Phase recovery in NMF for source separation: an insightful benchmark

Paul Magron, Roland Badeau, Bertrand David

 Télécom ParisTech, CNRS LTCI

April 22, 2015
 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

Time domain signals

\[ x_1 \xrightarrow{STFT} X_1 \xrightarrow{\cdot |^\alpha} V_1 \]
\[ + \]
\[ x_K \xrightarrow{STFT} X_K \xrightarrow{\cdot |^\alpha} V_K \]
\[ = \]
\[ \mathcal{X} \xrightarrow{STFT} \mathcal{X} \xrightarrow{\cdot |^\alpha} V \]

Generally \( V = |\mathcal{X}| \) or \( |\mathcal{X}|^2 \).

Assumption of an additivity property: \( V = \sum_{k=1}^{K} V_k \).
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

$\mathbf{H}$

$\mathbf{W}$

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:

<table>
<thead>
<tr>
<th>Mixture</th>
<th>Source 1</th>
<th>Source 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td><img src="image1.png" alt="Original Source 1" /></td>
<td><img src="image2.png" alt="Original Source 2" /></td>
</tr>
<tr>
<td>Estimated</td>
<td><img src="image3.png" alt="Estimated Source 1" /></td>
<td><img src="image4.png" alt="Estimated Source 2" /></td>
</tr>
</tbody>
</table>
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:** $\mathbb{R}^N \rightarrow S^{F \times T} \subset \mathbb{C}^{F \times T}$

Complex matrices $\mathbb{C}^{F \times T}$

Time domain signals $\mathbb{R}^N$

$X$

$\mathcal{F}(X)$

Consistent STFT

$STFT^{-1}$
Consistency-based approaches

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

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

- $\mathcal{F} = STFT \circ STFT^{-1}$.
- $\|\cdot\|$ 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 $V$.

![Diagram showing time domain signals $\mathbb{R}^N$ and complex matrices $\mathbb{C}^{F \times T}$](image)
**Griffin Lim** [Griffin and Lim, 1984]

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

**Diagram:**

- **Time domain signals $\mathbb{R}^N$**
- **Complex matrices $\mathbb{C}^{F \times T}$**
- **$STFT$**
- **$STFT^{-1}$**
- **Consistent STFT**
- **$\mathcal{F}(X^i)$**

- **Note:** The diagram illustrates the process of applying the Griffin-Lim algorithm, showing the transformation of time domain signals into complex matrices and back again.
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 $V$. 

![Diagram of Griffin Lim algorithm](image)
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).
   ⊕ Approximations on $\mathcal{I}$ allow fast computation.
NMF with phase estimation

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

\[ X(f, t) = \sum_k X_k(f, t) = \sum_k W_k(f) H_k(t) e^{i\phi_k(f, t)}. \]

- Joint estimation of magnitude and phase.
- Needs to be constrained, e.g. by enforcing the consistency [Le Roux et al., 2009].
  - The data dimension is no longer reduced.
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

\[
b_k(f, t) \sim \mathcal{N}(0, \sigma_k(f, t)) \text{ where } \sigma_k(f, t) = w(f, k)h(k, t)
\]

- Parameters estimation with EM algorithm or VBEM.
  - Naturally captures phase dependencies over time.
Outline

Overview of the compared methods

The benchmark
  Methodology
  Results
**Principle**

**Original Sources**
- Synthetic
- Real

**Estimated Sources**

**Source Separation Method**

**BSS EVAL**

**SDR**

**SIR**

**SAR**

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

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

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

- Blind separation with the HRNMF model provides slightly better results than with the other models.
- Best performance in the oracle benchmark.
HRNMF oracle results confirm it has the greatest potential.

HRNMF estimation does not improve the result of the initial KLNMF in the blind benchmark.
Dramatic reduction of blind source separation quality.

Oracle approach $\rightarrow$ this method has a high potential.
Consistency may not be an appropriate criterion for audio quality.

- Use model-based phase constraints.
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 🎶

<table>
<thead>
<tr>
<th>Wiener</th>
<th>HRNMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blind</td>
<td>🎶</td>
</tr>
<tr>
<td>Oracle</td>
<td>🎶</td>
</tr>
</tbody>
</table>
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 $\rightarrow$ 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).

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Initialization</th>
<th>SDR</th>
<th>SIR</th>
<th>SAR</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM</td>
<td>Random</td>
<td>5.3</td>
<td>6.4</td>
<td>14.3</td>
<td>379</td>
</tr>
<tr>
<td></td>
<td>ISNMF</td>
<td>15.0</td>
<td>21.2</td>
<td>17.0</td>
<td>376</td>
</tr>
<tr>
<td></td>
<td>KLNMF</td>
<td>17.0</td>
<td>22.2</td>
<td>18.7</td>
<td>377</td>
</tr>
<tr>
<td>VBEM</td>
<td>Random</td>
<td>1.4</td>
<td>2.8</td>
<td>11.1</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>ISNMF</td>
<td>16.9</td>
<td>25.3</td>
<td>17.7</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>KLNMF</td>
<td>16.9</td>
<td>24.5</td>
<td>17.8</td>
<td>0.89</td>
</tr>
</tbody>
</table>

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


