Report: Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals

Gauthier Multari gauthier.multari@ens-paris-saclay.fr
Yona Uzan yona.uz@gmail.com

January 22, 2024

1 Introduction and contributions

This report presents our work regarding the paper *Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals* [4]. This paper presents a method that is able to learn temporal waveforms as well as their associated spatial patterns. The paper focuses on electromagnetic brain signals recovered using two techniques, electroencephalography (EEG) and magnetic fields by magnetoencephalography (MEG). Patterns of neural activity at specific frequencies have traditionally been interpreted as sustained rhythmic oscillations, linked to cognitive processes such as attention, high-level visual processing, and motor control.

This paper proposes a departure from linear filters and traditional signal representations by introducing a multivariate convolutional sparse coding (CSC) algorithm to extract meaningful representations from recordings. When applied to EEG or MEG data, the algorithm not only captures typical temporal waveforms but also discerns spatial patterns, facilitating the localization of their origins in the brain. Leveraging alternated minimization and a greedy coordinate descent solver, the algorithm achieves state-of-the-art efficiency in processing extended time series. The method is applied to MEG data, showcasing its ability to recover biological artifacts. This approach unveils the presence of non-sinusoidal μ-shaped patterns, accompanied by topographic maps corresponding to the somatosensory cortex.

The method presented in the paper is implemented in the AlphaCSC library, available on GitHub [4]. We decided to explore the possibilities given by this package as well as the impact of the parameters on the results of the method, two aspects that were not discussed in depth by the authors. We first started by reproducing the results of the paper, then conducting some benchmarking of the method followed by experiments regarding the data through an analysis of univariate and multivariate models.

The breakdown of tasks for this work is as follows: Yona worked on retrieving the dataset in FIF format, filtering the data, comparing univariate and multivariate models, the influence of parameter modification on signal reconstruction, the manual comparison of their impact on the atoms learned in the CSC. Gauthier worked on the evaluation of the scaling of the method, the investigation of how to improve it. He then focused on extending the results of the paper using the MNE and alphaCSC libraries with the research of more optimal parameters for the coordinate descent and finnaly the computation of the reconstructed enveloppes of the signal as well as the plotting of the representation of the learned topomap. The reuse of the source code amounts to 30% because we have adapted it to our problem. We added univariate analysis and successfully reconstructed the first multivariate model. We made a comparison between filtered and unfiltered data for atom extraction. Attempts to automate parameter selection were made but failed.

2 Method

Model definition One of the most popular, and successful, approach for statistical analysis of complex data is to use models that estimate data-driven representations. The authors of the paper chose to focus on the Dictionnary learning models, which goal is to learn atoms that approximate data with sparse representations. An addition to this technique is to use shift-invariant atoms, they give better approximations of the signal than the usual bases such as Fourier or wavelets given that they are not limited to narrow frequency bands. The approach used in the paper is the Convolutional Sparse Coding (CSC). It is constructed by casting the problem as an optimisation problem by representing the signal as a sum of convolutions between the atoms and activation signals.

The formulation used in the paper is a multivariate adaptation with of the univariate shift-invariant sparse coding model from [3]. This multivariate version is defined as follows:

$$\begin{gathered}\min_{\{D_k\}_k,\{z_k^n\}k,n} \sum_{n=1}^{N} \frac{1}{2} \left\| X^n - \sum_{k=1}^{K} z_k^n * D_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1 \\ \text{s.t. } \|D_k\|_2^2 \leq 1,\, z_k^n \geq 0 \text{ and } \lambda > 0,\end{gathered} \tag{1}$$

with P the number of channel, $\{X^n\}_{n=1}^N \subset \mathbb{R}^{P\times T}$ the N observed multivariate signals, $\{D_k\}_{k=1}^K \subset \mathbb{R}^{P\times L}$ the spatio-temporal atoms we aim to learn and $\{z_k^n\}_{k=1}^K \subset \mathbb{R}^{P-L+1}$ the sparse activations associated with X^n. λ is the regularization parameter and therefore always positive. The model assumes that the coding signals z_k^n are sparse and positive. That means that we expect the signal to have few nonzero entries and that the temporal patterns are present each time with the same polarity.

The authors propose to adapt this model by adding a rank-1 constraint on the dictionnary, $D_k = u_k v_k^\top \in \mathbb{R}^{P\times L}$, with $u_k \in \mathbb{R}^P$ being the pattern over the channels and $v_k \in \mathbb{R}^L$ the pattern over time. Equation (1) then becomes:

$$\begin{gathered}\min_{\{u_k\}_k,\{v_k\}_k,\{z_k^n\}k,n} \sum_{n=1}^{N} \frac{1}{2} \left\| X^n - \sum_{k=1}^{K} z_k^n * (u_k v_k^\top) \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1 \\ \text{s.t. } \|u_k\|_2^2 \leq 1,\, \|v_k\|_2^2 \leq 1,\, z_k^n \geq 0,\end{gathered} \tag{2}$$

This constraint assumes that each source is linearly spread over all channels instantaneously with a constant topographic map (i.e the spatial localisation of the electrical potentials in the brain). This assumption yiels several advantages. The model should be more robust to the presence of independent noise over the different channels and it can help separating overlapping sources. Such overlapping sources often lead to high rank situations when the initial sources are rank-1.

Model estimation The minimisation problem of (2) is convex in each block of variables but not jointly. To solve this issue, the authors propose to use a block coordinate descent approach. The idea is to minimize the objective function alternatively over a block of the variables. In our case this means updating the activations z_k^n, the spatial patterns u_k and the temporal patterns v_k alternatively.

We first focus on the solving for the activations, called Z-Step in the paper. The goal is to retrieve the NK activation signals $z_k^n \in \mathbb{R}^{P-L+1}$ associated to the signal $X^n \in \mathbb{R}^{P\times T}$. To do so we need to solve the following ℓ_1-regularized optimization problem :

$$\min_{\substack{\{z_k^n\}k,n \\ z_k^n \geq 0}} \frac{1}{2}\left\| X^n - \sum_{k=1}^{K} z_k^n * D_k \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1 \tag{3}$$

Given that it is fully convex in z_k^n it can be solved. Multiple algorithms exist to do so (FISTA, ADMM, L-BFGS) but they can be very computationaly slow for long signals since they scale linearly on T. The authors propose to use a technique called Coordinate Descent.

The idea of the coordinate descent is to update the estimate of the solution one coordinate $z_k[t]$ at a time. Given that once restricted to one coordinate (3) has a closed form solution, it is possible to compute the optimal value $z'_k[t]$:

$$z'_k[t] = \max\left(\frac{\beta_k[t] - \lambda}{\|D_k\|_2^2}, 0\right), \quad \text{with} \quad \beta_k[t] = \left[D_k^{\Lsh} \tilde{*} \left(X^n - \sum_{l=1}^{K} z_l * D_l + z_k[t]e_t * D_k\right)\right][t] \tag{4}$$

where $e_t \in \mathbb{R}^{T-L+1}$ is the canonical basis vector, $D_k^{\Lsh}[t] = X[T+1-t]$ and $D \tilde{*} D' = \sum_{p=1}^{P} D_p * D'_p$.

When updating the coefficient $z_{k_0}[t_0]$ to the value $z'_{k_0}[t_0]$, β is updated with:

$$\beta_k^{(q+1)}[t] = \beta_k^{(q)}[t] + (D_{k_0}^{\Lsh} \tilde{*} D_k)[t - t_0](z_{k_0}[t_0] - z'_{k_0}[t_0]), \quad \forall (k,t) \neq (k_0, t_0). \tag{5}$$

There are different strategies to update the coordinate in order to find the optimal value. They are all a tradeoff between the computational cost and the improvements it provides. The authors propose to use the Locally greedy coordinate descent that is more efficient than the other methods when the z_k are sparser due to the fact that in this case, only few coefficients need to be updated in the signal, resulting in a low number of iterations. At each iteration q, we pick the coordinate to update from one of the M subsegments of the signal using a greedy strategy:

$$(k_0, t_0) = \underset{(k,t)}{\arg\max} |z_k[t] - z'_k[t]|, \quad m \equiv q (\bmod M) + 1 \tag{6}$$

We now focus on the solving for the atoms. To update the K spatial patterns u_k and v_k the problem to solve is:

$$\min_{\substack{\|u_k\|_2 \leq 1 \\ \|v_k\|_2 \leq 1}} \sum_{n=1}^{N} \frac{1}{2} \|X^n - \sum_{k=1}^{K} z_k^n * (u_k v_k^\top)\|_2^2 \tag{7}$$

This problem is once again nonconvex jointly but convex in each block of variable. To solve this issue, $\{u_k\}_k$ is optimized first, followed by $\{v_k\}_k$ using a projected gradient descent. The gradient relative to both block can be computed using the chain rule.

3 Data

The dataset used for the experiments is the sample dataset from the MNE-Python software [2]. The data was acquired with the Neuromag Vectorview system at MGH/HMS/MIT Athinoula A. Martinos Center Biomedical Imaging. EEG data from a 60-channel electrode cap was acquired along with with the MEG. The original MRI data set was acquired with a Siemens 1.5 T Sonata scanner using an MPRAGE sequence.

In this experiment, checkerboard patterns were presented to the subject into the left and right visual field, interspersed by tones to the left or right ear. The interval between the stimuli was 750 ms. Occasionally a smiley face was presented at the center of the visual field. The subject was asked to press a key with the right index finger as soon as possible after the appearance of the face.

We have worked on Convolutional Sparse Coding through the univariate and the multivariate models. In our context of signal processing or data analysis, "univariate" refers to a single variable or dimension. For example, a time series of a single channel EEG signal. A "multivariate" refers to data that involves multiple variables or dimensions such as a dataset with multiple time series recovered from multiple sensors measuring different aspects of a system.

We drew inspiration from the authors' experiments, which are available at [1]. However, the available code performs a multivariate convolutional sparse coding on gradiometer data and the authors do not retrieve the data in original format. We then followed the method presented by the authors in the paper: a comparison between univariate and multivariate CSC. Indeed, MNE-Python data structures are based around the FIF file format from Neuromag that Python can load. MNE-Python dataset detects different sensor types and handles each appropriately, then each channel signal is accessible. The data loaded is composed of 204 Gradiometers, 102 Magnetometers, 60 EEG (Electroencephalography signal), 9 Stimulus and 1 EOG (Electrooculogram), leading to 376 different channels available. We worked on the three first signals. In the current data set, signals were recorded with a sampling rate of 150 Hz, meaning their is a measurement from each electrode almost every 7 ms. The information given by the dataset reports that data was low-pass-filtered at 40 Hz during recording, which is very comfortably below the Nyquist frequency, given that the sampling rate was 150 Hz. A high-pass filter of 0.1 Hz was also applied, which would have the effect of removing very slow drift in the recording. Channels have a length of 41700 points, ie the record lasts 277.7 s.

Preprocessing We applied a low pass filter on EEG signals to avoid aliasing. This phenomenon causes high-frequency noise. Adjusting this frequency band enhances the signal, we have seen the importance of knowing the properties of the studied signal and the chosen detector.

Univariate model Beginning our experiments directly on multiple channels was quite busy because the format of the data is a little bit intricate. So, we began with testing on one channel. We made vary the number of atoms to learn with the parameter named K in the paper. We introduced the parameter overlap to see its influence on the atom learning. Overlapping windows help in capturing more information about the underlying signal but in practice on the dataset it did not improve the reconstruction.

Multivariate model With the rank-1 model, each atom is written as a product of two univariate vectors which rely on the channel and over time. The Batch algorithm for convolutional dictionary learning was used on multivariate model.

4 Results

The available tutorial published by the authors of the paper focuses on extracting µ-wave from the somato-sensory dataset. Our work then consisted in the exploitation of the method used in this tutorial and the enlargement to our database, the tools used and the algorithm developed. The tutorial develops the Batch Convolution Dictionary Learning algorithm for a multivariate data model extracted directly using load data library from alphacsc.datasets. Our work was then further developed, we extracted the data using the RAW file in the fif format as previously presented. We then varied some parameters such as the regularization $lambda$, the number of channels used P, the number of atoms K. The observation was manual because we did not manage to automate the loop.

The analysis of the results was manual with a finding on the displays of atoms and sometimes more applied. Indeed, we have not always been able to extract metrics on the evaluation of our method in order to have a critical opinion and some hindsight.

We computed Power Spectral Density on multivariate signals as in Figure 4 which tells how power of each signal (among EEG and MEG) is distributed across different frequencies. We can note that the signal's frequency content is high for frequencies lower than 40 Hz, frequency at which was applied the filter.

Then, the visualization of atoms for the univariate model emphasizes the importance and good mastery of the filter. The 7 highlights that variations in filtered data are more consequent but look alike more the signal. The difficulty here lies in the choice of the number of atoms when calculating the dictionary. Indeed, when we vary the number of atoms when learning the dictionary, one pattern returns and the others are noisy replicas cf. Figure 5. So, for gradiometers, magnetometers and EEGs, two atoms are enough to form the signal. This is an approximation for the univariate model.

However the visualization of atoms in multivariate model is not optimal. The authors of the paper insist that increasing the number of channels studied improves the filling or the overlap of the signal but our experiments do not. The atoms found in multivariate models are very unstable and seem to overlap, it may be because of overwriting vectors or due to poor graphical card. Figure 2 shows 3 of the 25 atoms in a simulation, 25 being parameter given in the paper for a test. Figure 6 shows 3 learned atoms from a dictionary of size 5. It is remarkable that the recovered atoms have a major overlap. It is also noteworthy that the displays for the multivariate model have an artifact with the vertical line that I could not explain or correct. Maybe our experiments and tests over parameters were wrongly applied, or the data used not adapted. Learning atoms on the univariate model presents interesting and questionable results to the naked eye. We varied the number of atoms K and the parameter of regularization λ.

Finally, we managed to overlay the reconstructed signal from sparse coding and the original multivariate MEG signal, which can be seen in Figure 3. Despite our uncertainty about the proper extraction and learning of atoms on the multivariate model, the reconstructed signal has the same overall shape as the original signal.

During the familiarization with the AlphaCSC library, we noticed a surprising behaviour of the parameter setting the number of CPU cores used for the computations. We decided to do a proper evaluation of the computation times to assess it. The tests were performed using an Intel Core i9 13900HX (8P cores + 16E Cores - 32 Threads) laptop CPU. The results can be found in Figure 8. The computational performances scales really well with the number of cores up until 4 cores, then keeps decreasing up until 8 cores before reaching a plateau. Upon further investigation we did not find an explanation for why the performance improvements slows after 4 jobs. The poor scaling after 8 jobs is likely due to the fact that the python multiprocess API does not use the "Efficient cores" properly. We did another test with a fixed amount of jobs (32) and increasing the number of splits of the signal as the authors claim it could help paralellization but we did not find a significant difference.

5 Appendix

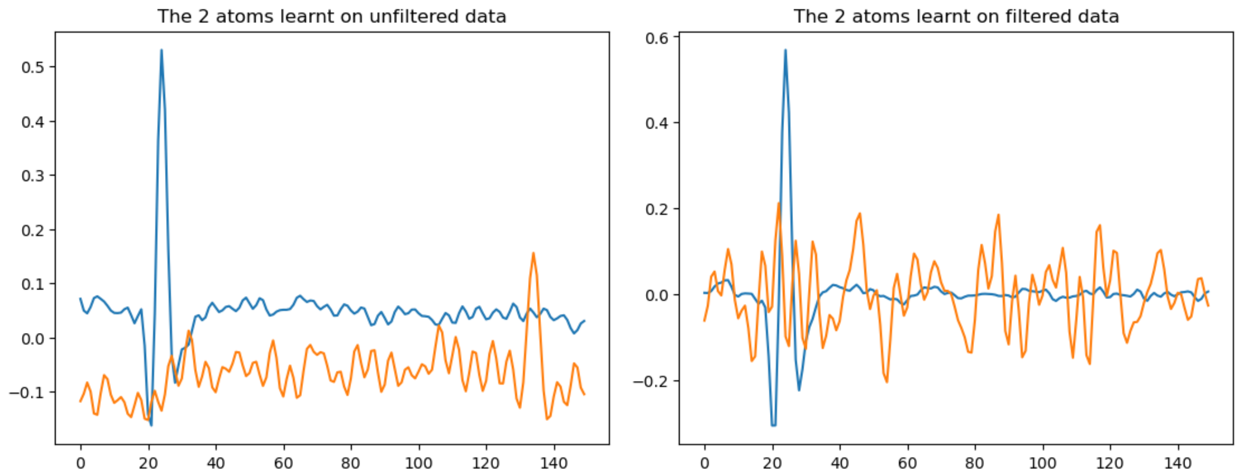

Figure 1: Comparison of atom extraction for filtered and unfiltered MEG data

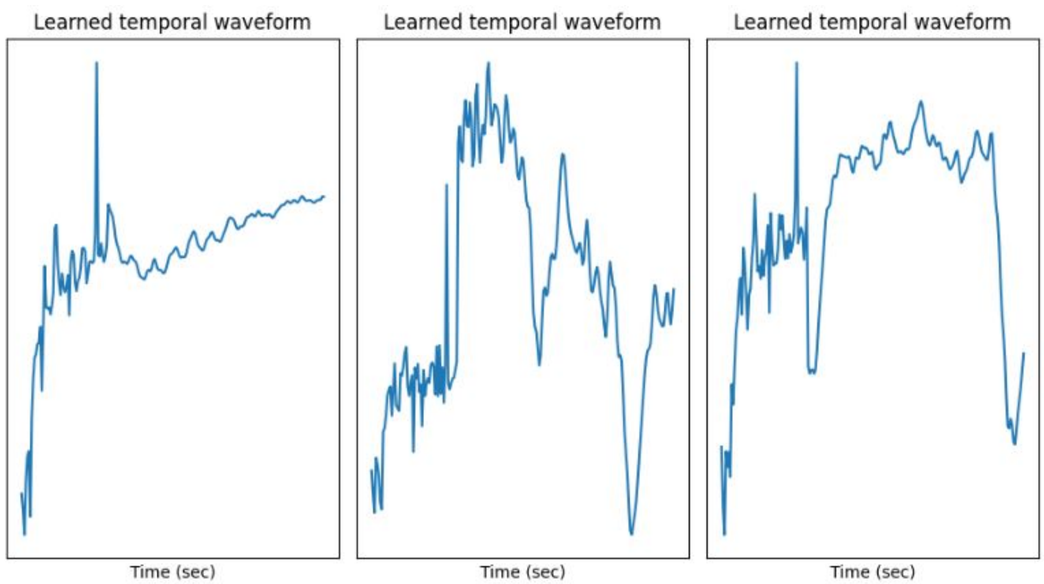

Figure 2: 3 of the 25 atoms learnt on multivariate data

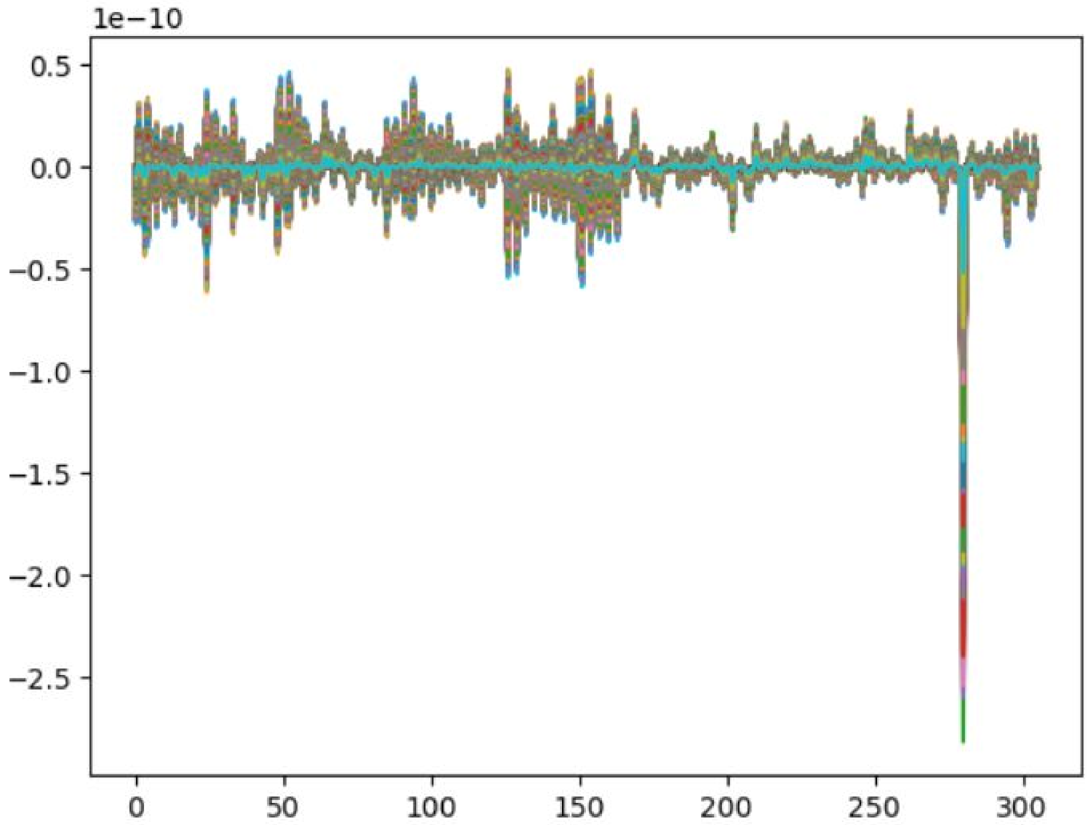

Figure 3: Reconstruction and original MEG signal

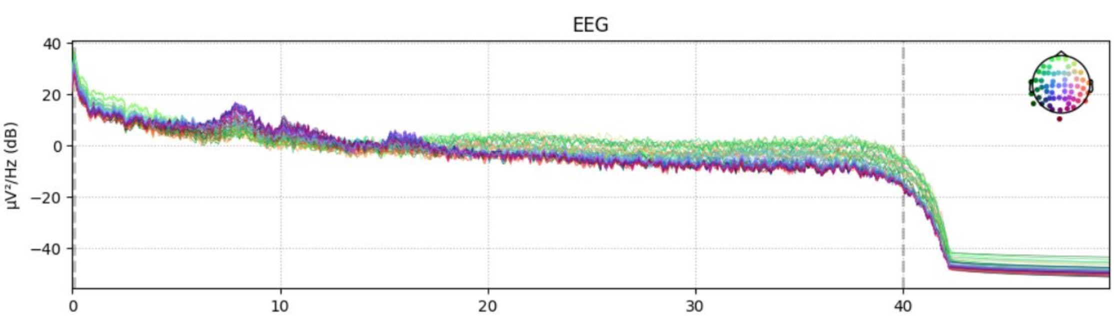

Figure 4: Power Spectral Density applied on EEG multivariate signal

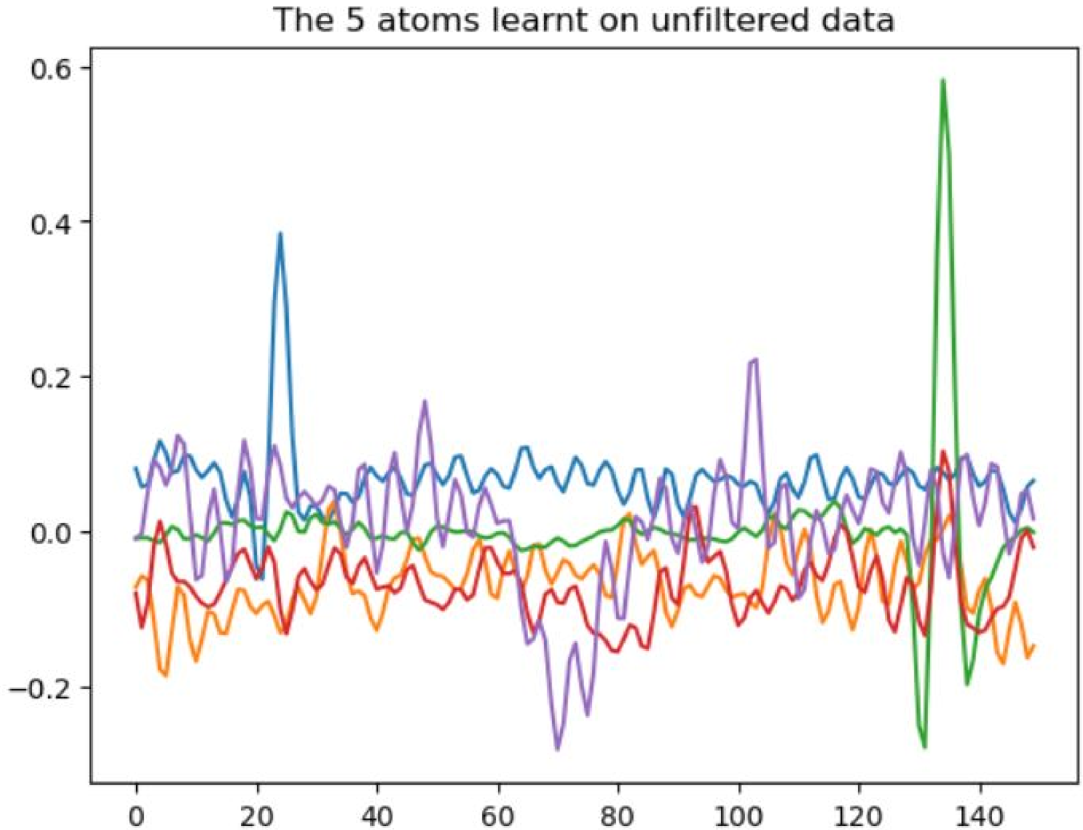

Figure 5: 5 atoms learnt on univariate model through CSC

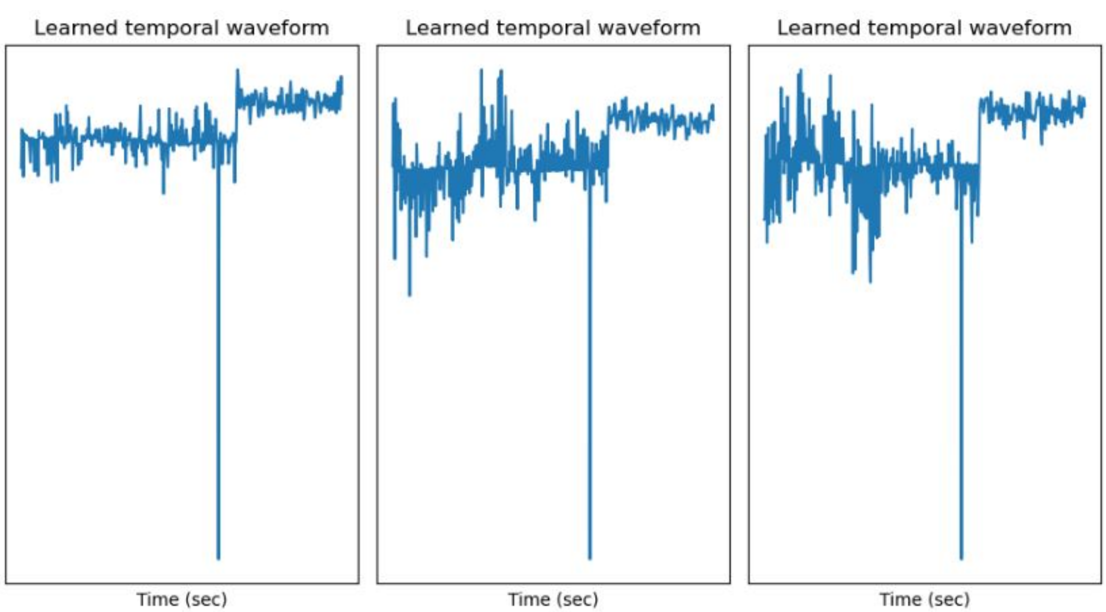

Figure 6: 3 atoms learnt among a dictionary of 5 on multivariate CSC processing

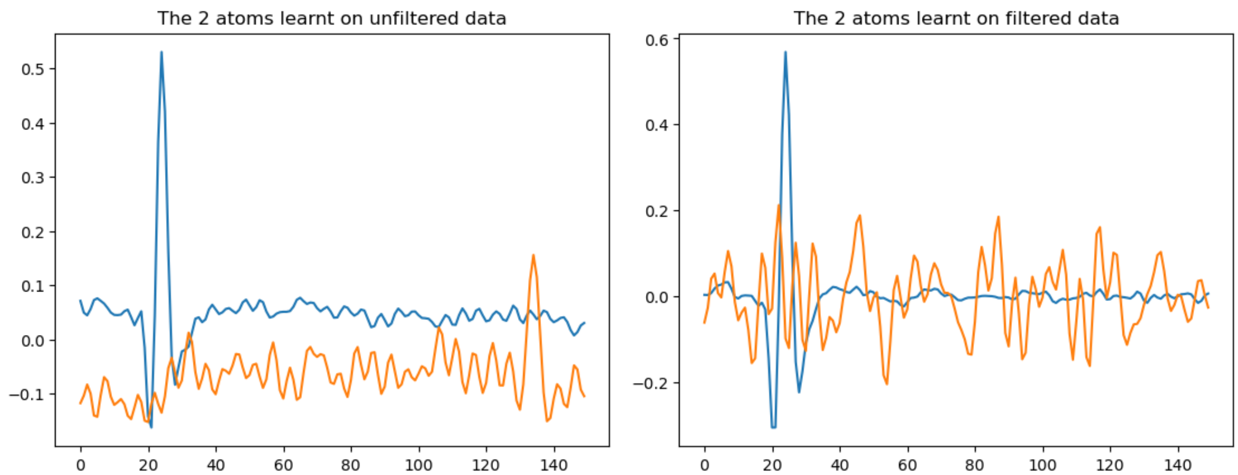

Figure 7: Comparison of atom extraction for filtered and unfiltered MEG data

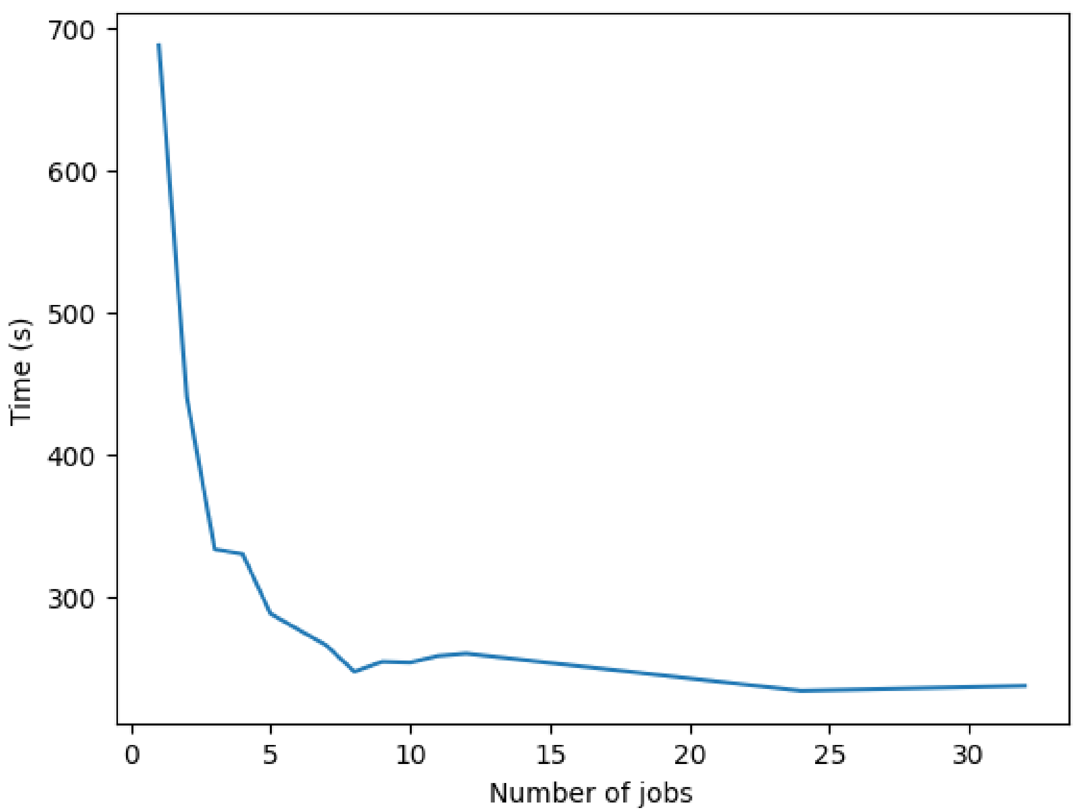

Figure 8: Computational scaling of the method with regards to the number of jobs

References

[1] Git. https://alphacsc.github.io/stable/auto_examples/multicsc/plot_somato_mu_waves.html#id1.

[2] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti S. Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(267):1–13, 2013.

[3] Roger Grosse, Rajat Raina, Heng Kwong, and Andrew Y. Ng. Shift-invariant sparse coding for audio classification. In *23rd Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 149–158. AUAI Press, 2007.

[4] T. D. La Tour, T. Moreau, M. Jas, and A. Gramfort. Multivariate convolutional sparse coding for electromagnetic brain signals. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 3292–3302, 2018.