Dictionary learning for M/EEG multidimensional data
Christos Papageorgakis, Sebastian Hitziger, Théodore Papadopoulo

To cite this version:
Christos Papageorgakis, Sebastian Hitziger, Théodore Papadopoulo. Dictionary learning for M/EEG multidimensional data. International conference on basic and clinical multimodal imaging (BACI), Sep 2015, Utrecht, Netherlands. hal-01243284

HAL Id: hal-01243284
https://hal.archives-ouvertes.fr/hal-01243284
Submitted on 14 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Dictionary learning for M/EEG multidimensional data

Christos Papageorgakis1,2, Sebastian Hitziger1, Théodore Papadopoulo1
1Inria Sophia Antipolis - Méditerranée, France
2Université Côte d’Azur, France

1. Jitter-adaptive dictionary learning model (JADL)

JADL is a dictionary learning framework.

1. Atom Selection: The best shifted versions of the atoms contained in the extended dictionary $D^S$ are selected, over all the channels, leading to a compressed dictionary $D_2$.

2. Dictionary update: (i) Sparse coding finding the coefficients $a^s_{ij}$ and the jitters $\Delta^s_{ij}$

- \[ s^0_{ij} = \arg\min_{s_{ij}} \|D_2 s_{ij} - x_{ij}\|_2 \] for each atom $s_{ij}$.
- Let an "unrolled" version of the dictionary $D$ be a dictionary $D^0$ containing all allowed shifts $(5 - |\Delta|)$ of all its atoms.
- The sparse coding problem is solved using a modification of least angle regression (LARS)[4] by reweighting the problem as follows:

- Block coordinate descent is used to iteratively solve the constrained minimization problem for each atom:

- \[ x_{ij} = \Delta^s_{ij} \in \mathbb{R}^{K \times D^S} \]

- \[ D^S = \{ d_{ij} \} : d_{ij} \in D, \Delta \in \Delta \]

- \[ D^0 = \{ d_{ij} \} : d_{ij} \in D, \Delta \notin \Delta \]

- \[ s^1_{ij} = \arg\min_{s_{ij}} \frac{1}{2} \|D_2 s_{ij} - x_{ij}\|_2^2 + \lambda \|s_{ij}\|_1 \]

- Atoms present in a signal can suffer from unknown time delays (jitter).

- Introducing new random jitters to the dictionary of $D$ is the number of channels of the EEG data, $x_i$ is the signal of channel $i$, and $d_{ij}$ is the $j$-th atom of the extended dictionary $D^S$.

- The results of our multi-channel algorithm show:

- A very good fit of the learned dictionary to the generated one.
- A good fit also in the case where the signals were contaminated by noise.

- The multi-dimensional approach is tested using real MEG and EEG data:

- Both algorithms are executed with the same signals, initial random dictionary and latency parameters.

- The multi-channel algorithm is executed using all the channels from the input data, while the single-channel algorithm is executed several times, each time using a different channel.

- The results of our multi-channel algorithm show:

- A very good fit of the learned dictionary to the generated one.
- A good fit also in the case where the signals were contaminated by noise.

- The multi-channel model is extended to the multi-channel approach based on the coefficients vectors obtained by the goodness of fit metric: 0.999, 0.999 and 0.999 instead of 0.992, 0.977 and 0.964 for the single-channel approach using the best channel and 0.939, 0.512, 0.512 using the worst channel.

- Goodness of fit metric: \[ \text{score} = mape|s_{ij}| \]

- where $a_i$ is a generated atom, $a_j$ is a learned atom and $\Delta$ is a shifted version of the learned atom, with shifts within the expected range $\Delta \in \delta$ and $i \in [0, k]$.

3. Synthetic data generation

- Generate an extended dictionary of $D^S$ signals:

- Introducing random jitters (from the set $\Delta = \{ -103 \text{ contiguous allowed shifts} \}$) to the dictionary's atoms.

- Select 3 source groups, each of them containing 3 neighboring sources.

- Each source group is associated to shifted versions of the same atom.

- Combine the generated signals with a lead field matrix $C$ computed from real EEG measurements [3].

- where $D \in \mathbb{R}^{C \times N}$ is the lead field matrix, $C \in \mathbb{R}^{C \times C}$ is the sources matrix, $C$, $Q$ and $N$ are the numbers of channels, sources and time samples respectively.

- Perform a two previous algorithms for M trials.

- Introducing new random jitters to the dictionary of $K = 3$ synthetic atoms.

- Generated clean $M$ EEG measurements of $C = 6$ channels, $M = 200$ trials and $N = 515$ time samples.

4. Results on lead field synthetic data

A comparison between the original and our multi-dimensional JADL model

- Both algorithms are executed with the same signals, initial random dictionary and latency parameters.

- The multi-channel algorithm is executed using all the channels from the input data, while the single-channel algorithm is executed several times, each time using a different channel.

- The results of our multi-channel algorithm show:

- A very good fit of the learned dictionary to the generated one.
- A good fit also in the case where the signals were contaminated by noise.

- The multi-channel model is extended to the multi-channel approach based on the coefficients vectors obtained by the goodness of fit metric: 0.999, 0.999 and 0.999 instead of 0.992, 0.977 and 0.964 for the single-channel approach using the best channel and 0.939, 0.512, 0.512 using the worst channel.

- Goodness of fit metric: \[ \text{score} = mape|a_i| \]

- where $a_i$ is a generated atom, $a_j$ is a learned atom and $\Delta$ is a shifted version of the learned atom, with shifts within the expected range $\Delta \in \delta$ and $i \in [0, k]$.

5. Results on real data

The multi-dimensional approach is tested using real MEG and EEG data:

- $C = 200$ channels.
- $M = 63$ trials.
- $N = 541$ time samples.
- Contaminated by ambient noise.

Input parameters:

- $S = 103$ contiguous allowed shifts.
- $K = 3$ atoms.

6. Conclusions

- The method shows superior performance and less noisy estimated waveforms compared to the original single-channel JADL framework, both on synthetic and real data.

- It is more robust to various levels of noise.

- Using the JADL framework allows one to deal with signal variations such as jitters which is difficult to do with standard methods such as PCA or ICA.

- Not having to select a “best” channel (as with the JADL method) is both a user simplification and allows the exploitation of all the available information for M/EEG trial by trial signal decomposition. This thus provides better estimations of waveforms in the dictionary.