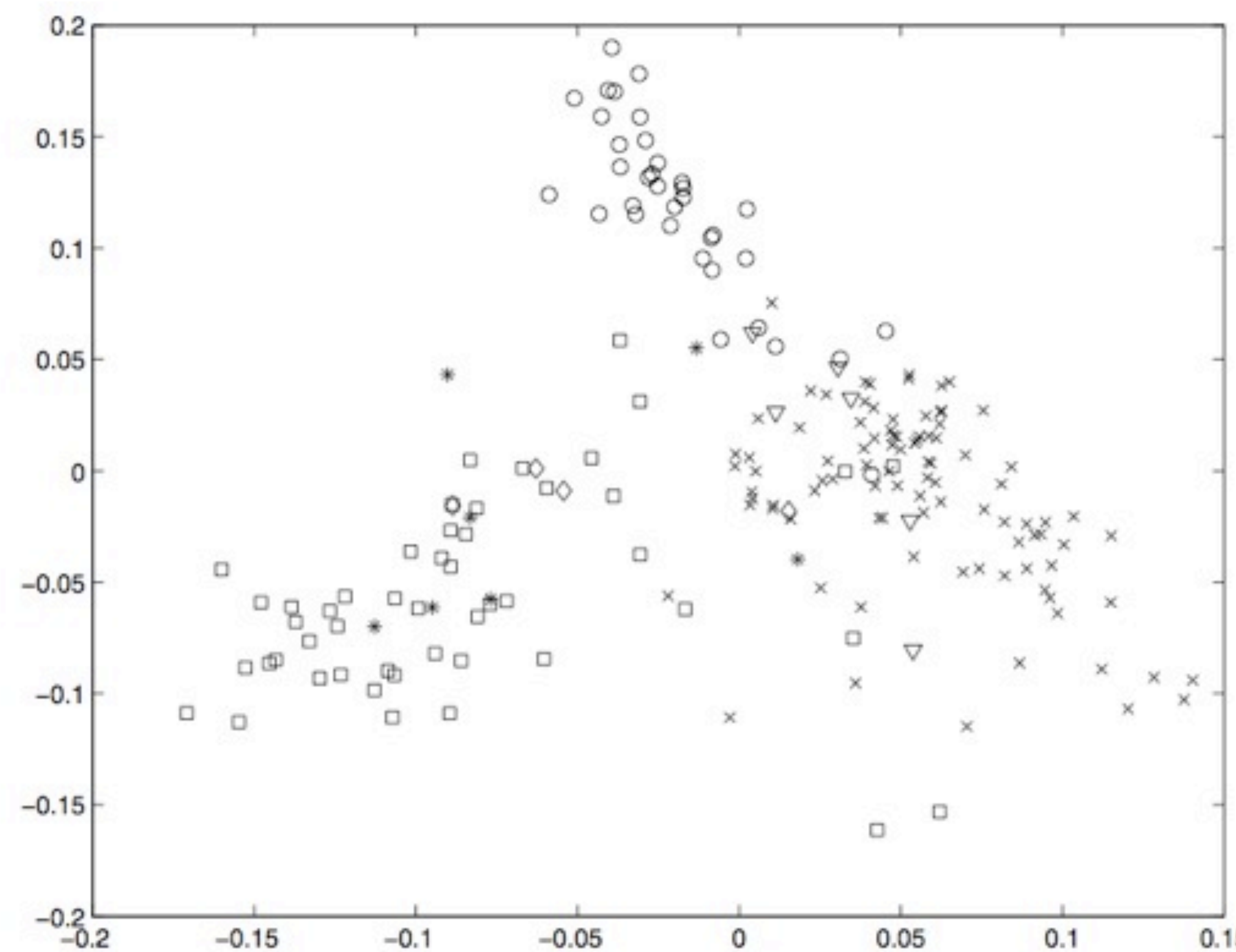
Kernel Canonical Correlation Analysis (kCCA)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

× Fundamental Rights

□ Political Rights

✳ Social Objectives



CC-2

CC-1

De Bie and Cristianini. Kernel methods for exploratory pattern analysis: a demonstration on text data. Springer Lecture Notes in Computer Science, 2004

Temporal kernel CCA (tkCCA)

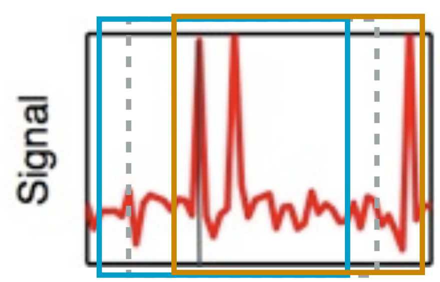
- If variables are coupled with delays
  - simultaneous samples will not be correlated
  - Standard (k)CCA will not find the right solution
- Solution
  - Shift one variable relative to the other
  - Maximise correlation for (a sum over) all relative time lags
  - ➡ (k)CCA finds *canonical variates and correlation*
  - ➡ tkCCA finds *canonical convolution and correlogram*

$$\underset{w_x(\tau), w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_{\tau} w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)$$

Temporal kernel CCA

$$\underset{w_x(\tau),w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_\tau w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)$$

$$\tilde{X} = \begin{bmatrix} \boxed{X_{\tau_1}} \\ X_{\tau_2} \\ \vdots \\ \boxed{X_{\tau_T}} \end{bmatrix}$$



Data is *embedded in its temporal context* by appending time shifted copies to each data point

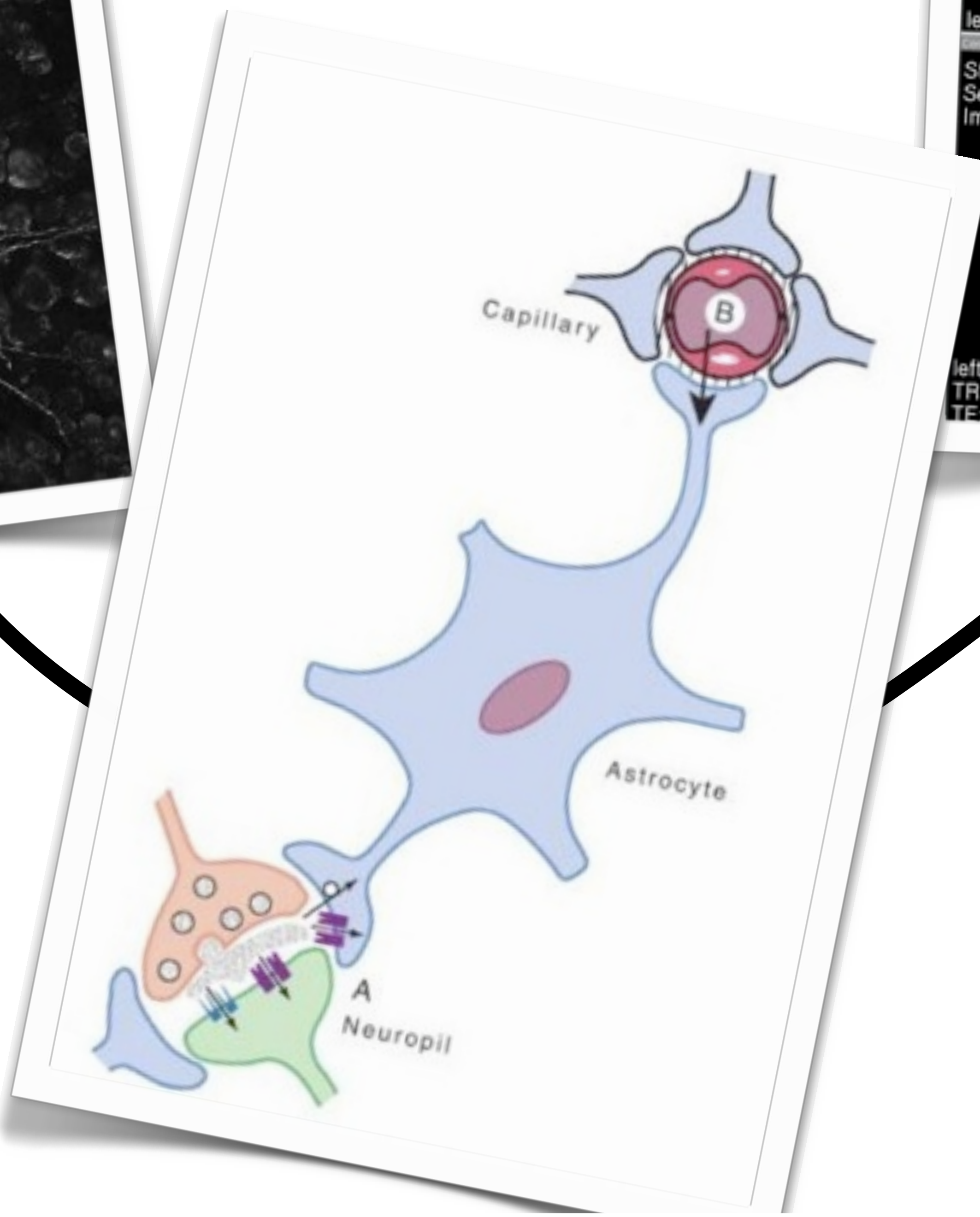
Temporal kernel CCA

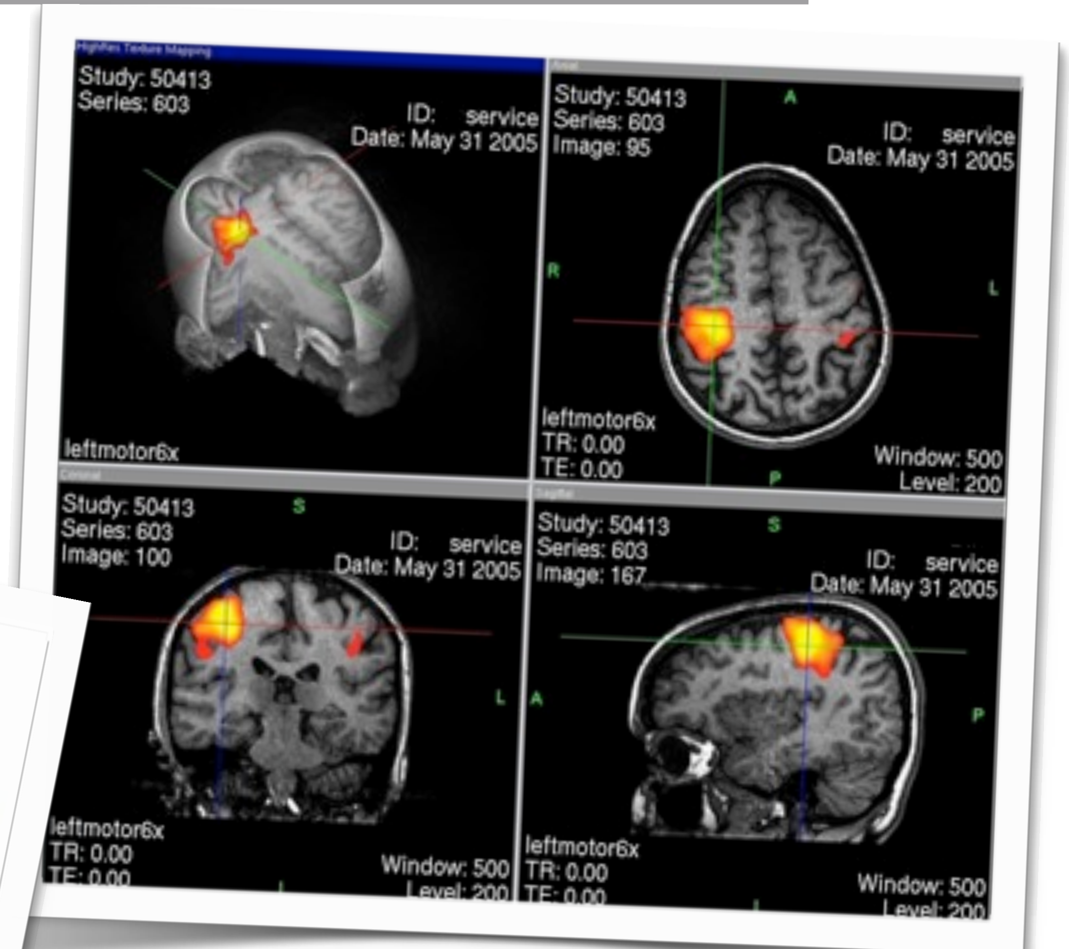
$$\underset{w_x(\tau), w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_\tau w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)$$

$$\tilde{X} = \begin{bmatrix} X_{\tau_1} \\ X_{\tau_2} \\ \vdots \\ X_{\tau_T} \end{bmatrix} \qquad \tilde{w}_x = \begin{bmatrix} w_x(\tau_1) \\ w_x(\tau_2) \\ \vdots \\ w_x(\tau_T) \end{bmatrix}$$

$$\Downarrow$$

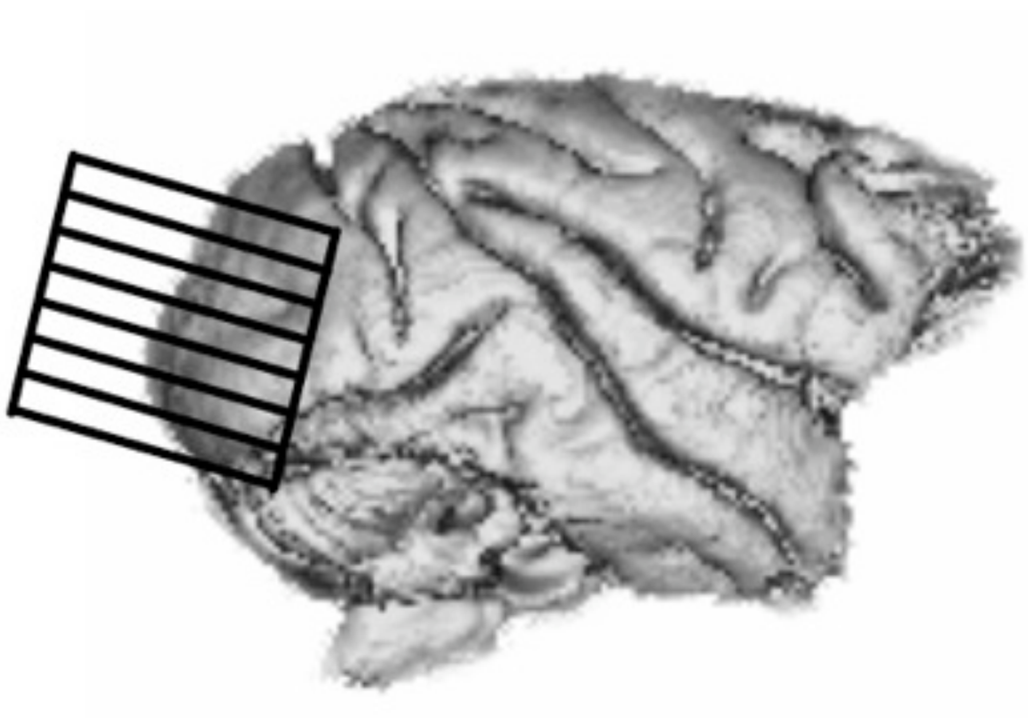
$$\underset{w_{\tilde{x}}, w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\tilde{w}_x^\top \tilde{X}, w_y^\top Y\right)$$

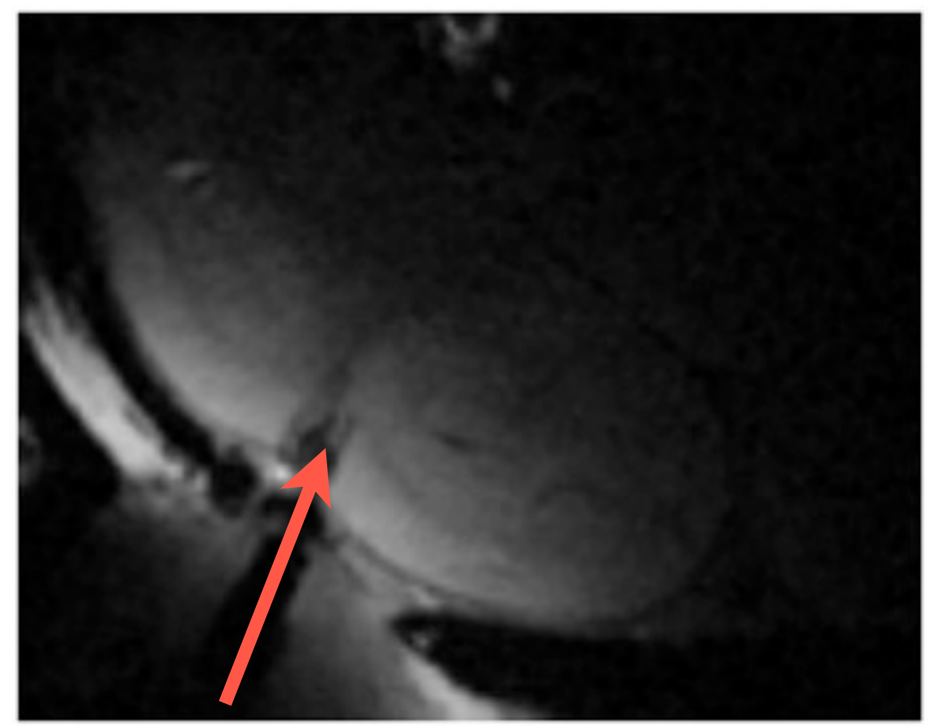

Capillary
B
Astrocyte
A
Neuropil


Study: 50413
Series: 603
ID: service
Date: May 31 2005
leftmotor6x
Study: 50413
Series: 603
Image: 95
ID: service
Date: May 31 2005
leftmotor6x
TR: 0.00
TE: 0.00
Window: 500
Level: 200
Study: 50413
Series: 603
Image: 100
ID: service
Date: May 31 2005
leftmotor6x
TR: 0.00
Window: 500
Study: 50413
Series: 603
Image: 167
ID: service
Date: May 31 2005
leftmotor6x
TR: 0.00
Window: 500

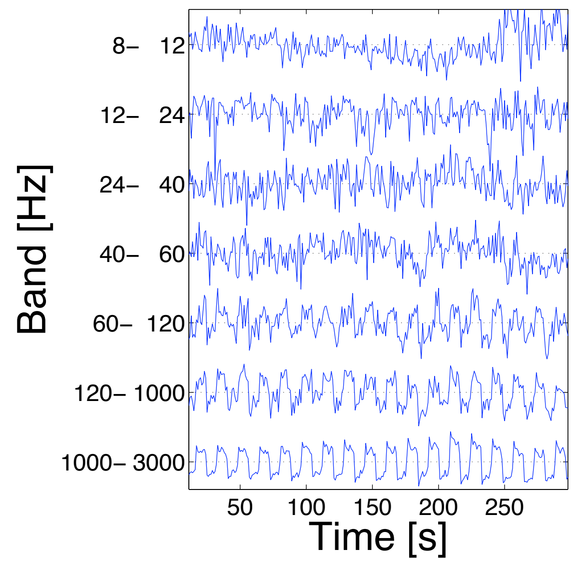
Simultaneous measurements of

- fMRI/ BOLD signal
- Intracortical neural activity

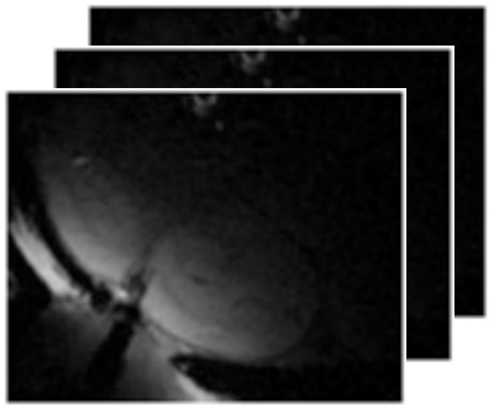




Spectrogram of neural activity



fMRI
Time series



Data Setting

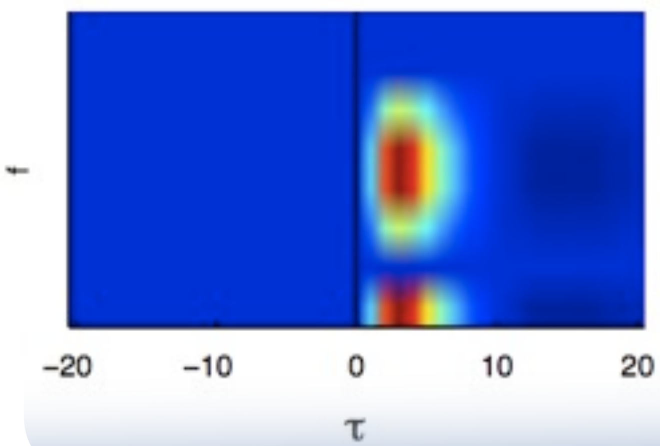
1. X and Y have different dimensions
2. Data is high-dimensional
3. Couplings are non-instantaneous

Appropriate Method

➡ CCA
➡ kCCA
➡ tkCCA

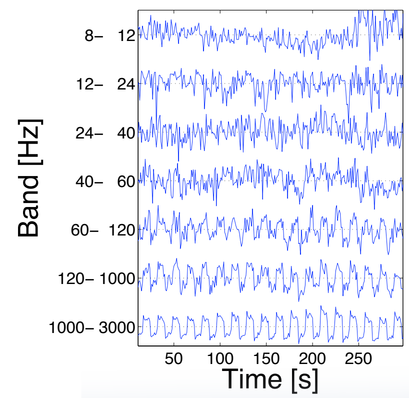
Temporal kernel CCA

$w_x(\tau)$

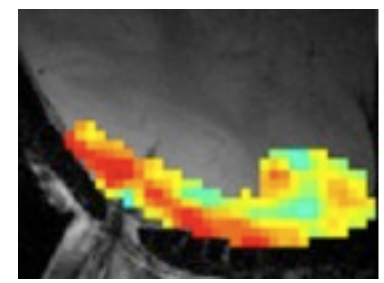


$\odot$

$x(t)$

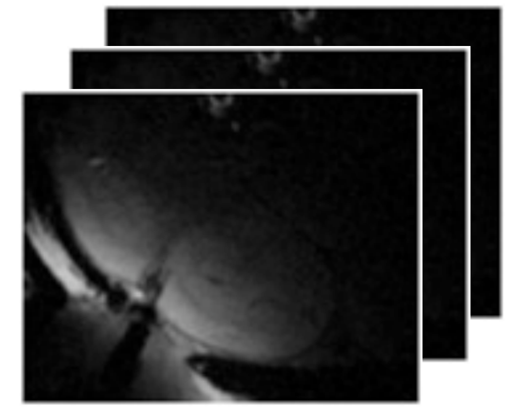


w_y



$\times$

$y(t)$



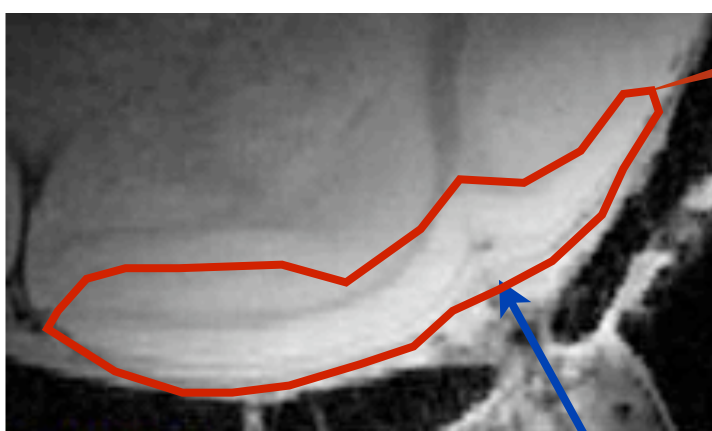
$$\underset{w_x(\tau), w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_\tau w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)$$

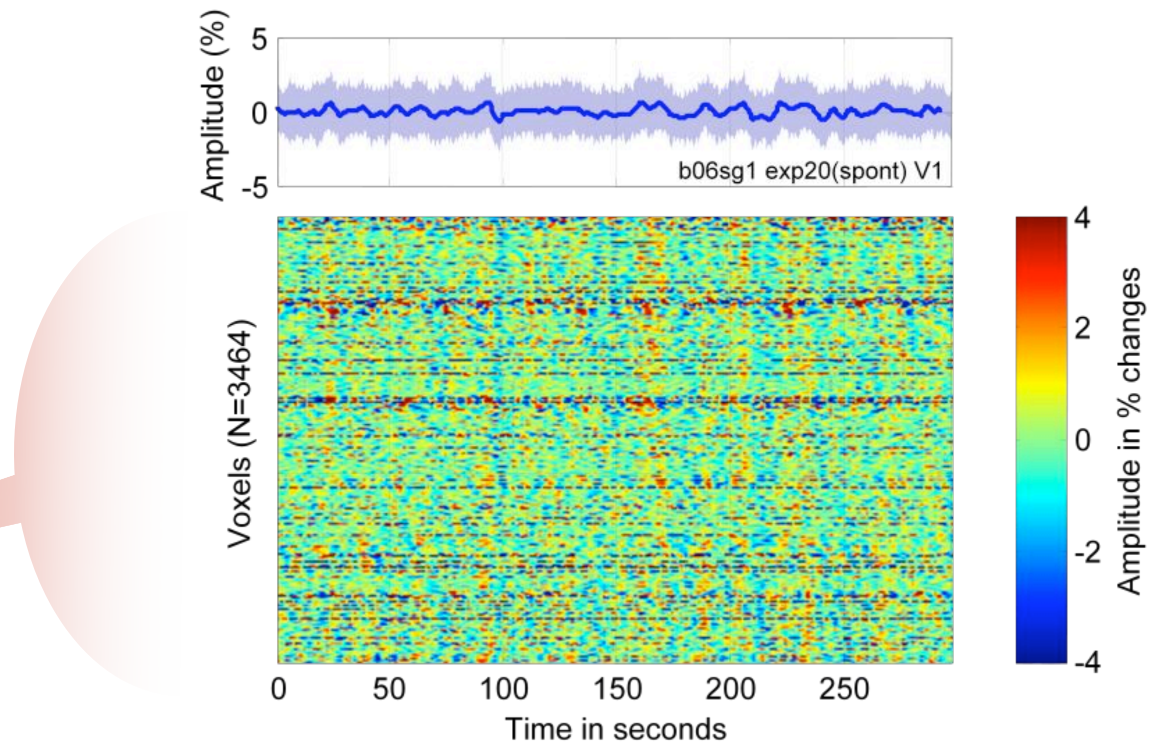
multivariate convolution of the neurophysiological signal with frequency dependent HRF

spatial weighting of voxels with activation pattern.

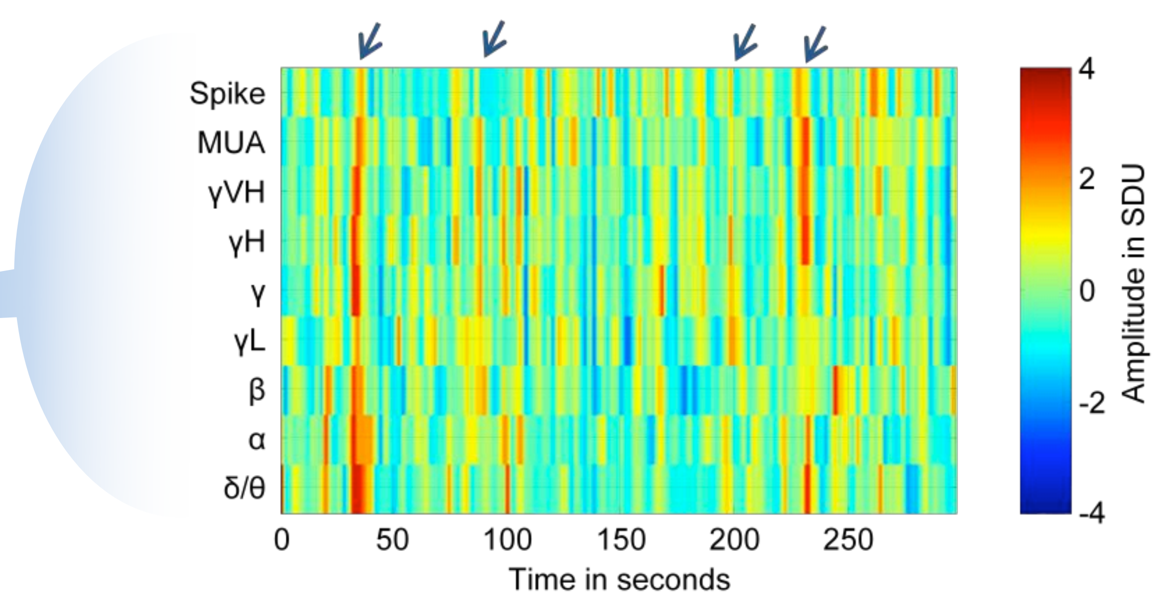
Raw Data During Spontaneous Activity

BOLD signal

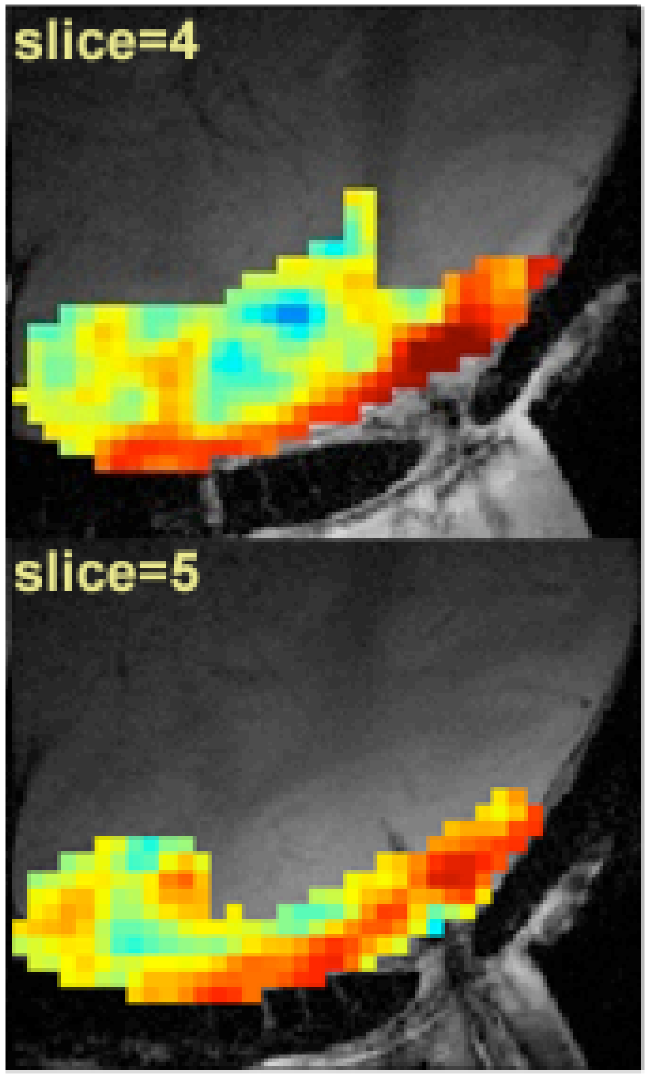




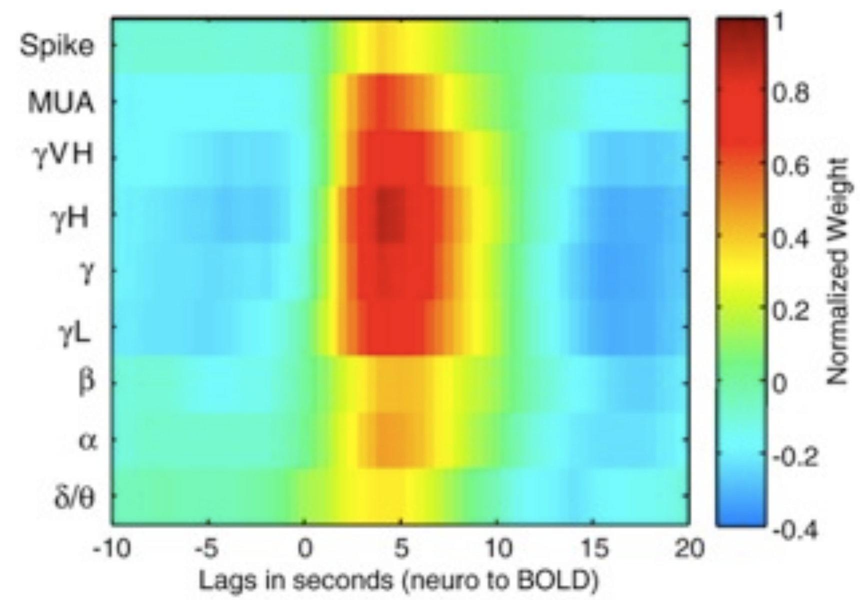
Neural signal



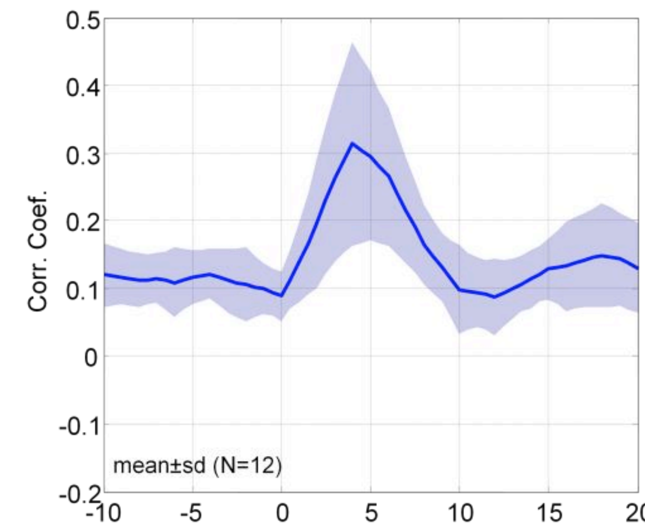
tkCCA Results: Spatial dependencies and HRF



Spatial Dependencies



Haemodynamic Response Function



Canonical Correlogram

Murayama et al., "Relationship between neural and haemodynamic signals during spontaneous activity studied with temporal kernel CCA", Magnetic Resonance Imaging, 2010

- CCA
  - finds projections for sets of variables that maximise correlation
- kernel CCA
  - extends CCA to non-linear dependencies
  - applicable to high dimensional data
- Temporal kernel CCA
  - extends kCCA to data with non-instantaneous correlations
  - computes multivariate convolution from one modality to another

References

H. Hotelling, "Relations between two sets of variates",
Biometrika 1936

T.W. Anderson, "An Introduction to Multivariate Statistical Analysis", 1958

Tijl De Bie and Nello Cristianini.
"Kernel methods for exploratory pattern analysis: a demonstration on text data".
Springer Lecture Notes in Computer Science, 2004

F. Bießmann, F.C. Meinecke, A. Gretton, A. Rauch, G. Rainer, N.K. Logothetis, K.-R. Müller
"Temporal Kernel CCA and its Application in Multimodal Neuronal Data Analysis"
Machine Learning Journal, 2010

Y. Murayama, F. Bießmann, F.C. Meinecke, K-R. Müller, M. Augath, A. Öltermann, N.K. Logothetis
"Relationship between neural and haemodynamic signals during spontaneous activity studied with tkCCA",
Magnetic Resonance Imaging, 2010

F. Bießmann and A. Harth.
"Analysing Dependency Dynamics in Web Data".
AAAI Spring Symposium, 2010