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

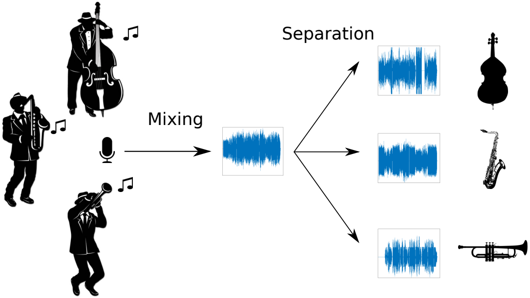
Outline

1 Background and problem setting

2 Complex NMF

3 Experimental results

Audio source separation

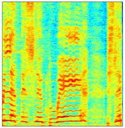


- Estimate constitutive sources that form a mixture;
- Applications: speech enhancement, augmented musical mixing...
- Challenges: Reduction of **interference** and **artifacts**.

General framework

In the STFT domain: $X = \sum_j S_j$.

Nonnegative representation | **Model / estimation** | **Filtering / phase recovery**

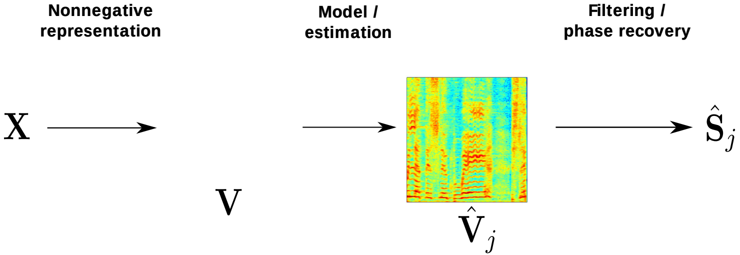
$\mathbf{X} \longrightarrow \mathbf{V} \longrightarrow$  $\longrightarrow \hat{\mathbf{S}}_j$

$\hat{\mathbf{V}}_j$

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

General framework

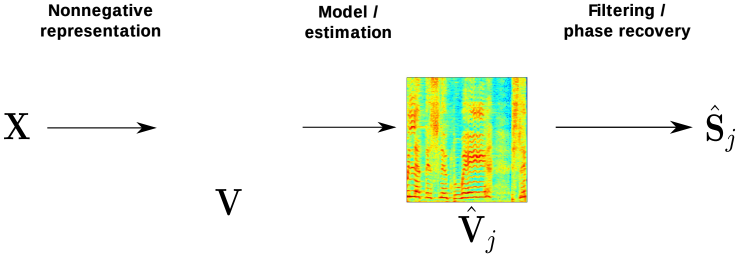
In the STFT domain: $\mathbf{X} = \sum_j \mathbf{S}_j$.



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

General framework

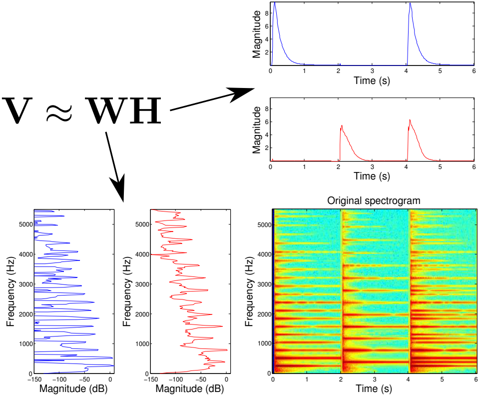
In the STFT domain: $X = \sum_j S_j$.



- 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:

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

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

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

- Optimization techniques (heuristic approach, majorize-minimization...) ! multiplicative updates rules.

NMF - Wiener filtering

$$\hat{S}_j = \frac{W_j H_j}{\sum_{k=1}^{J} W_k H_k} X$$

Source STFT (Estimated) — Mask — Mixture STFT

- -source = -mixture.

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

  Mixture Original Oracle Wiener

Complex NMF

Directly models the complex-valued STFTs:

$$\hat{X} = \sum_{j=1}^{J} \underbrace{W_j H_j}_{\text{NMF magnitude}} \underbrace{e^{i\phi_j}}_{\text{Phase}}$$

Estimated by minimizing $D_{EUC}(X; \hat{X})$.

- 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:

$$\hat{X} = \sum_{j=1}^{J} \underbrace{W_j H_j}_{\text{NMF magnitude}} \underbrace{e^{i\phi_j}}_{\text{Phase}}$$

Estimated by minimizing $D_{EUC}(X;\hat{X})$.

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:

$$s_j \quad N \quad (0; \ _j) \text{ with } \ _j = \begin{matrix} _j & c_j \\ c_j & _j \end{matrix}$$

- $_j$ = variance ! energy of the source.
- c_j = relation term ! non-uniformity of the phase.

Isotropic Gaussian sources

- Traditionally, circularly-symmetric ("isotropic") sources:

$$c_j = 0\ , \quad \text{uniform phase.}$$

- NMF variance: $_j = W_j H_j$.
- Estimation:

$$\text{maximum likelihood}, \quad \min D_{IS}(|X|^2; WH)$$

$$! \quad \text{ISNMF model.}$$

Anisotropic Gaussian model

Here, anisotropic sources:

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

The variance / relation terms are:

$$_j = (1 \qquad ^2) W_j H_j$$
$$c_j = \quad W_j H_j e^{i2 \quad _j}$$

- $_j$ = phase location parameter.
- = = functions of , quantify the non-uniformity of the phase.
  - = 0 ! $_j = [W_j H_j]$ and $c_j = 0$! ISNMF

Anisotropic Gaussian model

Inference

The likelihood is not tractable ! 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:

$$Q \stackrel{c}{=} \sum_{j=1}^{J} \log(W_j H_j) + \frac{P_j}{W_j H_j} \stackrel{c}{=} \sum_{j=1}^{J} D_{IS}(P_j ; W_j H_j)$$

with:

$$P_j \, / \, E_{post}\left(\underline{s}_j^H \, {}_j^{\,1} \, \underline{s}_j \right) > 0$$

- P_j is phase-aware;
- $= 0 \; ! \quad P_j = E_{post}\left(\underline{s}_j^H \underline{s}_j \right)$ = posterior power of s_j .

P_j = phase-corrected posterior power of the j-th source:

Phase-corrected posterior power

The EM functional rewrites:

$$Q \overset{c}{=} \sum_{j=1}^{J} \log(W_j H_j) + \frac{P_j}{W_j H_j} \overset{c}{=} \sum_{j=1}^{J} D_{IS}(P_j ; W_j H_j)$$

with:

$$P_j \,/\, E_{post}\; \underline{s}_j^H \;{}_j{}^{1} \underline{s}_j \; > 0$$

- P_j is phase-aware;
- $= 0 \;!\;\; P_j = E_{post}\; \underline{s}_j^H \underline{s}_j$ = posterior power of s_j .

P_j = phase-corrected posterior power of the j-th source:

Phase-corrected posterior power

The EM functional rewrites:

$$Q \overset{c}{=} \sum_{j=1}^{J} \log(W_j H_j) + \frac{P_j}{W_j H_j} \overset{c}{=} \sum_{j=1}^{J} D_{IS}(P_j ; W_j H_j)$$

with:

$$P_j \propto \mathbb{E}_{post}\left(\underline{s}_j^H \; {}_j^{\,1} \underline{s}_j \right) > 0$$

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

P_j = phase-corrected posterior power of the j-th source:

Phase-corrected posterior power

The EM functional rewrites:

$$Q \stackrel{c}{=} \sum_{j=1}^{J} \log(W_j H_j) + \frac{P_j}{W_j H_j} \stackrel{c}{=} \sum_{j=1}^{J} D_{IS}(P_j ; W_j H_j)$$

with:

$$P_j \; / \; E_{post} \; \underline{s}_j^H \; {}_j \, {}^{1} \underline{s}_j \; > \; 0$$

- P_j is phase-aware;
- $= 0 \; ! \; P_j = E_{post} \; \underline{s}_j^H \underline{s}_j$ = posterior power of s_j .

P_j = phase-corrected posterior power of the j -th source:

Introducing the beta-divergence

- Minimization of $D_{IS}(P_j; W_j H_j)$! Complex ISNMF".

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

- Heuristic ! no more convergence guarantee.
- Minimization of $D\ (P_j; W_j H_j)$! multiplicative updates.

! Complex NMF

Phase-aware decomposition of the data;

Uses a distortion metric adapted to audio.

Introducing the beta-divergence

- Minimization of $D_{IS}(P_j; W_j H_j)$! Complex ISNMF".

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

- Heuristic ! no more convergence guarantee.
- Minimization of $D\ (P_j; W_j H_j)$! multiplicative updates.

! Complex NMF

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:

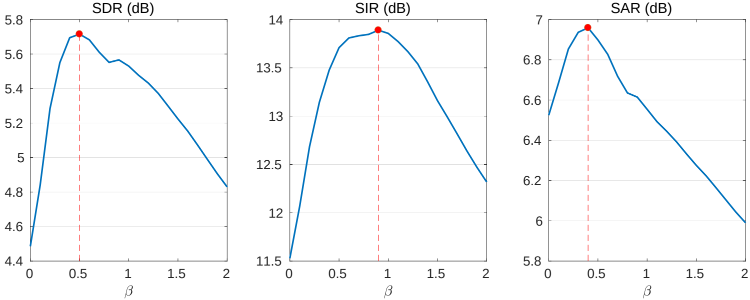
- Each excerpt is split into two signals of 10 seconds.
- The first is used for learning dictionaries W_j (k-means clustering with 50 basis per dictionary).
- The second is used for performing the separation.

Source separation quality:

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

Learning β

On the learning set:



- Trade-off between interference (= 0:9) and artifacts (= 0:4) reduction.
- = 0:5 ! best overall distortion reduction.

Comparison to other approaches

Baselines: NMF (also uses = 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 NMF	**5.5**	**12.6**	**7.2**

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

  ! Complex NMF 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 NMF;
- Alternative probabilistic model ! 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 NMF;
- Alternative probabilistic model ! 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 NMF;
- Alternative probabilistic model *!* 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 NMF;
- Alternative probabilistic model ! 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