
Institut
Mines-Telecom

# Phase recovery in NMF for source separation: an insightful benchmark

Paul Magron, Roland Badeau, Bertrand David

Télécom ParisTech, CNRS LTCI

April 22, 2015

# Context

Humans can focus on a specific part of a music excerpt.

- Source separation $\rightarrow$ Reproduction of this ability.

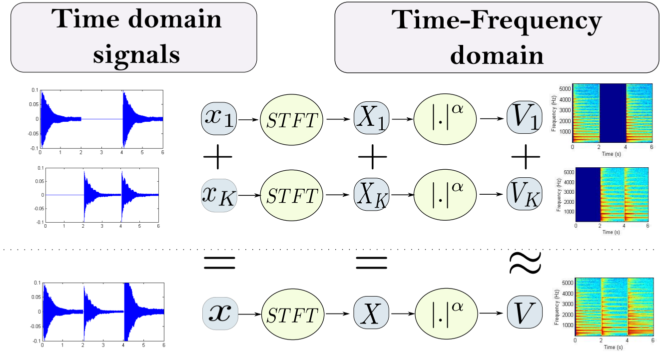
Approaches:

- Exploiting redundancies: PCA, ICA, sparse coding...
- Nonnegative Matrix Factorization (NMF) provides a decomposition intuitively interpretable.

NMF acts only on spectrograms:

- The phase needs to be reconstructed.
- Wiener filtering is commonly used.
- But it does not enforce *consistency*: the obtained complex-valued matrix is not the Short-Term Fourier Transform (STFT) of a time signal.

# Mixture model




- Generally $V = |X|$ or $|X|^2$.
- Assumption of an additivity property: $V = \sum_{k=1}^{K} V_k$.



TELECOM ParisTech

# NMF

NMF Model:

- $V \approx \hat{V}$ with $\hat{V} = WH$ [Lee and Seung, 1999].
- $W$ and $H$ are nonnegative matrices of rank $K \ll F, T$.

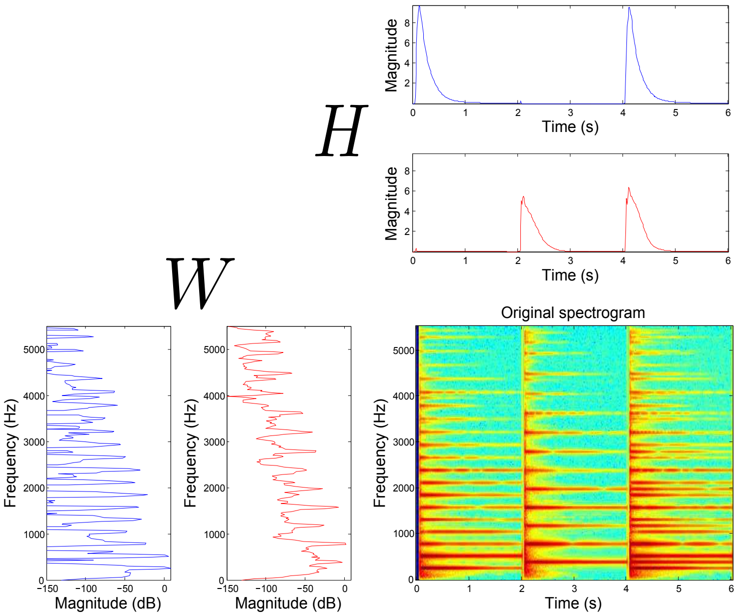
Estimation:

- Minimization of a cost function $D(V, WH)$.
- Popular choices:

  Euclidean distance,
  Kullback-Leibler divergence [Lee and Seung, 2001],
  Itakura-Saito divergence [Févotte et al., 2009].

- Multiplicative update rules.

# NMF


H
Magnitude
Time (s)
W
Frequency (Hz)
Magnitude (dB)
Original spectrogram

# Phase reconstruction

## Wiener filtering

Each estimated component is given the phase of the mixture:

$$X_k = \frac{W_k H_k}{\sum_{l=1}^{K} W_l H_l} X = \frac{\hat{V}_k}{\hat{V}} X.$$

Inaccurate when sources overlap in the Time-Frequency (TF) domain.

Example:

| | Mixture | Source 1 | Source 2 |
|---|---|---|---|
| Original | | | |
| Estimated | | | |

# Outline

Overview of the compared methods

- NMF + phase reconstruction algorithm
- NMF with phase estimation

The benchmark

- Methodology
- Results

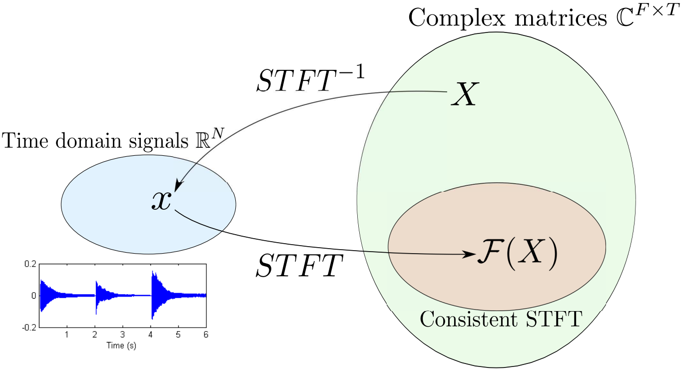
# Outline

Overview of the compared methods
- NMF + phase reconstruction algorithm
- NMF with phase estimation

The benchmark

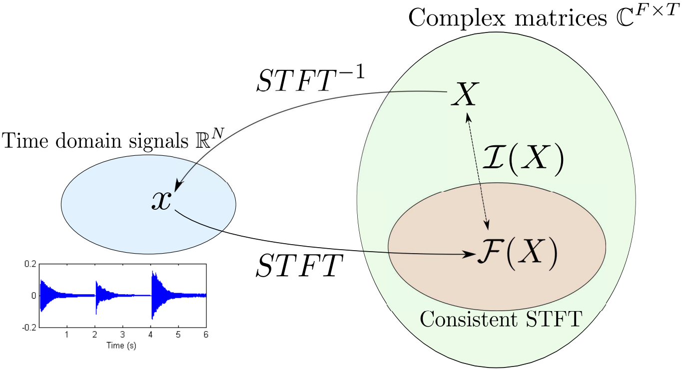
# Consistency-based approaches

**STFT**: $\mathbb{R}^N \to \mathbb{S}^{F\times T} \subset \mathbb{C}^{F\times T}$



# Consistency-based approaches

**STFT**: $\mathbb{R}^N \rightarrow \mathbb{S}^{F\times T} \subset \mathbb{C}^{F\times T}$



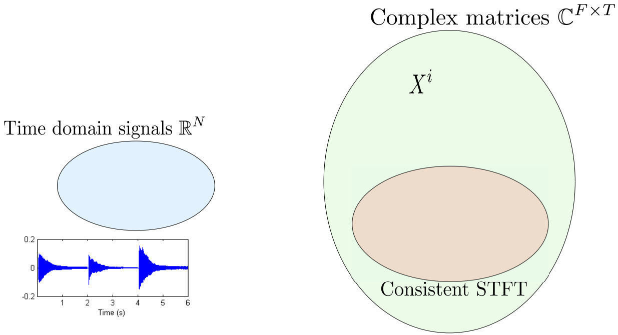

**Inconsistency**: $\mathcal{I}(X) = ||X - \mathcal{F}(X)||$ where:

- $\mathcal{F} = STFT \circ STFT^{-1}$.
- $||.||$ is the Euclidean norm.

# Consistency-based approaches

**Griffin Lim** [Griffin and Lim, 1984]

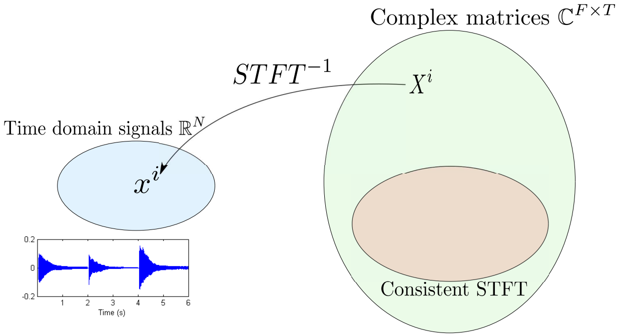
- Minimize $\mathcal{I}$ by iteratively applying $\mathcal{F}$.
- At each iteration, set the magnitude to its target value $V$.



# Consistency-based approaches

**Griffin Lim** [Griffin and Lim, 1984]

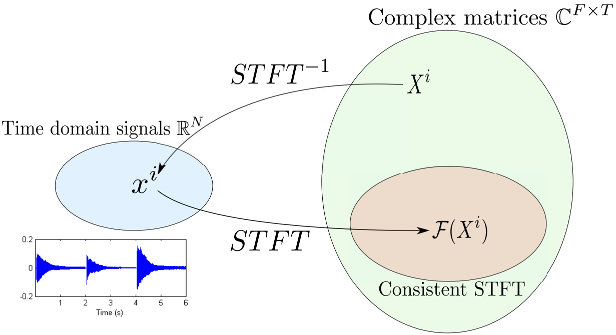
- Minimize $\mathcal{I}$ by iteratively applying $\mathcal{F}$.
- At each iteration, set the magnitude to its target value $V$.



# Consistency-based approaches

**Griffin Lim** [Griffin and Lim, 1984]

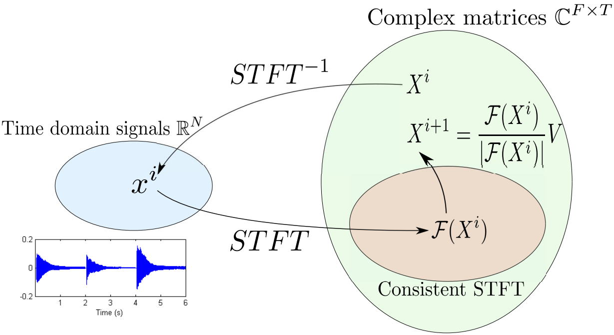
- Minimize $\mathcal{I}$ by iteratively applying $\mathcal{F}$.
- At each iteration, set the magnitude to its target value $V$.



# Consistency-based approaches

**Griffin Lim** [Griffin and Lim, 1984]

- Minimize $\mathcal{I}$ by iteratively applying $\mathcal{F}$.
- At each iteration, set the magnitude to its target value $V$.



# Consistency-based approaches

**Le Roux** [Le Roux et al., 2008]

1. Explicit calculation of $\mathcal{I}$.
2. Direct minimization of $\mathcal{I}$ (coordinate descent method).
   $\oplus$ Approximations on $\mathcal{I}$ allow fast computation.

# NMF with phase estimation

**Complex NMF** (CNMF) [Kameoka et al., 2009]
Mixture of complex sources:

$$X(f,t) = \sum_k X_k(f,t) = \sum_k W_k(f)H_k(t)e^{j\phi_k(f,t)}.$$

- Joint estimation of magnitude and phase.
- Needs to be constrained, e.g. by enforcing the consistency [Le Roux et al., 2009].

  $\ominus$ The data dimension is no longer reduced.

# NMF with phase estimation

**High Resolution NMF** (HRNMF) [Badeau and Plumbley, 2014]
Modeling each frequency band by means of AR filtering:

$$X_k(f,t) = b_k(f,t) + \sum_{p=1}^{P(k,f)} a_p(k,f) X_k(f,t-p),$$

with
$b_k(f,t) \sim \mathcal{N}(0, \sigma_k(f,t))$ where $\sigma_k(f,t) = w(f,k)h(k,t)$

- Parameters estimation with EM algorithm or VBEM.

  $\oplus$ Naturally captures phase dependencies over time.

# Outline

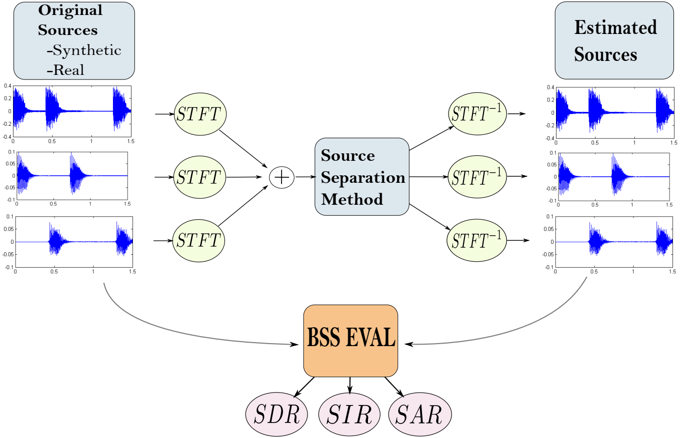
Overview of the compared methods

The benchmark

Methodology

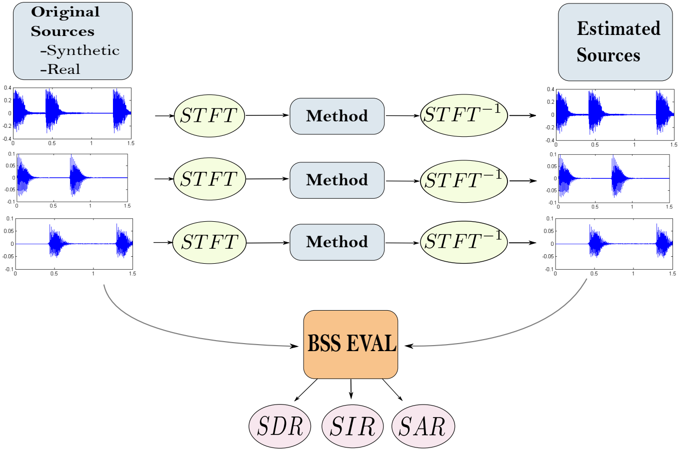
Results

# Principle




**Blind** benchmark: performance of the techniques in terms of source separation quality (BSS Eval [Vincent et al., 2006]).

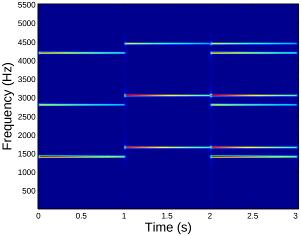
# Principle

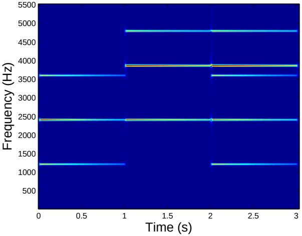



**Oracle** benchmark: best performance possible, potential of the methods.

# Datasets

- Mixtures of damped sinusoids (parameters are randomly defined) with or without TF overlap.




- Mixtures of piano notes (MAPS database [Emiya et al., 2010]).
- A MIDI audio excerpt (3 bass notes and 1 guitar chord).

# Protocol

Number of parameters

- HRNMF is used with AR filters of order 1.
- NMF: double frequency resolution.
- CNMF uses more parameters than the original data.

# Protocol

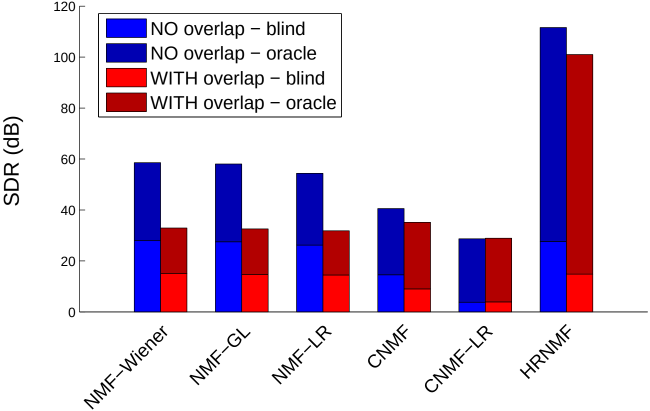
Number of parameters

- HRNMF is used with AR filters of order 1.
- NMF: double frequency resolution.
- CNMF uses more parameters than the original data.

Algorithms

- NMF with Kullback-Leibler (KL) divergence and MUR.
- HRNMF initialized with KL-NMF MUR and estimated with the VBEM algorithm.

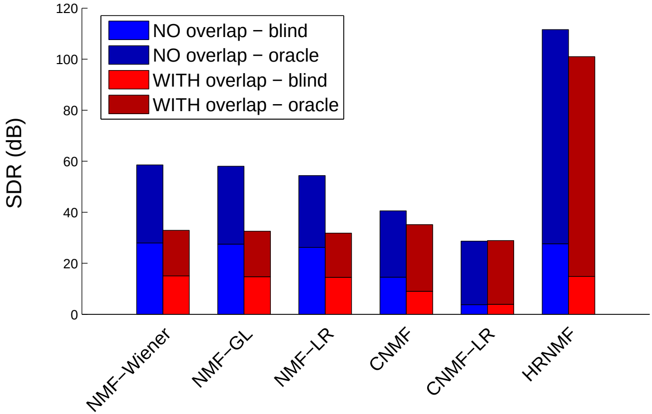
# Synthetic data




## Consistency

- **GL** and **LeRoux**: poor results in terms of audio quality.
- Slight decrease of SDR and SAR compared to **NMF-Wiener**.

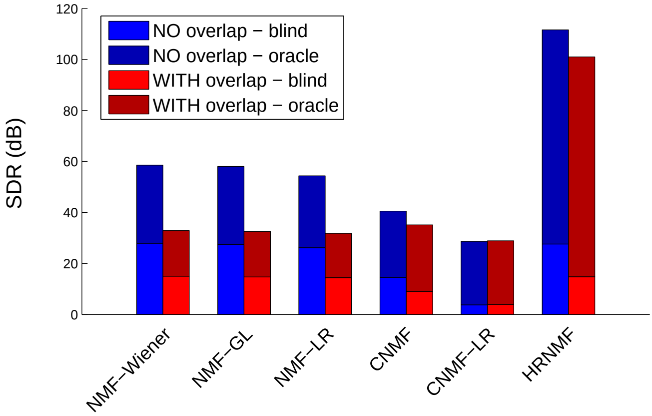
# Synthetic data




## Complex NMF

- **CNMF-LR** does not provide better results than **NMF-LR**.
- Requires much more memory for storing the phase fields.
- **CNMF** provides better results than **CNMF-LR**.

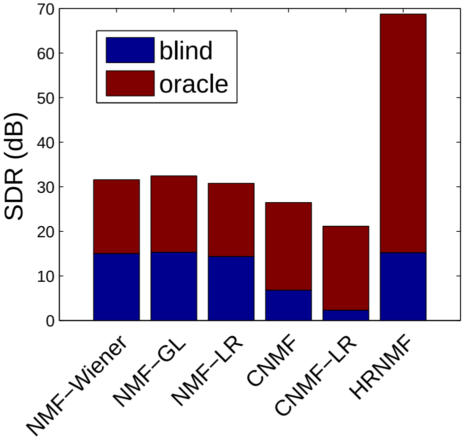
# Synthetic data




## HRNMF

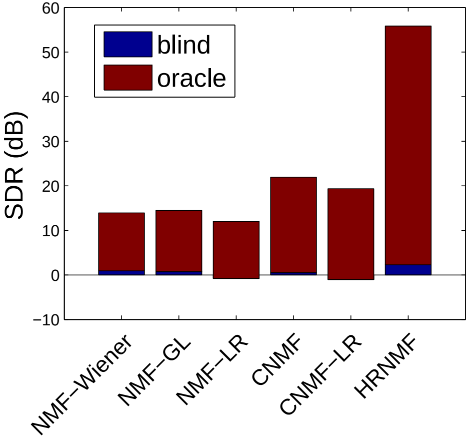
- Blind separation with the **HRNMF** model provides slightly better results than with the other models.
- Best performance in the oracle benchmark.

# Piano notes




- **HRNMF** oracle results confirm it has the greatest potential.
- **HRNMF** estimation does not improve the result of the initial KLNMF in the blind benchmark.

# MIDI excerpt




- Dramatic reduction of blind source separation quality.
- Oracle approach → this method has a high potential.

# Conclusions and future work

Consistency may not be an appropriate criterion for audio quality.

- Use model-based phase constraints.

# Conclusions and future work

Consistency may not be an appropriate criterion for audio quality.

- Use model-based phase constraints.

HRNMF is a promising model for the source separation task.

# Conclusions and future work

Consistency may not be an appropriate criterion for audio quality.

- Use model-based phase constraints.

HRNMF is a promising model for the source separation task.

Original: mixture and bass

| | Wiener | HRNMF |
|---|---|---|
| Blind | | |
| Oracle | | |

# Conclusions and future work

Consistency may not be an appropriate criterion for audio quality.

- Use model-based phase constraints.

HRNMF is a promising model for the source separation task.

- Oracle results → mostly effective when source separation is partially informed.
- Prior information on the sources, alternative estimation methods.


Thank you!

Webpage: `http://perso.telecom-paristech.fr/~magron/`

# HRNMF initialization and estimation algorithm

HRNMF requires a well-chosen initialization.
Mixtures of piano notes (MAPS).

| Algorithm | Initialization | SDR | SIR | SAR | Time (s) |
|---|---|---|---|---|---|
| EM | Random | 5.3 | 6.4 | 14.3 | 379 |
| | ISNMF | 15.0 | 21.2 | 17.0 | 376 |
| | KLNMF | 17.0 | 22.2 | 18.7 | 377 |
| VBEM | Random | 1.4 | 2.8 | 11.1 | 1.03 |
| | ISNMF | **16.9** | **25.3** | **17.7** | **0.95** |
| | KLNMF | **16.9** | **24.5** | **17.8** | **0.89** |

The best performance is obtained with KL-NMF and VBEM algorithm.

Badeau, R. and Plumbley, M. D. (2014).
Multichannel high resolution NMF for modelling convolutive mixtures or non-stationary signals in the time-frequency domain.
*IEEE Transactions on Audio Speech and Language Processing*, 22(11):1670–1680.

Emiya, V., Bertin, N., David, B., and Badeau, R. (2010).
MAPS - A piano database for multipitch estimation and automatic transcription of music.
Technical Report 2010D017, Télécom ParisTech, Paris, France.

Févotte, C., Bertin, N., and Durrieu, J.-L. (2009).
Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis.
*Neural computation*, 21(3):793–830.

Griffin, D. and Lim, J. (1984).
Signal estimation from modified short-time Fourier transform.

*IEEE Transactions on Acoustics, Speech and Signal Processing*, 32(2):236–243.

Kameoka, H., Ono, N., Kashino, K., and Sagayama, S. (2009).
Complex NMF: A new sparse representation for acoustic signals.
In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3437–3440, Taipei, Taiwan.

Le Roux, J., Kameoka, H., Vincent, E., Ono, N., Kashino, K., and Sagayama, S. (2009).
Complex NMF under spectrogram consistency constraints.
In *Proc. Acoustical Society of Japan Autumn Meeting*, Hukushima, Japan.

Le Roux, J., Ono, N., and Sagayama, S. (2008).
Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction.
In *Proc. ISCA Workshop on Statistical and Perceptual Audition (SAPA)*, pages 23–28, Brisbane, Australia.

Lee, D. D. and Seung, H. S. (1999).
Learning the parts of objects by non-negative matrix factorization.
*Nature*, 401(6755):788–791.

Lee, D. D. and Seung, H. S. (2001).
Algorithms for non-negative matrix factorization.
In Leen, T., Dietterich, T., and Tresp, V., editors, *Advances in Neural Information Processing Systems 13*, pages 556–562. MIT Press.

Vincent, E., Gribonval, R., and Févotte, C. (2006).
Performance measurement in blind audio source separation.
*IEEE Transactions on Speech and Audio Processing*, 14(4):1462–1469.