TAMPERE UNIVERSITY OF TECHNOLOGY

Towards complex nonnegative matrix factorization with the beta-divergence

Paul Magron, Tuomas Virtanen

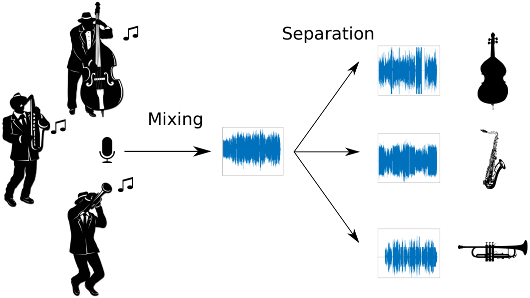
International Workshop on Acoustic Signal Enhancement (iWAENC)

18.09.2018

Outline

Outline

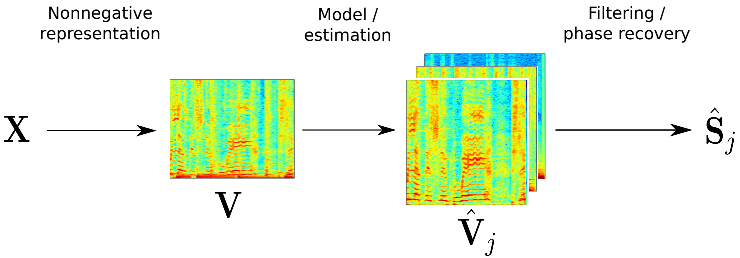
Audio source separation



- Estimate constitutive sources that form a mixture;
- Applications: speech enhancement, augmented musical mixing...
- Challenges: Reduction of **interference** and **artifacts**.

General framework

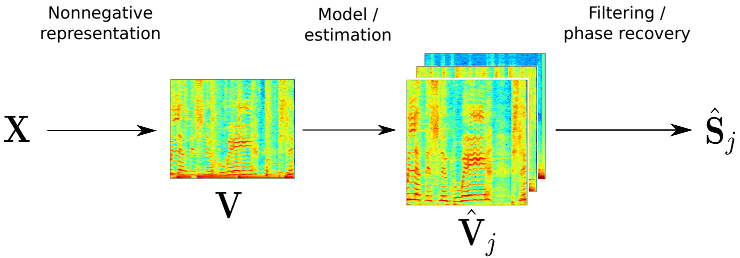
In the STFT domain: $\mathbf{X} = \sum_j \mathbf{S}_j$.



- Nonnegative representation: magnitude/power spectrogram;
- Spectrogram model: KAM, **NMF**, DNNs...
- Complex-valued STFTs retrieval: Wiener-like filtering...

General framework

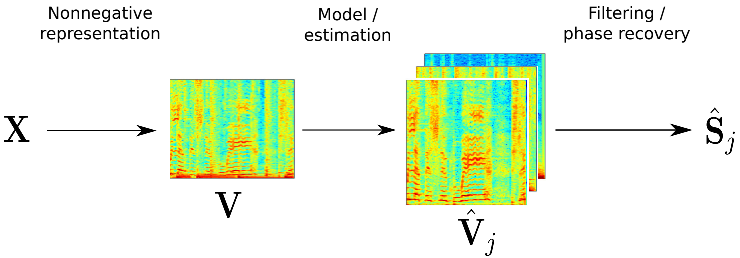
In the STFT domain: $\mathbf{X} = \sum_j \mathbf{S}_j$.



- Nonnegative representation: magnitude/power spectrogram;
- Spectrogram model: KAM, **NMF**, DNNs...
- Complex-valued STFTs retrieval: Wiener-like filtering...

General framework

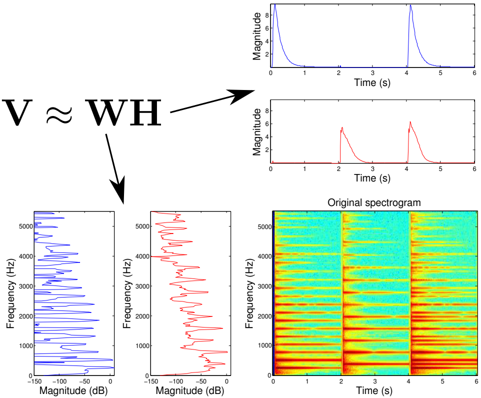
In the STFT domain: $\mathbf{X} = \sum_j \mathbf{S}_j$.



- Nonnegative representation: magnitude/power spectrogram;
- Spectrogram model: KAM, **NMF**, DNNs...
- Complex-valued STFTs retrieval: Wiener-like filtering...

Nonnegative matrix factorization

Find a factorization of a nonnegative matrix $\mathbf{V}$ (e.g., a spectrogram):



NMF - beta-divergences

- Estimation of NMF:

$$\min_{\mathbf{W},\mathbf{H}} D(\mathbf{V}, \mathbf{WH})$$

- Popular choices for D are the beta-divergences D_β:

$\beta = 2$	$\beta = 1$	$\beta = 0$
Euclidean distance	Kullback-Leibler (KL) divergence	Itakura-Saito (IS) divergence
Emphasis on high-energy components	$\leftarrow$ In between $\rightarrow$	Scale invariance

- Optimization techniques (heuristic approach, majorize-minimization...) $\rightarrow$ multiplicative updates rules.

NMF - Wiener filtering

$$\underbrace{\hat{\mathbf{S}}_j}_{\text{Source STFT (Estimated)}} = \underbrace{\frac{\mathbf{W}_j \mathbf{H}_j}{\sum_{k=1}^{J} \mathbf{W}_k \mathbf{H}_k}}_{\text{Mask}} \odot \underbrace{\mathbf{X}}_{\text{Mixture STFT}}$$

- Φ-source = Φ-mixture.

  ☹ Issues in sound quality when sources overlap in the TF domain:

  Mixture Original Oracle Wiener

Complex NMF

Directly models the complex-valued STFTs:

$$\hat{\mathbf{X}} = \sum_{j=1}^{J} \underbrace{\mathbf{W}_j \mathbf{H}_j}_{\text{NMFmagnitude}} \odot \underbrace{e^{i\boldsymbol{\Phi}_j}}_{\text{Phase}}$$

Estimated by minimizing $D_{\mathsf{EUC}}(\mathbf{X}, \hat{\mathbf{X}})$.

☺ Allows to incorporate phase constraints;

☹ Not straightforward to extend it to other beta-divergences.

→ A probabilistic model in which complex NMF can be extended to any beta-divergence.

Complex NMF

Directly models the complex-valued STFTs:

$$\hat{\mathbf{X}} = \sum_{j=1}^{J} \underbrace{\mathbf{W}_j \mathbf{H}_j}_{\text{NMF magnitude}} \odot \underbrace{e^{i\mathbf{\Phi}_j}}_{\text{Phase}}$$

Estimated by minimizing $D_{\mathsf{EUC}}(\mathbf{X}, \hat{\mathbf{X}})$.

- ☺ Allows to incorporate phase constraints;
- ☹ Not straightforward to extend it to other beta-divergences.

→ **A probabilistic model in which complex NMF can be extended to any beta-divergence.**

Outline

Gaussian model

Mixture in each TF bin:

$$x = \sum_{j=1}^{J} s_j$$

Gaussian sources:

$$s_j \sim \mathcal{N}(0, \Gamma_j) \text{ with } \Gamma_j = \begin{pmatrix} \gamma_j & c_j \\ \bar{c}_j & \gamma_j \end{pmatrix}$$

- γ_j = variance $\rightarrow$ energy of the source.
- c_j = relation term $\rightarrow$ non-uniformity of the phase.

Isotropic Gaussian sources

- Traditionally, circularly-symmetric (=**isotropic**) sources:

$$c_j = 0 \Leftrightarrow \text{uniform phase.}$$

- NMF variance: $\gamma_j = \mathbf{W}_j \mathbf{H}_j$.
- Estimation:

$$\text{maximum likelihood} \Leftrightarrow \min D_{\text{IS}}(|\mathbf{X}|^{\odot 2}, \mathbf{WH})$$

$$\rightarrow \text{ISNMF model.}$$

Anisotropic Gaussian model

Here, **anisotropic** sources:

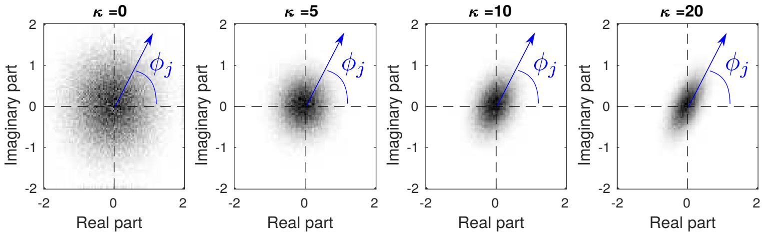
$$c_j \neq 0 \Leftrightarrow \text{non-uniform phase.}$$

The variance / relation terms are:

$$\gamma_j = (1 - \lambda^2)\mathbf{W}_j\mathbf{H}_j$$
$$c_j = \rho\mathbf{W}_j\mathbf{H}_j e^{i2\phi_j}$$

- ϕ_j = phase location parameter.
- λ/ρ = functions of κ, quantify the non-uniformity of the phase.
  - $\kappa = 0 \rightarrow \gamma_j = [\mathbf{W}_j\mathbf{H}_j]$ and $c_j = 0 \rightarrow$ ISNMF

Anisotropic Gaussian model



Inference

The likelihood is not tractable $\rightarrow$ Expectation-maximization (EM).

- Maximize a lower bound Q of the likelihood.
- Alternate between:
  - E-step: compute Q given the current parameter estimates.
  - M-step: maximize Q to update the parameters.

Phase-corrected posterior power

The EM functional rewrites:

$$Q \stackrel{c}{=} -\sum_{j=1}^{J} \log(\mathbf{W}_j\mathbf{H}_j) + \frac{\mathbf{P}_j}{\mathbf{W}_j\mathbf{H}_j} \stackrel{c}{=} -\sum_{j=1}^{J} D_{\mathsf{IS}}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j)$$

with:

$$\mathbf{P}_j \propto \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \Gamma_j^{-1} \underline{s}_j\right) > 0$$

- $\mathbf{P}_j$ is phase-aware;
- $\kappa = 0 \rightarrow \mathbf{P}_j = \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \underline{s}_j\right)$ = posterior power of s_j.

$\mathbf{P}_j$ = **phase-corrected posterior power** of the j-th source:

Phase-corrected posterior power

The EM functional rewrites:

$$Q \stackrel{c}{=} -\sum_{j=1}^{J} \log(\mathbf{W}_j\mathbf{H}_j) + \frac{\mathbf{P}_j}{\mathbf{W}_j\mathbf{H}_j} \stackrel{c}{=} -\sum_{j=1}^{J} D_{\mathsf{IS}}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j)$$

with:

$$\mathbf{P}_j \propto \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \Gamma_j^{-1} \underline{s}_j\right) > 0$$

- $\mathbf{P}_j$ is phase-aware;
- $\kappa = 0 \rightarrow \mathbf{P}_j = \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \underline{s}_j\right)$ = posterior power of s_j.

$\mathbf{P}_j$ = **phase-corrected posterior power** of the j-th source:

Phase-corrected posterior power

The EM functional rewrites:

$$Q \overset{c}{=} -\sum_{j=1}^{J} \log(\mathbf{W}_j\mathbf{H}_j) + \frac{\mathbf{P}_j}{\mathbf{W}_j\mathbf{H}_j} \overset{c}{=} -\sum_{j=1}^{J} D_{\mathsf{IS}}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j)$$

with:

$$\mathbf{P}_j \propto \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \Gamma_j^{-1} \underline{s}_j\right) > 0$$

- $\mathbf{P}_j$ is phase-aware;
- $\kappa = 0 \rightarrow \mathbf{P}_j = \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \underline{s}_j\right)$ = posterior power of s_j.

$\mathbf{P}_j$ = **phase-corrected posterior power** of the j-th source:

Phase-corrected posterior power

The EM functional rewrites:

$$Q \stackrel{c}{=} -\sum_{j=1}^{J} \log(\mathbf{W}_j\mathbf{H}_j) + \frac{\mathbf{P}_j}{\mathbf{W}_j\mathbf{H}_j} \stackrel{c}{=} -\sum_{j=1}^{J} D_{\mathsf{IS}}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j)$$

with:

$$\mathbf{P}_j \propto \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \Gamma_j^{-1} \underline{s}_j\right) > 0$$

- $\mathbf{P}_j$ is phase-aware;
- $\kappa = 0 \rightarrow \mathbf{P}_j = \mathbb{E}_{\text{post}}\left(\underline{s}_j^H \underline{s}_j\right)$ = posterior power of s_j.

$\mathbf{P}_j$ = **phase-corrected posterior power** of the j-th source:

Introducing the beta-divergence

- Minimization of $D_{\mathsf{IS}}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j) \rightarrow$ "Complex ISNMF".

Proposed approach: we replace the IS divergence with the beta-divergence.

- Heuristic $\rightarrow$ no more convergence guarantee.
- Minimization of $D_{\beta}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j) \rightarrow$ multiplicative updates.

$\rightarrow$ **Complex βNMF**

☺ Phase-aware decomposition of the data;

☺ Uses a distortion metric adapted to audio.

Introducing the beta-divergence

- Minimization of $D_{\mathsf{IS}}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j) \to$ "Complex ISNMF".

Proposed approach: we replace the IS divergence with the beta-divergence.

- Heuristic $\to$ no more convergence guarantee.
- Minimization of $D_{\beta}(\mathbf{P}_j, \mathbf{W}_j\mathbf{H}_j) \to$ multiplicative updates.

$\to$ **Complex βNMF**

☺ Phase-aware decomposition of the data;

☺ Uses a distortion metric adapted to audio.

Outline

Setup

Harmonic/percussive source separation task ($J = 2$ sources).

Dataset:

- DSD100: 100 music songs, split into learning/test sets;
- Each excerpt is 20 seconds long.

Supervised separation scenario:

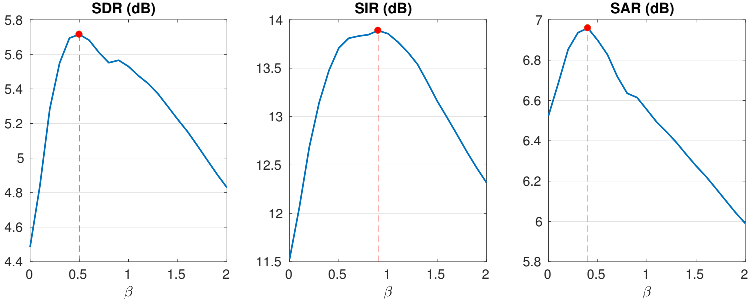
- Each excerpt is split into two signals of 10 seconds.
- The first is used for learning dictionaries $\mathbf{W}_j$ (k-means clustering with 50 basis per dictionary).
- The second is used for performing the separation.

Source separation quality:

- Signal-to-distortion/interference/artifact ratios (SDR, SIR, and SAR).

Learning β

On the learning set:



- Trade-off between interference ($\beta = 0.9$) and artifacts ($\beta = 0.4$) reduction.
- $\beta = 0.5 \rightarrow$ best overall distortion reduction.

Comparison to other approaches

Baselines: NMF (also uses $\beta = 0.5$) and complex Euclidean NMF.

Median results over the test dataset:

	SDR	SIR	SAR	
NMF	5.4	12.3	7.1	
Complex Euclidean NMF	2.0	8.6	3.5	
Complex βNMF	**5.5**	**12.6**	**7.2**	

- Complex βNMF outperforms its Euclidean counterpart;
- Slightly better results than NMF.

$\rightarrow$ Complex βNMF allows to incorporate phase constraints.

Conclusion

A novel probabilistic framework where complex NMF can be extended to any beta-divergence.

Future work:

- Better understanding of the phase-corrected posterior power;
- Phase-constrained complex βNMF;
- Alternative probabilistic model $\rightarrow$ convergence guarantees;
- Use DNNs instead of NMF for joint magnitude/phase estimation.

  cf. some other IWAENC papers!

Conclusion

A novel probabilistic framework where complex NMF can be extended to any beta-divergence.

Future work:

- Better understanding of the phase-corrected posterior power;
- Phase-constrained complex βNMF;
- Alternative probabilistic model $\rightarrow$ convergence guarantees;
- Use DNNs instead of NMF for joint magnitude/phase estimation.

  cf. some other IWAENC papers!

Conclusion

A novel probabilistic framework where complex NMF can be extended to any beta-divergence.

Future work:

- Better understanding of the phase-corrected posterior power;
- Phase-constrained complex βNMF;
- Alternative probabilistic model $\rightarrow$ convergence guarantees;
- Use DNNs instead of NMF for joint magnitude/phase estimation.

  cf. some other IWAENC papers!

Conclusion

A novel probabilistic framework where complex NMF can be extended to any beta-divergence.

Future work:

- Better understanding of the phase-corrected posterior power;
- Phase-constrained complex βNMF;
- Alternative probabilistic model $\rightarrow$ convergence guarantees;
- Use DNNs instead of NMF for joint magnitude/phase estimation.

  cf. some other IWAENC papers!

Thanks!

https://github.com/magronp/complex-beta-nmf