Imperial College London

Tensor decomposition for data mining from brain electrical responses

MEng Thesis

Divyansh Manocha

Supervisor: Professor Danilo Mandic
Electronic and Information Engineering, Imperial College London

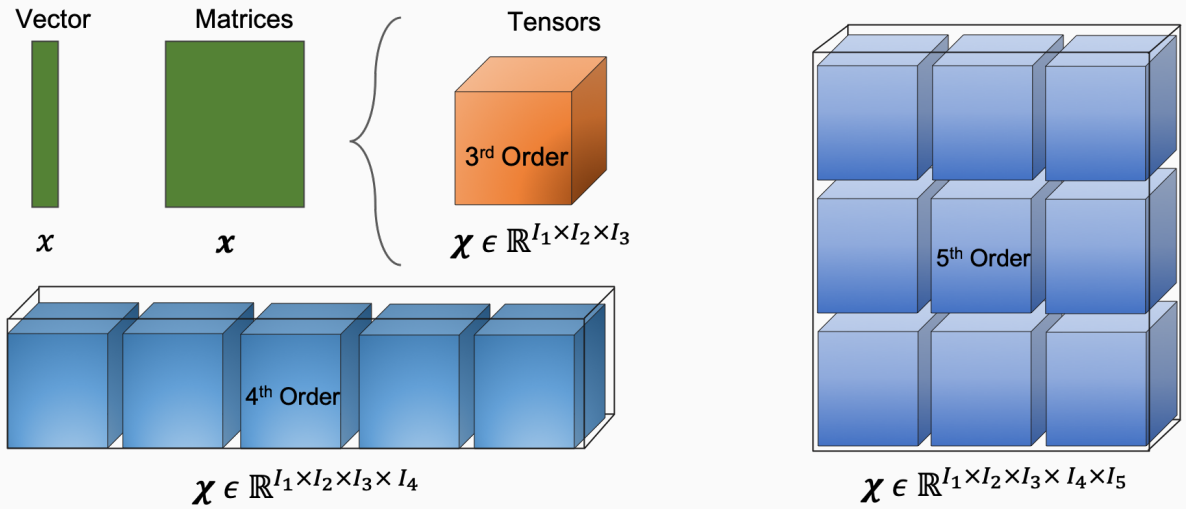
June 26, 2019

Table of contents

Background

From Matrices to Tensors

`Tensors` are multi-linear generalisations of arrays.



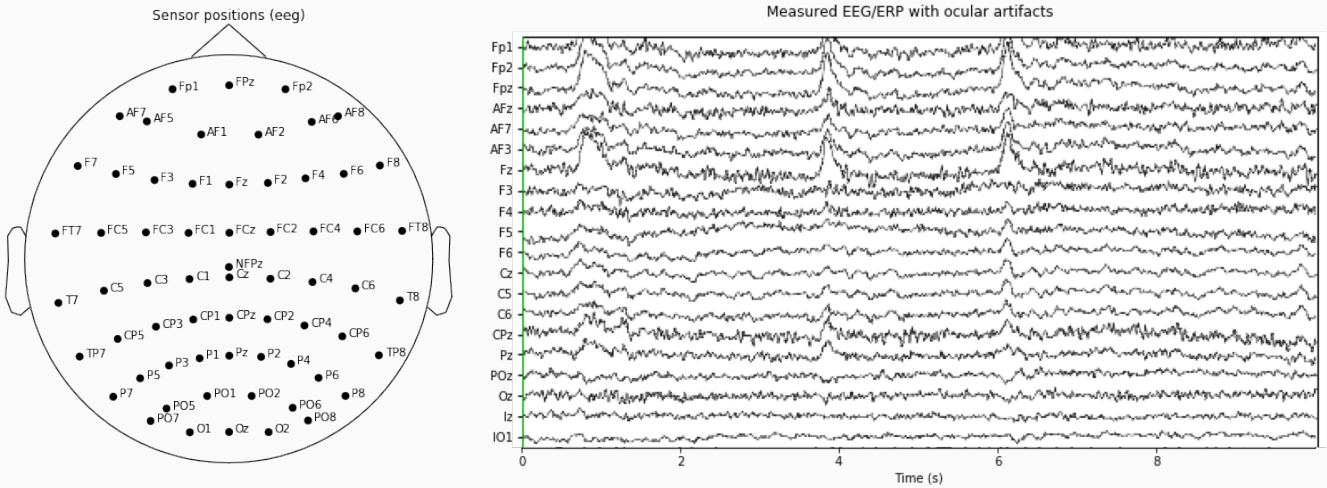
Generalising to tensors

- Data - natural representation
- No well defined rank, and to find one is a NP-hard problem
- There are *N modes* for a Tensor of order *N*

EEG

`ElectroEncephaloGram (EEG)` capture voltage fluctuations in the brain measured with the use of electrodes, in a non-invasive manner.

- High temporal resolution
- Unusual behaviour offers insights into neurological problems (e.g. epilepsy)



Sensor positions for EEG on the scalp

Problem 1: Separation

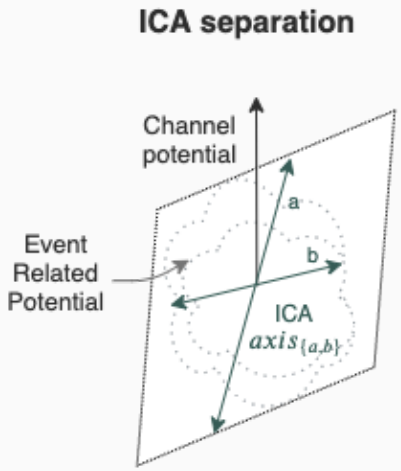
Superposition of neural activity, noise and undesirable distortions.

Noise made of biological or technical artifacts

Decomposition by ICA

Independent Component Analysis performs a full rank matrix factorization into statistically independent components

$$
\begin{aligned}
X &= B \cdot c \\
&= \sum_{i}^{R} c_i b_i^T \\
&= \sum_{i}^{R} Y_i
\end{aligned}
$$



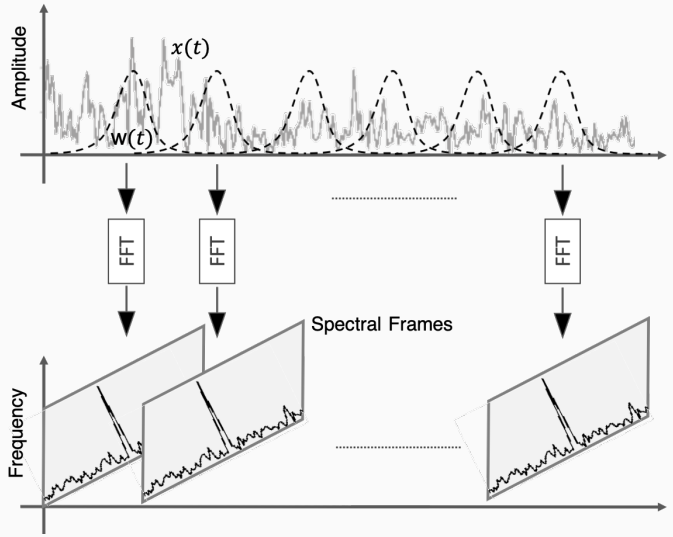
Assumptions:

- PDFs are not Gaussian
- Statistical independence

Works on artifacts that occur often and coherently, such as eye movements.

Short Time Fourier Transform

Temporal localisation of the Fourier spectrum $x(t)$ using the shifted window function of fixed duration and shape. Assumption: each segment is stationary (likely valid for small time windows)



STFT illustration

Fixed Time and Frequency resolution

$$STFT(t,f) = \int_{-\infty}^{\infty} \left(x(t)w^*(t-t')\right) e^{-2\pi jft} dt$$

$$Wf(s,u) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi^*(\frac{t-t'}{s}) dt \quad (1)$$

Wavelet transforms: the coefficients are functions of position and scale.

Notations and Definitions: EEG

Epoch

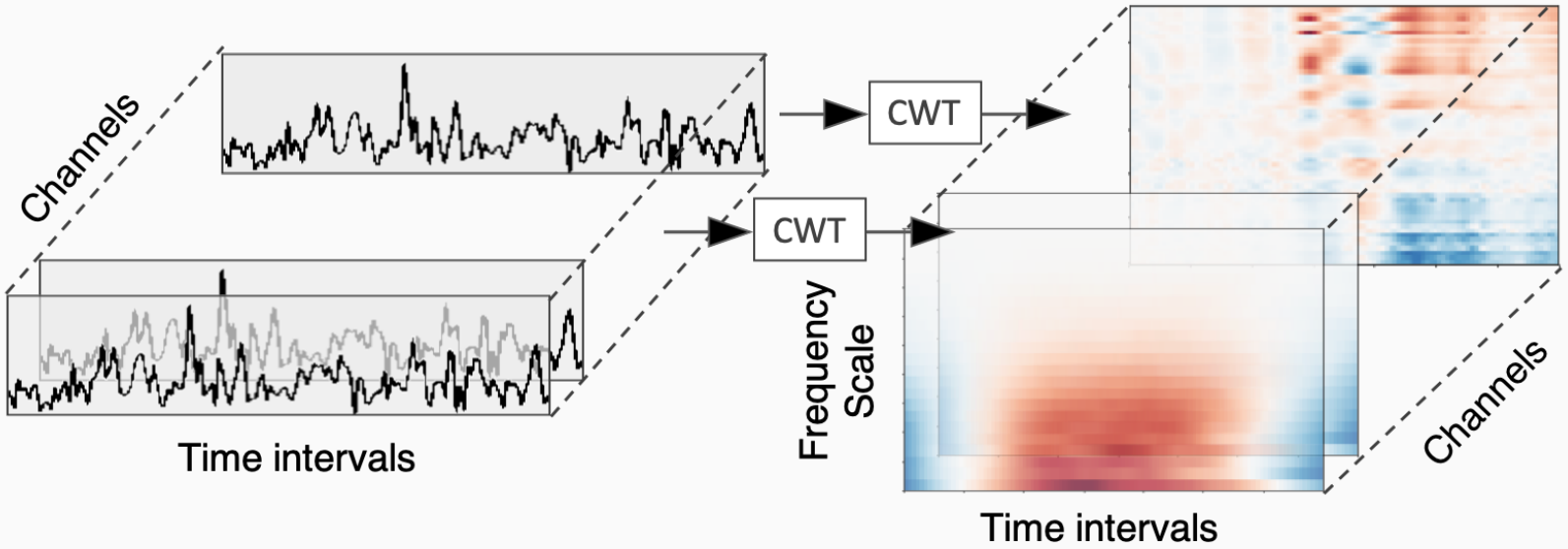
An extracted time window from the continuous signal, that is time locked relative to an event/baseline.

Traditional analysis uses of Event Related Potentials (ERPs).

ERPs

ERP components formed by averaging epochs.

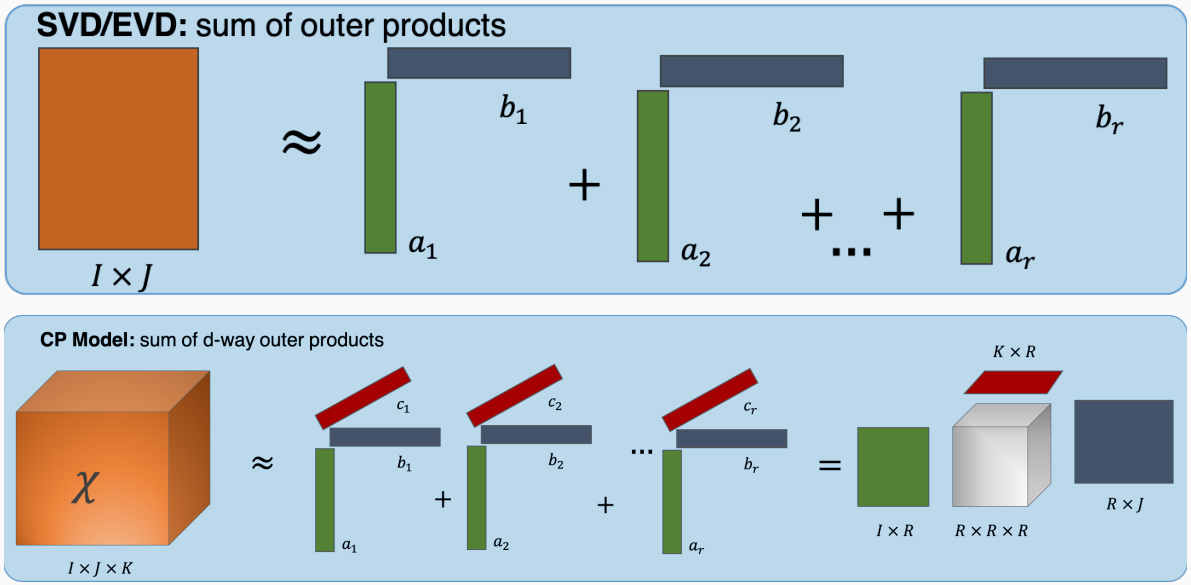
Small amplitudes of *'true'* signal $\implies$ EEG very sensitive to artifacts



EEG Tensors

Theoretical hypothesis

Decomposition methods



View 1 of extending matrix decompositions

CP Decomposition:

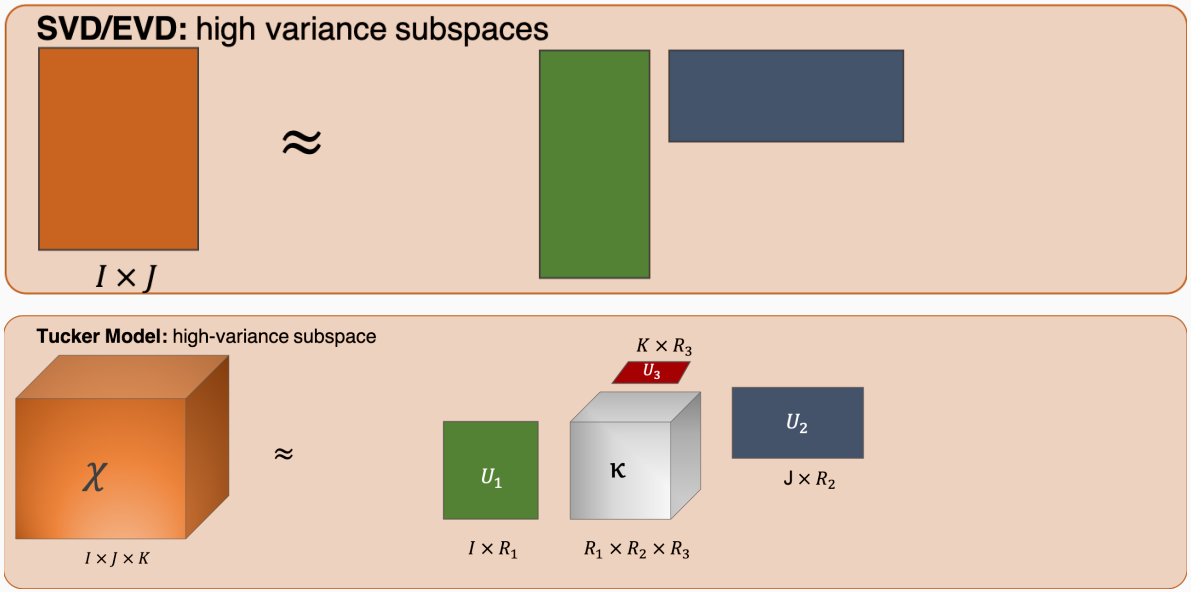
$$\mathcal{X}_{i_1,i_2,..,i_d} = u_1^{(d)} \otimes u_1^{(d-1)} \otimes \ldots \otimes u_1^{(1)} + \ldots + u_R^{(d)} \otimes u_R^{(d-1)} \otimes \ldots \otimes u_R^{(1)}$$
$$\mathcal{X} \approx \sum_{r=1}^{R} \lambda_r b_r^{(1)} \circ b_r^{(2)} \circ \ldots \circ b_r^{(d)} \tag{2}$$

Rank

Smallest number of rank one components that generate $\mathcal{X}$ as their sum

Alternating least square algorithm individually optimises each component

Decomposition methods



View two of extending matrix decompositions

Tucker Decomposition:

$$\begin{aligned}\chi &\approx \sum_{r_1=1}^{R_1} \dots \sum_{r_N=1}^{R_N} \kappa_{r_1,r_2\dots r_N} (u_{r_1}^{(1)} \circ u_{r_2}^{(2)} \circ \dots \circ u_{r_N}^{(N)}) \\ &= \boldsymbol{\kappa} \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_N U^{(N)} \\ &= [[\boldsymbol{\kappa}; U^{(1)}, U^{(2)}, \dots, U^{(N)}]]\end{aligned} \tag{3}$$

HOSVD or HOOI

Why EEG and Tensors

Properties

EEG signals are non-stationary, non-linear and not Gaussian in general
Spatially correlated to their neighbouring channels. Shown by Ille et al. (2002).

Parafac

For a particular channel p, $\mathcal{Y}$ can be written as shown in equation 4.

$$(\mathcal{Y}_i)_{channel\ p} = a_i \cdot c_{i,p} \cdot b_i^T \tag{4}$$

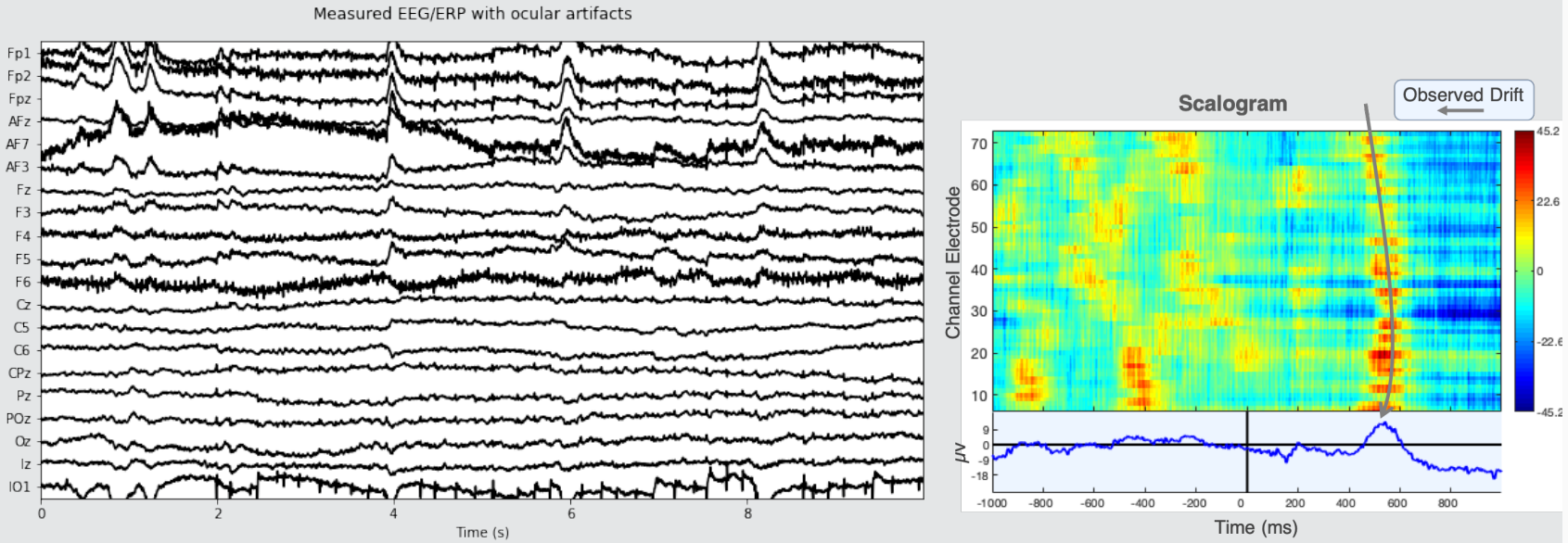
$a_i \cdot b_i^T$ *represents the time frequency distribution*

Observations

- Each channel will only vary by a scale factor: c_i
- Non-stationarity*(ish)* can now be represented if $a_i \cdot b_i^T$ is rank one
- Atoms from CPD must be rank one

Problem 2: Practical issue in measuring EEG

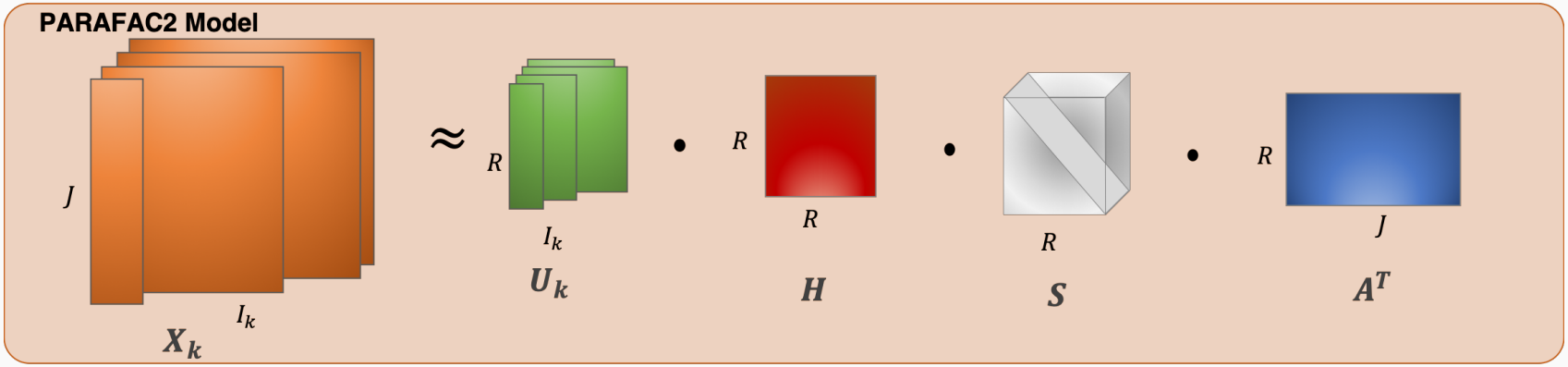
Moving temporal sources in the assumed synchronous channels (brain activity, sweating or muscle tension). Observable in scalograms.



(Left) Time series view of the channels from the real *reading* dataset. (Right) Generated scalogram of the dataset. Arrowed lines indicates asynchronicity.

Slow time shifts change zero level across channels considerably

Difficult to prevent at high frequency recordings from different spatial locations.



PARAFAC2

Definition: A generalised CPD model:

$$\hat{\mathcal{X}} = \sum_{i=1}^{R} a_i \circ (F_i \cdot diag(c_i)) \tag{5}$$

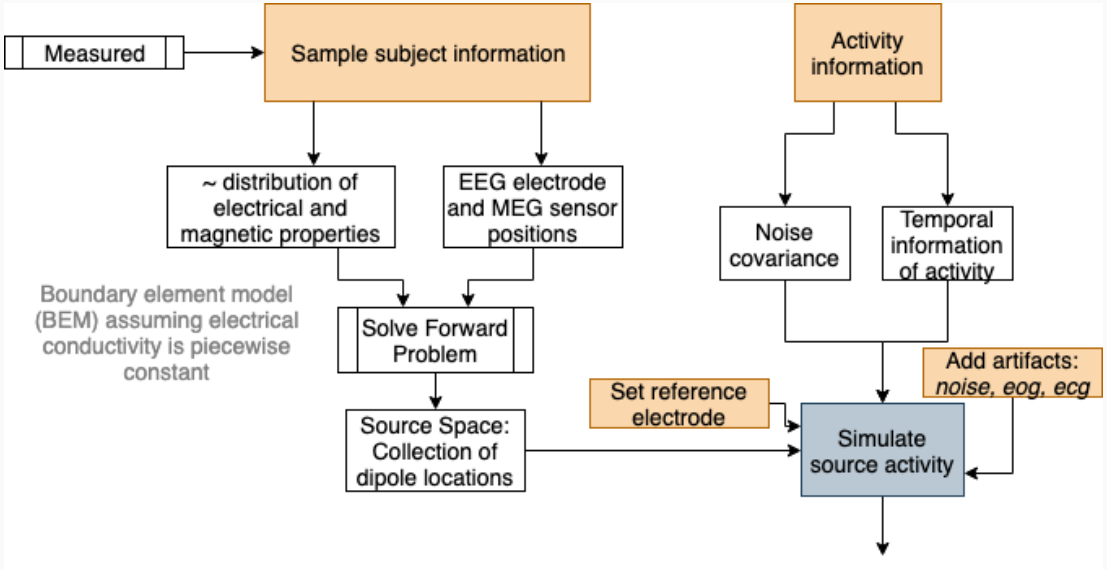
matrix $F_i \cdot diag(c_i)$ is new time-varying channel signature.

Allows evolution of channel signatures over time.

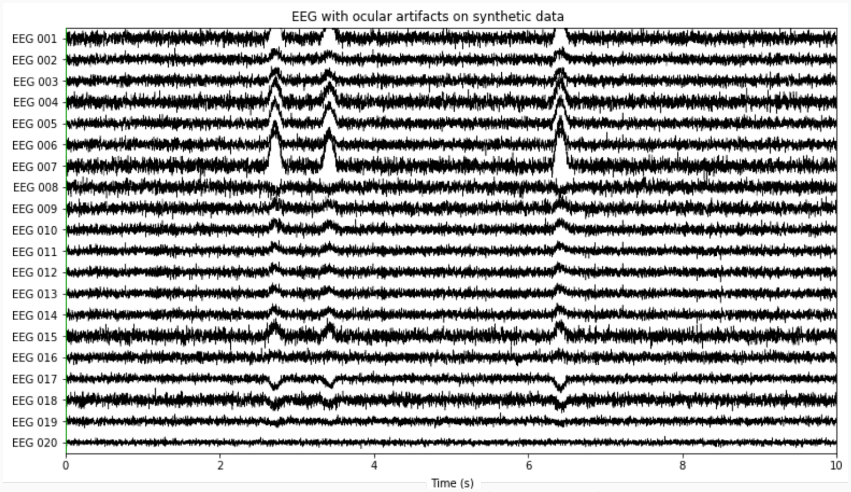
Measuring performance

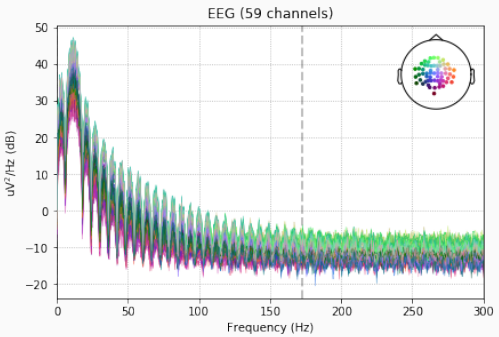
Synthetic Data

Multivariate Gaussian distribution using noise covariance from a real recording. Using MNE from Gramfort et al. (2014).



Simulation process



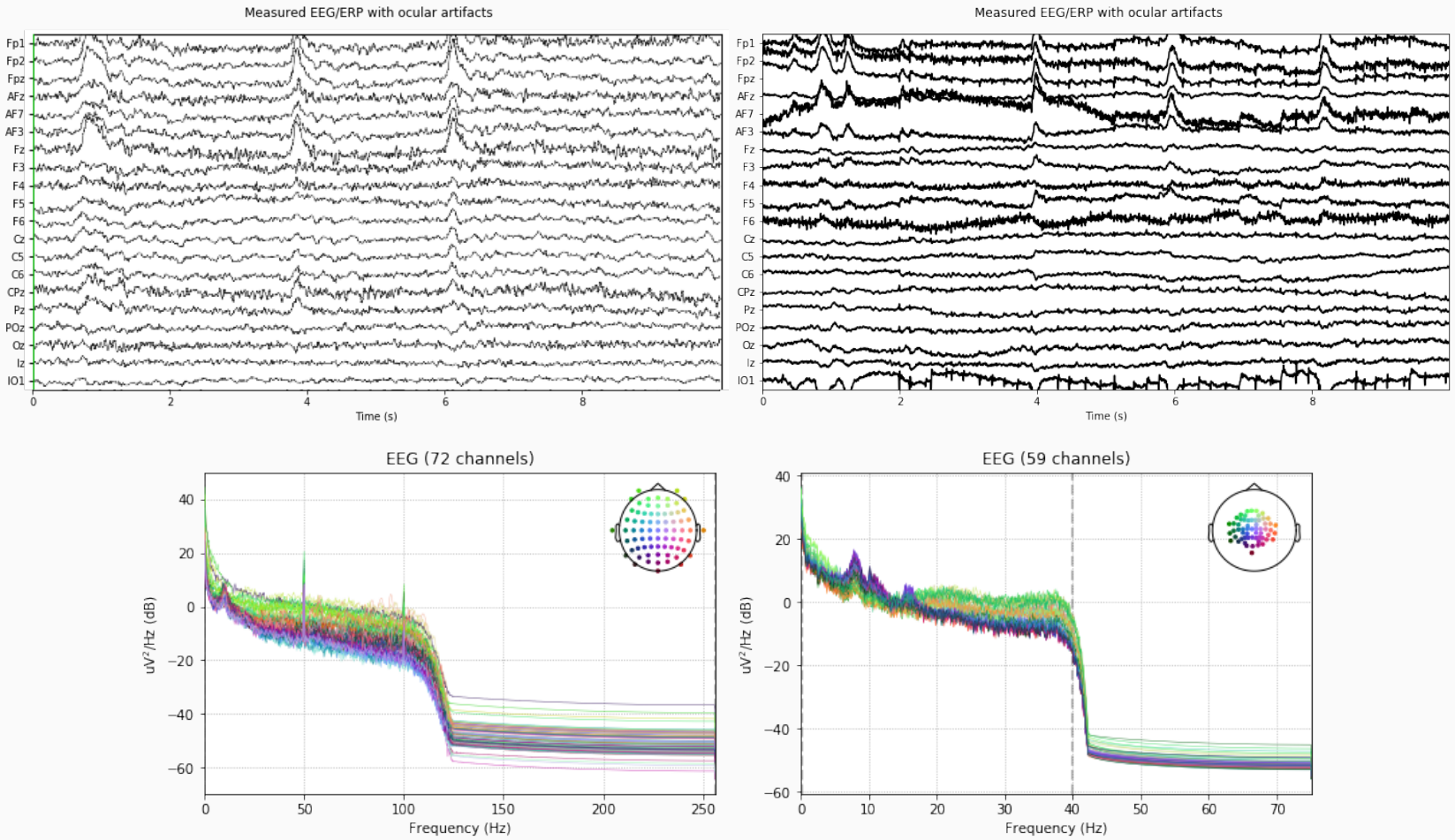


EEG, EOG (blinks) and Gaussian noise only

Real Data

Measured signal

Two datasets obtained from Henderson et al. (2015), Gramfort et al. (2014).
Need to extract only the neural activities.

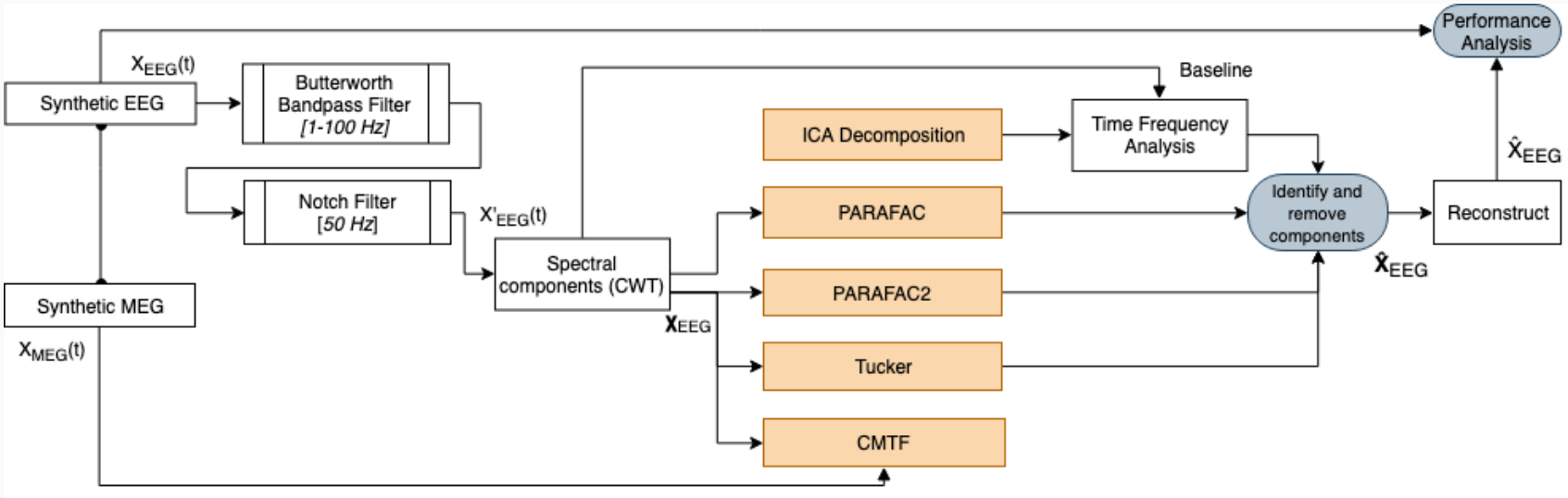


PSD

Artifact removal using Tensor Decomposition

Baseline ICA

ICA solves BSS explicitly. Empirical mode decomposition (as applied by Mandic et al. (2013)) considers each channel separately.



Artifact removal using decomposition methods

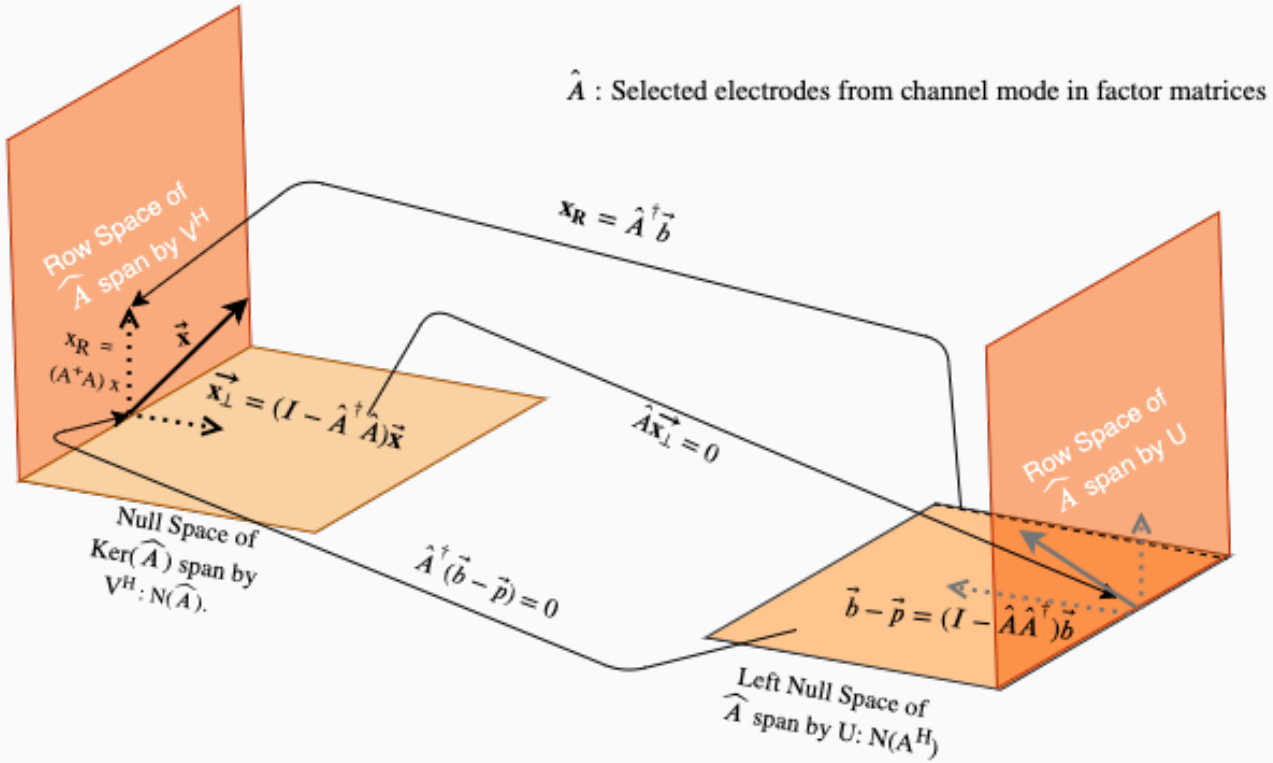
Known ground truth

- Correlation coefficient and relative mean squared error

Unknown ground truth

- Correlation between the reconstructed/clean signal with the original signal
- Power percentage change: channel view
- Entropy measures (turned out to be less useful)

Artifact Removal using Tensors

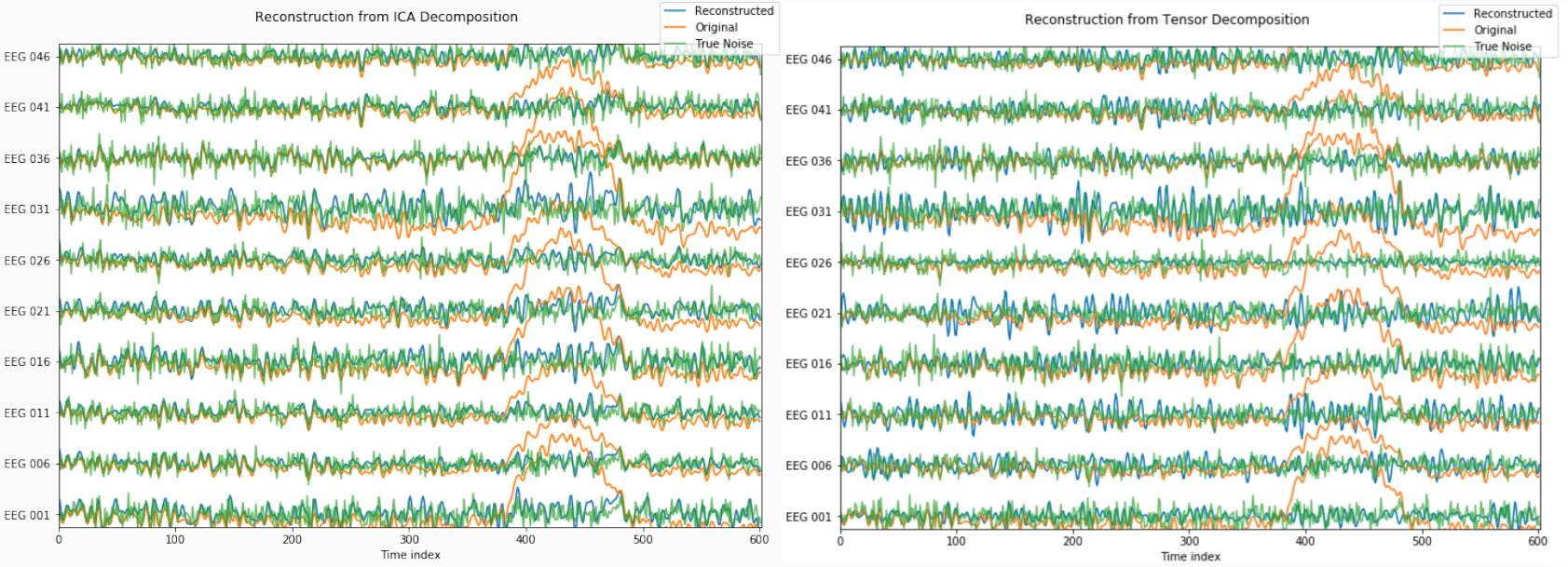


Projection onto the the null space of the artifact components, and reconstruction using $x_\perp$

$$\implies \boldsymbol{\chi}_{clean} = \boldsymbol{\kappa} \times_1 U^{(1)}(I - \hat{A}\hat{A}^\dagger) \times_2 U^{(2)} \ldots \times_N U^{(N)} \tag{6}$$

The isolation of the components corresponding to artifacts is conducted through qualitative examination in this study.

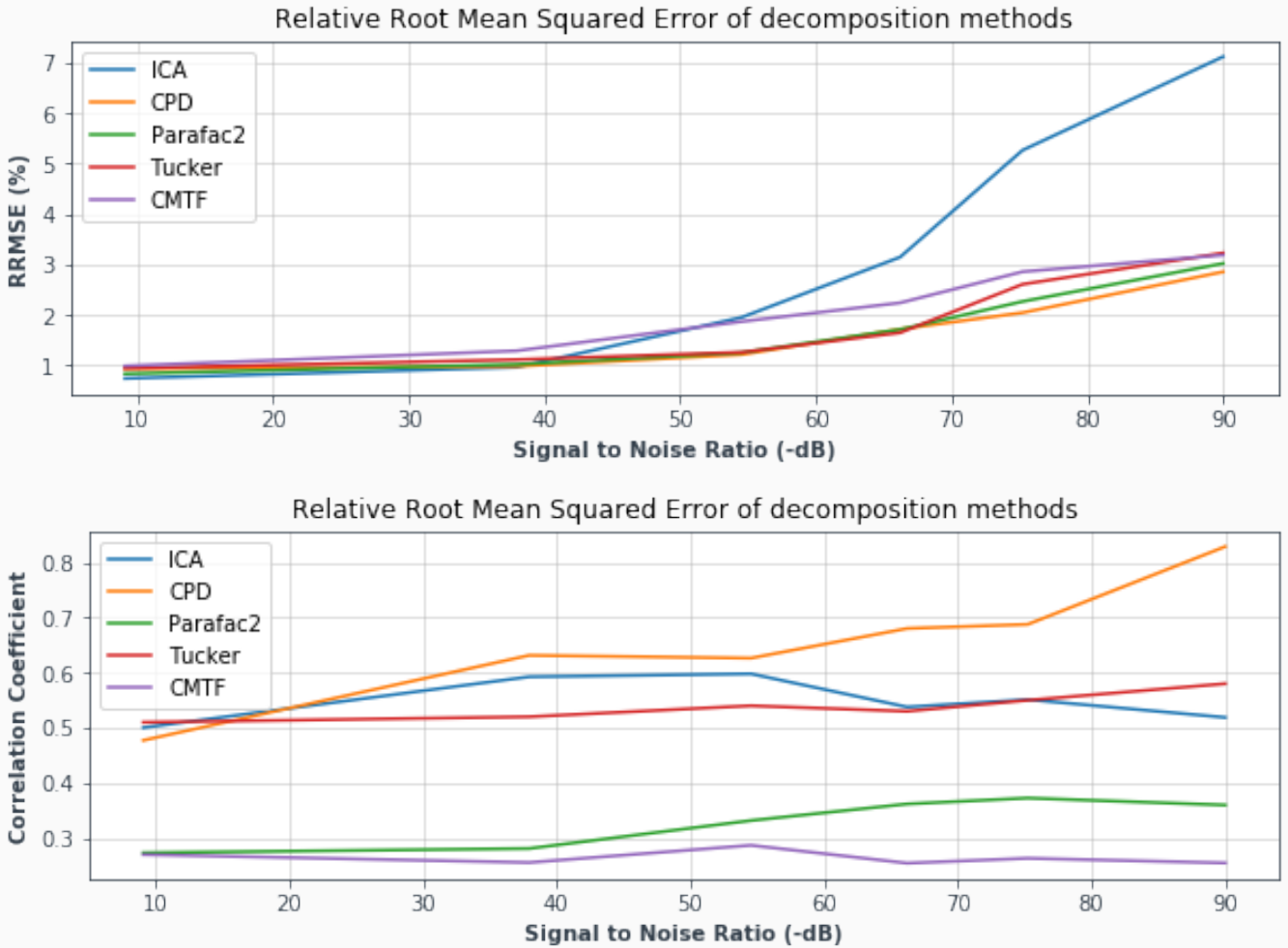
Examples



Reconstruction using decomposition based BSS

`Demonstrations on Jupyter Notebooks` available

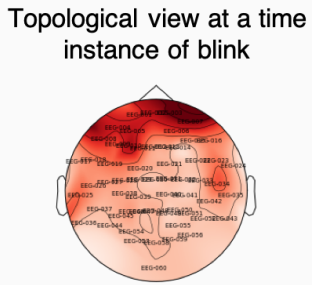
Results: Synthetic

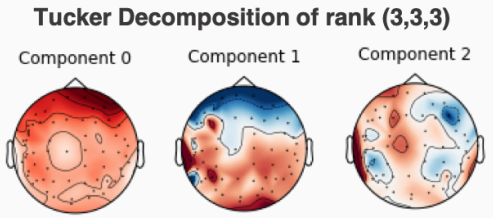


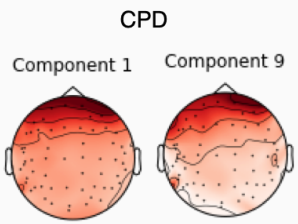
[Top] Relative Root Mean Squared Error (RRMSE) as on varying signal to noise ratios. [Bottom] Averaged Pearson's correlation coefficient on all electrodes over varying signal to noise ratios.

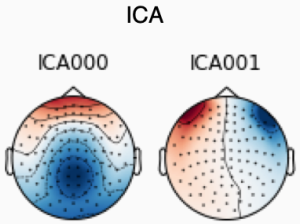
Results: Real

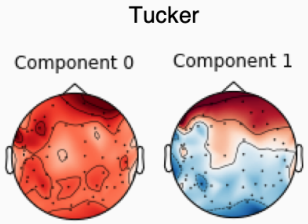
- Observation 1: Parafac and Parafac2 are better than Tucker - correlation
- Observation 2: Not in low rank cases

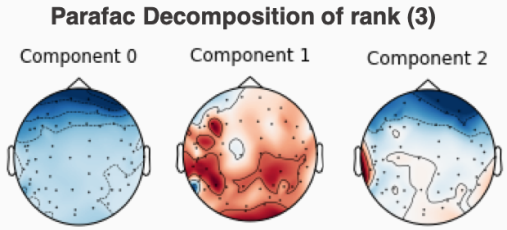


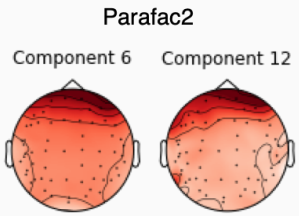










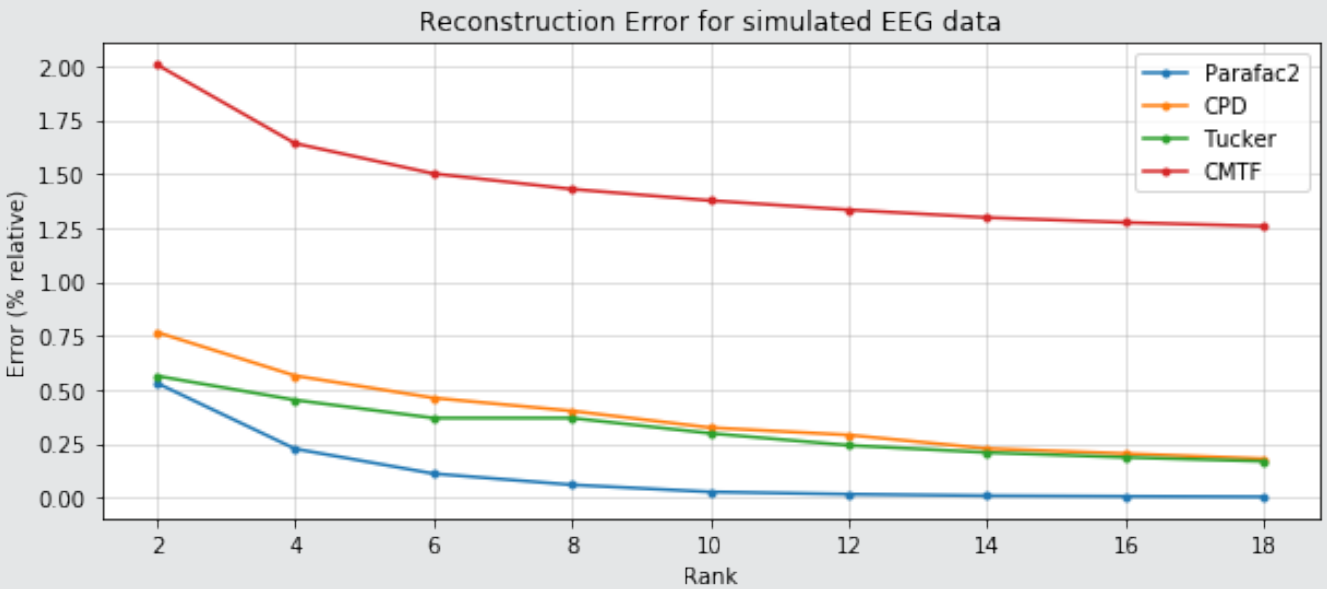


Improvements using Data fusion and optimisations

- Coupled analysis of heterogeneous data
- Simultaneous MEG, EEG factorisation

Results

CMTF considerably worse: directly losing 2% of neural information



Example for the MNE dataset

CMTF did not perform well on rrmse, correlation coefficient on synthetic data

Optimisations

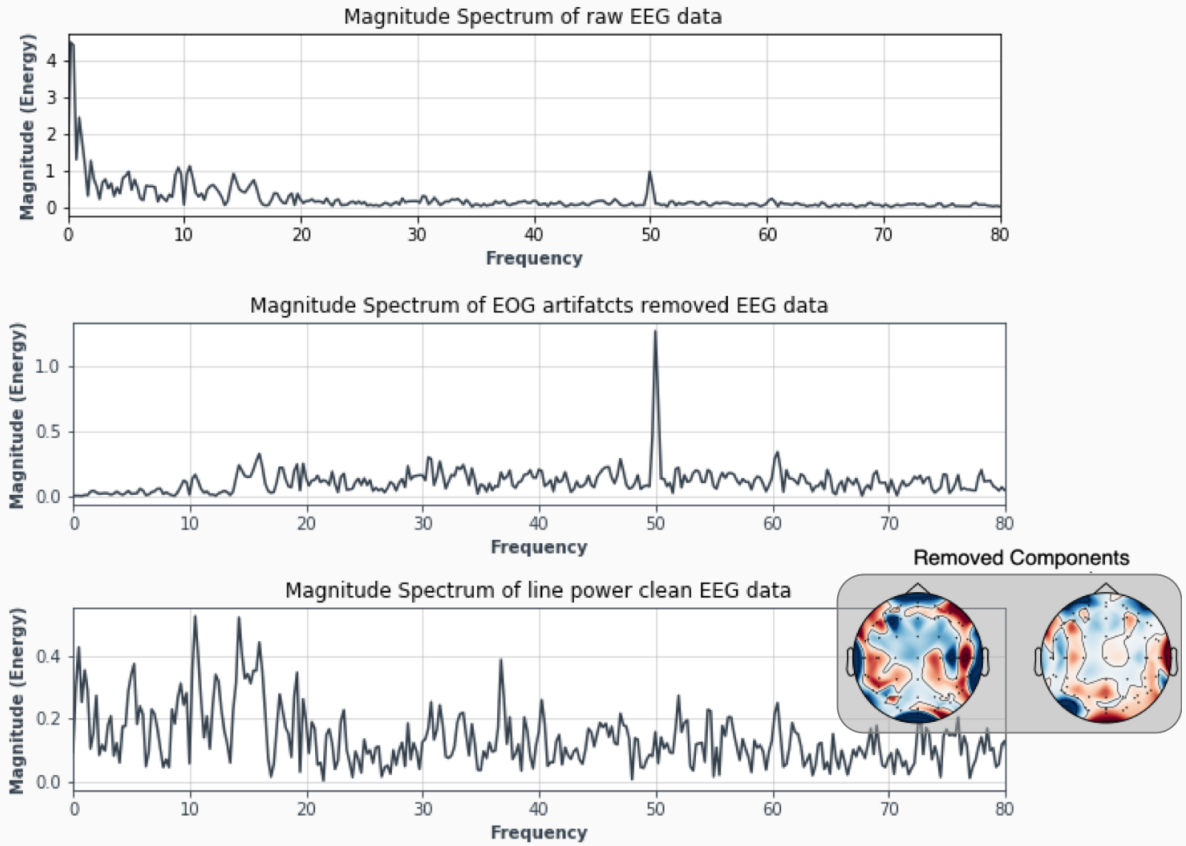
- Randomised CPD implementation (<2% time improvements)
- Core Consistency Diagnostics implementation (optimal rank ≈ 30)
- Performance optimisations (e.g. Randomised CPD)

Conclusion

A unified method

Why Tensors

Better performance using simultaneous time frequency analysis
But also insights into other, not so visually distinctive features. Consider line noise power.



The power (measured by RMS) was reduced from 103.5 to 7.93 in relative units.

Contributions

EEG with Tensor Decomposition

Theoretical justification of advantages, leading to PARAFAC2
Hypothesis Development

Artifact extraction

Describing a method, from a geometrical perspective, of removing artifacts in Parafac and Tucker family

Implementations to a tensor toolbox

RandomisedCPD, PARAFAC2, CMTF, CORCONDIA (CPD, TUCKER)
No other implementation of Parafac2, CORCONDIA available

Comparison

Developing quantification methods
Comparison: qualitative and quantitative

Questions?

`Tensors for EEG`: https://github.com/divyanshmanocha/EEGTensors

`HottBox`: https://github.com/hottbox/hottbox

Further work

- More efficient algorithms for large datasets
- In-depth analysis of higher order dimensions: more subjects and trials
- A probabilistic approach using Bayesian statistics to incorporate a priori information
- Bayesian Information Criteria (BIC) with Core Consistency Analysis
- Further analysis of time drift. Parafac on time drift corrected EEG better than Parafac2?

Wavelets or STFT

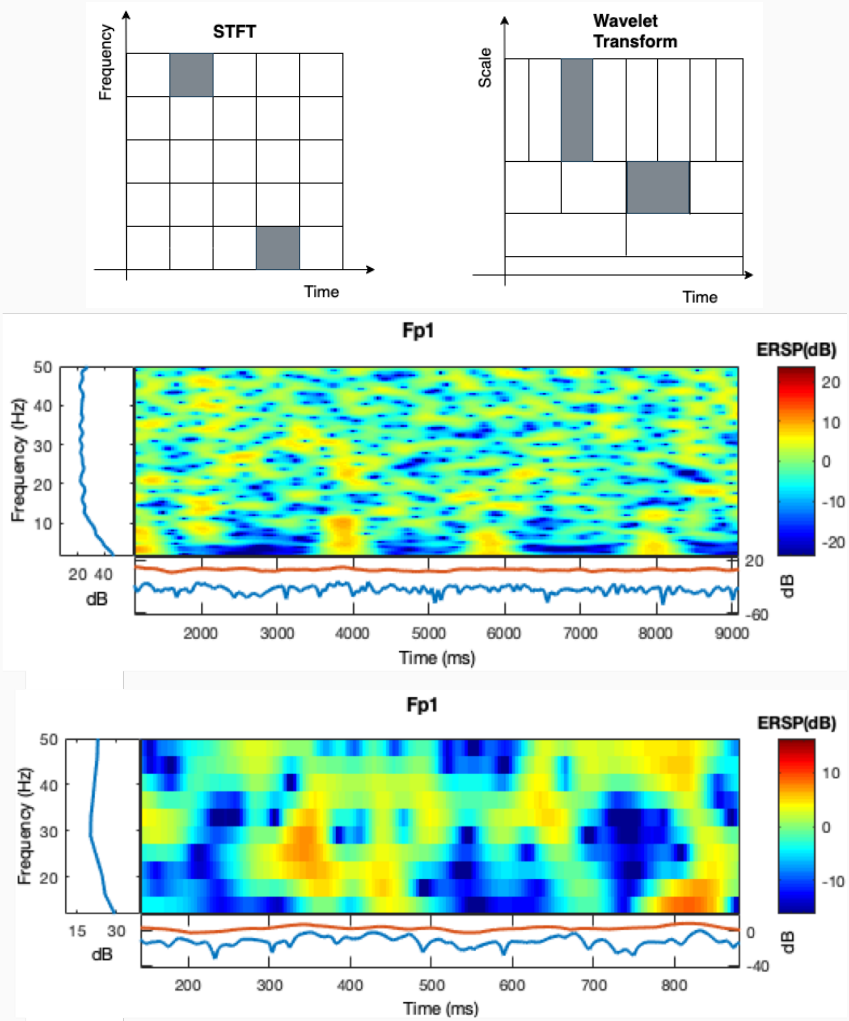


Figure 1: (a) Time and spectral resolution in STFT are fixed, whilst they can be altered in Wavelet Transforms. As can be seen for Wavelet Transforms, increasing the resolution of one decreases the resolution of the other. (b) Spectogram plot of FP1 channel from EEG with at two different time resolutions, demonstrating a clear difference in frequency resolution. The real (reading) dataset used in this project has been depicted with a morelet wavelet.

CORCONDIA implementation for rank selection

Randomised (Truncated) SVD for Parafac2 and Tucker. Core tensor $\boldsymbol{G}$ must be a superdiagonal array of ones (the identity I matrix in the two dimensional case). CORCONDIA simply exploits this by measuring the similarity between the superdiagonal core $\boldsymbol{T}$and least-squares fitted$\boldsymbol{G}$. In other words it is measuring the 'superdiagonality'.

$$\text{CORCONDIA} = 100 \left(1 - \frac{\sum_{i=1}^{R} \sum_{j=1}^{R} \sum_{k=1}^{R} (g_{ijk} - t_{ijk})^2)}{R} \right) \tag{7}$$

The results from the core consistency analysis are shown in Figure **??**. They suggest that, for the size of the dataset used, $\approx 24 - 32$ components for the Synthetic dataset, $27 - 32$ for the MNE dataset and $28 - 32$ for the real dataset are most appropriate.

Optimisations

Randomised CPD

Algorithm 7: CPRAND using ALS

Data: Tensor $\boldsymbol{\chi} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, Rank R, Samples S
Result: $A^{(1)} \in \mathbb{R}^{I_1 \times R}, A^{(2)} \in \mathbb{R}^{I_2 \times R}, \ldots, A^{(N)} \in \mathbb{R}^{N \times R}, \boldsymbol{\lambda} \in \mathbb{R}^{1 \times R}$

Initialise factor matrices $\boldsymbol{A^{(n)}}$ **repeat**
  for $n \leftarrow k$ ***to*** N **do**
    $\boldsymbol{S} \leftarrow SamplingOperator \in \mathbb{R}^{S \times \prod_{m \neq n} I_m}$
    $\boldsymbol{V_S} \leftarrow SampledKhatriRao(\boldsymbol{S}, \boldsymbol{A}^{(1)T}, \ldots, \boldsymbol{A}^{(n-1)}, \boldsymbol{A}^{(n+1)}, \ldots, \boldsymbol{A}^{(N)}$
    $\boldsymbol{X_S^{(T)}} \leftarrow \boldsymbol{S X_{(n)}^T}$
    $\boldsymbol{A}^{(n)} \leftarrow argmin_A \| \boldsymbol{V_S} bm A^T - \boldsymbol{X_S^T} \|_F$
    $\boldsymbol{\lambda} \leftarrow$ Norm of $\boldsymbol{A}^{(n)}$ columns
    Normalise $\boldsymbol{A}^{(n)}$ columns
  end
until *convergence criteria met*;
return $\boldsymbol{\lambda}, \boldsymbol{A}^{(N)}, \boldsymbol{A}^{(N-1)}, \ldots, \boldsymbol{A}^{(1)}$

References

A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen. Mne software for processing meg and eeg data. *NeuroImage*, 86:446 – 460, 2014. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2013.10.027. URL http://www.sciencedirect.com/science/article/pii/S1053811913010501.

J. Henderson, S. Luke, J. Schmidt, and J. Richards. Dataset 3: Natural reading, 2015. URL http://www2.hu-berlin.de/eyetracking-eeg/testdata.html#exampledata3.

N. Ille, P. Berg, and M. Scherg. Artifact correction of the ongoing eeg using spatial filters based on artifact and brain signal topographies. *Clin. Neurophysiol.*, 19:113–124, 05 2002. doi: 10.1097/00004691-200203000-00002.

D. P. Mandic, N. u. Rehman, Z. Wu, and N. E. Huang. Empirical mode decomposition-based time-frequency analysis of multivariate signals: The power of adaptive data analysis. *IEEE Signal Processing Magazine*, 30(6):74–86, Nov 2013. ISSN 1053-5888. doi: 10.1109/MSP.2013.2267931.