Blind Audio Source Separation with min.vol. β-divergence NMF

- Andersen Ang (Math et recherche opérationnelle UMONS, Belgium) → (Dept. Combinatorics and Optimization UWaterloo, Canada)
- Joint work with Valentin Leplat (UMONS → UCLouvain) and Nicolas Gillis (UMONS)

- [1] Leplat, V, Gillis, N., Ang, M.S., "*Blind audio source separation with minimum-volume betadivergence NMF*", IEEE Trans. Signal Processing 68, pp.3400-3410, May, 2020. DOI: 10.1109/TSP.2020.2991801
  - arxiv: 1907.02404
  - MATLAB: https://sites.google.com/site/nicolasgillis/publications
  - Youtube - Decomposition of El Doudou song: https://www.youtube.com/watch?v=1BrpxvpghKQ

What

- Single-channel

What

- Single-channel blind

What

- Single-channel blind source separation

What

- Single-channel blind source separation on audio data

- How

What

- Single-channel blind source separation on audio data

- How

  minimum-vol. Nonnegative Matrix Factorization in β-divergence,

What

- Single-channel blind source separation on audio data

- How

  minimum-vol. Nonnegative Matrix Factorization in β-divergence,

  a unsupervised model that

What

- Single-channel blind source separation on audio data

- How

  minimum-vol. Nonnegative Matrix Factorization in β-divergence,

  a unsupervised model that

  - with theoretical identifiability guarantee ☺

What

- Single-channel blind source separation on audio data

- How

  minimum-vol. Nonnegative Matrix Factorization in β-divergence,

  a unsupervised model that

  - with theoretical identifiability guarantee ☺
  - algorithm with theoretical convergence guarantee ☺

What

- Single-channel blind source separation on audio data

- How

  minimum-vol. Nonnegative Matrix Factorization in β-divergence,

  a unsupervised model that

  - with theoretical identifiability guarantee ☺
  - algorithm with theoretical convergence guarantee ☺
  - empirically found that can automatically select model order ☺

What

- Single-channel blind source separation on audio data

- How

  minimum-vol. Nonnegative Matrix Factorization in β-divergence,

  a unsupervised model that

  - with theoretical identifiability guarantee ☺
  - algorithm with theoretical convergence guarantee ☺
  - empirically found that can automatically select model order ☺
  - (the only drawback) not fast

    Improved in V. Leplat, N. Gillis and J. Idier, "Multiplicative Updates for NMF with β-Divergences under Disjoint Equality Constraints", October 2020; arXiv2010.16223

What

- Single-channel blind source separation on audio data

- How

  minimum-vol. Nonnegative Matrix Factorization in β-divergence,

  a unsupervised model that

  - with theoretical identifiability guarantee ☺
  - algorithm with theoretical convergence guarantee ☺
  - empirically found that can automatically select model order ☺
  - (the only drawback) not fast

    Improved in V. Leplat, N. Gillis and J. Idier, "Multiplicative Updates for NMF with β-Divergences under Disjoint Equality Constraints", October 2020; arXiv2010.16223

- The model actually also works for other applications.

What

- Given $x(t) = \sum_{k=1}^{K} s^{(k)}(t)$: observed recording in $\mathbb{R}^T$

- $s^{(k)}(t), k = 1, 2, \ldots, K$: source signals

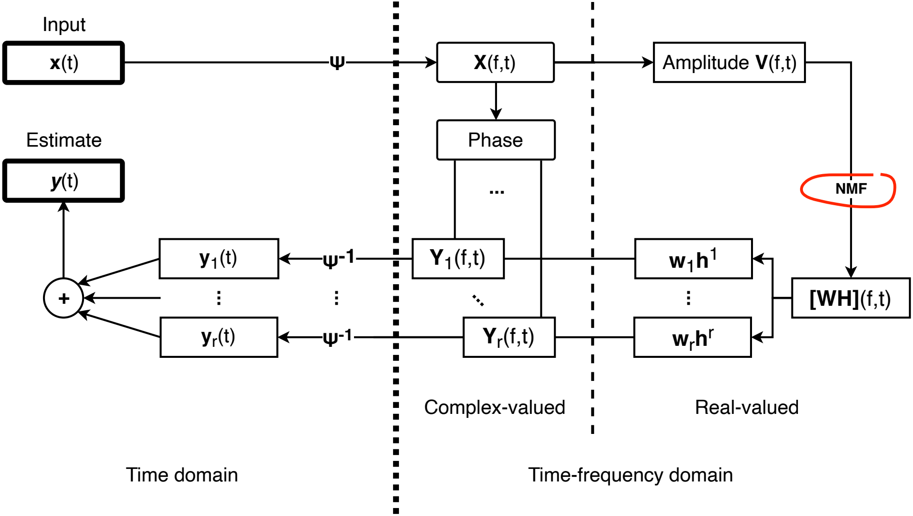
- Goal: find $s^{(k)}$ from $x(t)$

- $x(t) \overset{STFT}{\longrightarrow} X \in \mathbb{C}^{F \times T}$

- Amplitude spectrogram $V = |X| \in \mathbb{R}^{F \times T}$

- BSS: perform NMF on V, assuming
  - Each source $\iff$ each rank-1 component
  - No sound cancellation: NMF

The BSS pipeline



Minvol β-divergence NMF

$$\min_{W,H} \quad D_\beta(V|WH) + \lambda \log\det(W^T W + \delta I_r)$$

$$\text{s.t.} \quad H \geq 0$$

$$W \geq 0$$

$$W(:,j) \in \Delta^{r} \text{ for all } j$$

Minvol β-divergence NMF

β-divergence $d_{\beta=1}(x|y) = x\log\frac{x}{y} - x + y$

constant ≥ 0,

for lower bound

logdet volume

$$\min_{W,H} D_\beta(V|WH) + \lambda \log\det(W^T W + \delta I_r)$$

s.t. $H \geq 0$ Activation ≥ 0

$W \geq 0$ Basis profile ≥ 0

$W(:,j) \in \Delta^K$ for all j Basis normalization

Identifiability Theorem

- **Theorem 1** Assume $V = W^*H^*$ where rank$(V) = K$, $W^* \geq 0$, and H^* satisfies the *sufficiently scattered condition*, then the optimal sol. of
$$\min_{W \geq 0, H \geq 0} \det\left(W^T W\right) \text{ s.t. } V = WH,\ W^T e = e,$$
recovers (W^*, H^*) up to permutation and scaling.

- It is the first result of this type in the audio source separation literature.

- For the DEF of sufficiently scattered condition and the proof of theorem 1, see [1].

Algorithm to solve minvol β-NMF

- Propose an algo. to solve the minvol β-NMF.

$$\begin{aligned} \min_{W \geq 0, H \geq 0} \quad & D_\beta(V|WH) + \lambda \text{logdet}(W^\top W + \delta I_r) \\ \text{s.t.} \quad & H \geq 0, W \geq 0, W(:,j) \in \Delta^K \end{aligned}$$

- Idea: majorization-minimization (MM)

$$f(x) \leq g(x; \theta),$$

where f are the β-divergence and $\text{logdet}(W^\top W + \delta I)$; see [1].

- Objective function monotonically decrease → theoretical convergence.

Mary had a little lamb

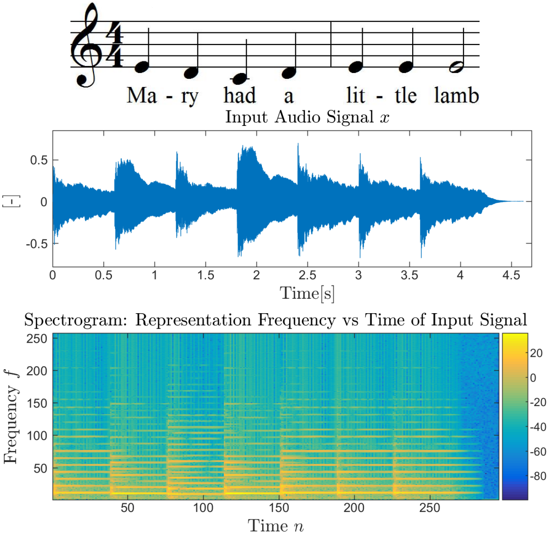


Figure: Three representations of the sample "Mary had a little lamb": (top) music score, (middle) time-domain signal x, and (bottom) log amplitude spectrogram (in dB).

Decomposing Mary had a little lamb

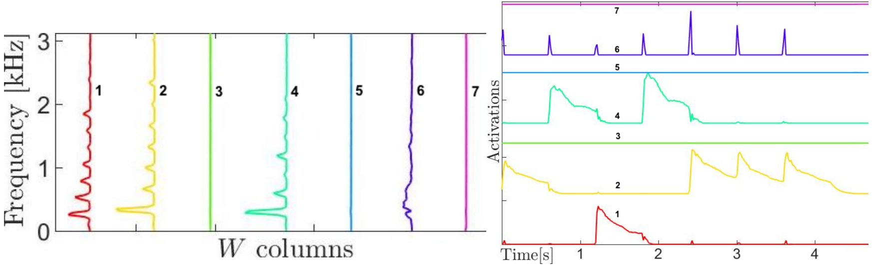


Figure: Minvol β-NMF applied to "Mary had a little lamb" amplitude spectrogram with $K = 7 > 3$. The sources 1,2,4 corresponds to the three notes, and source 6 corresponds to mechanical vibration of the piano.

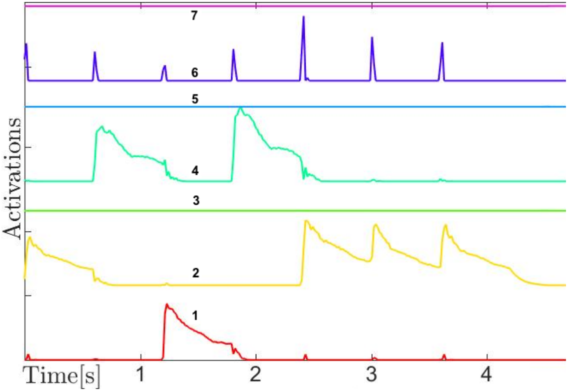
Validating the source estimates

The frequency peaks correspond to the theoretical values.

Notes / Octaves		1-lined	2-lined	3-lined
C	Theoretical	262	523	1046.5
	By NMF	250	531.3	1031
D	Theoretical	294	587	1175
	By NMF	281.3	593.8	1188
E	Theoretical	330	659	1318.5
	By NMF	343.8	656.3	1313

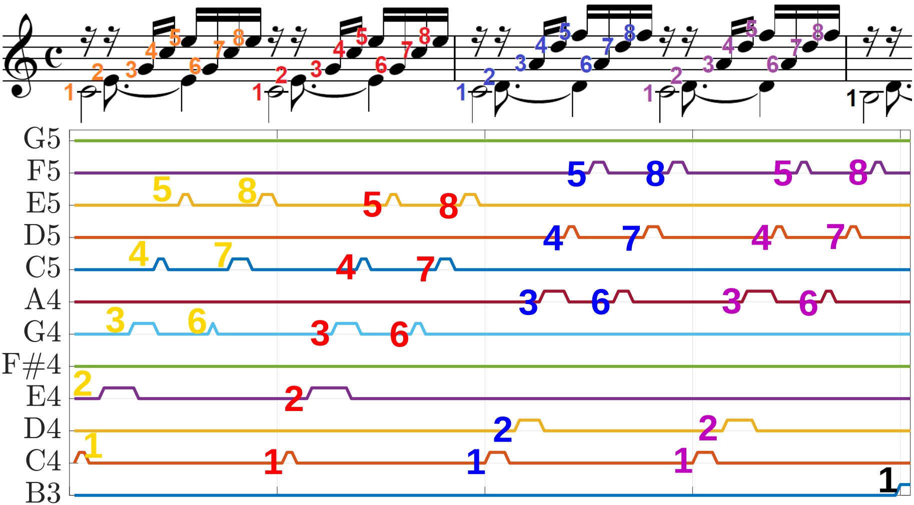
Table: Comparing frequency peaks (Hz) of the octaves obtained by minvol β-NMF

Automatic model order selection



- Note that factorization rank $= 7 > 3 =$ number of sources.
- Two source estimates are zero.
- 6: Hammer noise (of piano)

More complicated example: Prelude by J.S. Bach



* Rows of H here are threshold-ed to make it clear to view.

Last page - summary

[1] Leplat, V, Gillis, N., Ang, M.S., "*Blind audio source separation with minimum-volume betadivergence NMF*", IEEE Trans. Signal Processing 68, 2020.

- Minvol β-NMF
- Single-channel audio BSS
- Identifiability theorem
- MM algorithm with convergence guarantee
- Capacity of automatic model order selection

Slide, paper, code available: angms.science

End of document