From Convolutional Sparse Coding To *-NMF Factorization of Time-Frequency Coefficients

Jean-Baptiste Malagnoux, Matthieu Kowalski

FROM CONVOLUTIONAL SPARSE CODING TO *-NMF FACTORIZATION OF TIME-FREQUENCY COEFFICIENTS

Jean-Baptiste Malagnoux, Matthieu Kowalski

Inria, Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numériques
Gif-sur-Yvette, France

ABSTRACT

Convolutional Dictionary Learning (CDL) is a dictionary learning technique exploiting the translation invariance of elementary signals. In the time-frequency domain, the repetition of elementary frequency patterns can be exploited through the "nonnegative matrix factorization" (NMF) decompositions and extensions, such as semi or complex-NMF, of the spectrogram. We study the links between these two approaches here, and we show in particular that a signal which admits a Convolutive Sparse Coding decomposition admits time-frequency synthesis coefficients that can be decomposed in semi-NMF or complex-NMF. The different approaches are then compared experimentally on synthetic signals.

Index Terms— Convolutional Sparse Coding, Convolutional Dictionary Learning, Time-Frequency synthesis, Non-negative Matrix Factorization

1. INTRODUCTION

Sparse coding (SC) is a widely used unsupervised learning technique for signal representation [1]. This method decomposes a signal into a linear combination of a few elements, or atoms, of a pre-established dictionary, where most of the coefficients are zero. Initially, dictionaries were based on waveforms such as wavelets or time-frequency transforms [2]. However, using predefined dictionaries limits the ability of the SC to model more complex signals.

One solution is to use union of dictionaries [3, 4]. For example, time-frequency dictionaries with different window sizes can be combined to model unknown temporal and spectral structures while limiting Heisenberg's uncertainty principle. This allows for an increase in the diversity of dictionary elements and better represent the underlying structures of the signals. Although [5] proposes to learn the best possible combination of dictionaries, this method still requires a manual selection of the base dictionaries and does not allow optimizing the dictionary elements directly automatically.

This problem can be solved by adopting a "data-driven" approach, which automatically optimizes the dictionary elements from training data [6, 7]. However, dictionary learning with Sparse Coding requires a trade-off between representation quality and dictionary complexity. Indeed, a too-small dictionary will not make it possible to adequately represent the signals, while a too-large dictionary can lead to overfitting the data. Therefore, finding an optimal dictionary is an important problem in signal processing.

Convolutional Dictionary Learning (CDL), allied to Convolutional Sparse Coding (CSC), is a dictionary learning method that has been proposed in the context of music audio signals [8]. This method consists of learning a dictionary of (linear, time-invariant) filters. CDL has been successfully applied in various applications, such as automatic transcription [9] or biomedical signal analysis [10].

Another popular approach to signal analysis is non-negative matrix factorization (NMF) [11]. In the context of temporal signals, particularly music, the NMF approach consists of decomposing the spectrogram into a linear combination of spectra (of amplitude or power) and temporal profiles [12]. NMF is commonly used for audio signal analysis [13] and source separation [14]. Since NMF operates on the squared modulus of the spectrogram, various alternatives have been suggested to handle the phase, such as semi-NMF [15] or complex-NMF [16, 17], which directly operate in the time-frequency domain. Another option is the Low-Rank Time-Frequency Synthesis (LRTFS) [18], which utilizes Itakura-Saitao-based NMF as a prior for the time-frequency coefficients, enabling the direct synthesis of the signal in the time domain.

Contributions and organization of the article: After a brief reminder of the CDL and the NMF factorization of the Short Time Fourier transform (STFT) coefficients in Section 2, Theorem 1 establishes the equivalence, from a mathematical point of view, between the CDL methods and the semi/complex-NMF of synthesis coefficients of a STFT at maximum redundancy in time in Section 3. More particularly, we show that a signal constructed as the convolution of atoms and a sparse mapping can always be synthesized from factorizable time-frequency coefficients in the form of semi/complex-NMF of rank 1. We then demonstrate the equivalence between the optimization problems of the two approaches. Thus, we present the methods of resolution of these problems. Section 4 presents the experimental results of applying the CSC and NMF methods on an audio piano signal.

2. STATE OF THE ART

2.1. Convolutional Dictionnary Learning

Let $\mathbf{x} \in \mathbb{R}^{T_x}$, for all $k \in [\![1, K]\!]$, $\mathbf{d}_k \in \mathbb{R}^{T_d}$, $\mathbf{z}_k \in \mathbb{R}^{T_z}$. The linear convolutional model of order K of $\mathbf{x}$ writes [19]:

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{d}_k * \mathbf{z}_k + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathbb{R}^{T_x}$ is an assumed Gaussian white noise. The dictionary $\{\mathbf{d}_k\}$ comprises elementary atoms $\mathbf{d}_k$, which are activated at the different positions encoded in $\mathbf{z}_k$. These positions are assumed to be sparse, with each atom only appearing at a few locations. Thus, the

This work was supported by the French National Agency for Research through the BMWs project (ANR-20-CE45-0018).

estimation of the dictionary $\{\mathbf{d}_k\}$ and the coefficients $\{\mathbf{z}_k\}$ can be done using the CDL [20]:

$$\min_{\{\mathbf{z}_k\},\{\mathbf{d}_k\}} \frac{1}{2}\|\mathbf{x} - \sum_{k=1}^{K} \mathbf{d}_k * \mathbf{z}_k\|^2 + \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \quad \text{s.t. } \|\mathbf{d}_k\| \leq 1 \quad \forall k. \tag{2}$$

The CDL, therefore, amounts to solving a problem of convolutional sparse coding (estimating activations $\{\mathbf{z}_k\}$) and estimating the dictionary of filters $\{\mathbf{d}_k\}$. Several algorithms have been proposed, such as FastCSC [21] or distributed algorithms [19].

2.2. Time-Frequency decomposition and NMF

2.2.1. Gabor transform

Let $\mathbf{g} \in \mathbb{R}^{T_g}$ be a window. The STFT, or Gabor analysis coefficients, of a signal $\mathbf{x} \in \mathbb{R}^{T_x}$ is given for all $\tau \in [\![0, N-1]\!]$, $\nu \in [\![0, M-1]\!]$ by:

$$X[\tau,\nu] = \sum_{t=1}^{T_x} x[t]\bar{g}[t-a\tau]e^{-\frac{i2\pi\nu(t-a\tau)}{M}} = \langle \mathbf{x}, \varphi_{\tau,\nu}\rangle \tag{3}$$

with $\varphi_{\tau,\nu}[t] = g[t-a\tau]e^{\frac{i2\pi\nu(t-a\tau)}{M}}$. The parameter $a > 0$ is the size of the *jump* between two windows, such as $aT_g \leq N$, and $M \leq T_g$ is the number of points for the FFT. When the dictionary $\varphi_{\tau,\nu}$ forms a frame, there is an infinity of *synthesis coefficients* $\alpha[\tau,\nu]$ such that:

$$x[t] = \sum_{\tau=0}^{N-1}\sum_{\nu=0}^{M-1} \alpha[\tau,\nu]\varphi_{\tau,\nu}[t]\ . \tag{4}$$

We denote by Φ such a *synthesis operator*, that is, such that

$$\mathbf{x} = \Phi(\boldsymbol{\alpha})\ , \tag{5}$$

and we denote its adjoint the *analysis operator* Φ^* such that

$$\mathbf{X} = \Phi^*(\mathbf{x})\ ,\ \text{with } X[\tau,\nu] = \langle \mathbf{x}, \varphi_{\tau,\nu}\rangle\ . \tag{6}$$

2.2.2. *-NMF

Decomposition techniques based on non-negative matrix factorization (NMF) of time-frequency representations make it possible to exploit particular signal structures by identifying which frequency patterns are repeated over time. The first NMFs were applied to the amplitude ($p = 1$) or power ($p = 2$) spectrogram of the signal, ie:

$$V[\tau,\nu] = |X[\tau,\nu]|^p\ . \tag{7}$$

NMF looks for a decomposition of the form:

$$\mathbf{V} \simeq \mathbf{W}\mathbf{Z}^T\ ,\ \text{s.t. } \mathbf{W} \in \mathbb{R}_+^{MK}, \mathbf{Z} \in \mathbb{R}_+^{NK} \tag{8}$$

To better take into account the phase of the Gabor transform, other decompositions have been proposed, such as the semi-NMF [15]:

$$\min_{\mathbf{W}\in\mathbb{C}^{MK},\mathbf{Z}\geq 0} \|\mathbf{X} - \mathbf{W}\mathbf{Z}^T\|^2 \tag{9}$$

or, still with $\mathbf{W} \in \mathbb{C}^{M,K}$ and $\mathbf{Z} \geq 0$, the complex-NMF [16, 17] decomposes the Gabor transform in the form:

$$X[\tau,\nu] = \sum_{k=1}^{K} \mathbf{W}[\nu,k]\mathbf{Z}[\tau,k]e^{i\theta_k[\tau,\nu]}\ . \tag{10}$$

where $\theta_k[\tau,\nu] \in]-\pi,\pi]$ is a global phase term. Finally, the "Low-Rank Time-Frequency Synthesis" (LRTFS) [18] approach makes it possible to use the NMF with Itakura-Saito divergence (IS-NMF) [12] as a "prior" on the synthesis time-frequency coefficients: the global phase is preserved to reconstruct the signal, although the phase of each of the components is artificially reconstructed from the global phase.

3. CDL AND TIME-FREQUENCY SYNTHESIS

This section shows that synthesizing a signal using sparse convolutional dictionary learning or using NMF in the time-frequency domain is equivalent. The following theorem states this result for a signal with a single convolutional component.

Theorem 1. *Let $\mathbf{x} \in \mathbb{R}^{T_x}$. Let $\{\varphi_{\tau,\nu}\}_{\tau,\nu}$ be a Gabor dictionary as defined in Section 2.2.1 with $a = 1$ and $M \geq T_g$. Then, there is a filter $\mathbf{d} \in \mathbb{R}^{T_d}$, supp$(d) \subset$ supp(g), such that*

$$x[t] = (d * z)[t]\ ,\ \mathbf{z} \in \mathbb{R}^{T_z}$$

iff

$$\forall(\tau,\nu)\ \alpha[\tau,\nu] = \hat{w}(\nu)z(\tau)$$

with

$$\hat{w}(\nu) = \sum_{t=0}^{T_g-1} \frac{d[t]}{g[t]} e^{-\frac{i2\pi\nu t}{M}}$$

Proof. We set $w[t] = d[t]/g[t]$. As supp$(d) \subset$ supp(g), we have

$$x[t] = (z*d)[t] = \sum_{\tau=0}^{T_z-1} z[\tau]g[t-\tau]w[t-\tau] \tag{11}$$

$$= \sum_{\tau=0}^{T_z-1} z(\tau)g[t-\tau] \sum_{\nu=0}^{M-1} \hat{w}(\nu)e^{i\frac{2\pi\nu(t-\tau)}{M}} \tag{12}$$

$$= \sum_{\tau=0}^{T_z-1}\sum_{\nu=0}^{M-1} \hat{w}(\nu)z[\tau]g[t-\tau]e^{i\frac{2\pi\nu(t-\tau)}{M}} \tag{13}$$

$$= \sum_{\tau=0}^{T_z-1}\sum_{\nu=0}^{M-1} \alpha[\tau,\nu]g[t-\tau]e^{i\frac{2\pi\nu(t-\tau)}{M}} \tag{14}$$

∎

As a consequence of this theorem, the following corollary shows the equivalence between the CDL and the factorization of the Gabor coefficients at the synthesis

Corollary 1. *Let $\mathbf{x} \in \mathbb{R}^{T_x}$. Let $\{\varphi_{\tau,\nu}\}_{\tau,\nu}$ be a Gabor dictionary as defined in Section 2.2.1 with $a = 1$ and $M \geq T_g$. Let $\{\mathbf{d}_k\}_{k=1}^K$ be a dictionary of filters with, for all k, supp$(d_k) \subset$ supp(g). Then*

$$\min_{\mathbf{d}_k\in\mathbb{R}^{T_d},\mathbf{z}_k\in\mathbb{R}_+^{T_z}} \frac{1}{2}\|\mathbf{x} - \sum_k d_k * z_k\|^2 + \lambda\sum_k \|\mathbf{z}_k\|_1 = \min_{\boldsymbol{\alpha}\in\mathbb{C}^{M,N}} \frac{1}{2}\|\mathbf{x} - \Phi(\boldsymbol{\alpha})\|^2 + \lambda\|\mathbf{Z}\|_1 \tag{15}$$

$$s.t.\ \begin{cases} \alpha[\tau,\nu] = \sum_{k=1}^{K} \mathbf{W}[\nu,k]\mathbf{Z}[k,\tau]e^{i\theta_k[\tau,\nu]} \\ \mathbf{W}[\nu,k], \mathbf{Z}[k,\tau] \geq 0, \theta_k[\tau,\nu] \in]-\pi,\pi] \end{cases}$$

Proof. It suffices to apply the Theorem 1 to the elementary signals $\mathbf{x}_k = \mathbf{d}_k * \mathbf{z}_k$, and to notice

$$x_k[t] = (z_k * d_k)[t] \tag{16}$$

$$= \sum_{\tau=0}^{T_x-1} \sum_{\nu=0}^{M-1} \hat{w}_k(\nu) z_k[\tau] g[t-\tau] e^{i\frac{2\pi\nu(t-\tau)}{M}} \tag{17}$$

$$= \sum_{\tau=0}^{T_x-1} \sum_{\nu=0}^{M-1} e^{i\theta_k[\tau,\nu]} |\hat{w}_k(\nu)||z_k[\tau]| g[t-\tau] e^{i\frac{2\pi\nu(t-\tau)}{M}} \tag{18}$$

and take $\theta_k[\tau,\nu]$ such that $e^{i\theta_k[\tau,\nu]} = \frac{w_k(\nu)}{|w_k(\nu)|}\text{sgn}(z_k(\tau))$ ∎

In other words, a signal admits a CDL-type representation with positive activations if and only if it admits synthesis time-frequency coefficients that factorize in semi-NMF or complex-NMF. Moreover, the positivity constraint on the coefficients $\mathbf{z}_k$ and the activation matrix $\mathbf{Z}$ in Eq. (15) can be relaxed to have real activations and recover the classical CDL. When the jump between two windows is such that $a > 1$, this constrains the activation coefficients $\mathbf{z}_k[\tau] = 0$ for all the $\tau \neq at$, and thus forces the parsimony of the CDL.

The minimization of Eq. (15) can be done by a projected gradient descent, whose algorithm is given in Algorithm 1. The Algorithm 1 uses a semi-NMF decomposition with a multiplicative algorithm as proposed in [15], which has been adapted to complex matrices to minimize Eq. (15). Although it is also possible to replace the call to the semi-NMF by a complex-NMF, as proposed in [16, 17], to minimize Eq. (15), we have limited ourselves here at the semi-NMF.

Algorithm 1: Minimization of Eq. (15) by projected gradient descent and semi-NMF

Input: $t = 0$, $\boldsymbol{\alpha}^0 \in \mathbb{C}^{MN}$, $\mathbf{W}^0 \in \mathbb{C}^{MK}$, $\mathbf{Z} \in \mathbb{R}_+^{KN}$, $\lambda \geq 0$;
Output: $\boldsymbol{\alpha} \in \mathbb{C}^{MN}$, $\mathbf{W} \in \mathbb{C}^{MN}$, $\mathbf{Z} \in \mathbb{R}_+^{MN}$;
while *not converged* **do**
  $\boldsymbol{\alpha}^{t+1/2} = \boldsymbol{\alpha}^t + \Phi^*(\mathbf{x} - \Phi(\boldsymbol{\alpha}^t))$;
  $\mathbf{W}^{t+1}, \mathbf{Z}^{t+1} = \text{semi-NMF}(\boldsymbol{\alpha}^{t+1/2})$;
  $\boldsymbol{\alpha}^{t+1} = \mathbf{W}^{t+1}\mathbf{Z}^{t+1^T}$;
  $t = t + 1$
end

4. NUMERICAL RESULTS

In this section, we will compare the numerical results obtained from various NMF and CDL methods for two distinct sound types: a piano sound and an artificial sound consisting of frequency chirps (Fig. 1 and Fig. 2).

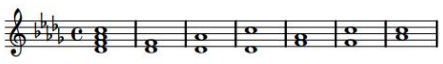


Fig. 1: piano partition

The piano sound showcases distinctive frequency motifs that correspond to individual notes, making it an intriguing case for evaluating the performance of the methods in accurately capturing and separating these specific frequency patterns. This piano song has already been used in several articles utilizing NMF decomposition, as referenced in [12].The piano composition involves four different notes initially played simultaneously, followed by each pair of notes played separately. The duration of the piano sound is approximately 15 seconds, sampled at a rate of 22050 Hz (consisting of 355,000 data points). Given that CDL resolution involves alternating gradient descent in the Fourier domain, while NMF methods are utilized in the Gabor domain, we aim to use small atom sizes ($M = 2048$) to mitigate edge effects. For the NMF algorithm, we employ a small Gabor window ($M = 2048$) and aim for a high overlap rate close to 100% ($a = 8$). In Fig. 3 we present the outcomes of the CDL and semi-NMF methods. It can be observed that the CDL method accurately identifies three of the four notes, although the fourth note is somewhat obscured by artifacts. Conversely, the semi-NMF method successfully captures all four notes, even though some of them may be separated across multiple motifs. It is important to note that reducing the overlap in the semi-NMF methods significantly compromises the quality of the reconstruction. If there is a need to decrease computational cost by reducing the overlap, employing the IS-NMF method proves to be more robust to low time redundancy, as illustrated in Fig. 3 (with an overlap of 50%).

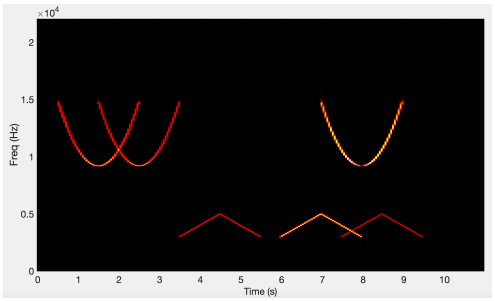
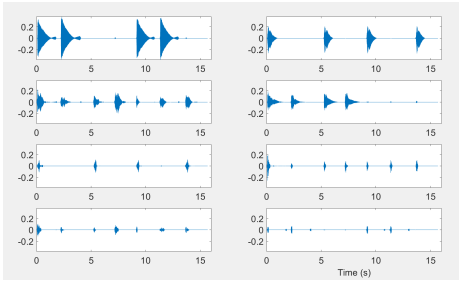


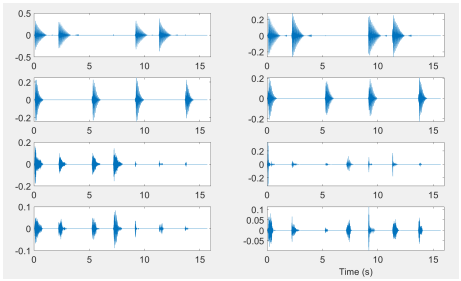
Fig. 2: Chirps signal spectrogram

The chirps signal consists in two non-stationary chirps each lasting 2 seconds. This choice of signal allowed us to assess the ability of the methods to recover non-stationary atoms, highlighting their effectiveness in capturing time-varying structures. The signal length is about 15 s sampled at 22050 Hz. The frequency non-stationarity of the motifs prompted the introduction of convolutive NMF as suggested in [22] and [23] within the same type of signal. However, it is worth noting that even with the introduction of convolutive NMF, the classical NMF and semi-NMF methods can still successfully identify the motifs as long as the Gabor window size exceeds the duration of the motifs. We compared the results obtained by the CDL approach (with a positive activation constraint) and the Algorithm 1. We represented the spectrogram of the signal in Fig. 2. To satisfy the assumptions of Theorem 1, we searched for atoms of size 2.5 seconds using the CDL approach. The algorithm in Algorithm 1 utilized an STFT with a window size of 2.5 seconds and an 80% overlap. Fig. 4 depicts the spectrograms of the estimated components obtained by each method. While the CDL approach provided better separation between the two types of chirps, the semi-NMF method is struggling to capture both types of elementary signals. This can be attributed to the utilization of an overlap of 80% (i.e., $a > 1$). It is important to note that the complex/semi-NMF methods exhibit limited robustness in scenarios with low temporal redundancy. Since Gabor windows are very large, stretching the overlap towards 100% has a huge computational cost. As for the piano song, if one wants to reduce the overlap in order to decrease computational cost, it can

(a) CDL



(b) Semi-NMF 99,6% overlap ($a = 8$)



(c) IS-NMF 50% overlap ($a = \frac{M}{2}$)

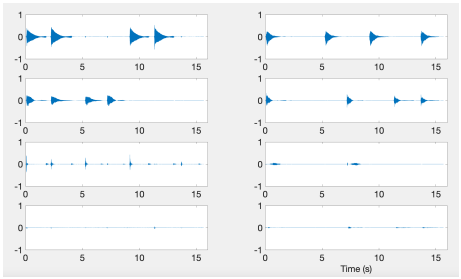
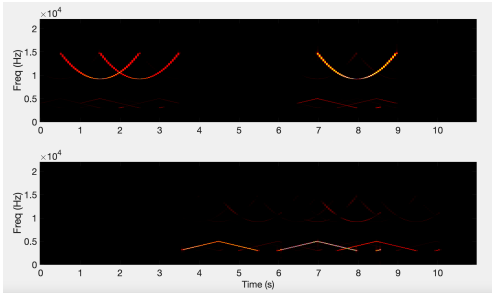
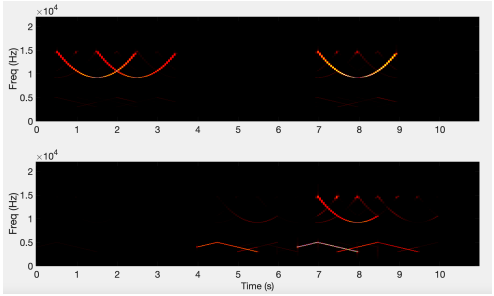


Fig. 3: Comparison of the piano notes estimated by the CDL, the Algorithm 1 and IS-NMF

(a) CDL



(b) Algorithm 1



(c) IS-NMF 50% overlap ($a = \frac{M}{2}$)

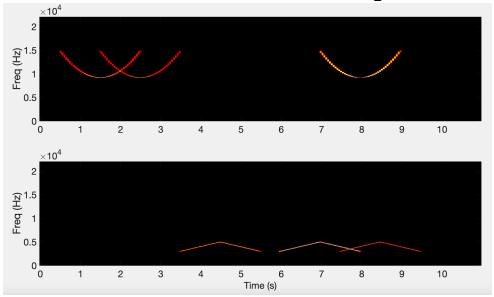


Fig. 4: Comparison of the spectrograms of the components estimated by the CDL, Algorithm 1 and the IS-NMF

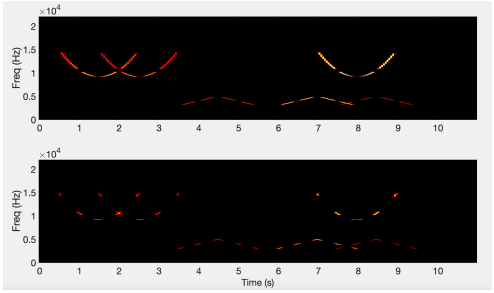


Fig. 5: Spectrograms of the components estimated by IS-NMF with 50 % overlap ($a = \frac{M}{2}$) and a small window ($M = 2048$)

be interesting to use the IS-NMF. It shows very good results with only a 50% overlap (Fig. 4) but still with a window size of 2.5 seconds. One could say it could be interesting to see the impact of a lower window size. In fact, as mentioned earlier in [22], [23], the convolutive-NMF is presented as the alternative to the classical NMF algorithm to reconstruct non-stationary elements. This statement is understandable as regards Fig. 5, but we have shown that the classical NMF algorithm can still reconstruct this kind of signals well if the window size is large enough. We also compared the results with an LRTFS-type approach. In practice, LRTFS gives similar results to CDL. Using an NMF in the LRTFS makes it more robust to phase changes and, therefore, the loss of temporal resolution due to a lower overlap.

5. CONCLUSION

We have demonstrated the theoretical equivalence between the "CDL" type approach and the semi-NMF (or complex-NMF) type decompositions of the time-frequency synthesis coefficients of a signal. In practice, the LRTFS or IS-NMF type decompositions give the best decompositions, although at the cost of a phase loss of the estimated time-frequency coefficients and a loss of temporal resolution compared to the CDL. Semi-NMF can be an alternative to CDL when the time-frequency redundancy is high. We possess experimental findings that lean towards confirming it. We will also investigate in more detail the theoretical links between LRTFS and CDL and the robustness of these approaches as a function of the window size compared to convolutional atoms.

6. REFERENCES

[1] Ron Rubinstein, Alfred M Bruckstein, and Michael Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1045–1057, 2010.

[2] Stéphane Mallat, *A wavelet tour of signal processing*, Elsevier, 1999.

[3] Laurent Daudet and Bruno Torrésani, "Hybrid representations for audiophonic signal encoding," *Signal Processing*, vol. 82, no. 11, pp. 1595–1617, 2002.

[4] J-L Starck, Y Moudden, J Bobin, M Elad, and DL Donoho, "Morphological component analysis," in *Wavelets XI*. SPIE, 2005, vol. 5914, pp. 209–223.

[5] Gabriel Peyré, Jalal Fadili, and Jean-Luc Starck, "Learning the morphological diversity," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 646–669, 2010.

[6] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, pp. 4311 – 4322, 2006.

[7] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro, "Online dictionary learning for sparse coding," in *ICML*, 2009, pp. 689–696.

[8] Roger Grosse, Rajat Raina, Helen Kwong, and Andrew Y Ng, "Shift-invariant sparse coding for audio classification," in *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, 2007, pp. 149–158.

[9] Andrea Cogliati, Zhiyao Duan, and Brendt Wohlberg, "Context-dependent piano music transcription with convolutional sparse coding," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 12, pp. 2218–2230, 2016.

[10] Mainak Jas, Tom Dupré la Tour, Umut Simsekli, and Alexandre Gramfort, "Learning the morphology of brain signals using alpha-stable convolutional sparse coding," *NeurIPS*, vol. 30, 2017.

[11] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[12] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Non-negative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[13] Paris Smaragdis and Judith C Brown, "Non-negative matrix factorization for polyphonic music transcription," in *WASPAA*. IEEE, 2003, pp. 177–180.

[14] Felix Weninger, Jonathan Le Roux, John R Hershey, and Shinji Watanabe, "Discriminative NMF and its application to single-channel source separation.," in *Interspeech*, 2014, pp. 865–869.

[15] Chris HQ Ding, Tao Li, and Michael I Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 1, pp. 45–55, 2008.

[16] Hirokazu Kameoka, Nobutaka Ono, Kunio Kashino, and Shigeki Sagayama, "Complex NMF: a new sparse representation for acoustic signals," in *ICASSP*. IEEE, 2009, pp. 3437–3440.

[17] Paul Magron and Tuomas Virtanen, "Complex ISNMF: a phase-aware model for monaural audio source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 20–31, 2018.

[18] Cédric Févotte and Matthieu Kowalski, "Low-rank time-frequency synthesis," *NeurIPS*, vol. 27, 2014.

[19] Thomas Moreau and Alexandre Gramfort, "Dicodile: Distributed convolutional dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 5, pp. 2426–2437, 2020.

[20] Vardan Papyan, Jeremias Sulam, and Michael Elad, "Working locally thinking globally: Theoretical guarantees for convolutional sparse coding," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5687–5701, 2017.

[21] Hilton Bristow, Anders Eriksson, and Simon Lucey, "Fast convolutional sparse coding," in *CVPR*, 2013, pp. 391–398.

[22] Paul D. O'Grady and Barak A. Pearlmutter, "Convolutive non-negative matrix factorisation with a sparseness constraint," in *2006 16th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, 2006, pp. 427–432.

[23] Paris Smaragdis, "Convolutive speech bases and their application to supervised speech separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 1–12, 2007.