Lévy NMF for robust nonnegative source separation

Paul Magron, Roland Badeau, Antoine Liutkus

IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)

17.10.2017
Source separation

- Problem: extract the $X_k, k \in \{1, ..., K\}$ from:

$$X = \sum_{k} X_k$$

- Nonnegative data: audio spectrograms, images, fluorescence spectra...

- Many methods: PCA, ICA, NMF...
Nonnegative matrix factorization

- Probabilistic approach: sources as latent variables;
- Maximum likelihood $\leftrightarrow$ Minimization of a cost function between $X$ and $WH$. 

\[ H \]
\[ W \]
Robustness

Traditional distribution are not *heavy-tailed* : no robustness to outliers.

→ Stable distributions :
  - Stability and robustness...
  - ... not a nonnegative support in general.

Goal : **A robust nonnegative data model for source separation**
Outline

1. Lévy NMF model
2. Parameter estimation
3. Experimental evaluation
1. Lévy NMF model

2. Parameter estimation

3. Experimental evaluation
Stable distributions

A family of **heavy-tailed** distributions
**Symmetric** $\alpha$-stable ($S\alpha S$) : $\beta = 0$.
**Stability** : a sum of stable variables is stable.

Special cases :
- Gaussian : $\alpha = 2$ and $\beta = 0$;
- Cauchy : $\alpha = 1$ and $\beta = 0$;
- Lévy : $\alpha = 1/2$ and $\beta = 1$;
Positive $\alpha$-stable distributions

In general, the support of the stable distributions is $\mathbb{R}$ (or $\mathbb{C}$). For $\beta = 1$ and $\alpha < 1$, the support is $[\mu; +\infty[$.

→ Positive $\alpha$-stable ($\mathcal{P}_\alpha \mathcal{S}$) distributions:

$$\mathcal{P}_\alpha \mathcal{S}(\sigma) = \mathcal{S}(\alpha, 0, \sigma, 1), \text{ with } \alpha < 1.$$ 

Lévy distribution ($\alpha=1/2$):

$$p(x \mid \sigma) = \begin{cases} \sqrt{\frac{\sigma}{2\pi}} \frac{1}{x^{3/2}} e^{-\frac{\sigma}{2x}} & \text{if } x > 0 \\ 0 & \text{else.} \end{cases}$$
Positive $\alpha$-stable distributions
Mixture model

- Nonnegative data $X \in \mathbb{R}^{F \times T}_+ : X = \sum_k X_k$.

- Independent Lévy coefficients:
  \[
  X_k(f,t) \sim \mathcal{L}(\sigma_k(f,t))
  \]
  \[
  \rightarrow X \sim \mathcal{L}(\sigma) \text{ with } \sqrt{\sigma} = \sum_k \sqrt{\sigma_k}.
  \]

- NMF on the dispersion parameters:
  \[
  \sqrt{\sigma} = WH,
  \]
  where $W \in \mathbb{R}^{F \times K}_+$ and $H \in \mathbb{R}^{K \times T}_+$.

  \[
  \rightarrow \text{Lévy NMF model.}
  \]
1 Lévy NMF model

2 Parameter estimation

3 Experimental evaluation
Maximum likelihood (ML)

Log-likelihood of the data:

\[
L(W, H) = \sum_{f,t} \log(p(X(f, t); \sigma(f, t)))
\]

\[
c = \frac{1}{2} \sum_{f,t} \log([WH](f, t)^2) - \frac{[WH](f, t)^2}{X(f, t)}
\]

\[
c = -\frac{1}{2} d_{IS}([WH]^\odot2, X),
\]

ML \leftrightarrow \text{Minimize the Itakura-Saito divergence between } [WH[^\odot2] \text{ and } X}
Heuristic approach

Decomposition of the gradient of $C$ w.r.t. $\theta$ ($= W$ or $H$) :

$$\frac{\partial C}{\partial \theta} = \nabla_\theta^+ - \nabla_\theta^-, \text{ with } \nabla_\theta^+ > 0 \text{ and } \nabla_\theta^- > 0.$$ 

Update rule :

$$\theta \leftarrow \theta \odot \frac{\nabla_\theta^-}{\nabla_\theta^+}$$

- No guarantee that the cost function is non-increasing;
- In practice, it is observed for many NMF models...
- ... but not for the Lévy case.
Majorize-Minimization

Auxiliary function $G$:

$$\forall (\theta, \bar{\theta}), \ C(\theta) \leq G_{\theta}(\theta), \text{ and } C(\bar{\theta}) = G_{\bar{\theta}}(\bar{\theta})$$

Update: $\theta^{(it+1)} = \arg \min_{\theta} G_{\theta^{(it)}}(\theta)$
Majorize-Minimization

- $G$ is obtained by using convexity inequalities;
- For Lévy NMF:

$$W \leftarrow W \odot \left( \frac{[WH]^{-1}H^T}{([WH] \odot X^{-1})H^T} \right)^{1/2}$$

and

$$H \leftarrow H \odot \left( \frac{W^T[WH]^{-1}}{W^T([WH] \odot X^{-1})} \right)^{1/2}$$

- Similar updates to the heuristic approach, with a power $1/2$.
- The cost function is non-increasing under these updates.
Lévy NMF vs. ISNMF

If $K = 1$ and $W(f) = 1 \forall f$:

$$H_{IS}(t) \leftarrow \frac{1}{F} \sum_{f} X(f, t), \quad H_{Lévy}(t) \leftarrow \sqrt{\frac{F}{\sum_{f} \frac{1}{X(f, t)}}}.$$  

- ISNMF $\rightarrow$ **Arithmetic** mean;
- Lévy NMF $\rightarrow$ **Harmonic** mean (and $\sqrt{\cdot}$).

If $F = 10$ and $X(f, t) = 1$ except for one entry: $X(f_0, t_0) = 10^8$:

$$H_{IS}(t_0) \leftarrow 10^7, \quad H_{Lévy}(t_0) \leftarrow 1.05.$$  

$\rightarrow$ **Robustness of Lévy NMF**
Source estimation

- Natural estimator: $\hat{X}_k = \mathbb{E}_{X_k|X}(X_k)$.

- For any $P_\alpha S$ distribution:

$$\hat{X}_k = \sum_l \frac{\sigma_k^\alpha}{\sigma_l^\alpha} \otimes X$$

→ **Generalized Wiener filtering**

- For Lévy NMF:

$$\hat{X}_k = \frac{W_k H_k}{\sum_l W_l H_l} \otimes X$$
1. Lévy NMF model

2. Parameter estimation

3. Experimental evaluation
Music spectrogram inpainting

- Data: 6 guitar pieces;
- The spectrograms are corrupted with impulsive noise;
- The models are learned on the corrupted data;
- The noise localization is unknown.
## Music spectrogram inpainting

<table>
<thead>
<tr>
<th>Method</th>
<th>Log(KL)</th>
<th>SDR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISNMF</td>
<td>9.0</td>
<td>-23.5</td>
</tr>
<tr>
<td>KLNMF</td>
<td>6.2</td>
<td>-8.9</td>
</tr>
<tr>
<td>Cauchy NMF</td>
<td>3.4</td>
<td>7.6</td>
</tr>
<tr>
<td>RPCA</td>
<td>3.6</td>
<td>7.4</td>
</tr>
<tr>
<td>Lévy NMF</td>
<td>3.2</td>
<td>9.2</td>
</tr>
<tr>
<td>Weighted ISNMF</td>
<td>3.8</td>
<td>4.5</td>
</tr>
</tbody>
</table>

- Bad results with classical NMFs (IS and KL);
- Lévy NMF compares with other methods.
Musical accompaniment enhancement

- Data: 50 music songs;
- Musical accompaniment is assumed well-represented by a low-rank NMF model;
- Voice is assumed to be similar to impulsive noise.
Conclusion

**A robust nonnegative data model**
- Many areas of application: data mining, applied physics...
- An extension of Wiener filtering to nonnegative data.

**Future work**
- MAP estimation: priors on the parameters (sparsity, temporal smoothness...);
- Generalization to $\mathcal{P}_\alpha S$ or inverse-Gamma distributions.