Nonnegative Matrix Factorization With Transform Learning

Dylan Fagot, Herwig Wendt and Cédric Févotte

CNRS, IRIT, University of Toulouse

Motivation

- Nonnegative matrix factorization (NMF) can be used to decompose a spectrogram $\mathbf{V} \in \mathbb{R}^{M \times N}$ into two nonnegative latent factors $\mathbf{W} \in \mathbb{R}^{M \times K}$ and $\mathbf{H} \in \mathbb{R}^{K \times N}$ which respectively encode spectral patterns (dictionary) and how these are mixed (activation).
- Results depend heavily on the time-frequency transform used for computing $\mathbf{V}$.
- **Can we learn a transform $\boldsymbol{\Phi}$ so that $\mathbf{V}$ can be well approximated using NMF?**

Baseline : IS-NMF

Audio data

$\mathbf{Y} \in \mathbb{R}^{M \times N}$: matrix that contains N adjacent and overlapping short-time M-wide frames of the sound sample y

IS-NMF with sparsity

Minimize

$$D(|\boldsymbol{\Phi}_{\mathrm{DCT}}\mathbf{Y}|^{\circ 2}|\mathbf{W}\mathbf{H}) + \lambda \frac{M}{K}||\mathbf{H}||_1$$

$$\text{s.t. } \mathbf{W} \geq 0, \mathbf{H} \geq 0, \ \forall k, ||\mathbf{w}_k||_1 = 1 \qquad (1)$$

with $D(\mathbf{A}|\mathbf{B}) = \sum_{ij}(a_{ij}/b_{ij} - \log(a_{ij}/b_{ij}) - 1)$ (Itakura-Saito divergence), factorization rank K

Transform Learning

Proposed TL-NMF (inspired from [1])

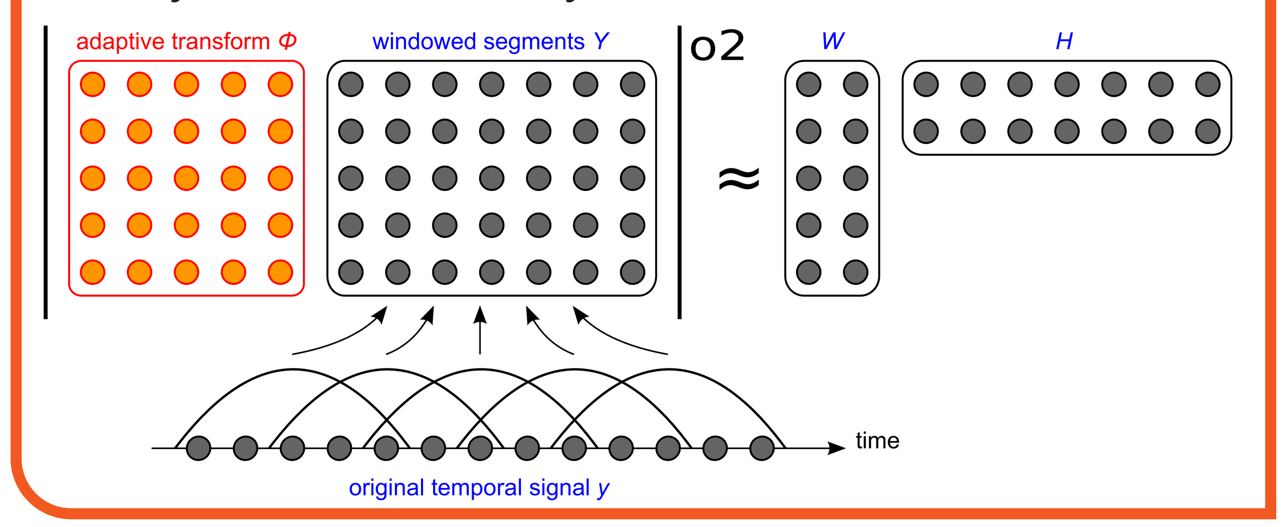
Minimize

$$C_\lambda(\boldsymbol{\Phi}, \mathbf{W}, \mathbf{H}) \overset{\text{def}}{=} D(|\boldsymbol{\Phi}\mathbf{Y}|^{\circ 2}|\mathbf{W}\mathbf{H}) + \lambda \frac{M}{K}||\mathbf{H}||_1$$

$$\text{s.t. } \mathbf{W} \geq 0, \mathbf{H} \geq 0, \ \forall k, ||\mathbf{w}_k||_1 = 1, \boldsymbol{\Phi}^T\boldsymbol{\Phi} = \mathbf{I}_M \qquad (2)$$

Orthogonal constraint on $\boldsymbol{\Phi}$

- Gently departs from $\boldsymbol{\Phi}_{\mathrm{DCT}}$
- Avoids singularity along with trivial solutions such as $(\boldsymbol{\Phi}, \mathbf{W}, \mathbf{H}) = (\mathbf{0}, \mathbf{0}, \mathbf{0})$
- Easy inversion for synthesis



Proposed Algorithm

Algorithm 1: TL-NMF

Input : $\mathbf{Y}, \tau, K, \lambda$
Output: $\boldsymbol{\Phi}$, $\mathbf{W}$, $\mathbf{H}$

Initialize $\boldsymbol{\Phi}$, $\mathbf{W}$ and $\mathbf{H}$ at random
while $\epsilon > \tau$ **do**
  Update $\mathbf{H}$
  Update $\mathbf{W}$
  Update $\boldsymbol{\Phi}$ (new)
  Compute stopping criterion ϵ
end

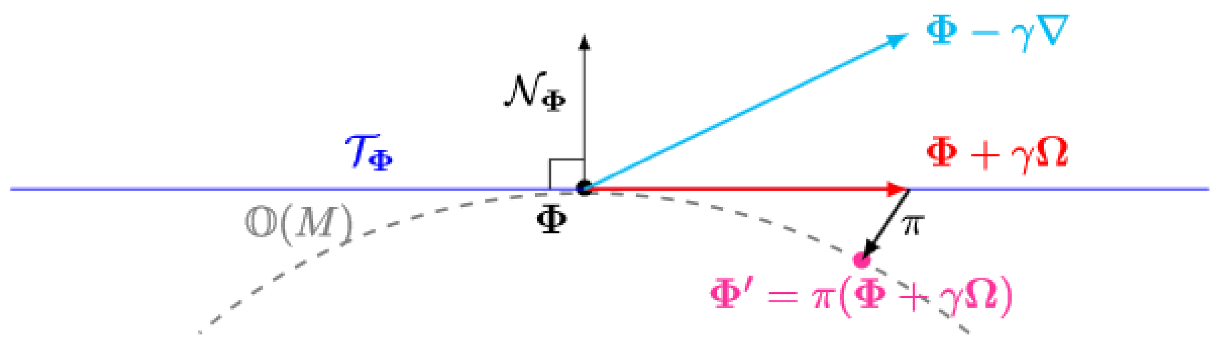
Update of H and W

Majoration-minimization leading to standard mutliplicative updates [2]

Update of $\boldsymbol{\Phi}$

Projected gradient descent onto the orthogonal matrices manifold following [3]

1) Compute gradient ∇ of the objective function
2) Compute natural gradient $\boldsymbol{\Omega} = \boldsymbol{\Phi}\nabla^T\boldsymbol{\Phi} - \nabla$
3) Find a suitable stepsize γ satisfying Armijo rule on the manifold
4) Update the transform via a projection onto the manifold as $\boldsymbol{\Phi} \leftarrow \pi(\boldsymbol{\Phi} + \gamma\boldsymbol{\Omega})$
5) Resolve sign ambiguity on $\boldsymbol{\Phi}$ by imposing its first column entries to be positive



References

[1] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE T. Signal Process.*, 2013.

[2] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Comput.*, 2011.

[3] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE T. Signal Process.*, 2002.

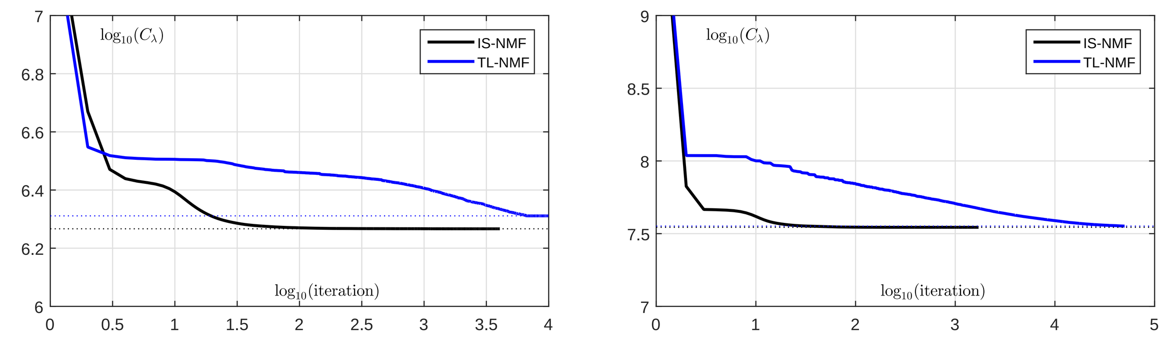
Music Decomposition

Setup

23 s long excerpt from *Mamavatu* by Susheela Raman using 50% overlapping 40 ms-long sine bell windows with factorization rank $K = 10$

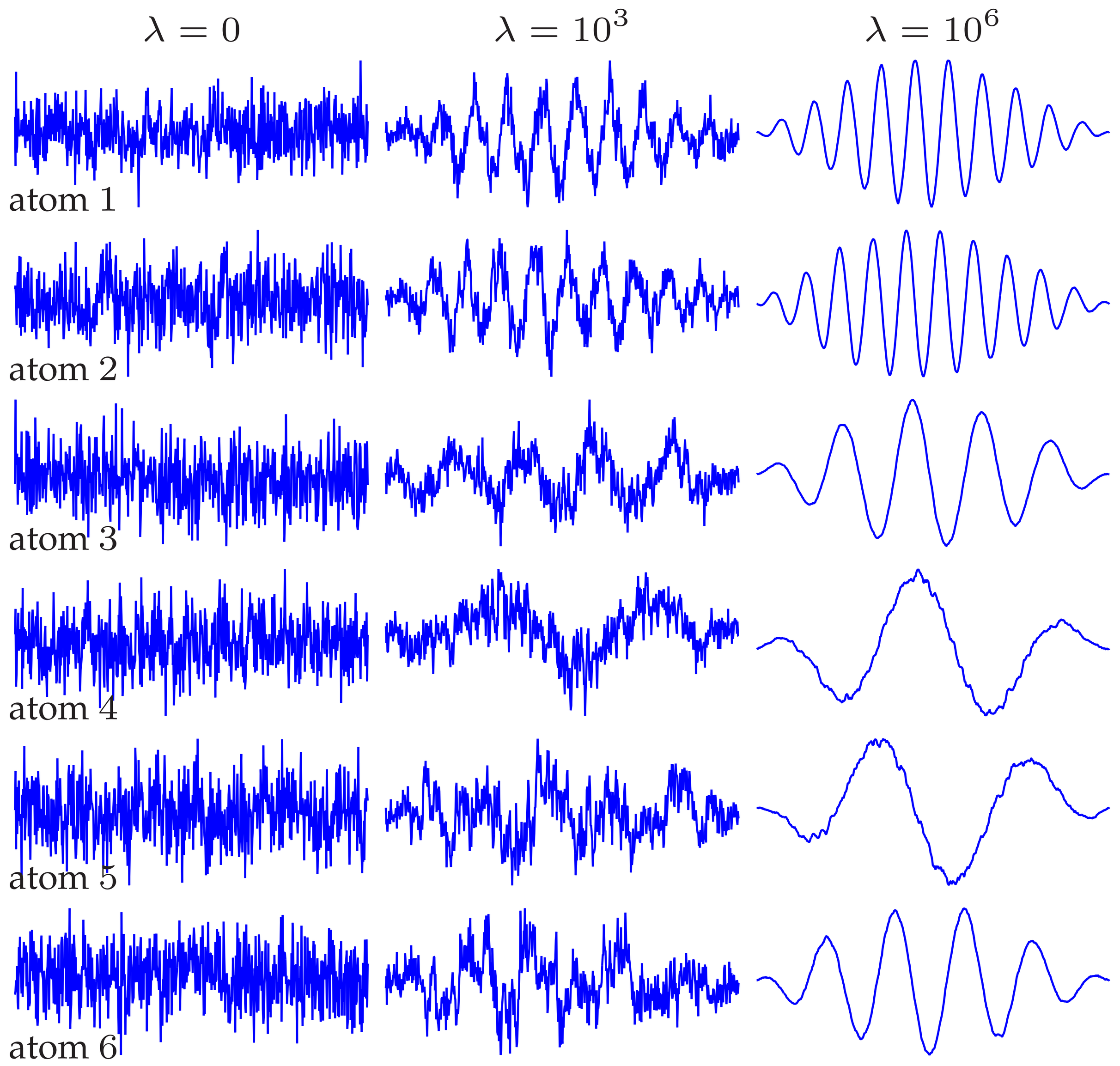
Results

$\lambda = 10^3$ $\lambda = 10^6$



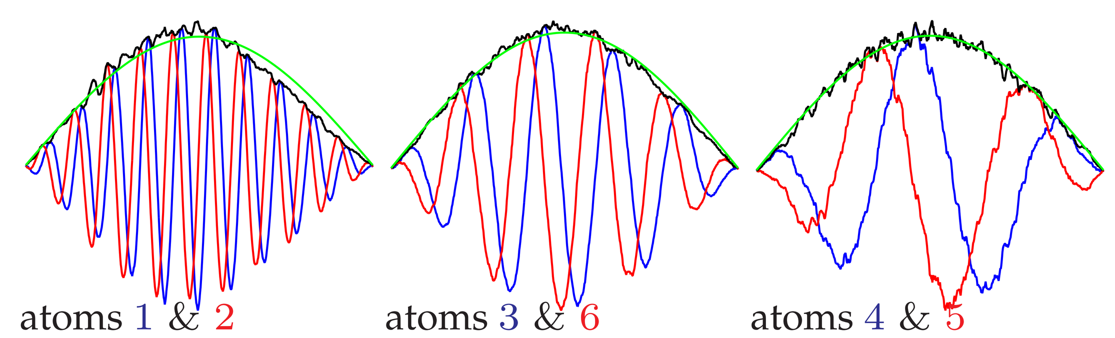
Values of C_λ w.r.t. iterations

TL-NMF reaches similar objective function values despite random initialization



Six most significant atoms learnt by TL-NMF from random initializations

Rows of $\boldsymbol{\Phi}$ become oscillatory and smoother as λ increases



Atoms form pairs in phase quadrature

Supervised Separation

Setup

Separate a sound sample as $y = \hat{y}_{\mathrm{sp}} + \hat{y}_{\mathrm{no}}$ based on the reference data

- $\mathbf{Y}_{\mathrm{sp}} \in \mathbb{R}^{M \times N_{\mathrm{sp}}}$, speech female speaker (21 s)
- $\mathbf{Y}_{\mathrm{no}} \in \mathbb{R}^{M \times N_{\mathrm{no}}}$, bus noise (30 s)

as

$$\mathbf{V} \approx \mathbf{W}_{\mathrm{sp}}\mathbf{H}_{\mathrm{sp}} + \mathbf{W}_{\mathrm{no}}\mathbf{H}_{\mathrm{no}} \qquad (3)$$

where $\mathbf{W}_{\mathrm{sp}} = |\boldsymbol{\Phi}_{\mathrm{DCT}}\mathbf{Y}_{\mathrm{sp}}|^{\circ 2}$, $\mathbf{W}_{\mathrm{no}} = |\boldsymbol{\Phi}_{\mathrm{DCT}}\mathbf{Y}_{\mathrm{no}}|^{\circ 2}$ and $\mathbf{H}_{\mathrm{sp}}, \mathbf{H}_{\mathrm{no}}$ are subject to a sparsity constraint.

Separation with TL-NMF

Minimize

$$C_\Lambda(\boldsymbol{\Phi}, \mathbf{H}_{\mathrm{sp}}, \mathbf{H}_{\mathrm{no}}) \overset{\text{def}}{=} \qquad (4)$$

$$D\left(|\boldsymbol{\Phi}\mathbf{Y}|^{\circ 2} \,\middle|\, |\boldsymbol{\Phi}\mathbf{Y}_{\mathrm{sp}}|^{\circ 2}\mathbf{H}_{\mathrm{sp}} + |\boldsymbol{\Phi}\mathbf{Y}_{\mathrm{no}}|^{\circ 2}\mathbf{H}_{\mathrm{no}}\right)$$

$$+ \lambda_{\mathrm{sp}} \frac{M}{N_{\mathrm{sp}}}||\mathbf{H}_{\mathrm{sp}}||_1 + \lambda_{\mathrm{no}} \frac{M}{N_{\mathrm{no}}}||\mathbf{H}_{\mathrm{no}}||_1$$

$$\text{s.t. } \boldsymbol{\Phi}^T\boldsymbol{\Phi} = \mathbf{I}, \mathbf{H}_{\mathrm{sp}} \geq 0, \mathbf{H}_{\mathrm{no}} \geq 0,$$

$\boldsymbol{\Phi}$ now appears in both sides of the divergence

Results

Sound sample generated by mixing a speech utterance with a bus noise at two different SNR

Comparison using BSS_eval metrics with baseline ($\hat{y}_{\mathrm{sp}} = \hat{y}_{\mathrm{no}} = y/2$) and IS-NMF with sparsity

$\lambda_{\mathrm{sp}} = \lambda_{\mathrm{no}} = \lambda$ was fixed manually

Method	SDR (dB)		SIR (dB)		SAR (dB)	
SNR = -10 dB	$\hat{y}_{\mathrm{sp}}$	$\hat{y}_{\mathrm{no}}$	$\hat{y}_{\mathrm{sp}}$	$\hat{y}_{\mathrm{no}}$	$\hat{y}_{\mathrm{sp}}$	$\hat{y}_{\mathrm{no}}$
Baseline	-9.50	10.00	-9.50	10.00	∞	∞
IS-NMF	-6.75	6.82	-5.00	**13.95**	**4.12**	7.93
TL-NMF	**1.73**	**12.29**	**13.44**	13.33	2.22	**19.20**
SNR = 0 dB	$\hat{y}_{\mathrm{sp}}$	$\hat{y}_{\mathrm{no}}$	$\hat{y}_{\mathrm{sp}}$	$\hat{y}_{\mathrm{no}}$	$\hat{y}_{\mathrm{sp}}$	$\hat{y}_{\mathrm{no}}$
Baseline	0.10	0.08	0.10	0.08	∞	∞
IS-NMF	1.73	0.69	3.06	5.32	**9.30**	3.65
TL-NMF	**6.50**	**5.81**	**12.11**	**9.16**	8.16	**9.00**

Conclusion

- Introduction of transform learning for NMF
- Proposal of a new block-coordinate descent algorithm
- TL-NMF automatically uncovers data-driven oscillatory atoms