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>
1. Jitter-adaptive dictionary learning model (JADL)

JADL is a dictionary learning framework.

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 multidimensional measurements such as M/EEG.

The algorithm solving the JADL problem, is based on an implementation in [2]

\[ \mathbf{D} = \arg\min_{\mathbf{D}} \sum_{i=1}^{K} \| \mathbf{x}_i - \mathbf{D} \mathbf{s}_i \|_2^2 \quad \text{s.t.} \quad \| \mathbf{D} \|_{\infty,2} \leq 1 \]

The dictionary learning for M/EEG multidimensional data

\[ \mathbf{D} = \{ \mathbf{D}_j \}_{j=1}^{K} \]

\[ \mathbf{D}_j = \{ \mathbf{d}_{ij} \}_{i=1}^{\text{dim}(\mathbf{s}_j)} \]

\[ \mathbf{s}_j = \mathbf{D}_j \mathbf{a}_j \]

\[ \mathbf{a}_j = \arg\min_{\mathbf{a}_j} \| \mathbf{x}_j - \mathbf{D}_j \mathbf{a}_j \|_2^2 \quad \text{s.t.} \quad \| \mathbf{D}_j \|_{\infty,2} \leq 1 \]

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 characteristics 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 decomposition. It thus allows for the exploitation of all the available information for M/EEG trial by trial decomposition.

5. Results on real data

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

\[ C = 200 \text{ channels} \]

\[ M = 63 \text{ trials} \]

\[ N = 541 \text{ time samples} \]

\[ \text{contaminated by ambient noise} \]

Input parameters:

\[ S = 103 \text{ contiguous allowed shifts} \]

\[ K = 3 \text{ atoms} \]

Figure 3: The single-channel (left) and the multi-channel method (right).

6. Conclusions

The synthetic data generation

1. Atom Selection: The best shifted versions of the atoms contained in the extended dictionary \( D^\ast \) are selected, over all the channels, leading to a compressed dictionary \( D_c \):

\[ d_{ij} - \arg\min_{d_{ij}} \| \mathbf{x}_i - \mathbf{D} \mathbf{a}_j \|_2^2 \]

where \( a_j \) is the signal of channel \( j \) and \( d_{ij} \) is the \( j \)-th atom of the extended dictionary \( D^\ast \).

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

The multi-channel approach based on the coefficients vectors obtained by the goodness of fit metric: 0.995, 0.998 and 0.999 instead of 0.990, 0.977 and 0.994 for the single-channel approach using the best channel and 0.936, 0.512, 0.512 using the worst channel.

A small but superior performance for the multi-channel approach when a medium or the worst channel is used.

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 decomposition. It thus allows for the exploitation of all the available information for M/EEG trial by trial decomposition.

Goodness of fit metric:

\[ \text{goodness of fit} = \sum_{i,j} | S_{ij} | \]

where \( a_j \) is a generated atom, \( a_j \) is a learned atom and \( a_{ij} \) is a shifted version of the learned atom, with shifts within the expected range \( r \in \Delta \) and \( i,j \in [0, K] \).

Figure 1: The generated (blue color) and learned (red color) dictionary using our model (left). The learned dictionary on contaminated signals by noise of SNR: 0.021 (right).

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 the worst channel is used.
- Using 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: 0.995, 0.998 and 0.999 instead of 0.990, 0.977 and 0.994 for the single-channel approach using the best channel and 0.936, 0.512, 0.512 using the worst channel.

With a generated atom, \( a_j \) is a learned atom and \( a_{ij} \) is a shifted version of the learned atom, with shifts within the expected range \( r \in \Delta \) and \( i,j \in [0, K] \).

Figure 2: The learned dictionaries by the single-channel method: using the best (left) and the worst (right) channel. Wrong recovered components are marked by the red ellipses.