



HAL
open science

nmf_audio_benchmark: Benchmarking NMF algorithms on audio tasks

Axel Marmoret

► **To cite this version:**

Axel Marmoret. nmf_audio_benchmark: Benchmarking NMF algorithms on audio tasks. Journée "Synthèse en traitement du signal audio numérique", GDR IASIS, Nov 2024, Paris, France. hal-04753826

HAL Id: hal-04753826

<https://hal.science/hal-04753826v1>

Submitted on 25 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Take home message

Introducing `nmf_audio_benchmark` [1], an open-source toolbox for benchmarking NMF algorithms and variants on audio tasks.

github.com/ax-le/nmf_audio_benchmark/

NMF - Nonnegative Matrix Factorization

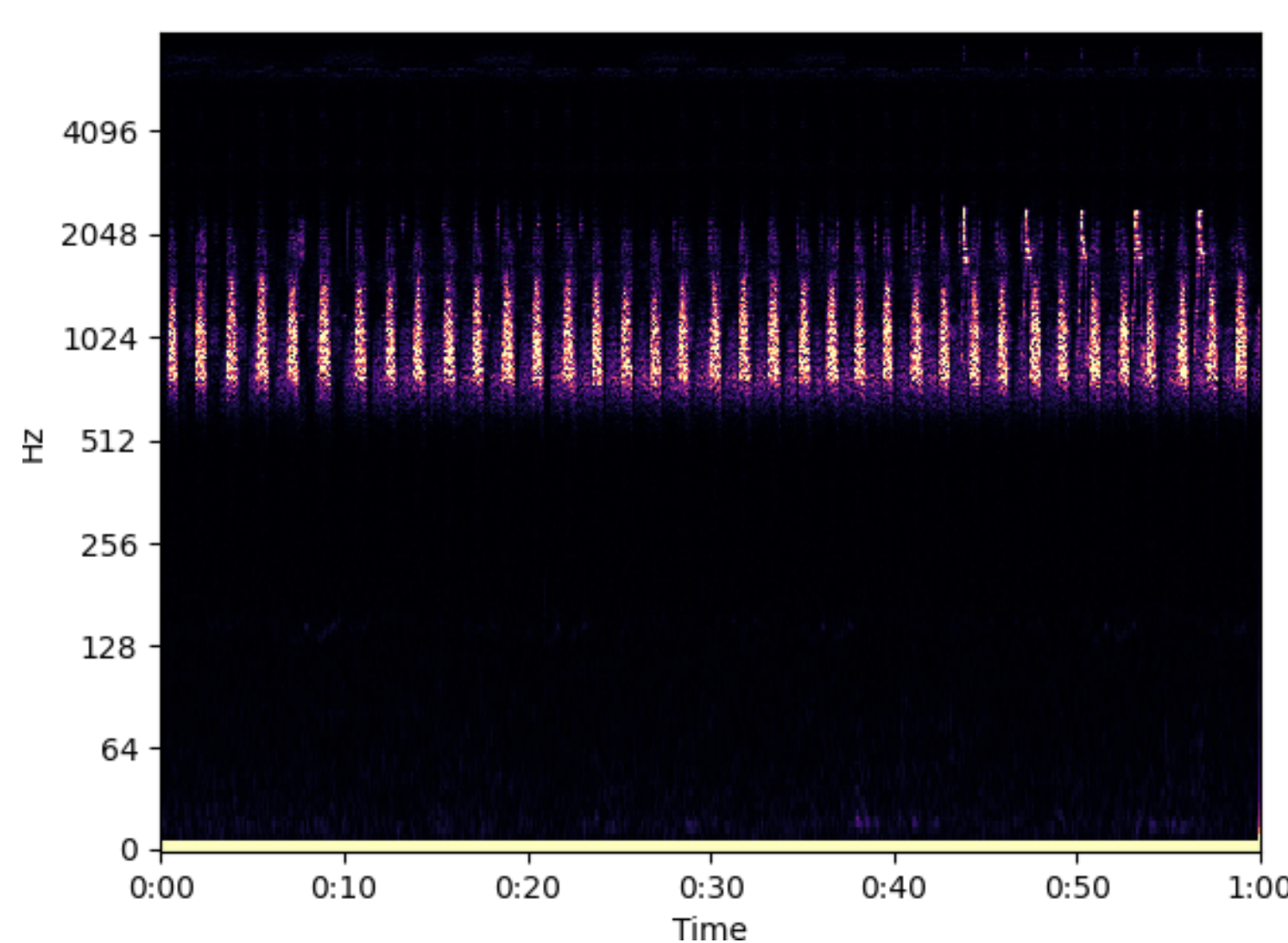
NMF is an unsupervised learning technique that factorizes a matrix V into two nonnegative matrices W and H , such that [2]:

$$V \approx WH \quad (1)$$

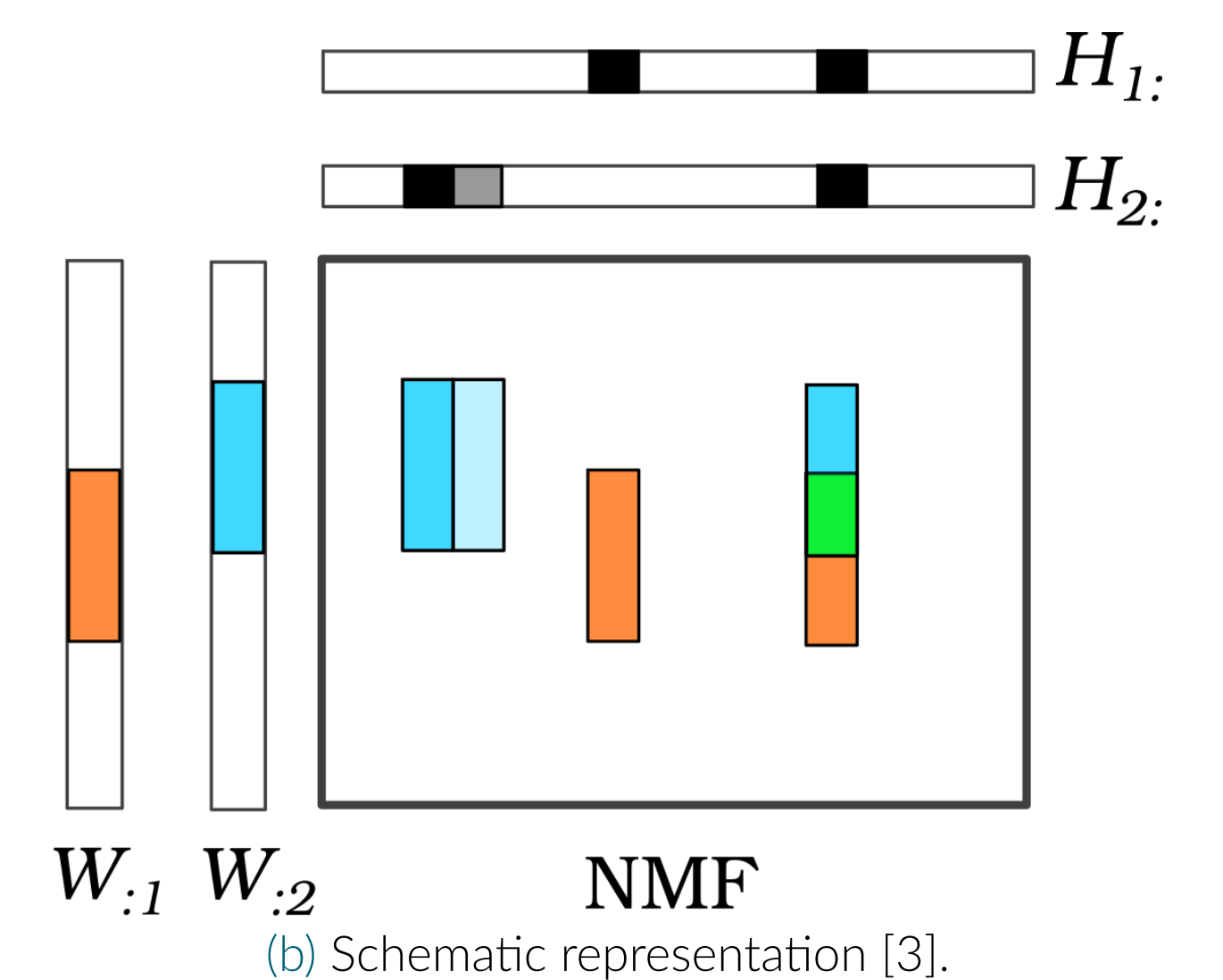
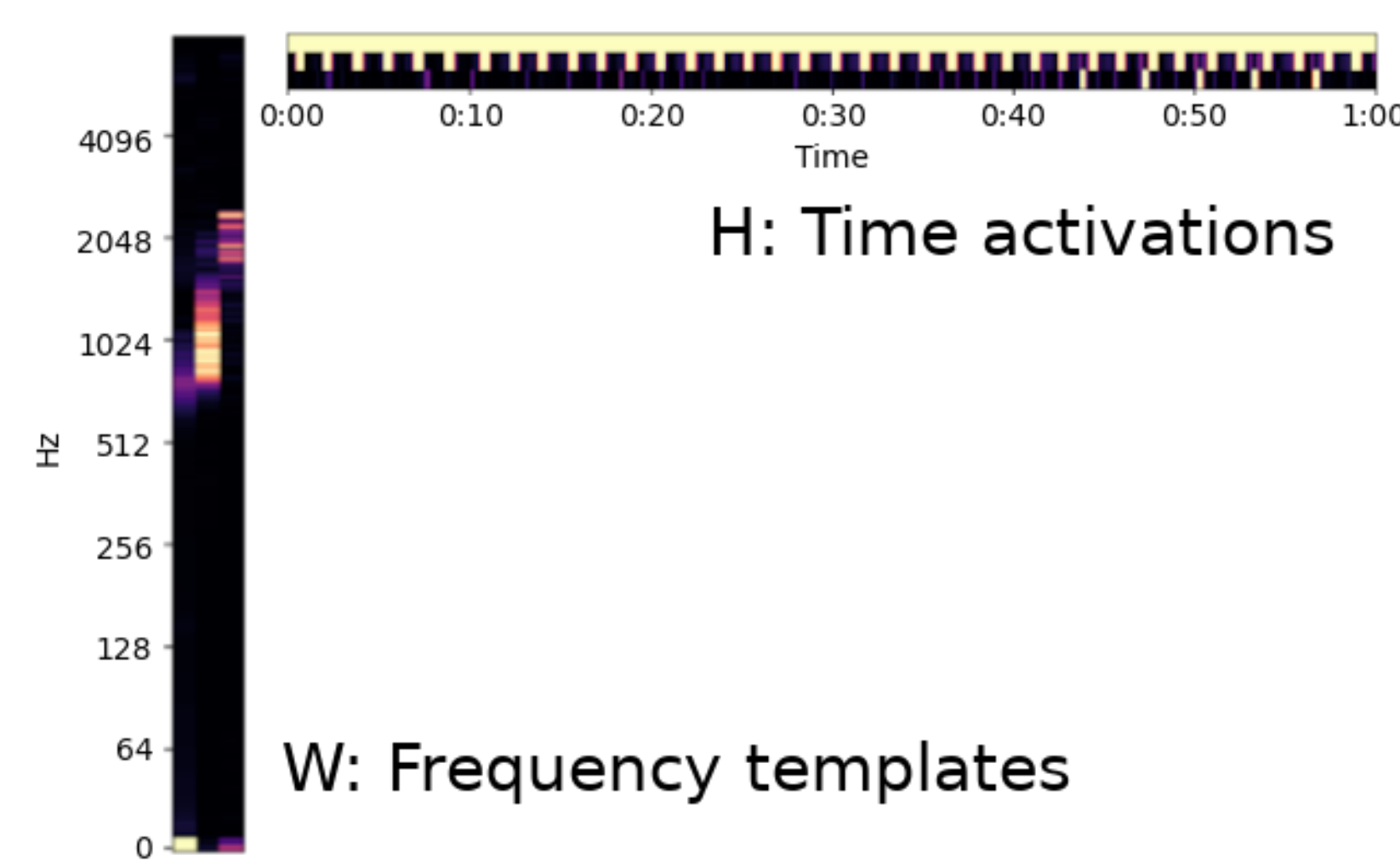
This is obtained by minimizing some distance function $d(\cdot)$ between V and WH :

$$\arg \min_{W, H \geq 0} d(V, WH) \quad (2)$$

For instance, using the Euclidean distance, the optimization problem becomes: $\arg \min_{W, H \geq 0} \sum_{i,j} (V_{ij} - (WH)_{ij})^2$



(a) NMF example.



Why this toolbox?

TL;DR: help researchers in MIR to easily use NMF / help researchers in numerical optimization to test their new NMF models on audio.

1. NMF models still constitute an active research domain

Goal: easy and consistent tool to evaluate a new the model.
In particular, avoid having to re-implement the evaluation process for a standard task.

2. Bridge (again) Audio and Low-Rank Factorization communities.

Goal: reinvigorate the use of NMF in audio processing (long-standing method [4, 5, 6], still used recently [3, 7, 8])

3. In some contexts, NMF is still relevant

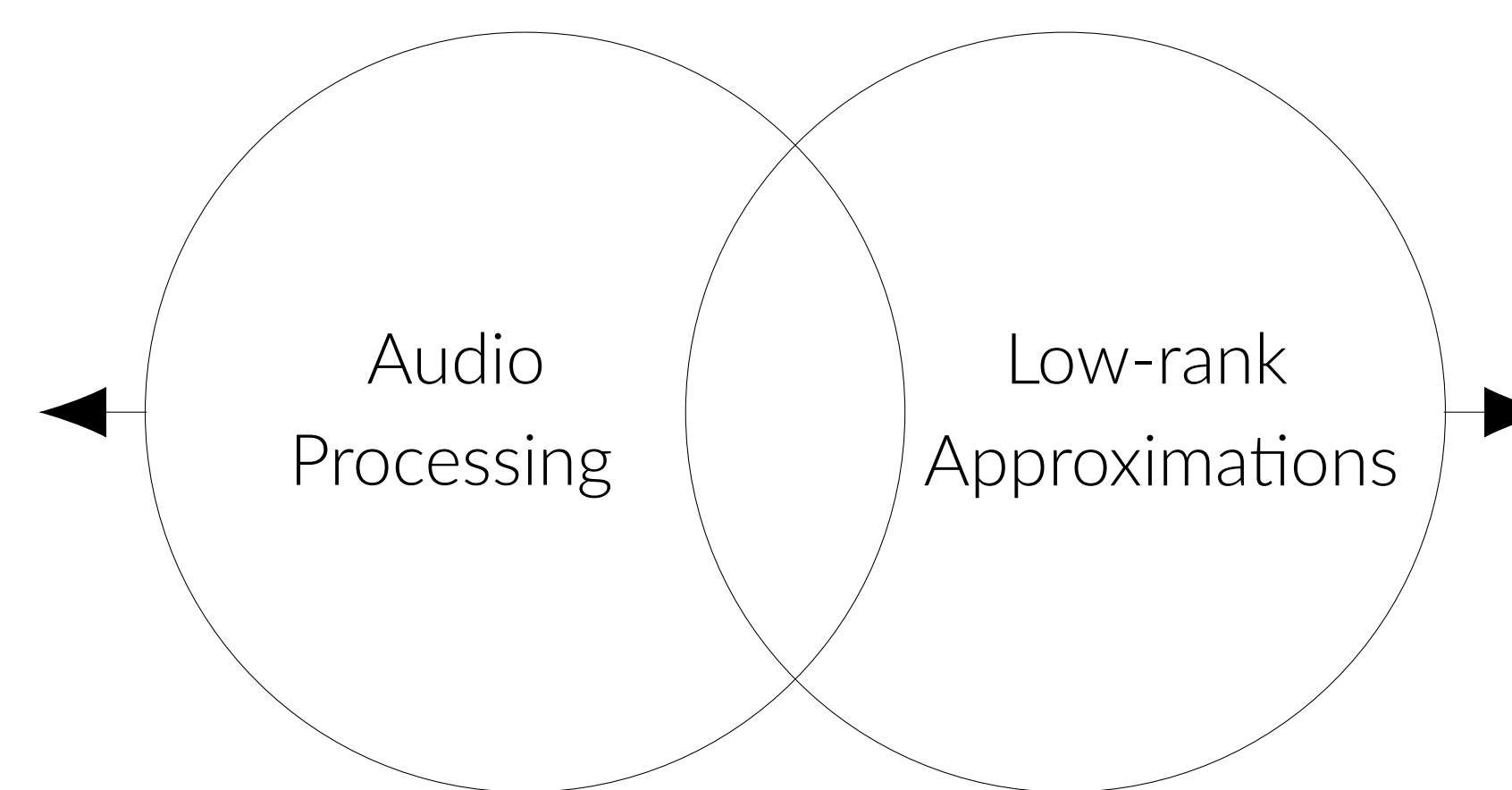
It may help when dealing with a **limited set of data** (common in Ecoacoustics), or “Out-of-distribution” data (NMF is unsupervised). In addition, NMF is a frugal and interpretable method!

What is the state of the toolbox now?

As of now, some tasks are already implemented in the toolbox:

- | | |
|----------------------------|---------------------------|
| MIR: | Ecoacoustics: |
| ▪ Music Transcription | ▪ Blind Source Separation |
| ▪ Blind Source Separation | ▪ Source Count |
| ▪ Music Structure Analysis | ▪ ... and more to come! |

For now, datasets include MAPS, MusDB18, SALAMI, and RWC Pop for MIR, and AnuraSet for Ecoacoustics.



For now, several NMF are implemented:

- $\arg \min_{W, H \geq 0} \|X - WH\|_2^2$ (Euclidean-NMF)
- $\arg \min_{W, H \geq 0} d_{KL}(X, WH)$ (KL-NMF)
- $\arg \min_{W, H \geq 0} d_{IS}(X, WH)$ (IS-NMF)

Details about KL- and IS-NMF in audio processing: [9].

Sometimes (soon!), Sparse-NMF [10] and minimum-volume NMF [11].

Toolbox details

The toolbox is fragmented in **algorithms**, where NMF models are defined, **tasks**, defining how to post-process W and H and defining metrics, and **datasets**, defining the dataloaders. The toolbox relies on the `doce` [12] package to run experiments. It is developed in Python.

Future work

- | | |
|--|--|
| Audio Processing: | Low-Rank Factorizations: |
| ▪ Enhanced tasks (in particular, the Transcription task). | ▪ Constrained models (minimum-volume, simplex-structured, sparse, orthogonal, ...) |
| ▪ Adding new tasks (ex: sound event detection). | ▪ Coupled NMF & tensor models (ex: CP decomposition for multichannel processing). |
| ▪ Expand datasets. | ▪ Computing NMF on GPUs instead of CPUs. |

Off-topic, for ATIAM students: Internship proposal

I propose an internship directly based on this toolbox for the ATIAM students that may be in this room.
The idea will be to use coupled NMF models in a multi-channel context, with different features (ex: factorizing simultaneously STFT, CQT, Mel, ...).
Come talk to me or reach me at axel.marmoret@imt-atlantique.fr if interested!

References

- [1] A. Marmoret, "nmf_audio_benchmark," 2024.
- [2] D. Lee and H. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [3] H. Wu, A. Marmoret, and J. E. Cohen, "Semi-supervised convolutive nmf for automatic music transcription," in *Proceedings of the 19th Sound and Music Computing Conference*, 2022.
- [4] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No. 03TH8684)*, pp. 177–180, IEEE, 2003.
- [5] N. Bertin, R. Badeau, and E. Vincent, "Enforcing harmonicity and smoothness in bayesian non-negative matrix factorization applied to polyphonic music transcription," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 538–549, 2010.
- [6] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 3, pp. 550–563, 2009.
- [7] A. Marmoret, J. E. Cohen, N. Bertin, and F. Bimbot, "Uncovering audio patterns in music with nonnegative Tucker decomposition for structural segmentation," in *International Society for Music Information Retrieval Conference (ISMIR)*, pp. 788–794, 2020.
- [8] J. Parekh, S. Parekh, P. Mozharovskiy, G. Richard, and F. d'Alché Buc, "Tackling interpretability in audio classification networks with non-negative matrix factorization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2024.
- [9] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.
- [10] J. E. Cohen and V. Leplat, "Efficient algorithms for regularized nonnegative scale-invariant low-rank approximation models," *arXiv preprint arXiv:2403.18517*, 2024.
- [11] V. Leplat, N. Gillis, and A. M. Ang, "Blind audio source separation with minimum-volume beta-divergence nmf," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3400–3410, 2020.
- [12] M. Lagrange, "doce."