Dictionary learning for M/EEG multidimensional data

Christos Papageorgakis, Sebastian Hitziger, Théodore Papadopoulo

Dictionary learning for M/EEG multidimensional data

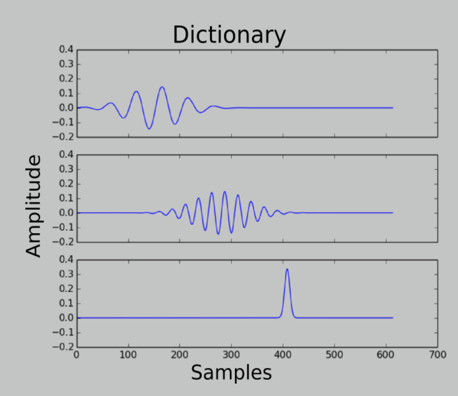
Christos Papageorgakis[1,2], Sebastian Hitziger[1], Théodore Papadopoulo[1]

[1]Inria Sophia Antipolis - Méditerranée, France
[2]Université Côte d'Azur, France

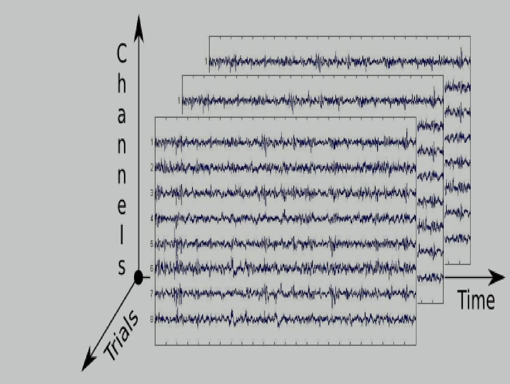
Sept. 1-5, 2015

BaCI

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 for 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 multidimensional measurements such as M/EEG.



1. Jitter-adaptive dictionary learning model (JADL)

JADL is a **dictionary learning framework**:

- Compensate for small **variations** in **latency** and phase of atoms d_i.
- Atoms learned by JADL are defined on the entire signal domain.

Hypotheses:

- The set of signals of interest $\{x_j\}_{j=1}^M$ can be generated by a dictionary.
- Atoms present in a signal can suffer from unknown time delays (jitter).

$$x_j = \sum_{i=1}^{K} a_{ij}\delta_{ij}(d_i) + \epsilon$$

K : #atoms, M : #trials
$a_{ij} \in \mathbb{R}$: coefficient
$\delta_{ij} \in \Delta$: shift operator
Δ : finite set of allowed shifts
d_i : atom
$D = \{d_i\}_{i=1}^K$: dictionary
ϵ : Gaussian noise

$$\min_{d_i, a_{ij}, \delta_{ij}} \sum_{j=1}^{M} \left(\frac{1}{2} \left\| x_j - \sum_{i=1}^{K} a_{ij}\delta_{ij}(d_i) \right\|_2^2 + \lambda \|a_j\|_1 \right), \quad s.t. \|d_i\|_2 = 1.$$

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 jitters $\{\delta_{ij}\}$

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

$$D^S = \{\delta(d) : d \in D, \delta \in \Delta\},$$
a matrix of dimension $N \times KS$.

The sparse coding problem is solved using a modification of least angle regression (LARS)[4] by rewriting the problem as follows:

$$a_j^S \leftarrow \operatorname{argmin} \frac{1}{2}\|x_j - D^S a_j^S\|_2^2 + \lambda \|a_j^S\|_1,$$
$$s.t. \ \|a_j^{S,i}\|_0 \leq 1, \ \ i = 1, \ldots, K.$$

Once an atom is chosen all its shifts are forbidden.

(ii) **Dictionary update**: finding the shapes $\{d_i\}$.

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

$$d_k = \operatorname*{argmin}_{d_k} \sum_{j=1}^{M} \frac{1}{2} \left\| x_j - \sum_{i=1}^{K} a_{ij}\delta_{ij}(d_i) \right\|_2^2$$
$$s.t. \|d_k\|_2 = 1$$

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.
- Learns 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 algorithm (LARS):

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_S.

$$d_j^S = \operatorname*{argmax}_{d_j^S \in D^S} \sum_{c=1}^{C} \left\| s_c \cdot d_j^S \right\|,$$

where C is the number of channels of the EEG data, s_c is the signal of channel c and d_j^S 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_{ijc} are computed using the compressed dictionary D_S 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

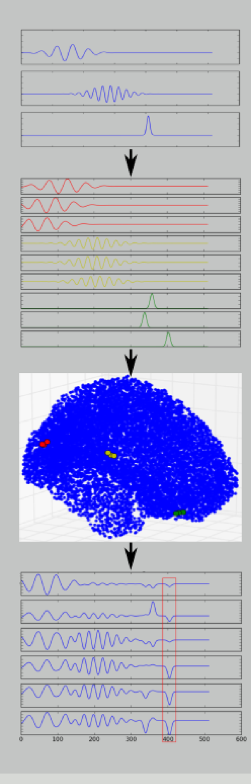
- Create a dictionary of $K = 3$ synthetic atoms.
- Generate an extended dictionary of 9 signals:
  - Introducing random jitters (from the set Δ of $S = 103$ contiguous allowed shifts) to the dictionary's atoms .
- Select 3 source groups, each of them containing 3 neighboring sources.
  - Each source 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 [3].

$$\mathscr{M} = G\mathscr{S},$$

where $\mathscr{M} \in \mathbb{R}^{C \times N}$ is the measurement matrix either MEG or EEG, $G \in \mathbb{R}^{C \times Q}$ is the lead field matrix, $\mathscr{S} \in \mathbb{R}^{Q \times N}$ is the sources matrix. C, Q and N 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.



References

[1] Hitziger, S., Clerc, M., Gramfort, A., Saillet, S., Bénar, C. and Papadopoulo, T. Jitter-adaptive dictionary learning-application to multi-trial neuroelectric signals. *arXiv preprint:1301.3611*, (2013).

[2] Mairal, J., Bach, F., Ponce, J. & G. Sapiro. Online learning for matrix factorization and sparse coding. *The Journal of Machine Learning Research*, 11, 19-60 (2010).

[3] Henson, R., Wakeman, D., Phillips & C., Tilman, L. Effective Connectivity between OFA and FFA during face perception: DCM of evoked MEG, EEG and fMRI. *Hum Brain Mapp*, (2013).

[4] Efron, B., Hastie, T, Johnstone, I. & Tibshirani, R. Least angle regression *The Annals of Statistics*, 32 (2):407-499, (2004).

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.

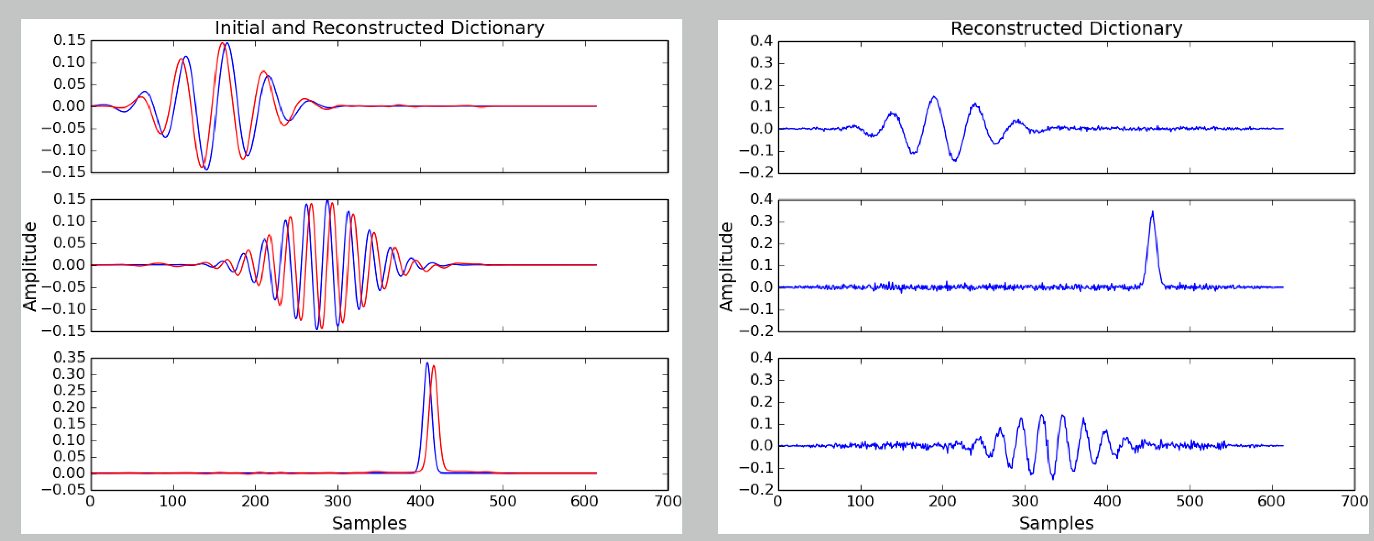


Figure 1 : The generated (blue color) and learned (red color) dictonary using our model (left). The learned dictionary on contaminated signals by noise of SNR : 0.021 (right).

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 worse channel is used.
- Cases where 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, 0.996 and 0.995 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: $cor_i = \max_{j,\tau} |a_i a_{j_\tau}|$,

where a_i is a generated atom, a_j is a learned atom and a_{j_τ} is a shifted version of the learned atom, with shifts within the expected range $\tau \in \Delta$ and $i, j \in [0, k]$.

SNR	SNR_{dB}	Atom1	Atom2	Atom3
0.804	-0.944	0.998	0.999	0.997
0.021	-16.700	0.993	0.973	0.983
0.001	-29.240	0.954	0.821	0.892
0.0002	-36.107	0.826	0.585	0.462

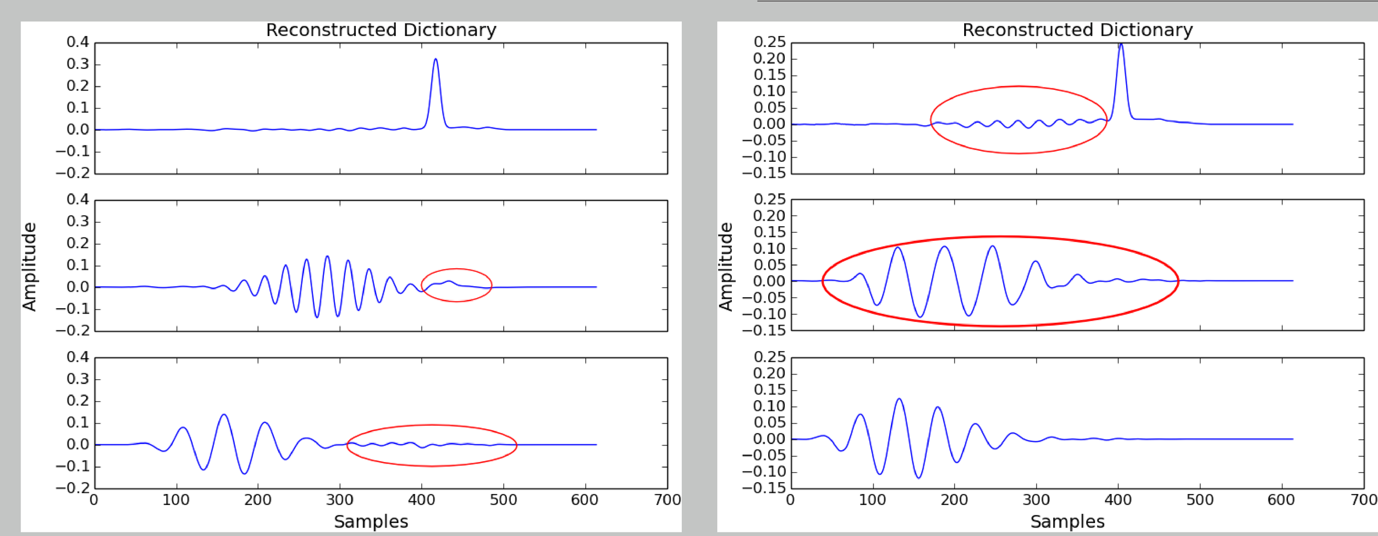


Figure 2 : The learned dictionaries by the single-channel method: using the best (left) and the worst (right) channel. Wrong recovered components are marked by the red ellipses.

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.

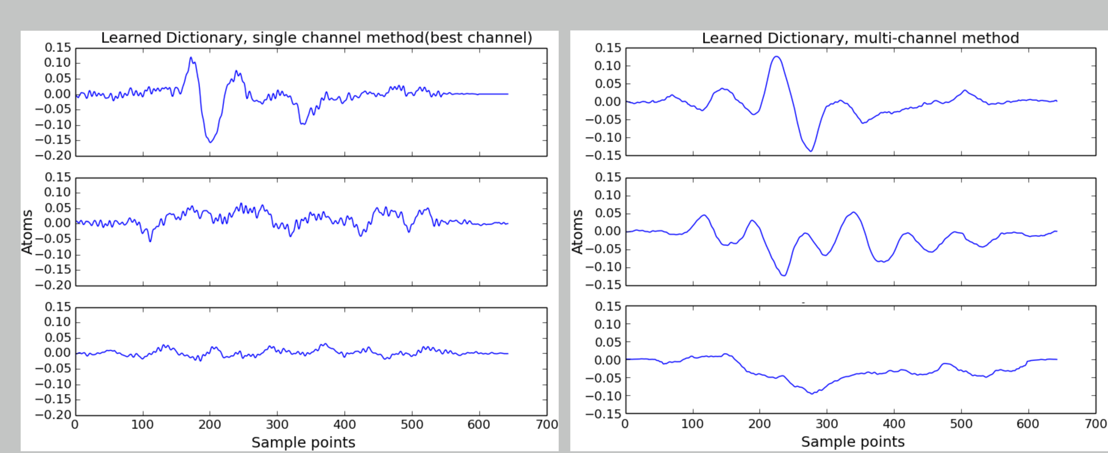


Figure 3 : The single-channel (left) and the multi-channel method (right).

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. It thus provides better estimations of waveforms in the dictionary.

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