Independent-Variation Matrix Factorization With Application to Energy Disaggregation

Simon Henriet, Umut Simsekli, Sérgio Santos, Benoît Fuentes, Gael Richard

Independent-Variation Matrix Factorization with Application to Energy Disaggregation

Simon Henriet, Umut Şimşekli, *Member, IEEE,* Sergio Dos Santos, Benoit Fuentes, Gaël Richard, *Fellow, IEEE*

***Abstract*—Matrix factorization techniques have proven to be useful in many unsupervised learning applications. Such techniques have been recently applied to Non Intrusive Load Monitoring (NILM), the process of breaking down the total electric consumption of a building into consumptions of individual appliances. While several studies addressed the NILM problem for small-scale buildings, only few studies considered the problem for large buildings, where the signals exhibit significantly different behavior. To overcome the unaddressed difficulties of processing high frequency current signals that are measured in large buildings, we propose a novel technique called Independent-Variation Matrix Factorization (IVMF), which expresses an observation matrix as the product of two matrices: the *signature* and the *activation*. Motivated by the nature of the current signals, it uses a regularization term on the temporal variations of the activation matrix and a positivity constraint, and the columns of the signature matrix are constrained to lie in a specific set. To solve the resulting optimization problem, we rely on an alternating minimization strategy involving dual optimization and quasi-Newton algorithms. The algorithm is tested against Independent Component Analysis (ICA) and Semi Nonnegative Matrix Factorization (SNMF) on a synthetic source separation problem and on a realistic NILM application for large commercial buildings. We show that IVMF outperforms competing methods and is particularly appropriate to recover positive sources that have a strong temporal dependency and sources whose variations are independent from each other.**

***Index Terms*—Dictionary Learning, Semi-Nonnegative Matrix Factorization, Independent Component Analysis, Total Variation, Non-Intrusive Load Monitoring.**

I. Introduction

In the context of energy efficiency, Non-Intrusive Load Monitoring (NILM) is the process of breaking down the total electric energy consumption of a building into consumptions of individual appliances, using a few sensors and a signal disaggregation algorithm [1]. The NILM task can be formulated as a source separation problem where the observed total consumption (*the mixture*) is the simple sum of individual consumptions (*the unknown sources*). Depending on the type of data available, different matrix factorization techniques have been developed to solve this problem ([2], [3], [4], [5]), either fully unsupervised or using a pre-learned dictionary.

 Simon Henriet, U. Şimşekli and Gaël Richard are with the LTCI, Télécom Paris, Institut Polytechnique de Paris, 75013, Paris, France. Simon Henriet, Sergio Dos Santos and Benoit Fuentes are with Smart Impulse, 75019, Paris, France. E-mail: {simon.henriet, umut.simsekli, gael.richard}@telecom-paris.fr, {simon.henriet, sergio.dos-santos, benoit.fuentes}@smart-impulse.com.

The objective of NILM is to recover power consumptions of individual devices (and/or group of devices) ($p_k(t)$) from the total power consumption measured from a building ($p_{\text{total}}(t)$) with:

$$p_{\text{total}}(t) = \sum_{k=1}^{K} p_k(t) \tag{1}$$

where $t \in [\![1, T]\!]$ is the time index of a voltage period and $k \in [\![1, K]\!]$ is the index of a particular device. Due to Kirchhoff's law, the same formula stands for the total current:

$$X_{\text{total}}(n, t) = \sum_{k=1}^{K} X_k(n, t) \tag{2}$$

where $n \in [\![1, N]\!]$ corresponds to the sampling index within a period t. Considering a pure periodical Voltage ($U(n, t) = u_0(n)$), the power and current are linked by:

$$p_k(t) \overset{def}{=} \frac{1}{N} \sum_{n=1}^{N} X_k(n, t) u_0(n), \quad \forall k. \tag{3}$$

In a previous work, we showed that individual current matrices can be accurately approximated by low rank matrices [6]. As an illustration, let us assume that X_k can be approximated as a rank-one matrix:

$$X_k(n, t) \approx S(n, k) A(k, t), \quad \forall k, \tag{4}$$

where, $S \in \mathbb{R}^{N \times K}$ is called the *signature* matrix and $A \in \mathbb{R}_+^{K \times T}$ is called the *activation*. Combining Equations (2) and (4) gives the *factorization* model:

$$X_{\text{total}}(n, t) \approx \sum_{k=1}^{K} S(n, k) A(k, t) \tag{5}$$

Using Equation (3), the total power $p_{\text{total}}(t) = \sum_k p_k(t)$ reads:

$$p_{\text{total}}(t) \approx \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} S(n, k) A(k, t) u_0(n) \tag{6}$$

$$= \sum_{k=1}^{K} A(k, t) \alpha(k) = \sum_{k=1}^{K} \hat{p}_k(t) \tag{7}$$

where $\alpha(k) = \sum_{n=1}^{N} S(n, k) u_0(n)$ is a normalization factor and $\hat{p}_k$ is the power estimation for device k.

The low-rank nature of the observation matrices motivated the use of matrix factorization techniques for NILM [6]. In this line of research, Nonnegative Matrix Factorization has been applied to NILM with a modified formulation due to a different kind of data (only access to total power consumption) [2], [3], [4]. However, these algorithms are based on power signals and thus are not able to benefit from the advantages brought by the current signals, such as incorporating high frequency information that is contained in the current waveforms.

In a recent study, Lange and Berges [5] proposed a Binary Matrix Factorization algorithm to disaggregate current signals. They used a real valued signature matrix for typical current waveforms of the devices and a binary matrix for their activations. Although their strategy obtains encouraging results on residential buildings (since in this case most of the devices induce binary consumptions), it also shows clear limitations in the NILM task for commercial buildings that are much larger than typical residential buildings, e.g. large offices, shopping malls, warehouses, hospitals. Indeed, the electric consumption in commercial buildings exhibits specific characteristics that need to be taken into account [7], [6]. For example, commercial buildings usually have equipments with smoothly varying load curves which depart from the binary hypothesis of [5]. On the other hand, while designing a NILM algorithm for large buildings, one needs to take into account the physical characteristics of the power consumption. First of all, the positivity of the power consumption is an obvious constraint for every *consuming* device. Secondly, as illustrated in [6], the 'few at a time' hypothesis should be considered, stating that the consumptions of only a few equipments can change at any time. This is an extension of the popular 'one at a time' hypothesis, which is only accurate for residential buildings [8]. Finally, due to the lack of individual device measurements, we preferred to rely on a fully unsupervised approach than to exploit pre-learned dictionaries.

Contributions: In this paper we develop a matrix factorization technique, called Independent-Variation Matrix Factorization (IVMF), that aims at addressing NILM in commercial buildings. The novelty of the method resides in the physically-inspired constraints and the regularization applied to the signature and activation matrices: (i) we set linear and quadratic inequalities constraints on the signature columns; (ii) we add a positivity constraint on the activations and a regularization over the time variation of the activation. While the proposed optimization problem accurately captures the underlying physical behavior of current signals, the proposed regularization function and the constraints render the problem highly non-trivial to be solved by using standard optimization algorithms. To solve the resulting constrained non-convex optimization problem, we develop an alternating minimization strategy involving dual optimization and a quasi-Newton algorithm. We finally investigate the behavior and performance of our algorithm compared to two related methods (Independent Component Analysis (ICA) and Semi Nonnegative Matrix Factorization (SNMF)) on a synthetic source separation problem and on a realistic NILM application for commercial buildings. We show that IVMF is particularly adapted to recover positive sources that exhibit a strong temporal dependency and sources whose variations are independent from each other. To the best of our knowledge this constitutes the first attempt to solve the NILM problem for commercial buildings using current measurements factorization.

II. Related work

In this section, we will briefly explain two related matrix factorization algorithms, SNMF and ICA. While these techniques can partially address our task, they cannot take advantage of all the properties of power consumption presented previously.

In SNMF, the observation matrix is approximated by the matrix product of two factors: $X \approx SA$, with a real valued S factor and a nonnegative A factor. SNMF can formally be written down as the following optimization problem [9]:

$$\begin{aligned} \underset{S,A}{\text{minimize}} \quad & \frac{1}{2}\|X - SA\|_{Fro}^2 \\ \text{subject to} \quad & A \geq 0. \end{aligned} \tag{8}$$

In practice, the activation rows exhibit high correlations which is contrary to the properties presented before. Another weakness is that the positivity constraints on A is not sufficient to ensure positivity of the estimated power consumption. As introduced in Equation (7), the normalizations $\alpha(k)$ also need to be positive.

1-D Total Variation (TV) regularization may be defined as the penalization of the absolute value of the variations of a signal [10]. For our activations, it reads: $\text{TV}(A) = \sum_k \sum_t |A(k,t+1) - A(k,t)|$. Seichepine et. al. have used this regularization on an NMF problem [11] to obtain piecewise constant activations. The problem reads:

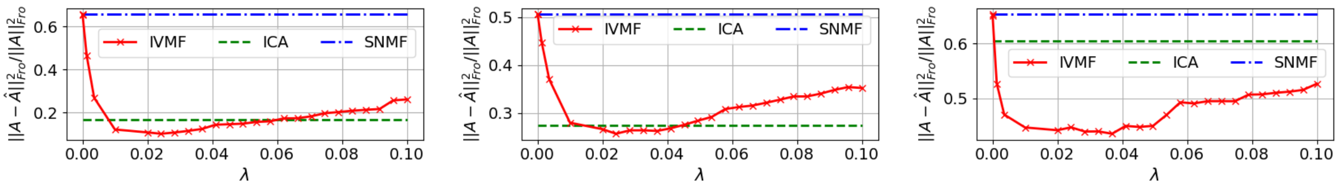
$$\begin{aligned} \underset{S,A}{\text{minimize}} \quad & \mathcal{D}(X, SA) + \lambda \text{TV}(WA) \\ \text{subject to} \quad & A \geq 0, S \geq 0. \end{aligned} \tag{9}$$

where $\mathcal{D}$ is an element-wise divergence (such as Kullback-Leibler) and W is a $k \times k$ diagonal matrix with values $W(k,k) = \|S_k\|_1$ that fix the inherent scaling ambiguity (S_k is the k^{th} column of S). Although this approach directly considers the variation of the activations, the positivity constraint on both the signature and activation matrices is not suitable for our problem.

ICA can also be viewed as a matrix factorization ($X = SA$) technique whose goal is to find activation rows (sources) as independent as possible. The FastICA implementation [12] can be expressed in our notation as:

$$\begin{aligned} \underset{S,A}{\text{maximize}} \quad & \|\text{E}\{G(A)\} - \mathbb{E}\{G(\nu)\}\|_2^2 \\ \text{subject to} \quad & X = SA \quad \text{and} \quad S^\top S = I. \end{aligned} \tag{10}$$

where X has been centered and whitened, I is the identity matrix, G is non linear function, $\mathbb{E}$ is the expectation, ν is a multivariate Gaussian variable with identity covariance matrix and E represents the mean (over the columns of a matrix). One important assumption of ICA is the time independence of the sources (i.e A_k our activation rows). It is then usual (see [13]) to apply ICA to a transformation of the data such that the assumption is fulfilled in this new domain. We will use this approach in the experiment section. The weakness of ICA in our problem resides in the positivity of the estimated consumption that would not be ensured.



(a) sparsity = 80 %

(b) sparsity = 0%

(c) noise = 1 + sparsity = 0%

Fig. 1. Influence of sparsity level, noise and regularization parameters: comparison of IVMF, ICA and SNMF.

III. Independent Variations Matrix Factorization

A. The proposed optimization problem

To overcome all these limitations, we extend (8) by introducing: (i) a specific regularization and a positivity constraint over the activations; (ii) linear and quadratic constraints on the signature matrix to ensure power estimation positivity and limit scaling ambiguities. The IVMF problem is given by:

$$\begin{aligned} \underset{S,A}{\text{minimize}} \quad & \frac{1}{2}\|X - SA\|^2_{Fro} + \lambda\mathcal{G}(A) \qquad (11) \\ \text{subject to} \quad & \|S_k\|^2_2 \leq 1, \qquad \forall k \in [\![1, K]\!], \\ & S_k^\top u_0 \geq \alpha_0, \qquad \forall k \in [\![1, K]\!], \\ & A \geq 0, \end{aligned}$$

where $\mathcal{G}(A) = \sum_k \sum_t G(A(k, t+1) - A(k,t))$ and G is a non-linear scalar function. Inspired by total variation and its absolute value, we chose G as a 'smooth absolute value': $G(x) = \sqrt{x^2 + \epsilon}$, where ϵ is a small positive constant. $\lambda > 0$ is the regularization parameter. The quadratic constraint on the columns of S enables us to fix the inherent scaling ambiguity in such factorization problems. Indeed, without this constraint, multiplying S by a scalar and dividing A by the same value would artificially decrease the penalization without changing the 'data fitting' term or the shape of the solution. The linear constraint on S_k enables us to ensure the positivity of the power estimation (with u_0 being a voltage measurement and α_0 being a fixed parameter). We provide a probabilistic interpretation of our problem in the supplementary document.

B. Parameter Estimation

We propose to use an alternating minimization strategy where S and A are updated iteratively [14]. Updating S results in a least squares problem under quadratic and linear constraints [15], [16] whereas updating A is a least squares optimization problem with a smooth penalization on variations and a positivity constraint. Due to space limitations, we provide the derivations of the update rules in the supplementary material and summarize the overall IVMF algorithm in Algorithm 1. Our algorithm uses Scipy's wrapper for L-BFGS-B ([17], [18]). The alternating optimization algorithm guarantees to decrease iteratively the cost function but since the problem is not convex in both variables, the reached solution can only be a local optimum. In such a situation, the solution strongly depends on the initialization of the algorithm, which will be discussed in Section IV. It can finally be noticed that the computational complexity of one update of S and one update of A are respectively $\mathcal{O}(NK^2)$ and $\mathcal{O}(NKT)$. The complexity of the algorithm is thus driven by the update of A (since $T \gg K$) but remains linear in time.

Algorithm 1: IVMF with alternating minimization

input: X, u_0, λ, I_{max} the maximal number of iterations

Initialize $S^{(0)}$, $A^{(0)}$, $\mu^{(0)}$ and $\nu^{(0)}$

for $i = 0$ *to* I_{max} **do**

  /* A-update, see Appendix B-B */

  $A^{(i+1)} = \text{L-BFGS-B}(A^{(i)}, \mathcal{G}, \nabla_A \mathcal{G})$

  /* S-update, see Appendix B-A */

  $\mu^{(i+1)}, \nu^{(i+1)} = \text{L-BFGS-B}(\mu^{(i)}, \nu^{(i)}, \mathcal{D}, \nabla_{\mu_k, \nu_k}\mathcal{D})$

  /* Solve the linear system */

  $S^{(i+1)} = (XA^{(i+1)^\top} + \frac{1}{2}u_0\nu^{(i+1)^\top})(A^{(i+1)}A^{(i+1)^\top} + 2M(\mu^{(i+1)}))^{-1}$

IV. Experiments

In this section we evaluate the performance of IVMF using both synthetic examples and a NILM dataset. We evaluate intrinsic properties of the algorithm such as the sensitivity to initialization and also show its capability to perform independent source separation. As natural competitors, we compare our method to ICA and SNMF.

A. Experiments on Synthetic Data

In this section, we simulate data that follows the factorization model: $X = SA + \Gamma$, where A is the matrix of independent and positive activations, S is a real mixing matrix and Γ is an additive white noise. The simulation of A is as follows: (i) draw independent and identically distributed Laplace variable corresponding to the variations of A, (ii) [optional] set a certain percentage of the values to zero to add true sparsity, (iii) cumulate along the time dimension to calculate A, (iv) add to all elements of A its minimum value to ensure positivity. The entries of S are simulated independently from a normal distribution and then adjusted so that the ℓ_2-norm of each column equals 1. Γ is also simulated from a normal distribution. Finally, X is deterministically computed from S, A and Γ.

In a first experiment, we fix the noise variance to 0, set the sparsity to 0.8 (80% of the variations of A are zero), simulate 100 datasets and then run ICA, SNMF and IVMF. We repeat this procedure with a sparsity of 0 (meaning that no variation is forced to be 0). Finally, we set the noise variance to 1.

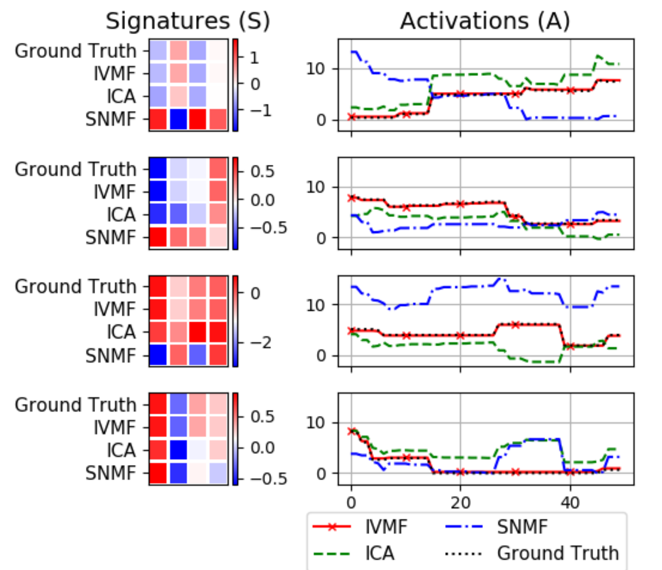


Fig. 2. Example of a *sparse* simulation: the result of IVMF, ICA and SNMF are presented component by component. The columns of S presented as a heatmap (left) and the rows of A as time series (right).

The size of the simulation is $K = N = 4$ and $T = 50$. For IVMF, we vary the λ parameter from 0 to 0.1. For ICA, we apply it to a transformation of the data that computes the time variations: $\tilde{X}(n,t) = X(n,t) - X(n,t-1)$. The optimal S_{ica} is then used to recover the optimal A_{ica} from the true data X with $A_{ica} = S_{ica}^{-1}X$. To estimate the performance of the algorithms we use the estimation error on A defined as: $E_A = \|A - \hat{A}\|_{Fro}^2 / \|A\|_{Fro}^2$, where $\hat{A}$ is the estimated matrix. To fix the multiplicative ambiguity we normalize the output of the different algorithms so that the ℓ_2-norm of the columns of S equals 1. We also reorder the rows of $\hat{A}$ by using a greedy algorithm to match each row of A with the closest row in $\hat{A}$.

Figure 1 presents the average error over the 100 simulations for the 3 generative procedures. It first shows that IVMF consistently outperforms SNMF. IVMF's performance also improve as λ decreases until an optimal value. For lower values the performance deteriorates until reaching those of SNMF when $\lambda \approx 0$. Figure 1(a) and 1(b) present the result of simulations without noise and IVMF outperforms ICA for λ values in the range $[0.01, 0.05]$. We may explain this by the fact that ICA does not take advantage of the positivity property of the signal and also that ICA enforces perfect decorrelation of the sources which in practice is not the case. We can notice that IVMF performs better on the sparse example than on the smooth one. Figure 1(c) shows that IVMF is more robust to additive noise than ICA. For the noisy setting IVMF's performance seems to be less sensitive to the choice of the regularization parameter λ.

The obtained decomposition of a sparse simulated example is illustrated on figure 2.

B. Experiment on the SHED dataset

In this second experiment, we consider a NILM dataset called SHED [6]. This dataset contains simulated current measurements for 8 commercial buildings and the corresponding individual power consumptions. These simulations have shown to be realistic for the NILM task [6]. We can also notice that the simulation process involves a factorization structure of the current data, which is more complex. We use a sample of this dataset corresponding to 1 day of current waveforms averaged at 5 minutes. The number of individual devices or categories ranges from 5 to 10 depending on the building. In this dataset, the voltage is periodic: $\forall\, t,\ U(n,t) = U(n,1)$.

We define the power error in the same manner as for the error on A in the previous section: $E_{\text{power}} = \sum_k \|p_k - \hat{p}_k\|_2^2 / \sum_k \|p_k\|_2^2$, where $p_k \in \mathbb{R}_+^T$ is the ground truth (given by the dataset) and $\hat{p}_k \in \mathbb{R}_+^T$ is the corresponding estimation.

The experiment consists in running the ICA, SNMF and IVMF algorithms with 50 different initializations. We chose to fix the value of K to 15, to make sure we have more sub-component than the true number of categories in the building. The dimensions of the problem are thus $N = 100$, $K = 15$ and $T = 288$. For IVMF, we chose to fix manually the λ parameter to a low value of 0.01 and α_0 to 0. The power error results given in Table I show that IVMF presents the best performance on all buildings but one (building 3) where no method seems to perform well (due to confusions between 2 categories). We also provide in the supplementary material some decomposition examples that illustrate the nice properties of the activation estimated by IVMF (positivity, time-variations sparsity, ability to detect the absence of activation when the number of components is greater than the number of categories to recover).

TABLE I
PERFORMANCE COMPARISON BETWEEN ICA, SNMF AND IVMF ON THE NILM TASK USING THE SHED DATASET.

Building	ICA		SNMF		IVMF	
	Avg.	Best	Avg.	Best	Avg.	Best
1	0.39	0.19	0.50	0.26	0.26	**0.12**
2	0.53	0.48	0.67	0.46	0.50	**0.39**
3	0.52	0.47	0.78	**0.41**	0.76	0.57
4	0.65	0.48	0.81	0.52	0.70	**0.31**
5	0.53	0.44	0.77	0.45	0.70	**0.43**
6	0.68	0.53	0.65	0.46	0.60	**0.45**
7	0.24	0.20	0.46	0.26	0.38	**0.19**
8	0.54	0.45	0.67	0.37	0.61	**0.32**

Avg (resp. Best) refers to the average (resp. minimal) error over 50 different runs of the algorithm.

V. CONCLUSION

We proposed a novel optimization problem and an associated optimization algorithm for the separation of time dependent sources whose variations exhibit independence. The proposed approach extends SNMF by introducing a physically-inspired regularizer over the variation of the sources and linear and quadratic constraints on the signature matrix. We have shown that our approach can achieve independent source separation in a NILM setting and outperforms its natural competitors on both toy example and realistic NILM data.

References

[1] C. Klemenjak and P. Goldsborough, "Non-intrusive load monitoring: A review and outlook," *arXiv preprint arXiv:1610.01191*, 2016.

[2] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," in *Advances in Neural Information Processing Systems*, 2010, pp. 1153–1161.

[3] A. Rahimpour, H. Qi, D. Fugate, and T. Kuruganti, "Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint," *IEEE Transactions on Power Systems*, 2017.

[4] S. Singh and A. Majumdar, "Deep sparse coding for non–intrusive load monitoring," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4669–4678, 2018.

[5] H. Lange and M. Bergés, "Bolt: Energy disaggregation by online binary matrix factorization of current waveforms," in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 2016, pp. 11–20.

[6] S. Henriet, U. Simsekli, B. Fuentes, and G. Richard, "A generative model for non-intrusive load monitoring in commercial buildings," *Energy and Buildings*, vol. 177, pp. 268 – 278, 2018.

[7] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers, "A comparison of non-intrusive load monitoring methods for commercial and residential buildings," *arXiv preprint arXiv:1408.6595*, 2014.

[8] G. W. Hart, "Nonintrusive appliance load monitoring," in *Proceedings of the IEEE*, 1992.

[9] C. H. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 1, pp. 45–55, 2010.

[10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[11] N. Seichepine, S. Essid, C. Févotte, and O. Cappe, "Piecewise constant nonnegative matrix factorization," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 6721–6725.

[12] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4, pp. 411–430, 2000.

[13] F. Feng and M. Kowalski, "Revisiting sparse ica from a synthesis point of view: Blind source separation for over and underdetermined mixtures," *Signal Processing*, vol. 152, pp. 165–177, 2018.

[14] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[15] R. Grosse, R. Raina, H. Kwong, and A. Ng, "Shift-invariant sparse coding for audio classification," in *Proceedings of the Twenty-third Conference on Uncertainty in Artificial Intelligence (UAI07)*, 2007.

[16] M. Jas, T. D. La Tour, U. Simsekli, and A. Gramfort, "Learning the morphology of brain signals using alpha-stable convolutional sparse coding," in *Advances in Neural Information Processing Systems*, 2017, pp. 1099–1108.

[17] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001. [Online]. Available: http://www.scipy.org/

[18] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.