# LOW RANK ACTIVATIONS FOR TENSOR-BASED CONVOLUTIONAL SPARSE CODING

*Pierre Humbert* [◇], *Student Member, IEEE* *Julien Audiffren*[◇] *Laurent Oudre* [†] *Nicolas Vayatis* [◇]

[◇] CMLA, ENS Paris-Saclay, CNRS, Université Paris-Saclay, 94235 Cachan, France
[†] L2TI, Université Paris 13, 93430 Villetaneuse, France

## ABSTRACT

In this article, we propose to extend the classical Convolutional Sparse Coding model (CSC) to multivariate data by introducing a new tensor CSC model that enforces sparsity and low-rank constraint on the activations. The advantages of this model are threefold. First, by using tensor algebra, this model takes into account the underlying structure of the data. Second, this model allows for complex atoms but enforces fewer activations to decompose the data, resulting in an improved summary (dictionary) and a better reconstruction of the original multivariate signal. Third, the number of parameters to be estimated are greatly reduced by the low-rank constraint. We exhibit the associated optimization problem and propose a framework based on alternating optimization to solve it. Finally, we evaluate it on both synthetic and real data.

***Index Terms* –** Dictionary learning, sparse coding, tensor method, low rank activations.

## 1. INTRODUCTION

In recent years, dictionary learning and convolutional sparse coding techniques (CSC) have been successfully applied to a wide range of topics, [1, 2], image restoration [3], and signal processing [4]. The main idea behind these representations is to conjointly learn a dictionary containing the patterns observed in the signal, and sparse activations that encode the temporal or spatial locations where these patterns occur. However, while previous works in dictionary learning have mainly focused on the study of univariate signals or images [5], in many real settings, data are multivariate and therefore better encoded by a tensor structure [6].

In this article, we propose to extend the classical CSC model to multivariate data by introducing a new tensor CSC model, referred to as Kruskal Convolutional Sparse Coding (K-CSC), that enforces sparsity and low-rank constraint on the activations. The idea of enforcing low-rank constraints for CSC is not novel: Rigamonti et al. [7] and Sironi et al. [8] used the idea of separable filters for learning low-rank atoms in order to improve computational runtime. However, in several applicative contexts, the low-rank structure naturally appears in the activations rather that in the atoms/dictionary.

Figure 1 displays an example of two spectrograms obtained from a stereo music recording. Both spectrograms exhibit a low-rank structure, which is here transferred into the activations tensors rather than into the observed patterns. In other words, although the time-frequency atoms may be complex (and thus without a low-rank structure), the activations (i.e. the time/frequency/channel positions where these atoms appear) clearly exhibit a low-rank structure. Build upon these observations, K-CSC enforces low-rank and sparsity constraints on the activation tensors rather than on the atoms/dictionary. We provide a framework called Alternating K-CSC (AK-CSC), based on an alternating procedure, that permits to estimate the K-CSC model. Finally, we evaluate it experimentally using synthetic and real tensor data, illustrating the advantages of our method.

**Notations and preliminaries** We introduce the following notations. The symbol $\circ$ refers to the outer product. The symbol $\star_{1,\cdots,p}$ refers to the multidimensional discrete convolutional operator between scalar valued functions where the subscript indices are the dimension involved. When the signal is unidimensional, $\star_1$ reduces to the 1-D discrete convolutional operator denoted by $\star$.

We now introduce the main tensor algebra concepts for this paper. Please refer to [9] for a more in-depth introduction on the topic.
For any $\boldsymbol{\mathcal{X}} \in \mathbb{X} \triangleq \mathbb{R}^{n_1} \times \ldots \times \mathbb{R}^{n_p}$, there exist $R > 0$, and, $\boldsymbol{x}_r^{(i)} \in \mathbb{R}^{n_i}$, $1 \leq i \leq p$, $1 \leq r \leq R$, such that

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \boldsymbol{x}_r^{(1)} \circ \cdots \circ \boldsymbol{x}_r^{(p)} \; . \tag{1}$$

The smallest $R$ for which such decomposition exists is called the *Canonical Polyadic rank* of $\boldsymbol{\mathcal{X}}$ (CP-rank($\boldsymbol{\mathcal{X}}$) or rank($\boldsymbol{\mathcal{X}}$) for short), and in this case equation 1 is referred to as the CP decomposition of $\boldsymbol{\mathcal{X}}$. The *Kruskal operator* $[\![ \;\cdot\; ]\!]$ is then defined as

$$[\![ \boldsymbol{X}_1, \cdots, \boldsymbol{X}_p ]\!] \triangleq \sum_{r=1}^{R} \boldsymbol{x}_r^{(1)} \circ \cdots \circ \boldsymbol{x}_r^{(p)},$$

where $\boldsymbol{X}_i = \left[\boldsymbol{x}_1^{(i)}, \ldots, \boldsymbol{x}_R^{(i)}\right] \in \mathbb{R}^{n_i \times R}$, $1 \leq i \leq p$.

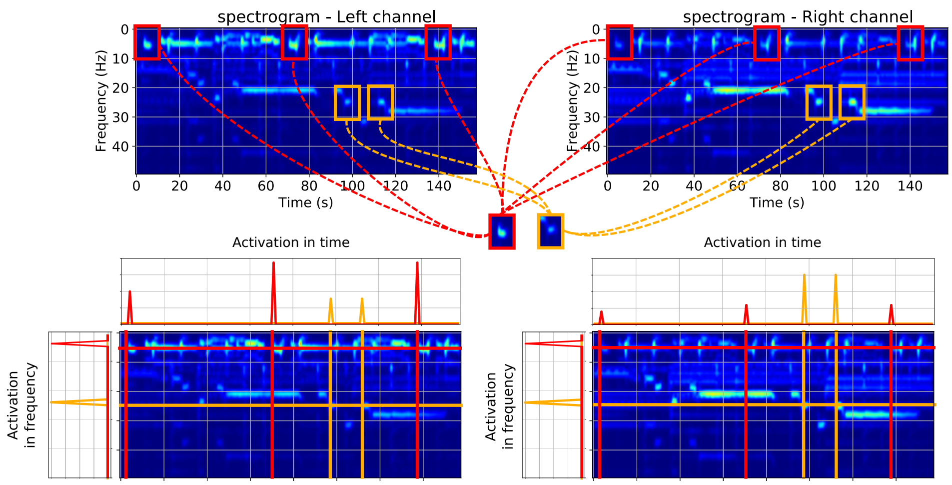



**Fig. 1**: On top, spectrograms of a stereo audio jazz signal. In red and orange, two atoms with their corresponding low-rank activations in the same color.

## 2. CONVOLUTIONAL SPARSE CODING WITH LOW-RANK ACTIVATIONS

### 2.1. Kruskal Convolutional Sparse Coding model

Let $\boldsymbol{\mathcal{Y}} \in \mathbb{Y} \triangleq \mathbb{R}^{n_1 \times \cdots \times n_p}$ be a multidimensional signal i.e a tensor of order $p > 0$, and $\boldsymbol{\mathcal{D}}_1, \cdots, \boldsymbol{\mathcal{D}}_K$ in $\mathbb{D} \triangleq \mathbb{R}^{w_1 \times \cdots \times w_p}$ a collection of $K$ multidimensional atoms such that $\forall i, 1 \leq w_i \leq n_i$. The ***Kruskal Convolutional Sparse Coding model*** (**K-CSC**) is defined as

$$\boldsymbol{\mathcal{Y}} = \sum_{k=1}^{K} \boldsymbol{\mathcal{D}}_k \star_{1,\cdots,p} \underbrace{\boldsymbol{\mathcal{Z}}_k}_{\text{CP-rank} \leq R} + \boldsymbol{\mathcal{E}} \,, \tag{2}$$

where a) $\forall 1 \leq k \leq K$, $\boldsymbol{\mathcal{Z}}_k \in \mathbb{Z} \triangleq \mathbb{R}^{m_1 \times \cdots \times m_p}$ (with $m_i = n_i - w_i + 1$) are sparse activation tensors with CP-rank lower than $R$, with $R$ small, and b) $\boldsymbol{\mathcal{E}} \in \mathbb{Y}$ is an additive (sub)gaussian noise, whose every component are independent and centered. Here, activations are sparse tensors which specify where atoms are placed in the input tensors of order $p$.

The advantages of this model are threefold. First, by using a tensor framework, this model takes into account the underlying structure of the data. Second, this models allows for complex atoms but enforces fewer activations to decompose the data, resulting in an improved summary (dictionary) and a better reconstruction of the original multivariate signal. Third, the number of parameters to be estimated are greatly reduced by the low-rank constraint.

With specific choices on the parameters or on the dimension values, the K-CSC model reduces to well-known CSC problems. It can therefore be seen as a generalization of several approaches in the litterature. For vector-valued atoms and signals ($p = 1$), our model reduces to the univariate Convolutional Dictionary Learning model (CDL) [10]. If $p > 1$ and $R = +\infty$ (i.e. no low-rank constraint), our model reduces to the multivariate CDL model also referred as multi-channel CDL [11]. If $p = 2$, $R = 1$ and $w_2 = 1$ (i.e. vector-valued atoms), our model reduces to the one from [7] with $R = 1$.

### 2.2. Model estimation

The estimation of $(\boldsymbol{\mathcal{D}}_k)_k$ and $(\boldsymbol{\mathcal{Z}}_k)_k$ of the **K-CSC** model (2) can be performed by solving the following optimization problem

$$\underset{(\boldsymbol{\mathcal{D}}_k) \in \mathcal{S}, (\boldsymbol{Z}_{k,\ell})}{\arg\min} \left\| \boldsymbol{\mathcal{Y}} - \sum_{k=1}^{K} \boldsymbol{\mathcal{D}}_k \star_{1,\cdots,p} [\![ \boldsymbol{Z}_{k,1}, \cdots, \boldsymbol{Z}_{k,p} ]\!] \right\|_F^2 \tag{3}$$

$$+ \sum_{\ell} \boldsymbol{\alpha}_\ell \sum_{k} \| \boldsymbol{Z}_{k,\ell} \|_1 \,, \tag{4}$$

where $\mathcal{S}$ is a set of constraints such that $\forall k$, $\|\boldsymbol{\mathcal{D}}_k\|_F \leq 1$ .

This problem relies on three important constraints.

- **The low-rank constraint** is embedded by the use of the *Kruskal operator* $[\![ \quad \cdot \quad ]\!]$.
- **The unit-ball constraints** on $\boldsymbol{\mathcal{D}}_k$ induced by the Frobenius norm permitting to avoid the scaling indeterminacy between the dictionary and the activations.
- **The sparsity contraints** on the activation tensors induced by the $\ell_1$ norms. Here, the sparsity constraint induces the sparsity of each element of the CP-decomposition for every activation tensor independently. Hence, the sparsity in each mode is controlled without the influence of the other modes i.e. the regularization (and not the objective function) is separable in each $(\boldsymbol{Z}_{k,\ell})_\ell$.

### 2.3. Solving the optimization problem

The minimization problem (3) is not convex due to the rank constraint. However, with respect to each $Z$ block

$([\boldsymbol{Z}_{k,i}, \cdots, \boldsymbol{Z}_{k,i}])_i$, and $([\boldsymbol{\mathcal{D}}_1, \cdots, \boldsymbol{\mathcal{D}}_K])$, the objective function is convex and the regularization is separable. Consequently, we use a block-coordinate strategy to minimize it 1) by freezing $\boldsymbol{\mathcal{D}}$ and all except one $Z$ block at a time ($\boldsymbol{\mathcal{Z}}$-step) 2) by freezing only the activation tensor ($\boldsymbol{\mathcal{D}}$-step). We recast each subproblem of the $\boldsymbol{\mathcal{Z}}$-step as a CSC problem which can therefore be solved with conventional algorithms. Each step of this framework called ***Alternating Kruskal Convolutional Sparse Coding*** (**AK-CSC**) is detailed in the following.

**Activations update, $\boldsymbol{\mathcal{Z}}$-step.** To solve the $\boldsymbol{\mathcal{Z}}$-step, we assume that the dictionary $\boldsymbol{\mathcal{D}}$ is fixed and we iteratively solve the problem where all mode except the $\ell$-th one of each activation tensor are constant, for $\ell$ varying between 1 and $p$. In other words, for each value of $\ell$, we solve the problem

$$\underset{\boldsymbol{Z}_{1,\ell},\cdots,\boldsymbol{Z}_{K,\ell}}{\arg\min} \left\| \boldsymbol{\mathcal{Y}} - \sum_{k=1}^{K} \boldsymbol{\mathcal{D}}_k \star_{1,\cdots,p} [\![ \boldsymbol{Z}_{k,1}, \cdots, \boldsymbol{Z}_{k,p} ]\!] \right\|_F^2 \quad (5)$$

$$+ \alpha_\ell \sum_{k=1}^{K} \|\boldsymbol{Z}_{k,\ell}\|_1 \,. \quad (6)$$

The CP decomposition is known to be unique when it satisfies the Kruskal condition [12], but only up to permutation of the normalized factor matrices. The scaling indeterminacy can be handled by a ridge-based penalization (e.g. $\sum_{\ell=1}^{p} \beta_\ell \sum_{k=1}^{K} \|\boldsymbol{Z}_{k,\ell}\|_F^2$, $(\beta_1, \cdots, \beta_p) \succeq 0$) to equation 3 (see [13] and [14]). We omit it in the following equations for clarity. The minimizers up to a permutation are isolated equivalent minimizers, and thus do not negatively impact the optimization [13].

**Proposition 1.** The first term of the minimization problem (5) is equal to

$$\sum_{c=1}^{C} \left\| \tilde{\boldsymbol{\mathcal{Y}}}_{:,c} - \sum_{s=1}^{S} \tilde{\boldsymbol{\mathcal{D}}}_{s,:,c} \star \boldsymbol{z}_s^{(\ell)} \right\|_2^2 ,$$

where $\boldsymbol{z}_s^{(\ell)} = \left[z_{1,1}^{(\ell)}, \cdots, z_{1,R}^{(\ell)}, z_{2,1}^{(\ell)}, \cdots, z_{K,R}^{(\ell)}\right]$ and $\tilde{\boldsymbol{\mathcal{D}}}$ is a third order tensor in $\mathbb{R}^{S \times w_\ell \times C}$ with $C = \prod_{i=1; i\neq\ell}^{p} n_i$, $S = KR$ and with elements given by

$$\tilde{\boldsymbol{\mathcal{D}}}_{k;r,j_1,i_2\cdots,i_p} = \left( \boldsymbol{\mathcal{D}}_{k;j_1,:,\cdots,:} \star_{2,\cdots,p} z_{k,r}^{(2)} \circ \cdots \circ z_{k,r}^{(p)} \right)_{i_2,\cdots,i_p} .$$

From the previous proposition, it is now clear that in equation 5, each subproblem of the $\boldsymbol{\mathcal{Z}}$-step is *a CSC with multichannel dictionary filters and single-channel activation maps* [11]. This problem has received only limited attention for many more than three channels ($C$=3) while in our reformulation, $C$ is very large. Fortunately, Garcia et al. [5] proposed algorithms which turn out to be scalable regarding values of $C \gg 3$. In particular, solvers based on Fast Iterative Shrinkage-Thresholding (FISTA) have proven to be effective in this case.

**Dictionary update, $\boldsymbol{\mathcal{D}}$-step.** Given the $K$ activation tensors $(\boldsymbol{\mathcal{Z}}_k)_k$, the dictionary update aims at improving how the model reconstructs $\boldsymbol{\mathcal{Y}}$ by solving

$$\underset{\forall k, \boldsymbol{\mathcal{D}}_k \in \mathbb{D}, \quad \|\boldsymbol{\mathcal{D}}_k\|_F \leq 1}{\arg\min} \left\| \boldsymbol{\mathcal{Y}} - \sum_{k=1}^{K} \boldsymbol{\mathcal{D}}_k \star_{1,\cdots,p} [\![ \boldsymbol{Z}_{k,1}, \cdots, \boldsymbol{Z}_{k,p} ]\!] \right\|_F^2 . \quad (7)$$

This step presents no significant difference with existing methods. This problem is smooth and convex and can be solved using classical algorithms [4, 15, 16].

## 3. EXPERIMENTS

In this section we evaluate our tensor-based dictionary learning framework AK-CSC on both synthetic and real-world datasets. The $\boldsymbol{\mathcal{Z}}$-step and $\boldsymbol{\mathcal{D}}$-ste are solved with FISTA. We compare AK-CSC (with FISTA) to state-of-the-art dictionary learning algorithms based on FISTA and on the Alternating Direction Method of Multipliers (ADMM) all implemented in `Sporco` [17], a `Python` package for convolutional sparse representations. Presented results correspond to the tensor regression task for which we report the Root Mean Square Error (RMSE).

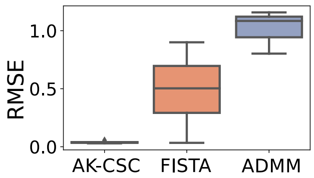
All experiments are made on a personal laptop through `Linux/Ubuntu` with 4-core 2.5GHz Intel CPUs using `Tensorly` [18] (for tensor algebra in `Python`), `Sporco` [17] and standard python libraries.

### 3.1. Synthetic data

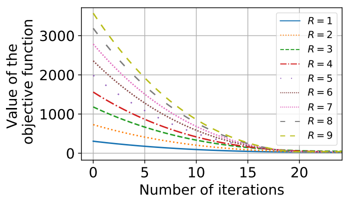
This first experiment aims at comparing the proposed AK-CSC framework to standard approaches and to study the influence of the hyperparameters of the method. To that end, we used synthetic data for which the ground truth is available.

**Data.** We propose to compare the performances of the methods on 10 totally independent synthetic input signals. For each one of them, we consider an input signal $\boldsymbol{\mathcal{Y}}$ in $\mathbb{R}^{10\times10\times10}$ constructed as follows. First, the $K = 5$ atoms of size $\mathbb{R}^{3\times3\times3}$ are drawn according to a standard Gaussian distribution. The sparse activations associated to each atom are tensor in $\mathbb{R}^{8\times8\times8}$ of maximal CP-rank $R^* = 4$: they are drawn from a Bernoulli-Uniform distribution with Bernoulli parameter equal to $0.1$, and range of values in $[-2, 2]$. Finally, we generate the input tensor according to the model (2).

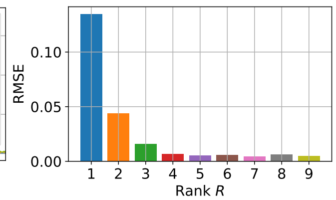
**Results.** For all simulations, the true dictionary is given and only the CSC task is evaluated. First, we compare our method with CSC algorithms based on FISTA and ADMM when the true rank $R^* = 4$ is supposed to be known. To highlight the capacity of our method to learn better representations with fewer activations (more sparsity), we first set $\alpha_1 = \alpha_2 = \alpha_3 = 0.015$ in order to reach the true number of non-zero coefficients (around 10%). Then, we adjust the regularization for the two other methods to obtain the same level of sparsity as the one reach by AK-CSC. Figure




(a)




(b)




(c)

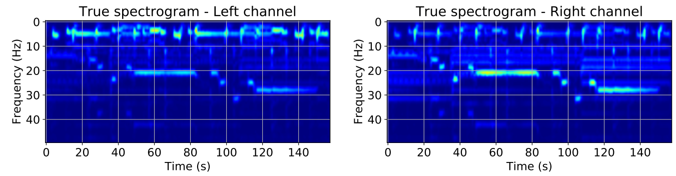
**Fig. 2**: (a) Boxplots of the RMSE across 10 independent input signals for AK-CSC (with FISTA), and the two state-of-the-arts solvers based on FISTA and ADMM. (b, c) Influence of the rank on the algorithms.

2(a) displays boxplots of the RMSE across the 10 independent signals. As expected, the RMSE is greater than FISTA and ADMM. In average the RMSE is $0.02(\pm0.01)$ for AK-CSC, $0.48(\pm0.31)$) for FISTA, and $1.01(\pm0.15)$) for ADMM.

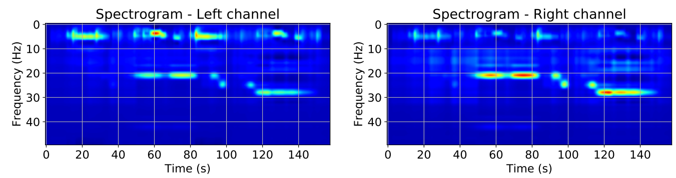
In a second time, we study the influence of the rank parameter $R$ on the reconstruction performances. We illustrate the convergence of AK-CSC by plotting in Figure 2 (b) the loss from equation (3) as a function of the rank parameter $R$. The convergence is achieved around 20 iterations regardless of the rank (full loop of the alternating procedure). Furthermore, as depicted in Figure 2 (c), an over estimation of the rank does not increase nor decrease the reconstruction error. Finally, note that, the error after convergence always decreases until we reach the true rank $R_*$.

### 3.2. Real data

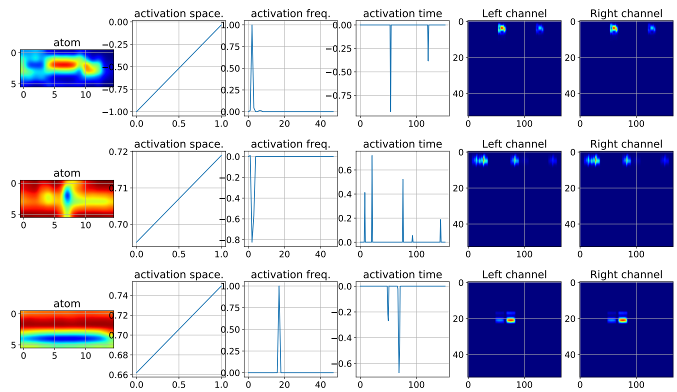
**Audio signal – 3rd order tensor with low-rank structure.** In this experiment, we consider a stereo signal of 6 seconds recorded at 8000Hz. For each signal (one per channel), we compute the short-time Fourier transform in order to obtain its spectrogram. Window size is set to 512 samples with 50% overlap : only the first 50 bins have been conserved ($0-781.25$ Hz). The data consists in a third order tensor $\boldsymbol{\mathcal{Y}}$ in $\mathbb{R}^{2\times50\times158}$. We use AK-CSC to reconstruct $\boldsymbol{\mathcal{Y}}$ using $K = 15$ frequency-time atoms of size $(w_1, w_2, w_3) = (1, 6, 15)$ (i.e. atoms of 0.5 seconds with 93.75 Hz bandwith). The maximal CP-rank of each associated activation is set to $R = 1$. Hyperparameters are set in order to bring enough sparsity while not deteriorating the reconstruction. Atoms and associated activations returned by AK-CSC are displayed Figure 3. Since in this audio signal the different instruments play at different frequency we can isolate them: the first two atoms of the Figure 3(c) correspond to the drum and the third one to the guitar. On the 15 possible atoms, AK-CSC only uses 10 of them to reconstruct the signal (setting the others to zero). Hence, our algorithm returns an improved dictionary with less atoms.




(a) Spectrograms of the true signal.




(b) Spectrograms of the resulting reconstruction when $R = 1$.




(c) Atoms with the activation in each mode.

**Fig. 3**: Results on an stereo audio signal when $R = 1$. Three relevant atoms are displayed in (c). For each of them, from left to right, the atom, its activations in Stereo 1 or 2, in frequency, and in time. The last two figures displayed the spectrograms when only one atom is considered.

## 4. CONCLUSION

In this paper, we introduced K-CSC, a new multivariate convolutional sparse coding model that enforces both sparsity and low-rankness on the activations tensors. We enumerate the advantages of this model and motivate it with an example on an audio signal. We provided a framework based on an alternate minimization called AK-CSC. We show that each subproblem is a particular CSC problem on which we can use off-the-shelf algorithms. Finally, we showed that when using a FISTA solver, AK-CSC achieves good performances on both synthetic and real data.

## 5. REFERENCES


[1] Julien Mairal, Jean Ponce, Guillermo Sapiro, Andrew Zisserman, and Francis R Bach, "Supervised dictionary learning," in *Advances in neural information processing systems*, 2009, pp. 1033–1040.

[2] Ke Huang and Selin Aviyente, "Sparse representation for signal classification," in *Advances in neural information processing systems*, 2007, pp. 609–616.

[3] Michal Aharon, Michael Elad, and Alfred Bruckstein, "$rmk$-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[4] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online learning for matrix factorization and sparse coding," *Journal of Machine Learning Research*, vol. 11, no. Jan, pp. 19–60, 2010.

[5] Cristina Garcia-Cardona and Brendt Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[6] Hua Zhou, Lexin Li, and Hongtu Zhu, "Tensor regression with applications in neuroimaging data analysis," *Journal of the American Statistical Association*, vol. 108, no. 502, pp. 540–552, 2013.

[7] Roberto Rigamonti, Amos Sironi, Vincent Lepetit, and Pascal Fua, "Learning separable filters," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2754–2761.

[8] Amos Sironi, Bugra Tekin, Roberto Rigamonti, Vincent Lepetit, and Pascal Fua, "Learning separable filters," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 1, pp. 94–106, 2014.

[9] Tamara G Kolda and Brett W Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[10] Brendt Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 301–315, 2016.

[11] Brendt Wohlberg, "Convolutional sparse representation of color images," in *2016 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. IEEE, 2016, pp. 57–60.

[12] Joseph B Kruskal, "Rank, decomposition, and uniqueness for 3-way and n-way arrays," *Multiway data analysis*, pp. 7–18, 1989.

[13] Evrim Acar, Daniel M Dunlavy, and Tamara G Kolda, "A scalable optimization approach for fitting canonical tensor decompositions," *Journal of Chemometrics*, vol. 25, no. 2, pp. 67–86, 2011.

[14] Pentti Paatero, "A weighted non-negative least squares algorithm for three-way 'parafac'factor analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 223–242, 1997.

[15] Florence Yellin, Benjamin D Haeffele, and René Vidal, "Blood cell detection and counting in holographic lens-free imaging by convolutional sparse dictionary learning and coding," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 650–653.

[16] Rakesh Chalasani, Jose C Principe, and Naveen Ramakrishnan, "A fast proximal method for convolutional sparse coding," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2013, pp. 1–5.

[17] Brendt Wohlberg, "Sporco: A python package for standard and convolutional sparse representations," in *Proceedings of the 15th Python in Science Conference, Austin, TX, USA*, 2017, pp. 1–8.

[18] Jean Kossaifi, Yannis Panagakis, Anima Anandkumar, and Maja Pantic, "Tensorly: Tensor learning in python," *Journal of Machine Learning Research*, vol. 20, no. 26, pp. 1–6, 2019.