Nonnegative Tucker Decomposition with Beta-divergence for Music Structure Analysis of Audio Signals

A. Marmoret[1], F. Voorwinden[1], V. Leplat[2], J.E. Cohen[3], F. Bimbot[1]

[1]Univ Rennes, Inria, CNRS, IRISA, France, [2]Center for Artificial Intelligence Technology (CAIT), Skoltech, Moscow, Russia, [3]CNRS, CREATIS, Villeurbanne France - axel.marmoret@irisa.fr

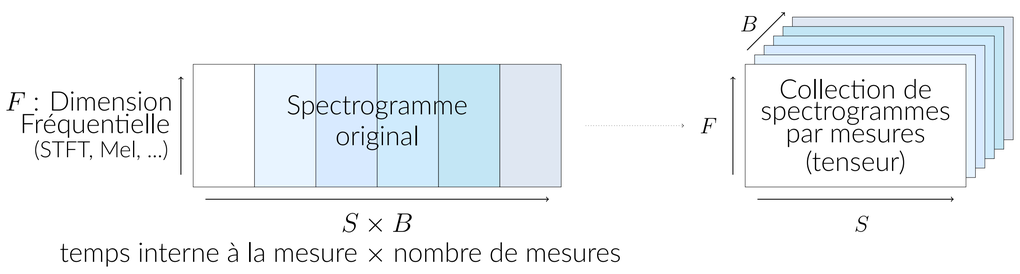
Résumé du poster

Ce poster présente une technique de factorisation appelée décomposition en Tucker nonnégatif (NTD), et un algorithme pour la calculer, optimisée selon la β-divergence [1].

La NTD permet d'extraire des patterns audios dans une musique, à l'échelle de la mesure [2]. Ces patterns peuvent être utilisés pour l'analyse musicale et la recomposition, ainsi que comme une représentation de mi-niveau, permettant notamment d'estimer la segmentation structurelle.

Dans le cadre de l'analyse de la musique, la β-divergence est montrée plus pertinente que la norme euclidienne utilisée dans un travail antérieur [1].

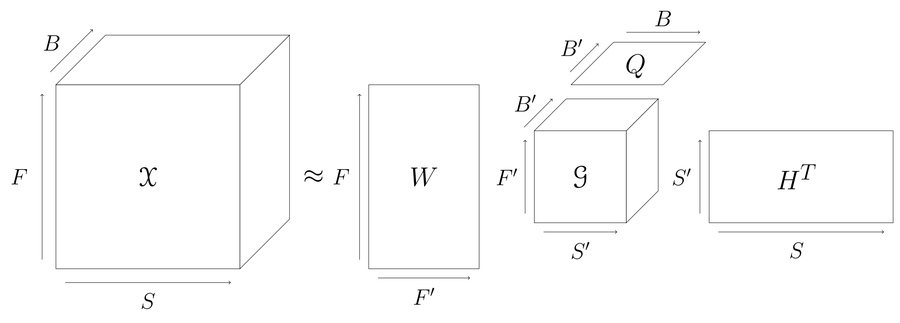
Etude à l'échelle de la mesure



Décomposition en Tucker nonnégatif (NTD)

Modèle NTD théorique :

NTD : $\mathcal{X} \approx \mathcal{G} \times_1 W \times_2 H \times_3 Q$



Pour chaque élément: $\mathcal{X}(f,s,b) \approx \sum_{f',s',b'=1}^{F',S',B'} \mathcal{G}(f',s',b') W(f,f') H(s,s') Q(b,b')$

NTD en pratique :

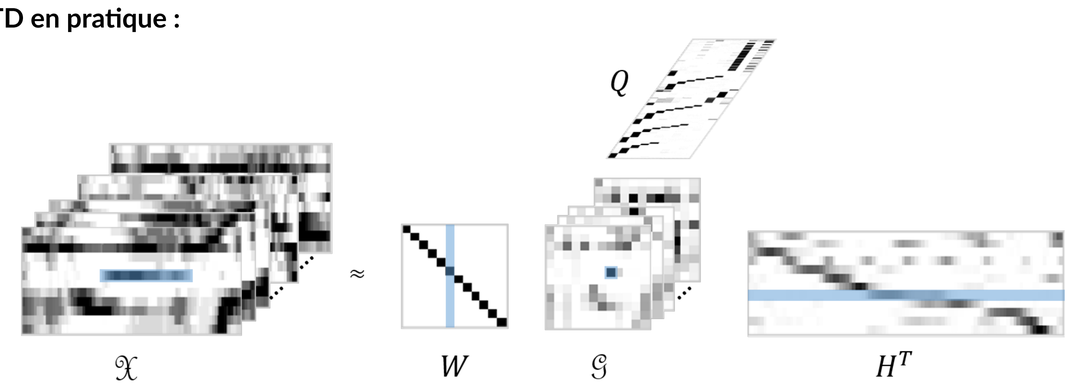


Figure. NTD sur le chromagramme de "Come Together" ($F'=12, S'=12$ et $B'=10$).

β-divergence

$$d_\beta(x|y) = \begin{cases} \frac{x}{y} - \log(\frac{x}{y}) - 1 & \beta = 0 \quad \text{Itakura-Saïto- (IS-) divergence} \\ x\log(\frac{x}{y}) + (y-x) & \beta = 1 \quad \text{Kullback-Leibler- (KL-) divergence} \\ \frac{x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}}{\beta(\beta-1)} & \beta \in \mathbb{R}\backslash\{0,1\} \end{cases} \quad (1)$$

NTD : problème d'optimisation

- Problème d'optimisation β-**NTD** (d_β est la β-divergence relative à chaque élément) :

$$\underset{W\geq 0, H\geq 0, Q\geq 0, \mathcal{G}\geq 0}{\arg\min} \quad d_\beta(\mathcal{X} | \mathcal{G} \times_1 W \times_2 H \times_3 Q)$$

L'algorithme [1] utilise les règles MU [3], revisitées et optimisées pour l'algèbre tensoriel.

Algorithm 1: Une boucle de β-NTD

Input: $\mathcal{X}, \mathcal{G}, W, H, Q, \epsilon, \beta$

Output: $\mathcal{G}, W, H, Q$

$V = (\mathcal{G} \times_2 H \times_3 Q)_{(1)}$

$W \leftarrow \max\left(W \cdot \left(\frac{\left[(WV)^{\cdot(\beta-2)} \cdot \mathcal{X}_{(1)}\right] V^\intercal}{(WV)^{\cdot(\beta-1)} V^\intercal}\right)^{\cdot\gamma(\beta)}, \epsilon\right)$ (Analogue pour H et Q)

```
/* Cette ligne utilise la propriété 𝒢_(1) (H ⊗ Q)^⊺ = (𝒢 ×_2 H ×_3 Q)_(1), qui permet de
   réduire fortement la complexité de l'algorithme                                  */
```

$\mathcal{N} = (\mathcal{G} \times_1 W \times_2 H \times_3 Q)^{\cdot(\beta-2)} \cdot \mathcal{X}$

$\mathcal{D} = (\mathcal{G} \times_1 W \times_2 H \times_3 Q)^{\cdot(\beta-1)}$

$\mathcal{G} \leftarrow \max\left(\mathcal{G} \cdot \left(\frac{\mathcal{N} \times_1 W^\intercal \times_2 H^\intercal \times_3 Q^\intercal}{\mathcal{D} \times_1 W^\intercal \times_2 H^\intercal \times_3 Q^\intercal}\right)^{\cdot\gamma(\beta)}, \epsilon\right)$

```
/* De même, utilisation de propritétés des produits tensoriels pour réduire la
   complexité, voir [1] pour le détail.                                             */
```

- On peut définir de même **Euc-NTD**, par rapport à la norme euclidienne [2] :

$$\underset{W\geq 0, H\geq 0, Q\geq 0, \mathcal{G}\geq 0}{\arg\min} \quad \|\mathcal{X} - \mathcal{G} \times_1 W \times_2 H \times_3 Q\|_2^2$$

Pattern musical

Aspect théorique :

Le produit $W\mathcal{G}_{::b'}H^T$ définit des "patterns musicaux" : spectrogrammes à l'échelle de la mesure.

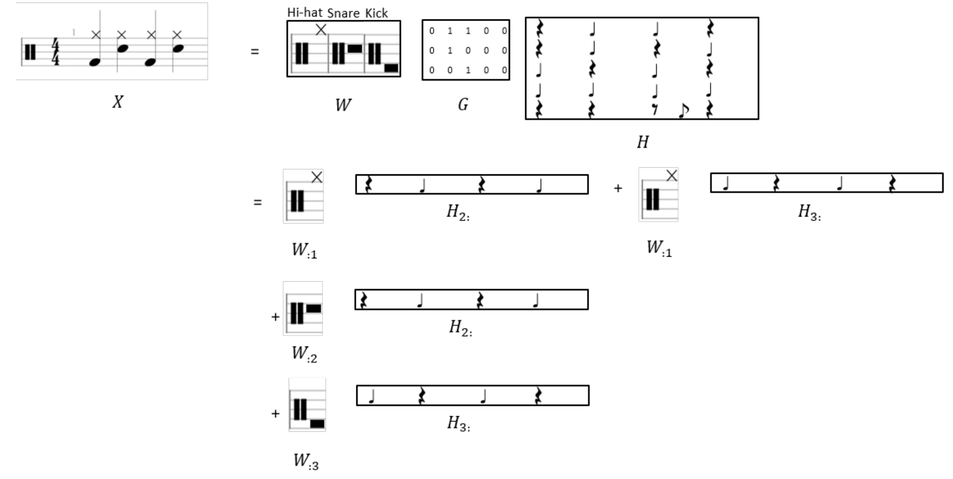


Figure. Un pattern musical théorique (partition de batterie).

Pattern musical en pratique :

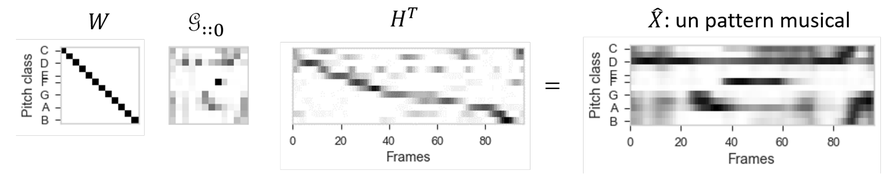


Figure. Un pattern musical, en pratique (chromagramme) : une combinaison linéaire des colonnes de W (extrême-gauche, information fréquentielle) et de H (centre-droit, information rythmique) est définie par la première tranche de $\mathcal{G}$ (centre-gauche), résultant en un chromagramme d'une mesure.

En estimant la phase, on peut écouter les résultats de la décomposition : https://ax-le.github.io/resources/examples/ListeningNTD.html

NTD pour la segmentation structurelle

Matrice Q^T : patterns musicaux comme descripteurs par mesure.

Chaque spectrogramme, à l'échelle de la mesure, se décompose : $\mathcal{X}_{::b} \approx \sum_{b'=1}^{B'} Q(b,b') W \mathcal{G}_{::b'} H^T$

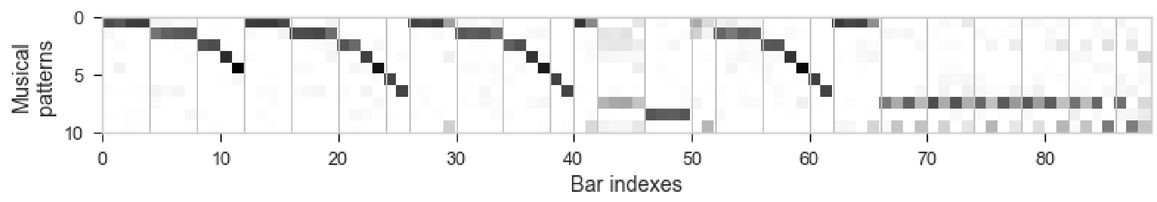


Figure. Matrice Q^T pour "Come Together". Lignes grises : annotations de segmentation.

Performances en segmentation structurelle (RWC Pop).

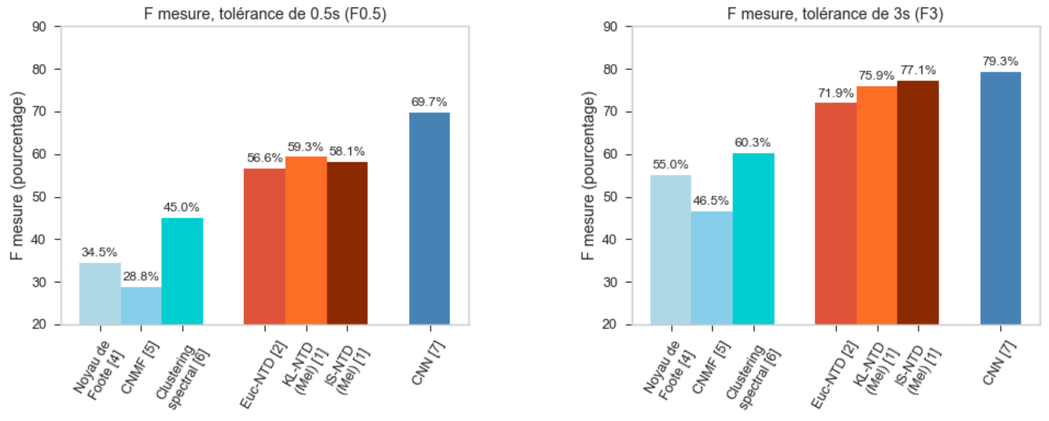


Figure. Comparaison des performances de segmentation (F mesures) avec l'état-de-l'art, respectivement [4, 5, 6] (algorithmes non-supervisés) et [7] (réseau de neurones supervisé).

Articles, code et notebooks

Voir les articles [1, 2]. Le code est en accès libre, avec des tutoriels et notebooks d'expériences :

Code NTD : **https://gitlab.inria.fr/amarmore/nonnegative-factorization**
Code de traitement et expériences : **https://gitlab.inria.fr/amarmore/musicntd**

Références

[1] A. Marmoret, F. Voorwinden, V. Leplat, J. E. Cohen, and F. Bimbot, "Nonnegative tucker decomposition with beta-divergence for music structure analysis of audio signals," *arXiv preprint arXiv:2110.14434*, 2021.

[2] A. Marmoret, J. Cohen, N. Bertin, and F. Bimbot, "Uncovering audio patterns in music with nonnegative Tucker decomposition for structural segmentation," in *ISMIR*, pp. 788–794, 2020.

[3] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[4] J. Foote, "Automatic audio segmentation using a measure of audio novelty," in *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proc. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*, vol. 1, pp. 452–455, IEEE, 2000.

[5] O. Nieto and T. Jehan, "Convex non-negative matrix factorization for automatic music structure identification," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 236–240, IEEE, 2013.

[6] B. McFee and D. Ellis, "Analyzing song structure with spectral clustering.," in *ISMIR*, pp. 405–410, 2014.

[7] T. Grill and J. Schlüter, "Music boundary detection using neural networks on combined features and two-level annotations.," in *ISMIR*, pp. 531–537, 2015.