

# Du convolutive sparse coding à la factorisation \*-NMF des coefficients temps-fréquence

Jean-Baptiste MALAGNOUX Matthieu KOWALSKI

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

**Résumé** – Le *Convolutional Dictionary Learning* (CDL) est une technique d'apprentissage de dictionnaire exploitant l'invariance par translation de signaux élémentaires. Dans le domaine temps-fréquence, la répétition de motifs fréquentiels élémentaires est exploitée à travers les décompositions de type "factorisation en matrice non négatives" (NMF), et ses dérivées, du spectrogramme. On étudie ici les liens entre ces deux approches, et on montre en particulier qu'un signal qui admet une décomposition de type CSC admet des coefficients de synthèse temps-fréquence qui peuvent se décomposer sous la forme de semi-NMF ou complex-NMF. Les différentes approches sont alors comparées expérimentalement sur un signal jouet.

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

## 1 Introduction

Le codage parcimonieux, ou "sparse coding" (SC), est une technique d'apprentissage non supervisée largement utilisée pour la représentation des signaux [19]. Cette méthode consiste à décomposer un signal en une combinaison linéaire de quelques éléments, ou atomes, d'un dictionnaire préétabli, où la plupart des coefficients sont nuls. Initialement, les dictionnaires étaient basés sur des formes d'ondes telles que les ondelettes ou les transformées temps-fréquence [15]. Toutefois, l'utilisation de dictionnaires prédéfinis limite la capacité du SC à modéliser des signaux plus complexes.

Une solution est l'utilisation d'union de dictionnaires [5, 21]. Par exemple, des dictionnaires temps-fréquence avec différentes tailles de fenêtre peuvent être combinés pour modéliser à la fois des structures temporelles et spectrales inconnues, tout en limitant le principe d'incertitude d'Heisenberg. Cela permet d'augmenter la diversité des éléments du dictionnaire et de mieux représenter les structures sous-jacentes des signaux. Bien que [18] propose d'apprendre la meilleure combinaison possible de dictionnaires, cette méthode nécessite toujours une sélection manuelle des dictionnaires de base et ne permet pas d'optimiser automatiquement les éléments du dictionnaire directement.

On peut résoudre ce problème en adoptant une approche "data-driven", qui consiste à optimiser automatiquement les éléments du dictionnaire à partir de données d'entraînement [1, 14]. Cependant, l'apprentissage de dictionnaire avec Sparse Coding nécessite un compromis entre la qualité de la représentation et la complexité du dictionnaire. En effet, un dictionnaire trop petit ne permettra pas de représenter les signaux de manière satisfaisante, tandis qu'un dictionnaire trop grand peut conduire à une sur-apprentissage (overfitting)

des données. Par conséquent, la recherche d'un dictionnaire optimal est un problème important en traitement de signal.

Le Convolutional Dictionary Learning (CDL), allié au Convolutional Sparse Coding (CSC), est une méthode d'apprentissage de dictionnaire qui a été proposée dans le contexte des signaux audio de musique [9]. Cette méthode consiste à apprendre un dictionnaire de filtres de convolution, qui sont capables de capturer des motifs temporels dans les signaux et exploite la notion d'invariance par translation de ces motifs ainsi que leur parcimonie. Le CDL a été appliqué avec succès dans diverses applications, telles que la transcription automatique [4] ou l'analyse de signaux biomédicaux [10].

Une autre approche populaire pour l'analyse de signaux est la factorisation en matrices non négatives (NMF) [12]. Dans le contexte des signaux temporels, et en particulier les musiques, l'approche NMF consiste à décomposer le spectrogramme en une combinaison linéaire de spectres (d'amplitude ou de puissance) et de profils temporels [7]. La NMF est couramment utilisée pour l'analyse de signaux audio [20] et la séparation de sources [22]. Elle permet de découvrir des motifs fréquentiels récurrents dans les signaux audio, tels que les notes de musique ou les harmoniques de la voix.

**Contributions et organisation de l'article** Après un bref rappel du CDL et de la factorisation NMF des coefficients STFT dans la Section 3, nous établissons l'équivalence entre les méthodes CDL et la semi/complex-NMF de coefficients de synthèse d'une transformée de Fourier à court terme (STFT) à redondance maximal en temps d'un point de vue mathématique dans la Section 3. Plus particulièrement, nous montrons qu'un signal construit comme la convolution d'atomes et d'un mapping parcimonieux peut toujours se synthétiser à partir de coefficients temps-fréquence factorisable sous forme de semi/complex-NMF de rang 1. Nous démontrons ensuite l'équivalence entre les problèmes d'optimisation des deux ap-

Ce travail est supporté par l'Agence Nationale de la Recherche à travers le projet BMWs (ANR-20-CE45-0018)

proches ainsi nous présentons les méthodes de résolution de ces problèmes. La [Section 4](#) présente les résultats expérimentaux obtenus à partir de l'application des méthodes CSC et NMF sur un signal audio de piano.

## 2 Etat de l'art

On rappelle ici les principes du CDL. On présente alors la transformée de Gabor discrète telle que présentée dans [2], puis on expose brièvement les décompositions NMF.

### 2.1 CDL

Soit  $\mathbf{x} \in \mathbb{R}^{T_x}$ , pour tout  $k \in \llbracket 1, K \rrbracket$ ,  $\mathbf{d}_k \in \mathbb{R}^{T_d}$ ,  $\mathbf{z}_k \in \mathbb{R}^{T_z}$ . Le modèle convolutionnel linéaire à l'ordre  $K$  de  $\mathbf{x}$  s'écrit [16] :

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

où  $\mathbf{n} \in \mathbb{R}^{T_x}$  est un bruit supposé blanc gaussien. Le dictionnaire  $\{\mathbf{d}_k\}$  est composé d'atomes élémentaires  $\mathbf{d}_k$  qui s'activent aux différentes positions encodées dans  $\mathbf{z}_k$ . Ces positions sont supposées parcimonieuses, chaque atome n'apparaissant qu'à quelques localisations. Ainsi, l'estimation du dictionnaire  $\{\mathbf{d}_k\}$  et des coefficients  $\{\mathbf{z}_k\}$  peut se faire à l'aide du CDL [17] :

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

Le CDL revient donc à résoudre un problème de convolutive sparse coding (estimation des activations  $\{\mathbf{z}_k\}$ ) et l'estimation du dictionnaire de filtres  $\{\mathbf{d}_k\}$ . Plusieurs algorithmes ont été proposés, comme le FastCSC [3] ou des algorithmes distribués [16].

### 2.2 STFT et NMF

**Transformée de Gabor.** Soit  $g \in \mathbb{R}^{T_g}$  une fenêtre. La STFT, ou coefficients d'analyse de Gabor, d'un signal  $\mathbf{x} \in \mathbb{R}^{T_x}$  est donnée pour tout  $\tau \in \llbracket 0, N-1 \rrbracket$ ,  $\nu \in \llbracket 0, M-1 \rrbracket$  par :

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

avec  $\varphi_{\tau, \nu}[t] = g[t-a\tau]e^{\frac{i2\pi\nu t}{M}}$ . Le paramètre  $a > 0$  est la taille du *saut* entre deux fenêtres, tel que  $aT_g \leq N$ , et  $M \leq T_g$  est le nombre de points pour la FFT. Lorsque le dictionnaire  $\varphi_{\tau, \nu}$  forme un frame, il existe une infinité de *coefficients de synthèse*  $\alpha[\tau, \nu]$  tels que :

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

On note l'opérateur de synthèse  $\Phi$  tel que  $\mathbf{x} = \Phi(\boldsymbol{\alpha})$  et son adjoint l'opérateur d'analyse tel que  $\mathbf{X} = \Phi^*(\mathbf{x})$ .

**\*-NMF** Les techniques de décomposition basées sur les factorisations en matrices non négatives (NMF) de représentations temps-fréquences permettent d'exploiter des structures particulières des signaux, en repérant quels motifs fréquentiels se répètent au cours du temps. Les premières NMF ont été appliquées au spectrogramme d'amplitude ( $p = 1$ ) ou de puissance ( $p = 2$ ) du signal, c'est-à-dire :  $V[\tau, \nu] = |X[\tau, \nu]|^p$ . La NMF cherche une décomposition de la forme :

$$\mathbf{V} \simeq \mathbf{W}\mathbf{Z}^T, \text{ s.t. } \mathbf{W} \in \mathbb{R}_+^{MK}, \mathbf{Z} \in \mathbb{R}_+^{NK} \quad (5)$$

Afin de mieux prendre en compte la phase de la transformée de Gabor, d'autres décompositions ont été proposées, comme la semi-NMF [6] :

$$\min_{\mathbf{W} \in \mathbb{C}^{MK}, \mathbf{Z} \geq 0} \|\mathbf{X} - \mathbf{W}\mathbf{Z}^T\|^2 \quad (6)$$

ou bien, toujours avec  $\mathbf{W} \in \mathbb{C}^{M,K}$  et  $\mathbf{Z} \geq 0$ , la complex-NMF [11, 13] décompose la transformée de Gabor sous la forme :

$$X[\tau, \nu] = \sum_{k=1}^K \mathbf{W}[\nu, k] \mathbf{Z}[\tau, k] e^{i\theta_k[\tau, \nu]}. \quad (7)$$

où  $\theta_k[\tau, \nu] \in ]-\pi, \pi]$  est un terme de phase global. Enfin, l'approche "Low Rank Time-Frequency Synthesis" (LRTFS)[8] permet d'utiliser la NMF avec divergence d'Itakura-Saito (IS-NMF)[7] comme "prior" sur les coefficients temps-fréquence de synthèse : la phase globale est préservée afin de reconstruire le signal, bien que la phase de chacune des composantes soit artificiellement reconstruite à partir de la phase globale.

## 3 CDL et synthèse temps-fréquence

Cette section montre que synthétiser un signal en utilisant l'apprentissage de dictionnaire convolutif parcimonieux ou en utilisant la NMF dans le domaine temps-fréquence est équivalent. Le théorème suivant énonce ce résultat pour un signal avec une seule composante convolutive.

**Théorème 1.** Soit  $\mathbf{x} \in \mathbb{R}^{T_x}$ . Soit  $\{\varphi_{\tau, \nu}\}_{\tau, \nu}$  un dictionnaire de Gabor défini comme en [Section 2.2](#) avec  $a = 1$  et  $M \geq T_g$ . Alors, il existe un filtre  $\mathbf{d} \in \mathbb{R}^{T_d}$ ,  $\text{supp}(\mathbf{d}) \subset \text{supp}(g)$ , tel que

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

$$\text{ssi } \forall(\tau, \nu) \alpha[\tau, \nu] = e^{i\frac{2\pi\nu\tau}{M}} \hat{w}(\nu)z(\tau)$$

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

*Démonstration.* On pose  $w[t] = d[t]/g[t]$ . Comme  $\text{supp}(\mathbf{d}) \subset \text{supp}(g)$ , on a

$$x[t] = (z * \mathbf{d})[t] = \sum_{\tau=0}^{T_x-1} z[\tau]g[t-\tau]w[t-\tau] \quad (8)$$

$$= \sum_{\tau=0}^{T_x-1} z(\tau)g[t-\tau] \sum_{\nu=0}^{M-1} \hat{w}(\nu) e^{i\frac{2\pi\nu(t-\tau)}{M}} \quad (9)$$

$$= \sum_{\tau=0}^{T_x-1} \sum_{\nu=0}^{M-1} e^{i\frac{2\pi\nu\tau}{M}} \hat{w}(\nu) z[\tau] g[t-\tau] e^{i\frac{2\pi\nu t}{M}} \quad (10)$$

$$= \sum_{\tau=0}^{T_x-1} \sum_{\nu=0}^{M-1} \alpha[\tau, \nu] g[t-\tau] e^{i\frac{2\pi\nu t}{M}} \quad (11)$$

■

En conséquence de ce théorème, le corollaire suivant montre l'équivalence entre le CDL et la factorisation des coefficients de Gabor à la synthèse

**Corollaire 1.** Soit  $\mathbf{x} \in \mathbb{R}^{T_x}$ . Soit  $\{\varphi_{\tau,\nu}\}_{\tau,\nu}$  un dictionnaire de Gabor défini comme en Section 2.2 avec  $a = 1$  et  $M \geq T_g$ . Soit  $\{\mathbf{d}_k\}_{k=1}^K$  un dictionnaire de filtres avec pour tout  $k$   $\text{supp}(\mathbf{d}_k) \subset \text{supp}(g)$ . Alors

$$\min_{\mathbf{d}_k, z_k \geq 0} \frac{1}{2} \|\mathbf{x} - \sum_k \mathbf{d}_k * z_k\|^2 + \lambda \sum_k \|\mathbf{z}_k\|_1 = \quad (12)$$

$$\min_{\mathbf{W}, \mathbf{Z} \geq 0} \frac{1}{2} \|\mathbf{x} - \Phi(\mathbf{W}\mathbf{Z}^T)\|^2 + \lambda \|\mathbf{Z}\|_1 = \quad (13)$$

$$\min_{\alpha} \frac{1}{2} \|\mathbf{x} - \Phi(\alpha)\|^2 + \lambda \|\mathbf{Z}\|_1 \quad (14)$$

$$\text{s.t.} \begin{cases} \alpha[\tau, \nu] = \sum_{k=1}^K \mathbf{W}[\nu, k] \mathbf{Z}[k, \tau] e^{i\theta_k[\tau, \nu]} \\ \mathbf{W}[\nu, k], \mathbf{Z}[k, \tau] \geq 0, \theta_k[\tau, \nu] \in ]-\pi, \pi[ \end{cases} \quad (15)$$

*Démonstration.* Il suffit d'appliquer le Théorème 1 aux signaux élémentaires  $\mathbf{x}_k = \mathbf{d}_k * \mathbf{z}_k$ , et de prendre  $\mathbf{W}[\nu, k] = \hat{w}_k(\nu)$  pour la première égalité. Pour la seconde égalité, il suffit de remarquer

$$x_k[t] = (z_k * \mathbf{d}_k)[t] \quad (16)$$

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

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

et prendre  $\theta_k[\tau, \nu]$  tel que  $e^{i\theta_k[\tau, \nu]} = e^{i\frac{2\pi\nu\tau}{T}} \frac{w_k(\nu)}{|w_k(\nu)|} \text{sgn}(z_k(\tau))$

Autrement dit, un signal admet une représentation de type CDL avec activations positives si et seulement s'il admet des coefficients temps-fréquence de synthèse qui se factorisent sous forme de semi-NMF ou de complexe-NMF. De plus, la contrainte de positivité sur les coefficients  $\mathbf{z}_k$  et la matrice d'activation  $\mathbf{Z}$  dans Eq. (13) peut être relâchée pour avoir des activations réelles et retrouver le CDL classique. Lorsque le saut entre deux fenêtres est tel que  $a > 1$ , cela revient à contraindre les coefficients d'activation  $\mathbf{z}_k[\tau] = 0$  pour tous les  $\tau \neq at$ , et ainsi à forcer la parcimonie du CDL.

La minimisation de Eq. (13) et Eq. (14) peut se faire par une descente de gradient projeté, dont l'algorithme est donné en Algorithme 1. L'Algorithme 1 fait appel à une décomposition semi-NMF avec un algorithme multiplicatif comme proposé dans [6], que l'on a adapté aux matrices complexes pour minimiser Eq. (13). Bien qu'il soit tout à fait possible de remplacer l'appel à la semi-NMF par une complexe-NMF, comme proposé dans [11, 13], pour minimiser Eq. (14), nous nous sommes limités ici à la semi-NMF.

## 4 Résultats numériques

On compare les résultats obtenus par l'approche CDL (avec contrainte d'activation positive) et l'Algorithme 1 sur un signal de 15 s échantillonné à 22050 Hz construit à l'aide de

**Algorithme 1 :** Minimisation de Eq. (13) par descente de gradient projeté et semi-NMF

---

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

---

deux composantes de type "chirp", chacune d'une durée de 2 s. On représente le spectrogramme de ce signal avec une fenêtre d'une longueur de 100 ms et un overlap entre fenêtre de 50% sur la Fig. 1. Afin de respecter les hypothèses du Théorème 1, on cherche des atomes de taille 2.5 s à l'aide du CDL. L'Algorithme 1 utilise une STFT avec une fenêtre de taille 1 s et un overlap de 80% (ce qui revient à "forcer" à 0 certaines activations dans le CDL). Sur la Fig. 2, on représente les spectrogrammes des composantes estimées par chacune des méthodes, en utilisant les mêmes paramètres que sur la Fig. 1. Si le CDL permet de mieux séparer les deux types de chirps, la semi-NMF est tout de même capable de capturer les deux types de signaux élémentaires. Nous avons aussi comparé les résultats par rapport à une approche de type LRTFS. En pratique, la LRTFS donne des résultats similaires au CDL. L'utilisation d'une NMF dans la LRTFS permet d'être plus robuste aux changements de phase et donc à la perte de résolution temporelle due à un moindre overlap entre les fenêtres.

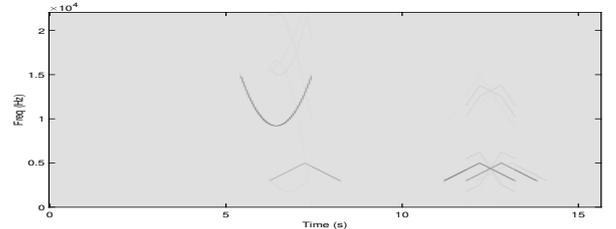


FIGURE 1 : Spectrogramme du signal utilisé

Nous avons aussi comparé les décompositions obtenues sur un signal de piano comme dans [7, 8] : un accord de 4 notes est suivi de toutes les combinaisons possibles de 2 notes. Ici, les fenêtres utilisées pour les décompositions temps-fréquences ont une durée de 100 ms avec un overlap de 50%, et sont donc beaucoup plus courtes que la durée des notes présentes d'une durée d'environ 2 s. Les approches de type LRTFS/IS-NMF donnent les meilleurs résultats (chaque note est bien isolée dans une composante), tandis que le CDL a tendance à mélanger certaines notes entre les composantes. L'approche basée sur la semi-NMF donne des résultats très similaires au CDL sur le son de piano lorsque la redondance temporelle est forte. Les dictionnaires NMF et semi-NMF sont initialisés en utilisant une décomposition en valeurs singulières (SVD) appliquée respectivement au spectrogramme et aux coefficients d'analyse du signal. Cette approche d'initialisation basée sur une décomposition de rang faible semble produire ici des résultats plus robustes que d'autres méthodes d'initialisation comme l'aléatoire.

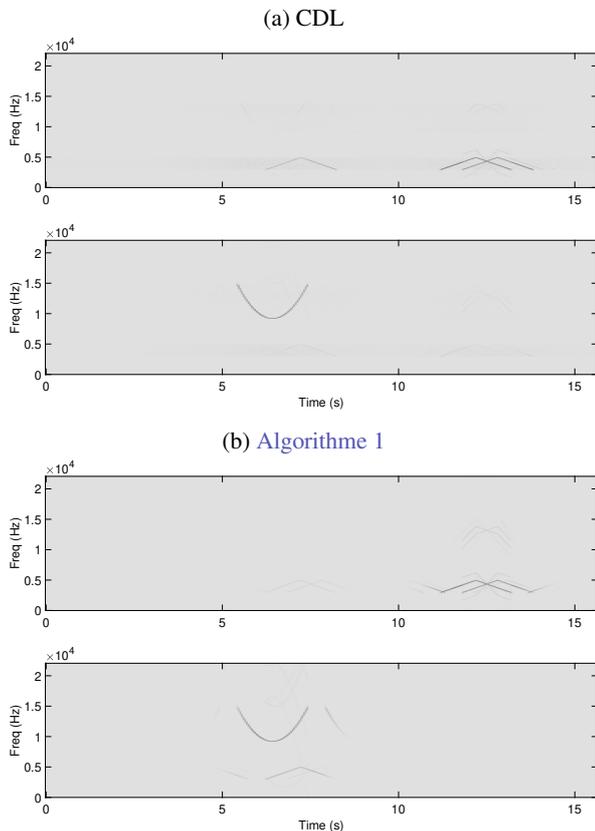


FIGURE 2 : Comparaison des spectrogrammes des composantes estimées par le CDL et l' [Algorithme 1](#)

## 5 Conclusion

Nous avons démontré l'équivalence théorique entre l'approche de type "CDL" et les décompositions de type semi-NMF (ou complex-NMF) des coefficients de synthèse temps-fréquence d'un signal. En pratique, les décompositions de type LRTFS ou IS-NMF donnent les meilleures décompositions, bien qu'au prix d'une perte de phase des coefficients temps-fréquence estimés et d'une perte de résolution temporelle par rapport au CDL. La semi-NMF peut s'avérer être une alternative au CDL lorsque la redondance temps-fréquence est élevée. Dans les travaux futurs, il sera important de confirmer expérimentalement ces premiers résultats. Nous étudierons également plus en détail les liens théoriques entre le LRTFS et le CDL, ainsi que la robustesse de ces approches en fonction de la taille de la fenêtre par rapport aux atomes convolutifs.

## Références

- [1] Michal AHARON, Michael ELAD et Alfred BRUCKSTEIN : K-SVD : an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54:4311 – 4322, 2006.
- [2] Peter BALAZS, Monika DÖRFLER, Matthieu KOWALSKI et Bruno TORRÉSANI : Adapted and adaptive linear time-frequency representations : a synthesis point of view. *IEEE Signal Processing Magazine*, 30(6):20–31, 2013.
- [3] Hilton BRISTOW, Anders ERIKSSON et Simon LUCEY : Fast convolutional sparse coding. *In CVPR*, pages 391–398, 2013.
- [4] Andrea COGLIATI, Zhiyao DUAN et Brendt WOHLBERG : Context-dependent piano music transcription with convolutional sparse coding. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(12):2218–2230, 2016.
- [5] Laurent DAUDET et Bruno TORRÉSANI : Hybrid representations for audiophonic signal encoding. *Signal Processing*, 82(11):1595–1617, 2002.
- [6] Chris HQ DING, Tao LI et Michael I JORDAN : Convex and semi-nonnegative matrix factorizations. *IEEE transactions on pattern analysis and machine intelligence*, 32(1):45–55, 2008.
- [7] Cédric FÉVOTTE, Nancy BERTIN et Jean-Louis DURRIEU : Nonnegative matrix factorization with the itakura-saito divergence : With application to music analysis. *Neural computation*, 21(3):793–830, 2009.
- [8] Cédric FÉVOTTE et Matthieu KOWALSKI : Low-rank time-frequency synthesis. *NeurIPS*, 27, 2014.
- [9] Roger GROSSE, Rajat RAINA, Helen KWONG et Andrew Y NG : Shift-invariant sparse coding for audio classification. *In Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence*, pages 149–158, 2007.
- [10] Mainak JAS, Tom Dupré la TOUR, Umut SIMSEKLI et Alexandre GRAMFORT : Learning the morphology of brain signals using alpha-stable convolutional sparse coding. *NeurIPS*, 30, 2017.
- [11] Hirokazu KAMEOKA, Nobutaka ONO, Kunio KASHINO et Shigeki SAGAYAMA : Complex NMF : a new sparse representation for acoustic signals. *In ICASSP*, pages 3437–3440. IEEE, 2009.
- [12] Daniel D LEE et H Sebastian SEUNG : Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788–791, 1999.
- [13] Paul MAGRON et Tuomas VIRTANEN : Complex ISNMF : a phase-aware model for monaural audio source separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1):20–31, 2018.
- [14] Julien MAIRAL, Francis BACH, Jean PONCE et Guillermo SA-PIRO : Online dictionary learning for sparse coding. *In ICML*, pages 689–696, 2009.
- [15] Stéphane MALLAT : *A wavelet tour of signal processing*. Elsevier, 1999.
- [16] Thomas MOREAU et Alexandre GRAMFORT : DicoDile : Distributed convolutional dictionary learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2426–2437, 2020.
- [17] Vardan PAPYAN, Jeremias SULAM et Michael ELAD : Working locally thinking globally : Theoretical guarantees for convolutional sparse coding. *IEEE Transactions on Signal Processing*, 65(21):5687–5701, 2017.
- [18] Gabriel PEYRÉ, Jalal FADILI et Jean-Luc STARCK : Learning the morphological diversity. *SIAM Journal on Imaging Sciences*, 3(3):646–669, 2010.
- [19] Ron RUBINSTEIN, Alfred M BRUCKSTEIN et Michael ELAD : Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.
- [20] Paris SMARAGDIS et Judith C BROWN : Non-negative matrix factorization for polyphonic music transcription. *In WASPAA*, pages 177–180. IEEE, 2003.
- [21] J-L STARCK, Y MOUDDEN, J BOBIN, M ELAD et DL DONOHO : Morphological component analysis. *In Wavelets XI*, volume 5914, pages 209–223. SPIE, 2005.
- [22] Felix WENINGER, Jonathan LE ROUX, John R HERSHEY et Shinji WATANABE : Discriminative NMF and its application to single-channel source separation. *In Interspeech*, pages 865–869, 2014.