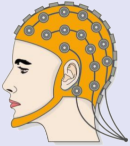
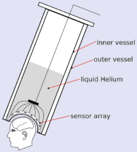
Multivariate Convolutional Sparse Coding for Electromagnetic Brain Signals

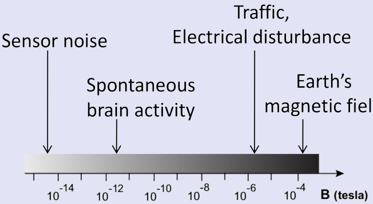
Dupré La Tour T., **TM**, Mainak J., Gramfort A.
INRIA Saclay

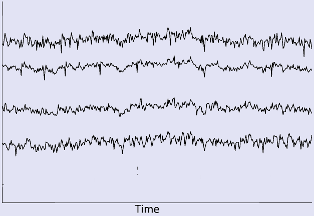
Studying the brain activty with Electromagnetic brain signal

- Brain activity produces electromagnetic activity
- This can be measured with EEG or MEG









Goal: learn representation from neural data

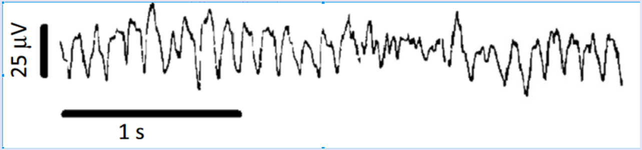
Many studies rely on Fourier or wavelet analyses, with fixed representation:

- Easy interpretation,
- Standard analysis *e.g.* canonical bands alpha, beta or theta.

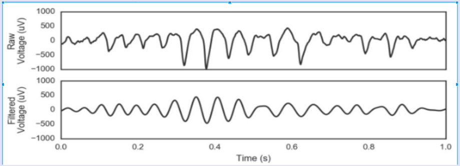
Goal: learn representation from neural data

However,

- Some brain rhythms are not sinusoidal, *e.g.*mu-waves



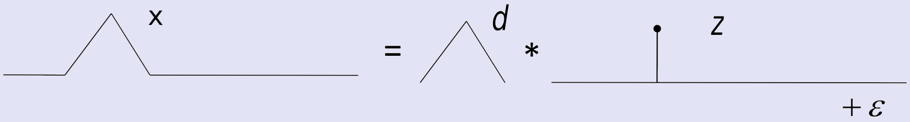
- Filtering degrade waveforms



$\Rightarrow$ Can we do better with data-driven approach?

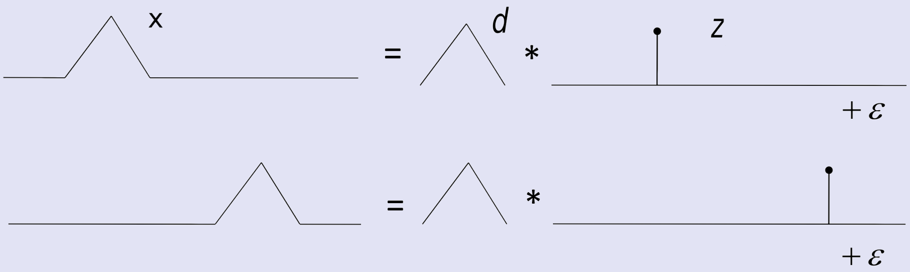
Extracting shift invariant patterns

Key idea: we are not interested in the time-shift of the patterns



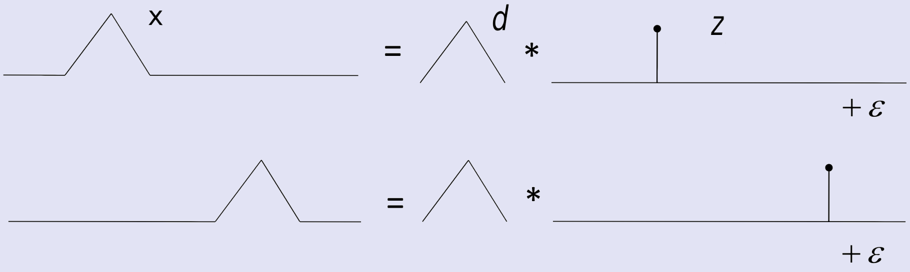
Extracting shift invariant patterns

Key idea: we are not interested in the time-shift of the patterns



Extracting shift invariant patterns

Key idea: we are not interested in the time-shift of the patterns

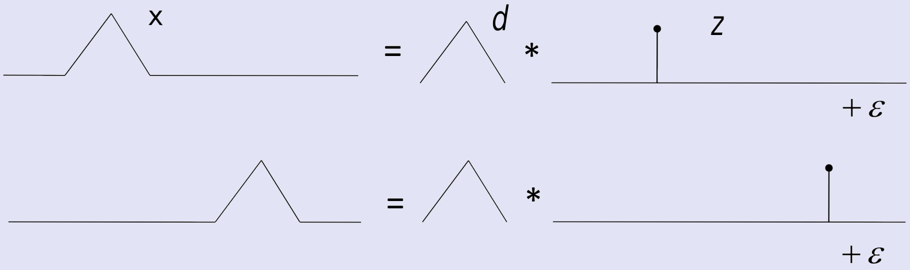


Extending this to K patterns:

$$x = \sum_{k=1}^{K} d_k * z_k + \mathcal{E}$$

Extracting shift invariant patterns

Key idea: we are not interested in the time-shift of the patterns



Extending this to K patterns:

$$x = \sum_{k=1}^{K} d_k * z_k + \mathcal{E}$$

$\Rightarrow$ Convolutional representation

Hypothesis: patterns are not present everywhere in the signal. They are localized in time.

$\Rightarrow$ Sparse activation signals z

Convolutional sparse coding problem (CSC) [Grosse et al., 2007]
For a set of N univariate signals x^n, solve

$$\min_{d_k, z_k^n} \sum_{n=1}^{N} \frac{1}{2} \| x^n - \sum_{k=1}^{K} z_k^n * d_k \|_2^2 + \lambda \sum_{k=1}^{K} \| z_k^n \|_1$$

Extra hypothesis: the activation z_k^n are positive.

How to extend CSC to multivariate signals?

We can just use multivariate convolution,

$$\underbrace{X[t]}_{\in\mathbb{R}^P} = \sum_{k=1}^{K} (z_k * D_k)[t] = \sum_{k=1}^{K}\sum_{\tau=1}^{L} z_k[t-\tau]\,\underbrace{D_k[\tau]}_{\in\mathbb{R}^P}$$

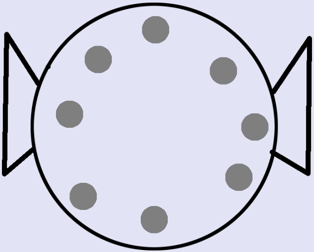
with:

- X a multivariate signal of length T in $\mathbb{R}^P$
- D_k a multivariate signal of length L in $\mathbb{R}^P$
- z_k a univariate activation signal of length $\widetilde{T} = T - L + 1$

However, this model does not account for the physics of the problem.

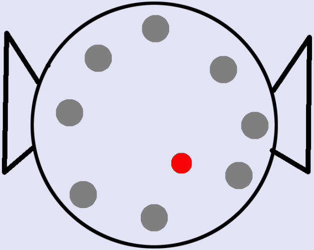
EM wave diffusion

- Recording here with 8 sensors



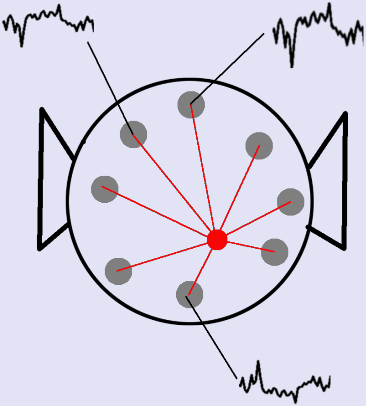
EM wave diffusion

- Recording here with 8 sensors
- EM activity in the brain



EM wave diffusion

- Recording here with 8 sensors
- EM activity in the brain
- The electric field is spread **linearly** and **instantaneously** over all sensors (Maxwell equations)



Multivariate CSC with rank-1 constraint

Idea: Impose a rank-1 constraint on the dictionary atoms D_k

To make the problem tractable, we decided to use auxiliary variables u_k and v_k *s.t.* $D_k = u_k v_k \top$.

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

Here,

- $u_k \in \mathbb{R}^P$ is the spatial pattern of our atom
- $v_k \in \mathbb{R}^L$ is the temporal pattern of our atom

Optimization strategy

The problem 1 is not jointly convex in z_k^n, u_k and v_k but it is convex in each block of coordinate.

We can use a block coordinate descent, aka alternate minimization, to converge to a local minima of this problem. The 3 following steps are applied alternatively:

- **Z-step:** given a fixed estimate of the atom, compute the activation signal z_k^n associated to each signal X^n.
- **u-step:** given a fixed estimate of the activation and temporal pattern, update the spatial pattern u_k.
- **v-step:** given a fixed estimate of the activation and spatial pattern, update the temporal pattern v_k.

Z-step: Locally greedy coordinate descent (LGCD)

N independent problem such that

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

This problem is convex in z_k and can be solved with different techniques:

- Greedy CD [Kavukcuoglu et al., 2010]
- Fista [Chalasani et al., 2013]
- ADMM [Bristow et al., 2013]
- L-BFGS [Jas et al., 2017]

$\Rightarrow$ These methods can be slow for long signals as the complexity of each iteration is at least linear in the length of the signal.

Z-step: Locally greedy coordinate descent (LGCD)

For the Greedy Coordinate Descent, only 1 coordinate is updated at each iteration: [Kavukcuoglu et al., 2010]

1. The coordinate $z_{k_0}[t_0]$ is updated to its optimal value $z'_{k_0}[t_0]$ when all other coordinate are fixed:

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

   with $\beta_k[t] = \left[D_k^{\Lsh} * \left(X - \sum_{l=1}^K z_l * D_l + z_k[t]e_t * D_k\right)\right][t]$

2. The updated coordinate is chosen greedily by maximizing $|z_k[t] - z'_k[t]|$

For each coordinate update, it is possible to maintain the value of β with $\mathcal{O}(KL)$ operations.

Z-step: Locally greedy coordinate descent (LGCD)

We propose to use the LGCD method proposed in Moreau et al. (2018) which is an extension of GCD. With LGCD, the coordinate is instead chosen greedily on one of M contiguous sub-segments of the signal $\mathcal{C}_m$,

$$\mathcal{C}_m = [\![1, K]\!] \times [\![(m-1)\widetilde{T}/M, m\widetilde{T}/M]\!]$$

With this strategy, if $M = \lfloor \widetilde{T}/(2L-1) \rfloor$ the complexity of the coordinate selection and the update of β are the same. The overall iteration complexity is $\mathcal{O}(KL)$ instead of (KT).

If very few coefficients are non-zero in z_k, only few iteration are needed and thus this strategy can be very efficient.

Z-step: Locally greedy coordinate descent (LGCD)

Algorithm 1: Locally greedy coordinate descent (LGCD)

Input : Signal X, atoms D_k, number of segments M, stopping parameter $\epsilon > 0$, z_k initialization

Initialize $\beta_k[t]$

repeat

  for $m = 1$ **to** M **do**

    Compute $z'_k[t] = \max\left(\frac{\beta_k[t]-\lambda}{\|D_k\|_2^2}, 0\right)$ for $(k,t) \in \mathcal{C}_m$

    Choose $(k_0, t_0) = \underset{(k,t)\in\mathcal{C}_m}{\operatorname{argmax}} |z_k[t] - z'_k[t]|$

    Update β

    Update the current point estimate $z_{k_0}[t_0] \leftarrow z'_{k_0}[t_0]$

until $\|z - z'\|_\infty < \epsilon$;

.

D-step: solving for the atoms

We use the projected gradient descent with an Armijo backtracking line-search [Wright and Nocedal, 1999] for both u-step and v-step for

$$\min_{\substack{\|u_k\|_2 \leq 1 \\ \|v_k\|_2 \leq 1}} E(\{u_k\}_k, \{v_k\}_k) \triangleq \sum_{n=1}^{N} \frac{1}{2} \|X^n - \sum_{k=1}^{K} z_k^n * (u_k v_k^\top)\|_2^2 \quad . \qquad (2)$$

One important computation trick is for fast computation of the gradient.

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

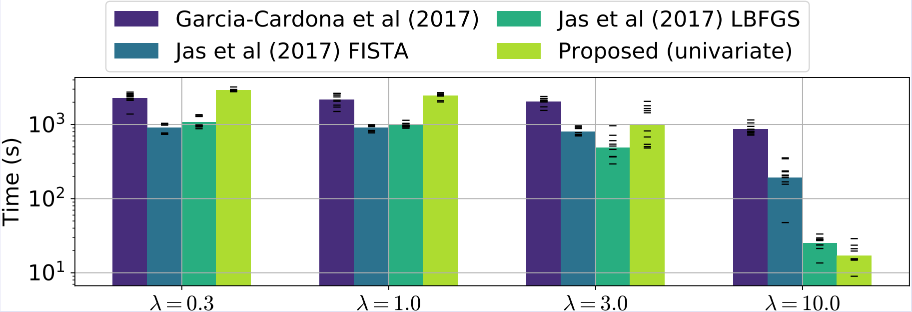
Computing $\nabla_{D_k} E(\{u_k\}_k, \{v_k\}_k)$ can be done efficiently

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

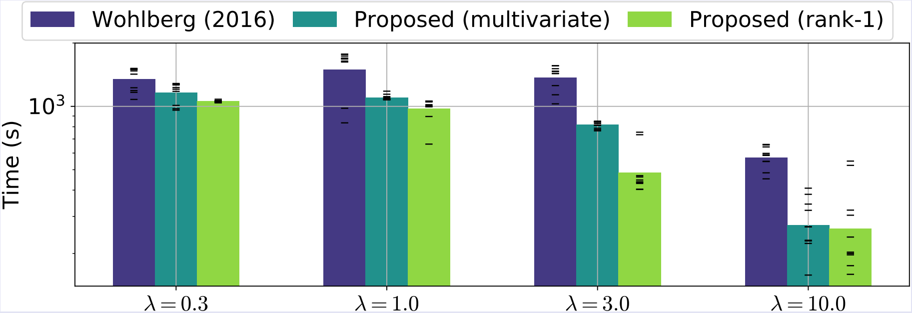
Experiments

Good time to wake-up if you got lost in the previous section!

Comparison with univariate methods on somato dataset with $T = 134,700$, $K = 8$ and $L = 128$

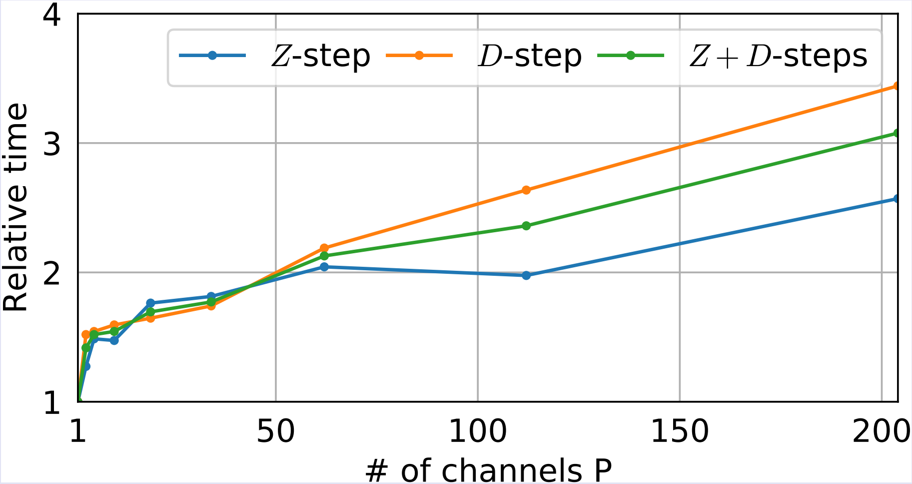


Comparison with multivariate methods on somato dataset with
$T = 134,700$, $K = 8$, $P = 5$ and $L = 128$



Good scaling in the number of channels P

Scaling relative to P on somato dataset with $T = 134,700$, $K = 2$, and $L = 128$



Pattern recovery

Test the pattern recovery capabilities of our method on simulated data,

$$X^n = \sum_{k=1}^{2} z_k * (u_k v_k^\top) + \mathcal{E}$$

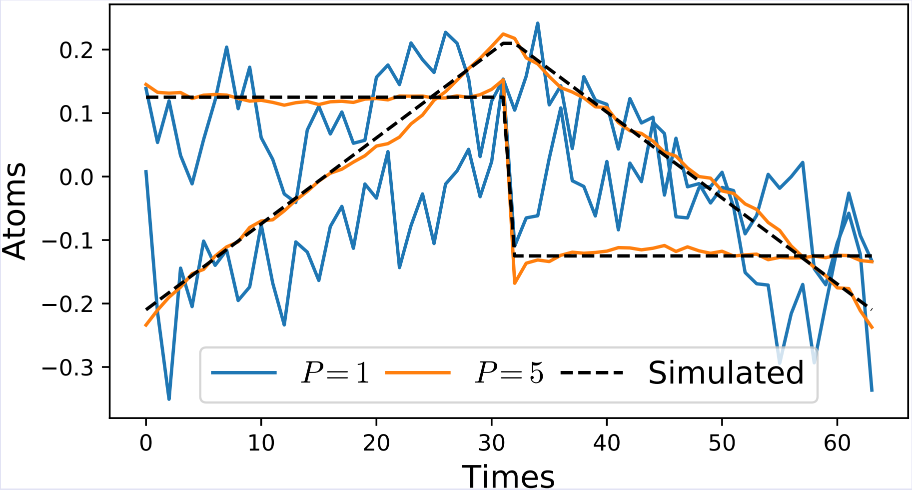
where (u_k, v_k) are chosen patterns of rank-1 and the activated coefficient $z_k^n[t]$ are drawn uniformly and their value are uniform in $[0, 1]$.

The noise $\mathcal{E}$ is generated as a gaussian white noise with variance σ.

We set $N = 100$, $L = 64$ and $\widetilde{T} = 640$

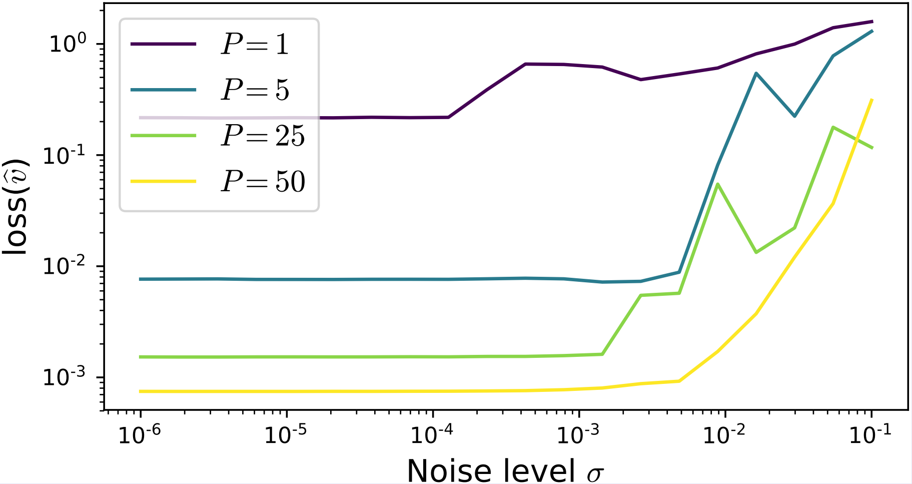
Pattern recovery

Patterns recovered with $P = 1$ and $P = 5$. The signals were generated with the two simulated temporal patterns and with $\sigma = 10^{-3}$.



Pattern recovery

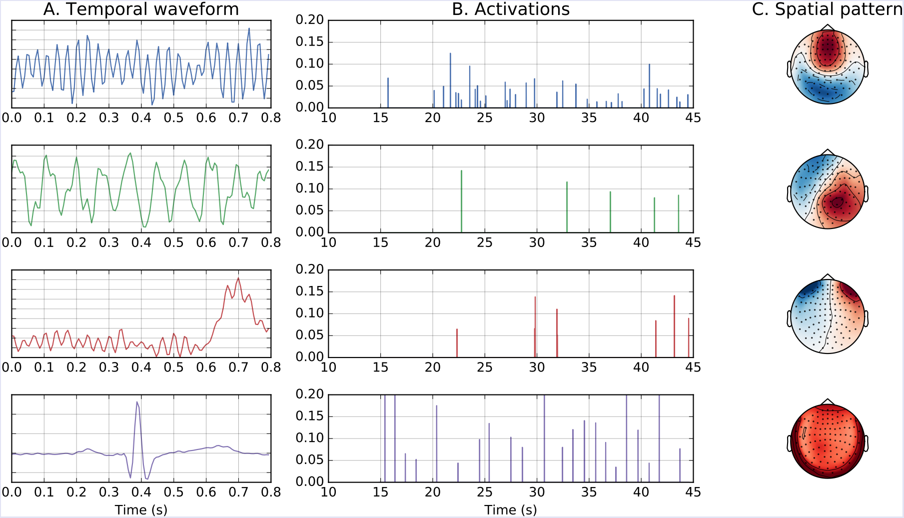
Evolution of the recovery loss with σ for different values of P. Using more channels improves the recovery of the original patterns.



Experiments on MEG data

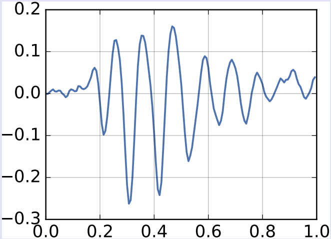
Even better time to wake-up!

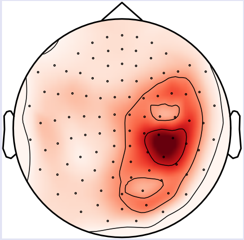
A selection of temporal waveforms of the atoms learned on the MNE sample dataset.

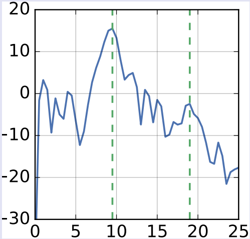


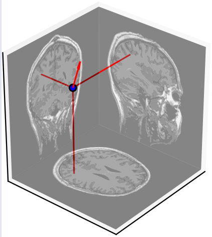
MNE somatosensory data

Atoms revealed using the MNE somatosensory data. Note the non-sinusoidal comb shape of the mu rhythm.









Conclusion

- We proposed a model for multivariate CSC with rank-1 constraint. This model makes sense for different type of data.
- We proposed a fast algorithm to solve the optimization problem involved in this model.
- We demonstrated numerically the performance of our algorithm on both simulated and real datasets.
- We illustrated the benefit of such method to study electromagnetic signals form recorded from brain activity.

Questions?

Reference

Bristow, H., Eriksson, A., and Lucey, S. (2013). Fast convolutional sparse coding. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 391–398, Portland, OR, USA.

Chalasani, R., Principe, J. C., and Ramakrishnan, N. (2013). A fast proximal method for convolutional sparse coding. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–5, Dallas, TX, USA.

Grosse, R., Raina, R., Kwong, H., and Ng, A. Y. (2007). Shift-Invariant Sparse Coding for Audio Classification. *Cortex*, 8:9.

Jas, M., Dupré La Tour, T., Şimşekli, U., and Gramfort, A. (2017). Learning the Morphology of Brain Signals Using Alpha-Stable Convolutional Sparse Coding. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1–15, Long Beach, CA, USA.

Kavukcuoglu, K., Sermanet, P., Boureau, Y.-l., Gregor, K., and Lecun, Y. (2010). Learning Convolutional Feature Hierarchies for Visual Recognition. In *Advances in Neural Information Processing Systems (NIPS)*, pages 1090–1098, Vancouver, Canada.

Wright, S. and Nocedal, J. (1999). *Numerical optimization*. Science Springer.