

# Séparation de sources par factorisation en matrices non-négatives *informée*

## 1 Introduction

Le problème séparer les différentes sources qui composent un signal de musique dans le cas où l'on dispose uniquement d'un signal monophonique est un sujet de recherche qui suscite un grand intérêt depuis plusieurs années. Ce problème difficile a de nombreuses applications telles que la transcription automatique des morceaux de musique [1], la reconnaissance d'instruments [2], ou l'estimation de la mélodie [3].

De nombreux éléments contribuent à faire de la séparation de sources un problème difficile. On peut citer parmi eux la complexité des sources musicales qui interfèrent dans les mélanges polyphoniques, les conditions variables d'enregistrement, le bruit, les effets digitaux appliqués pour les enregistrements en studio, etc.

Une approche populaire au problème de séparation de sources est d'utiliser une technique de Factorisation en matrices non-négatives (en anglais *Non-Negative Matrix Factorization* NMF). La NMF est une technique de factorisation qui permet d'expliquer les données par un petit nombre d'objets représentatifs. On se propose dans ce projet d'utiliser cette technique pour faire de la séparation de sources en musique.

## 2 Rappels théoriques sur la NMF

Mathématiquement, on formule la NMF de la façon suivante. Soit  $V \in \mathbb{R}_+^{M \times N}$  une matrice non-négative, c'est-à-dire dont tous les coefficients sont positifs ou nuls, de taille  $M \times N$  (dans les applications en musique  $V$  sera très souvent le spectrogramme d'amplitude). La factorisation en matrices non-négatives approxime  $V$  par  $\tilde{V}$  de la manière suivante :

$$\tilde{V} = WH \tag{1}$$

avec  $W \in \mathbb{R}_+^{M \times K}$  et  $H \in \mathbb{R}_+^{K \times N}$ , et où  $K$  est le rang de factorisation, généralement choisi tel que  $K(M + N) \ll MN$ .

La matrice  $W$  est appelée *dictionnaire*. Les vecteurs colonnes de  $W$  sont appelés *templates* ou *atomes*. La matrice  $H$  est appelée *matrice d'activation* : la  $k^{\text{ème}}$  ligne de  $H$  correspond en effet à l'activation du  $k^{\text{ème}}$  atome. Voir la figure 1 pour un exemple de décomposition d'un morceau de musique.

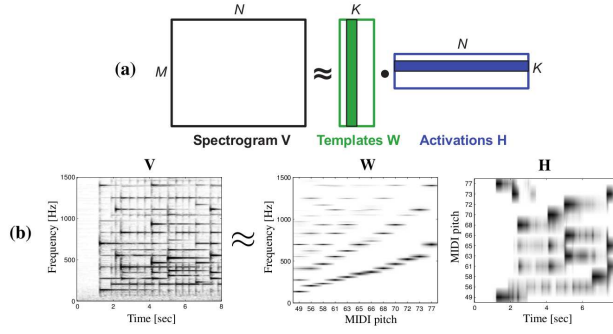


FIGURE 1 – Exemple de factorisation d’un spectrogramme d’amplitude d’un extrait du Prélude de Chopin Op. 28 No. 15, adapté de [5].

L’approximation de l’équation (2) est généralement quantifiée au moyen d’une divergence entre  $V$  et  $WH$ . Un algorithme de NMF est donc un algorithme de minimisation de cette divergence.

$$\min_{W, H \geq 0} D(V|W \cdot H)$$

L’article [6] propose un algorithme avec des règles de mise à jour multiplicatives qui permettent de converger vers un minimum local. Les matrices  $W$  et  $H$  sont optimisées de manière successive. Cet algorithme est très couramment utilisé en raison de sa grande simplicité d’implémentation et de la rapidité de calcul des itérations. Ces règles peuvent s’écrire de la manière suivante (notation matricielle) :

– Avec la divergence euclidienne :

$$H \leftarrow H \odot \frac{W^T \cdot V}{W^T(WH)} \quad W \leftarrow W \odot \frac{V \cdot H^T}{(WH) \cdot H^T}$$

– Avec la divergence de Kullback-Leibler :

$$H \leftarrow H \odot \frac{W^T \cdot \frac{V}{W \cdot H}}{W^T \cdot J} \quad W \leftarrow W \odot \frac{\frac{V}{W \cdot H} \cdot H^T}{J \cdot H^T}$$

L’opérateur  $\odot$  dénote le produit de Hadamard (multiplication point par point),  $J \in \mathbb{R}_+^{M \times N}$  dénote une matrice unitaire, et la division se fait point par point.

La procédure globale d’optimisation se fait de la manière suivante :

1. Initialiser les entrées  $W$  et  $H$  avec des valeurs aléatoires (positives)
2. Mettre à jour  $W$
3. Mettre à jour  $H$
4. Répéter de manière itérative les étapes 2 et 3 jusqu’à convergence

Reconstruction des sources séparées : étant donné un mélange de deux sources

$$X = S_1 + S_2, \quad X, S_1, S_2 \in \mathbb{C}^{M \times N} \quad (2)$$

On calcule la décomposition NMF du spectrogramme d'amplitude :

$$V = |X| \approx WH = W_1H_1 + W_2H_2 \quad (3)$$

On estime chaque source à l'aide de masques :

$$\hat{S}_1 = \frac{W_1H_1}{W_1H_1 + W_2H_2} \odot X \quad \hat{S}_2 = \frac{W_2H_2}{W_1H_1 + W_2H_2} \odot X \quad (4)$$

### 3 Travail demandé

Le but de ce projet est d'implémenter une méthode de séparation de sources basée sur la NMF *informée*, c'est-à-dire qui prend en compte de l'information musicale afin d'obtenir des résultats de séparation de sources qui sont physiquement et musicalement interprétables. On reprendra l'idée développée dans l'article [4] qui propose d'utiliser une partition MIDI pour initialiser l'algorithme, non pas avec matrices  $W$  et  $H$  aléatoires, mais avec des matrices qui reflètent la partition. Voir aussi l'article [5] pour plus de détails.

On étudiera les variantes suivantes (voir la Figure 2).

- Initialisation aléatoire de  $W$  et  $H$ .
- Initialisation avec des templates harmonique de  $W$  et initialisation aléatoire de  $H$ .
- Initialisation aléatoire de  $W$  et initialisation avec la position des notes midi pour  $H$ .
- Initialisation avec des templates harmonique de  $W$  et initialisation avec la position des notes midi pour  $H$ .
- Initialisation avec des templates harmonique, et la position des onsets de  $W$  et initialisation avec la position des notes midi pour  $H$ .

### 4 Outils

On utilisera pour l'algorithme de base de NMF le code du TP fait en cours (ne pas prendre en compte le terme de pénalité supplémentaire qui favorise la parcimonie de  $H$ ). Dans ce projet, il s'agit de modifier l'initialisation de  $W$  et  $H$ .

Pour récupérer l'information des fichiers MIDI, on pourra utiliser la toolbox MIDI, qu'on trouvera à l'adresse suivante : <https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/miditoolbox/>

*Remarque* Il faudra faire attention au réglage des différents paramètre. En particulier la nature de la décomposition dépend du choix du rang  $K$ , c'est à dire du nombre de lignes dans  $H$  et  $W$  respectivement.

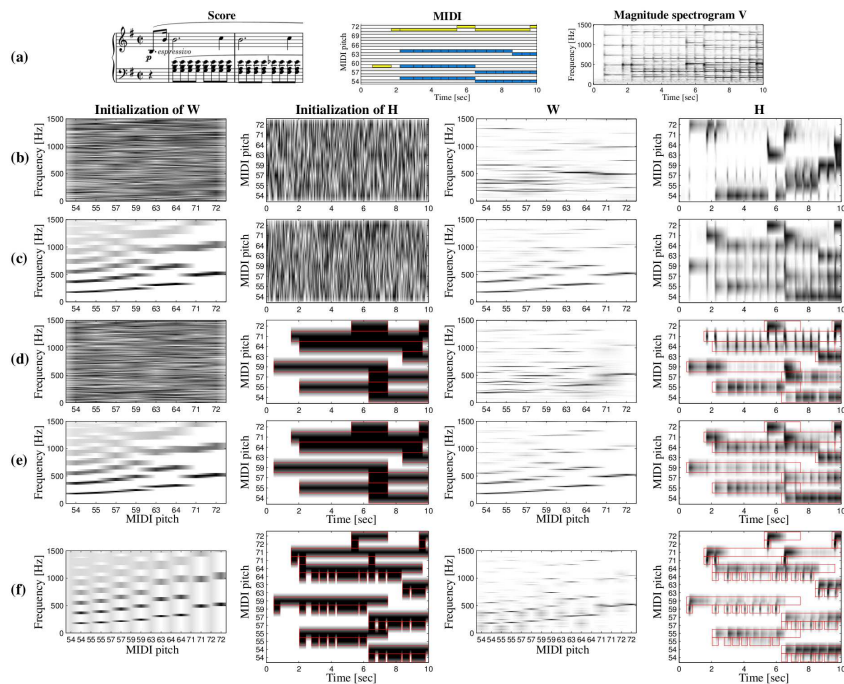


FIGURE 2 – Exemple de factorisation d’un spectrogramme d’amplitude d’un extrait du Prélude de Chopin Op. 28 No. 4, avec différentes initialisations, adapté de [4].

## 5 Morceaux de test

On donne en pièce jointe de ce projet deux extraits de chorals de Bach à quatre voix (Soprano, Alto, Tenor, Basses), joués par 4 instruments (violon, clarinette, saxophone et besson). On dispose des pistes séparées pour chaque instrument ainsi que du fichier midi correspondant<sup>1</sup>.

On pourra tester l’algorithme sur d’autres morceaux de différents types (par exemple piano seul, morceaux avec de la batterie), et comparer les résultats en fonction de l’instrumentation, du style de musique etc. On trouvera facilement sur Internet des fichiers MIDI correspondant à l’audio sélectionné. Vérifier en écoutant les fichiers MIDI qu’ils sont “acceptables”.

## Références

- [1] Plumbley, M.D. and Abdallah, S.A. and Bello, J.P. and Davies, M.E. and Monti, G. and Sandler, M.B. “Automatic music transcription and audio source separation”. In *Cybernetics and Systems*, 33(6), pp 603–627, (2002).
- [2] Heittola, T. and Klapuri, A.P. and Virtanen, T. “Musical instrument recognition in polyphonic audio using source-filter model for sound separation”. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp 327–332, Kobe, Japan, 2009.

1. Ces fichiers sont extraits du Bach10 dataset <http://music.cs.northwestern.edu/data/Bach10.html>.

- [3] Durrieu, J.L. and Richard, G. and David, B. and Févotte, C. “Source/filter model for unsupervised main melody extraction from polyphonic audio signals”. *IEEE Trans. Audio, Speech, Language Process.*, 18(3) :564–575, 2010.
- [4] Ewert, S. and Müller, M. “Using score-informed constraints for NMF-based source separation”. *Proc. ICASSP, Kyoto, Japan*, 2012.
- [5] Ewert, S. and Müller, M. “Score-Informed Source Separation for Music Signals”. In *Multimodal Music Processing, Schloss Dagstuhl–Leibniz-Zentrum für Informatik*, pp 73–94, Dagstuhl, Germany, 2012.
- [6] Lee, D.D. and Seung, H.S. “Algorithms for non-negative matrix factorization”. In MIT Press, *Advances in Neural Information Processing Systems (Eds.)*, vol.13, pp 556–562, 2000.