Complex NMF under phase constraints based on signal modeling

Application to audio source separation

Paul Magron, Roland Badeau, Bertrand David

LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

March 23, 2016
Source separation
NMF
Phase retrieval

\[ H \]

\[ W \]

Original spectrogram
Wiener filtering commonly used
Issues when the sources overlap in the TF domain.

How can we improve phase reconstruction in NMF-based source separation?

Wiener $\rightarrow$ CNMF-\(\phi\)
Outline

Phase reconstruction in NMF

Proposed Model

Experimental results
- Non negative data: magnitude spectrogram $|X|$
- $|X| \approx WH = \sum_k W_k H_k$
\[
    \hat{X}_k = \frac{W_k H_k}{\sum_{l=1}^{K} W_l H_l} \odot X
\]

- \( \phi \)-source = \( \phi \)-mixture
- Issues in sound quality when sources overlap in the TF domain
- \( \hat{X}_k \neq \text{STFT of a } \hat{x}_k(n) \)
Consistency-based approaches

- Find a $\hat{X}_k$ that is close to a STFT
- Griffin & Lim, 84 (iterative)
- Magron, Icassp 2015, Consistency $\not\Rightarrow$ sound quality
Proposed model / Key ideas

Our approach

- Phase constraints based on time signal properties
- Complex NMF (CNMF) framework [Kameoka, 2009]

2 novelties

- Phase unwrapping
- Repetition of audio events
Complex NMF (CNMF) [Kameoka, 2009]

- Mixture model:

\[
\hat{X}(f, t) = \sum_{k=1}^{K} W(f, k)H(k, t) e^{i\phi_k(f,t)}
\]  

- Estimation by minimization of

\[
\sum_{f,t} |X(f, t) - \hat{X}(f, t)|^2 + \sigma_s 2 \sum_{k,t} H(k, t)^p
\]

Distance \(D(X, \hat{X})\)

Sparsity penalty \(C_s(H)\)
Proposed model / Phase unwrapping

1. For each source, onset frames are detected $\rightarrow \{ T_k \}$

2. Each source is modeled as a $\sum$ of sines:
   - frequency peaks are estimated with QIFFT
   - each channel $f$ is assigned to one sine frequency $\nu_k(f)$
   - the phase in channel $f$ is mainly governed by $\nu_k(f)$

3. Phase unwrapping in channel $f$:
   
   $$\Delta \phi_k(f, t) = 2\pi S\nu_k(f),$$

Unwrapping cost function:

$$C_u(\phi) = \sum \sum |X(f, t)|^2 |e^{i\Delta \phi_k(f, t)} - e^{2i\pi S\nu_k(f)}|^2$$
Two onset signals are equal up to a gain factor and a delay:

\[ X(f, t_m) \approx X(f, t_0) \rho e^{i \lambda(m)f}, \text{ with } \lambda(m) = \frac{2\pi \eta(m)}{F}. \]

\[ \phi(f, t) \approx \psi(f) + \lambda(t)f. \]

- \( \phi(f, t) \): phase within an onset frame
- \( \psi(f) \): reference phase
- \( \lambda(t)f \): offset
Model of repeated audio events

Repetition cost function:

\[ C_r(\phi, \psi, \lambda) = \sum_{f,k} \sum_{t \in \Omega_k} |X(f, t)|^2 |e^{i\phi_k(f,t)} - e^{i\psi_k(f)} e^{i\lambda_k(t)} f| \]
CNMF under phase constraints

Complete cost function:

\[ C(\theta) = D(X, \hat{X}) + \sigma_u C_u(\phi) + \sigma_r C_r(\phi, \psi, \lambda) + \sigma_s C_s(H) \]

- The variables are \( \theta = \{ W, H, \phi, \psi, \lambda \} \);
- \( \sigma_u, \sigma_r \) and \( \sigma_s \) are prior weights which promote the constraints.

Model estimation:
Minimization of \( C(\theta) \).

- Coordinate descent method \( \rightarrow \) Iterative procedure.
- Convergence is not guaranteed but observed in practice.
Protocol & datasets

- Synthetic mixtures of sinusoids;
- Mixtures of piano notes (MAPS database);
- $Fs = 11025$ Hz;
- The STFT uses a 46 ms-long Hann window and 75 % overlap.

Methods:

- **NMF-W**: 30 iterations of KLNMF + Wiener filtering;
- **CNMF**: 10 iterations of CNMF without phase constraints;
- **CNMF-$\phi$**: 10 iterations of CNMF with phase constraints;

Score:

- BSS Eval [Vincent, 2006] $\rightarrow$ SDR, SIR and SAR.
Influence of the weights

- **Sparsity**: \( p = 1 \) and \( \sigma_s = \|X\|^2 K^{-(1-p/2)} 10^{-5} \).
- **Sinusoids**:

  ![Graphs showing SDR, SIR, and SAR for different \( \sigma_u \) and \( \sigma_r \) values]

- **Piano notes**:

  ![Graphs showing SDR, SIR, and SAR for different \( \sigma_u \) and \( \sigma_r \) values]
Influence of the weights

- With $\sigma_r = 0.2$:

$$\rightarrow (\sigma_u, \sigma_r) = (0.2, 0.2)$$ for robustness and higher scores.
Source separation

Reconstruction of a B2 piano note partial from a mixture made up of two piano notes (E2 and B2):

![Graph showing reconstruction results]
## Source separation

Separation results:

<table>
<thead>
<tr>
<th>Data</th>
<th>Method</th>
<th>SDR</th>
<th>SIR</th>
<th>SAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synthetic sinsuoids</td>
<td>NMF-W</td>
<td>12.1</td>
<td>17.5</td>
<td>14.1</td>
</tr>
<tr>
<td></td>
<td>CNMF</td>
<td>12.0</td>
<td>14.6</td>
<td>16.1</td>
</tr>
<tr>
<td></td>
<td>CNMF-φ</td>
<td>14.0</td>
<td>20.7</td>
<td>15.4</td>
</tr>
<tr>
<td>Piano notes</td>
<td>NMF-W</td>
<td>12.9</td>
<td>23.3</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>CNMF</td>
<td>13.5</td>
<td>20.0</td>
<td>14.8</td>
</tr>
<tr>
<td></td>
<td>CNMF-φ</td>
<td>14.0</td>
<td>24.0</td>
<td>14.6</td>
</tr>
</tbody>
</table>

- Improved interference rejection.
- Slight increase of SDR.
100 songs (rock, pop, electro...) from the Demixing Secret Database;
The optimal weights are learned on 50 songs;
Source separation is performed on the other 50.
Significant increase in interference rejection;
The trade-off between SDR, SIR and SAR highly depends on the weights values.
Source separation - Realistic data

Mixture | Voice
--- | ---
Original | NMF-W | CNMF | CNMF-φ | CNMF-φ

- SDR
- SIR
- SAR

<table>
<thead>
<tr>
<th>Decibels</th>
<th>NMF-W</th>
<th>CNMF</th>
<th>CNMF-φ</th>
<th>CNMF-φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIR</td>
<td>13</td>
<td>12</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>SAR</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

- Complex NMF under phase constraints based on signal modeling

March 23, 2016 | ICASSP 2016 | Complex NMF under phase constraints based on signal modeling
Conclusion

Complex NMF with signal model-based phase constraints

- A promising approach for separating overlapping sources in the TF domain;
- Better interference rejection than traditional Wiener filtering or unconstrained CNMF;
- The repetition constraints does not significantly improve the results.

Further work

- High sensitivity to the weight parameters;
- Optimization scheme is not efficient
  → New formulation of the problem: probabilistic framework.
Thank you!

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