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
Introduction

Signals obtained from magneto- or electroencephalography (M/EEG) are very noisy and inherently multi-dimensional, i.e. provide a vector of measurements at each single time instant. To cope with noise, researchers traditionally acquire measurements over multiple repetitions (trials) and average them to classify various patterns of activity. This is not optimal because of trial-to-trial variability (waveform variation, jitters). The jitter-adaptive dictionary learning method (JADL [1]) has been developed to better handle this variability (with a particular emphasis on jitters). JADL is a data-driven method that learns a dictionary (prototype pieces) from a set of signals, but is currently limited to a single channel, which restricts its capacity to work with very noisy data such as M/EEG. We propose an extension to the jitter-adaptive dictionary learning method, that is able to handle multi-dimensional measurements such as M/EEG.

1. Jitter-adaptive dictionary learning model (JADL)

JADL, is a dictionary learning framework. A main stay of machine learning and phase of atoms of d.
- Atoms learned by JADL are defined on the entire signal domain.
- Hypotheses:
  - The set of signals of interest \( x_i \) can be generated by a dictionary.
  - Atoms present in a signal can suffer from unknown time delays (jitter).
- The solution of the sparse coding problem is solved using a modification of least angle regression (LARS) [4] by rewriting the problem as follows:

\[
\sum_{i=1}^{K} \left( x_i - \sum_{j=1}^{r} a_{ij} d_j + r \right)^2 + \lambda \sum_{i=1}^{K} \left| a_{ij} \right| \]

\( a_{ij} \) are the coefficients of the dictionary atoms \( d_j \) with \( i \) the variable index and \( j \) the atom index.

The algorithm solving the JADL problem, is based on an implementation in [2] for common dictionary learning, which iteratively alternates between:

(i) Sparse coding: finding the coefficients \( a_{ij} \) and the jitter \( \delta_{ij} \).

Let an "unrolled" version of the dictionary \( D \) be a dictionary \( D^D \) containing all allowed shifts \((S = \{\Delta\})\) of all its atoms:

\[
D^D = \{ (i,j) | d \in D \in \Delta \}
\]

A very good fit of the learned dictionary to the generated one.

\[
D_i = \{ (i,j) | d \in D \} \text{ dictionary}
\]

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

Atoms present in a signal can suffer from unknown time delays (jitters). JADL is a data-driven method that learns a dictionary (prototype pieces) from a set of signals, but is currently limited to a single channel, which restricts its capacity to work with very noisy data such as M/EEG. We propose an extension to the jitter-adaptive dictionary learning method, that is able to handle multi-dimensional measurements such as M/EEG.

2. Our modified JADL model

We propose an extension to the jitter-adaptive dictionary learning method, that:
- Is able to handle multidimensional measurements such as M/EEG.
- A dictionary over M/EEG recordings that have the same waveform and jitter over all the channels in a single trial.
- Is still able to account for different jitters across trials.

Significant modifications are applied to the original JADL framework, especially in:

- The solution of sparse coding problem (i) by least angle regression (LARS) [4]:
  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 \( D_C \):

\[
D^S = \{ (i,j) | d \in D \} = \{ (i,j) | d \in D \}
\]

where \( C \) is the number of channels of the EEG data, \( s_i \) is the signal of channel \( s_i \) and \( d_j \) is the \( j \)-th atom of the extended dictionary \( D^S \).

2. Standard LARS sparse coding over the channels for the current atom set. During this step the multi-dimensional coefficients \( a_{ij} \) are computed using the compressed dictionary \( D_C \) selected by the previous step and the multi-channel signals for the given trial.

The dictionary update problem (ii) is also slightly modified to treat the measurements corresponding to the different channels as additional trials.

3. Synthetic data generation

- Create a dictionary of \( K = 3 \) synthetic atoms.
- Generate an extended dictionary of 9 signals:
  - Introducing random jitters (from the set \( \Delta \) of \( S = 103 \) 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 \( G \) computed from real EEG measurements [5].

\[
G = D^S \cdot F \cdot C \cdot Q \cdot R
\]

where \( G \in \mathbb{R}^{N \times C} \) is the measurement matrix either MEG or EEG, \( C \in \mathbb{R}^{C \times 0} \) is the lead field matrix, \( F \in \mathbb{R}^{F \times F} \) is the sources matrix, \( N, C, Q, R \) are the numbers of channels, sources and time samples respectively.
- Perform the above procedure 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.

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.

The comparison of our multi-channel approach with the single-channel algorithm showed:
- Similar results when the best channel is used by the single-channel algorithm.
- Worse results for the single-channel algorithm when a medium or worse channel is used.
- A good fit when the single-channel algorithm is unable to recover correctly all the atoms of the dictionary used to generate the signals.

A small but superior performance for the multi-channel approach based on the coefficients vectors obtained by the goodness of fit metric: 0.999, 0.996 and 0.995 instead of 0.992, 0.977 and 0.964 for the single-channel approach using the best channel and 0.936, 0.512, 0.512 using the worst channel.

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 variabilities 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.

References

[4] Christos Papageorgakis, BaCI

http://team.inria.fr/athena/ christos.papageorgakis@inria.fr