Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals

Tom Dupré La Tour[1] Thomas Moreau[2] Mainak Jas[1] Alexandre Gramfort[2]

[1]LTCI, Télécom ParisTech, Université Paris Saclay, Paris, France,
[2] Parietal, INRIA, Université Paris Saclay, Saclay, France

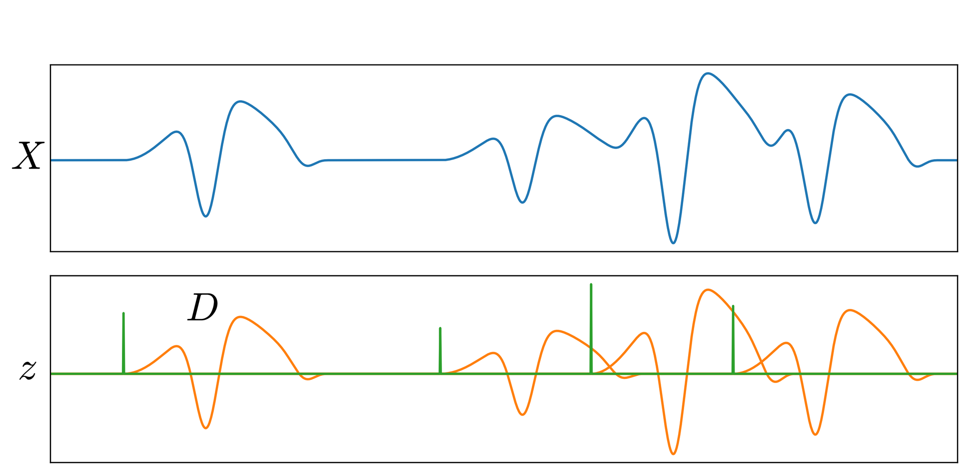
Code available at: `https://alphacsc.github.io`

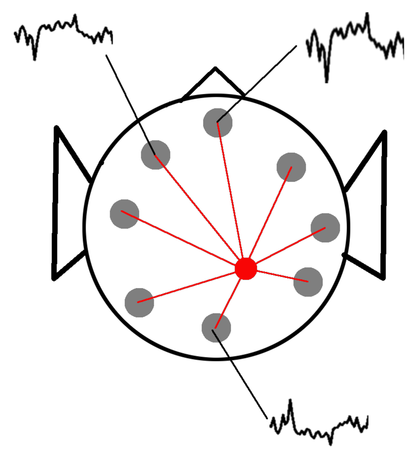
1. Convolutional Sparse Coding (CSC)

Convolutional linear model :

$$X = \sum_{k=1}^{K} z_k * D_k + \mathcal{E}, \qquad \mathcal{E} \sim \mathcal{N}(0, \sigma I) \tag{1}$$

with signal $X \in \mathbb{R}^{P\times T}$ (P sensors and T samples), K patterns $D_k \in \mathbb{R}^{P\times L}$ (duration L) and activations $z_k \in \mathbb{R}^{\widetilde{T}}$ such that $\widetilde{T} = T - L + 1$.





Multivariate CSC

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

Multivariate CSC with rank-1 constraint

$$\min_{u_k, v_k, z_k^n} \sum_{n=1}^{N} \frac{1}{2} \left\| X^n - \sum_{k=1}^{K} z_k^n * (u_k v_k^\top) \right\|_2^2 + \lambda \sum_{k=1}^{K} \|z_k^n\|_1 \,, \tag{3}$$
$$\text{s.t.} \quad \|u_k\|_2^2 \leq 1 \,,\ \|v_k\|_2^2 \leq 1 \text{ and } z_k^n \geq 0 \,.$$

$\leftarrow$ One source in the brain is spread linearly and instantaneously over all sensors. The rank-1 hypothesis is particularly suited for MEG signals.

2. Z-step: solving for the activations

The Z-step solves (2) or (3) for a fixed dictionary. We solve it using

Greedy Coordinate Descent (GCD):

- Optimization problem for one coordinate has a close form:

$$z'_k[t] = \max\left(\frac{\beta_k[t] - \lambda}{\|D_k\|_2^2}, 0\right) \tag{4}$$

  with $\beta_k[t] = \left[D_k^{\Lsh} * \left(X - \sum_{l=1}^{K} z_l * D_l + z_k[t] e_t * D_k\right)\right][t]$.
- Greedily update coefficient $(k_0, t_0) = \operatorname{argmax} |Z_k[t] - Z'_k[t]|$

Locally Greedy Coordinate Descent (LGCD): [Moreau et al., 2018]

- Select the best coordinate locally, on one of M contiguous segments $\mathcal{C}_m$:

$$(k_0, t_0) = \operatorname*{argmax}_{(k,t)\in\mathcal{C}_m} |Z_k[t] - Z'_k[t]|$$

- For $M = \lfloor \widetilde{T}/(2L-1) \rfloor$, the computational complexity of choosing the coefficient matches the complexity of performing the update.
- It is efficient when the updates are weakly dependent and when the solution is sparse.

3. D-step: solving for the atoms

The D-step solves (2) or (3) for a fixed activations. We solve it using

Projected Gradient Decent (PGD):

- Separate minimization over $\{u_k\}_k$ and $\{v_k\}_k$.
- The step size is set using a Armijo backtracking line-search.

Function and gradients computations:

- The gradient relatively to a full atom $D_k = u_k v_k^\top \in \mathbb{R}^{P\times L}$:

$$\nabla_{D_k} E = \sum_{n=1}^{N} (z_k^n)^\top * \left(X^n - \sum_{l=1}^{K} z_l^n * D_l \right) = \Phi_k - \sum_{l=1}^{K} \Psi_{k,l} * D_l \,, \tag{5}$$

  where $\Phi_k \in \mathbb{R}^{P\times L}$ and $\Psi_{k,l} \in \mathbb{R}^{2L-1}$ are constant during a D-step and can be precomputed.
- The gradients relatively to u_k and v_k are obtained using the chain rule:

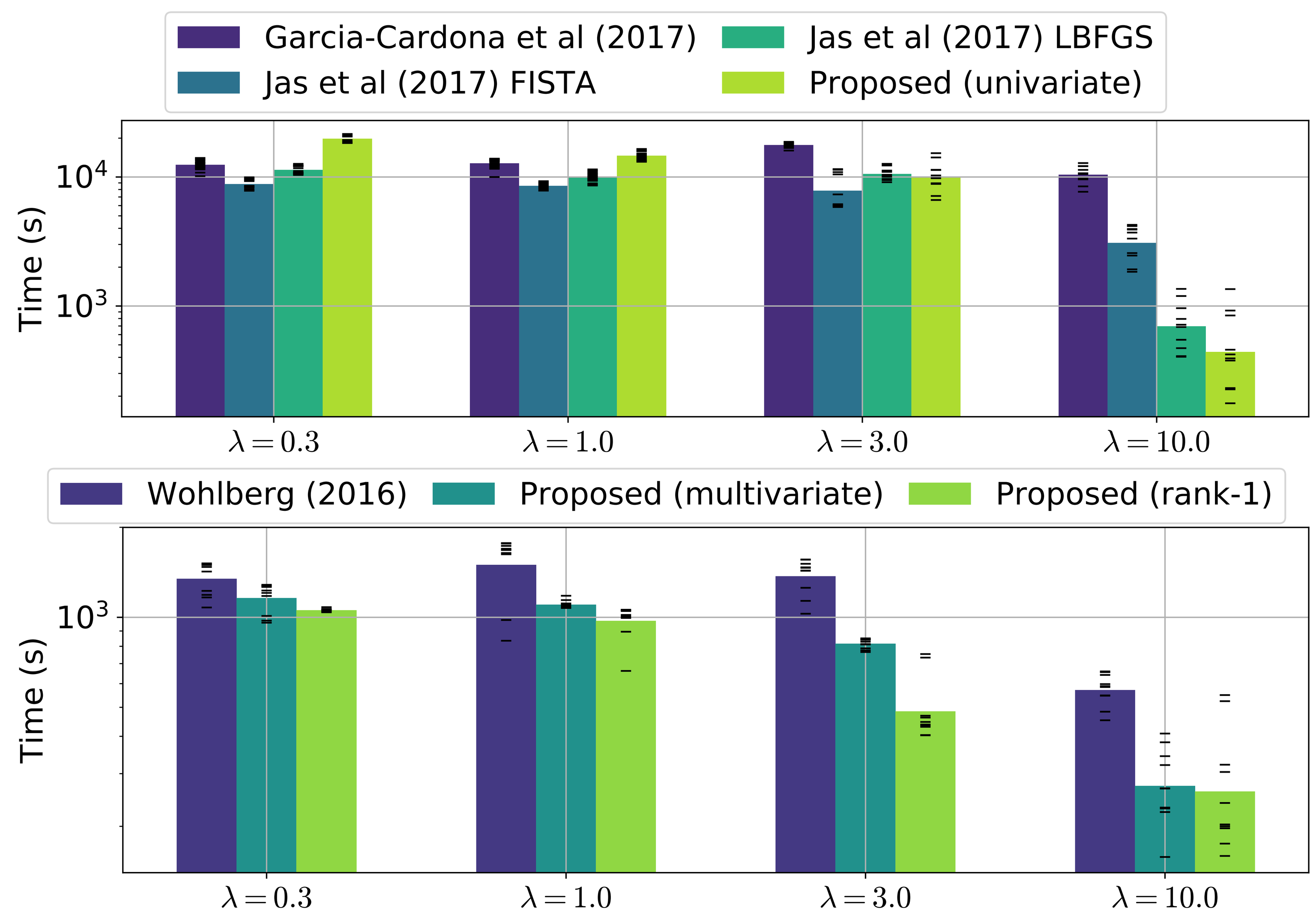
$$\nabla_{u_k} E = (\nabla_{D_k} E) v_k \quad \in \mathbb{R}^P \,, \tag{6}$$
$$\nabla_{v_k} E = u_k^\top (\nabla_{D_k} E) \quad \in \mathbb{R}^L \,, \tag{7}$$

- E can be computed, up to a constant term C, with the following:

$$E = \sum_{k=1}^{K} u_k^\top (\nabla_{D_k} E) v_k + C \,. \tag{8}$$

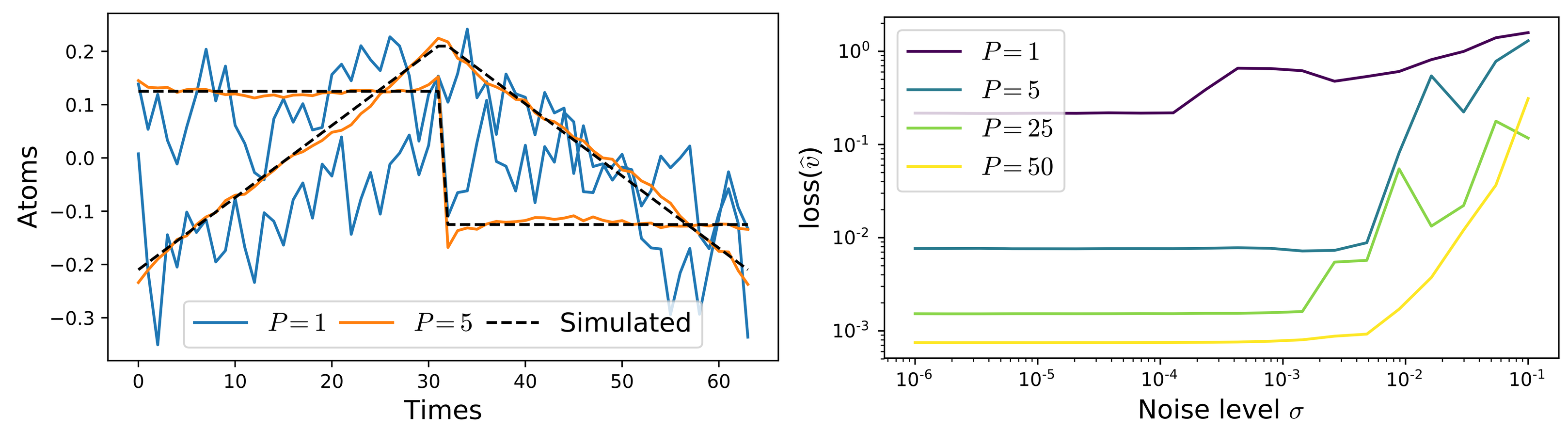
$\Rightarrow$ Using this pre-computation helps scaling with the number of channels as the computation are in $\mathcal{O}(P+L)$ instead of $\mathcal{O}(PL)$.

4. Multivariate speed benchmark



Speed benchmark for univariate-CSC (*top*) and multivariate-CSC (*bottom*).

5. Simulated Signals

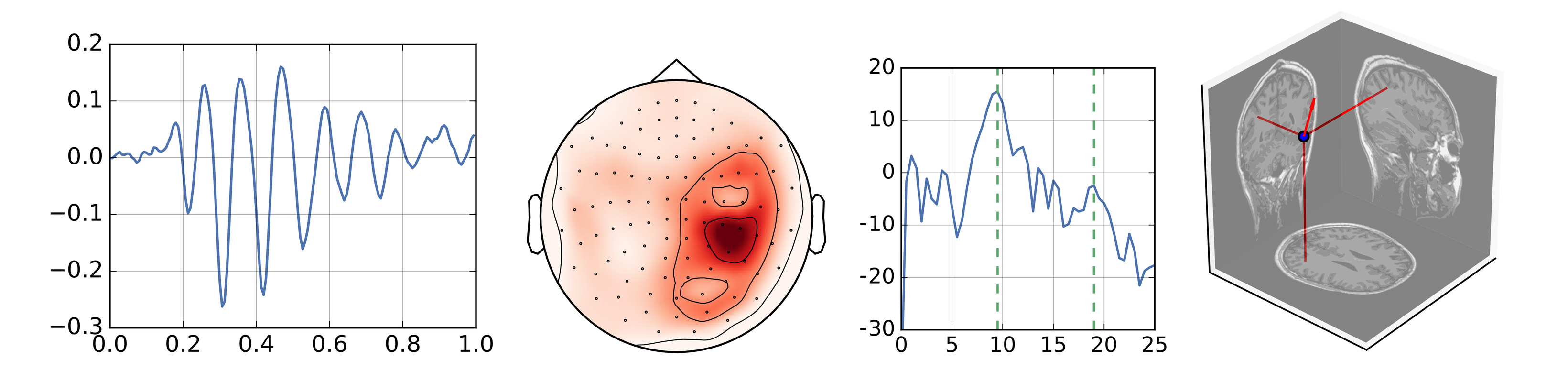


- Signals are generated following (1) over P channels.
- Recovered temporal patterns $\widehat{v}_k$ are evaluated using:

$$\text{loss}(\widehat{v}) = \min_{s\in\mathfrak{S}(K)} \sum_{k=1}^{K} \min\left(\|\widehat{v}_{s(k)} - v_k\|_2^2, \|\widehat{v}_{s(k)} + v_k\|_2^2\right) \,.$$

$\Rightarrow$ More channels improves the pattern recovery as it disentangling super-imposed patterns.

6. Experimental Signals



(a) Temporal waveform (b) Spatial pattern (c) PSD (dB) (d) Dipole fit

Atom learned using the MNE-somatosensory dataset. The learned temporal pattern illustrate mu-waveforms described for instance in [Cole and Voytek, 2017].