TAMPERE UNIVERSITY OF TECHNOLOGY

Lévy NMF

Un modèle robuste de séparation de sources non-négatives

Paul Magron, Roland Badeau, Antoine Liutkus

XXVIe colloque Gretsi

8 septembre 2017

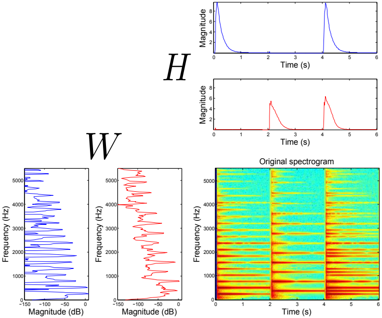
Séparation de sources

- Problème : extraire les X_k, $k \in \{1, ..., K\}$ qui forment un mélange :

$$X = \sum_k X_k$$

- Données non-négatives : spectrogrammes audio, images, spectres de fluorescence...
- De nombreuses méthodes : PCA, ICA, NMF...

Factorisation en matrices non-négatives



- Approches probabilistes : les sources sont des variables aléatoires ;
- Maximum de vraisemblance $\leftrightarrow$ Minimisation d'une fonction d'erreur entre X et WH.

Robustesse

Les distributions traditionnelles n'ont pas une queue lourde : pas de robustesse aux valeurs aberrantes.

$\rightarrow$ Lois stables :

- Stabilité et robustesse...
- ... pas non-négatives en général.

But : **Un modèle de données non-négatives robuste pour la séparation de sources**

Sommaire

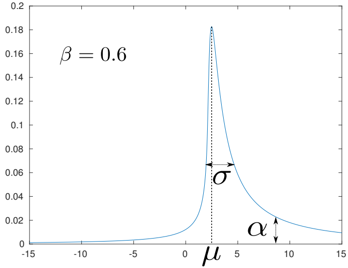
1 Modèle de Lévy NMF

2 Estimation des parametères

3 Évaluation expérimentale

Distributions stables

Distributions à **queue lourde**
Lois **Symmetriques** α-stables (SαS) : $\beta = 0$.
Stabilité : une somme de variables stables est stable.



Cas particuliers :

- Gaussien : $\alpha = 2$ et $\beta = 0$;
- Cauchy : $\alpha = 1$ et $\beta = 0$;
- Lévy : $\alpha = 1/2$ et $\beta = 1$;

Distributions Positives α-stables

En général, le support des lois stables est $\mathbb{R}$ (ou $\mathbb{C}$).

Pour $\beta = 1$ et $\alpha < 1$, le support $[\mu; +\infty[$.

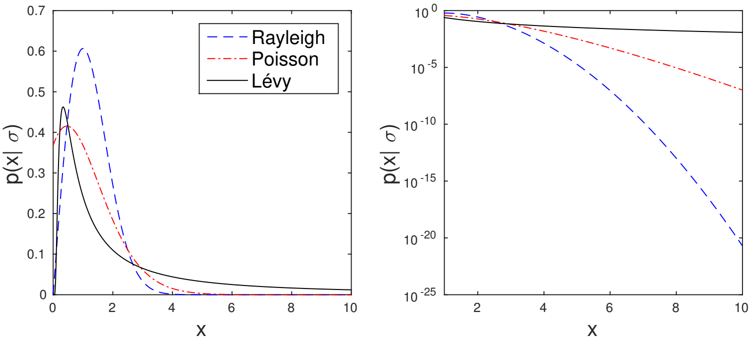
$\rightarrow$ Lois **Positive α-stables** (PαS) :

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

Loi de Lévy (α=1/2) :

$$p(x \mid \sigma) = \begin{cases} \sqrt{\dfrac{\sigma}{2\pi}} \dfrac{1}{x^{3/2}} e^{-\frac{\sigma}{2x}} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases}$$

Distributions Positives α-stables



Modèle de mélange

- Données non-négatives $X \in \mathbb{R}_+^{F \times T}$: $X = \sum_k X_k$.
- Coefficients TF indépendants et de Lévy :

$$X_k(f,t) \sim \mathcal{L}(\sigma_k(f,t))$$

$$\rightarrow X \sim \mathcal{L}(\sigma) \text{ avec } \sqrt{\sigma} = \sum_k \sqrt{\sigma_k}.$$

- NMF sur les paramètres de dispersion :

$$\sqrt{\sigma} = WH,$$

où $W \in \mathbb{R}_+^{F \times K}$ et $H \in \mathbb{R}_+^{K \times T}$.

$\rightarrow$ Modèle de **Lévy NMF**.

1 Modèle de Lévy NMF

2 Estimation des parametères

3 Évaluation expérimentale

Maximum de Vraisemblance (MV)

Estimation MV des paramètres W et H :

$$
\begin{aligned}
L(W,H) &= \sum_{f,t} \log(p(X(f,t);\sigma(f,t))) \\
&\overset{c}{=} \frac{1}{2}\sum_{f,t} \log([WH](f,t)^2) - \frac{[WH](f,t)^2}{X(f,t)} \\
&\overset{c}{=} -\frac{1}{2} d_{IS}([WH]^{\odot 2}, X),
\end{aligned}
$$

MV $\leftrightarrow$ Minimiser la divergence d'Itakura-Saito entre $[WH]^{\odot 2}$ et X

Approche heuristique

Décomposition du gradient de $\mathcal{C}$ par rapport à θ $(= W$ ou $H)$:

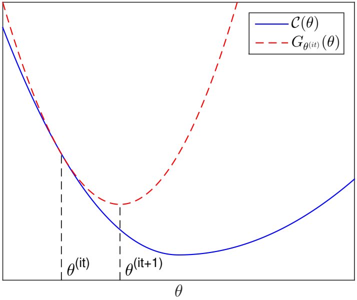
$$\frac{\partial \mathcal{C}}{\partial \theta} = \nabla_\theta^+ - \nabla_\theta^-, \text{ avec } \nabla_\theta^+ > 0 \text{ et } \nabla_\theta^- > 0.$$

Mise à jour :

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

- Pas de garantie de décroissance de la fonction de coût ;
- En pratique, cela marche pour de nombreux modèles NMF...
- ... mais pas pour Lévy.

Majoration-Minimisation



- Fonction auxiliaire G :

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

- Mise à jour $\theta^{(it+1)} = \arg\min_{\theta} G_{\theta^{(it)}}(\theta)$

Majoration-Minimisation

- G est obtenue grâce à des inégalités de convexité ;
- Pour Lévy NMF :

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

et

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

- Mises à jour similaires que par l'heuristique, avec exposant $1/2$.
- Garantie de décroissance de la fonction de coût.

Lévy NMF vs. ISNMF

Si $K = 1$ et $W(f) = 1 \; \forall f$:

$$H_{\mathsf{IS}}(t) \leftarrow \frac{1}{F} \sum_f X(f,t), \; H_{\mathsf{Lévy}}(t) \leftarrow \sqrt{\frac{F}{\sum_f \frac{1}{X(f,t)}}}.$$

- ISNMF $\rightarrow$ moyenne **arithmétique** ;
- Lévy NMF $\rightarrow$ moyenne **harmonique** (et $\sqrt{}$).

Si $F = 10$, et $X(f,t) = 1$ sauf pour une entrée : $X(f_0, t_0) = 10^8$:

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

$\rightarrow$ **Robustesse de Lévy NMF**

Estimation des sources

- Estimateur naturel : $\hat{X}_k = \mathbb{E}_{X_k|X}(X_k)$.
- Pour toute distribution PαS :

$$\hat{X}_k = \frac{{\sigma_k}^{\alpha}}{\sum_l {\sigma_l}^{\alpha}} \odot X \tag{1}$$

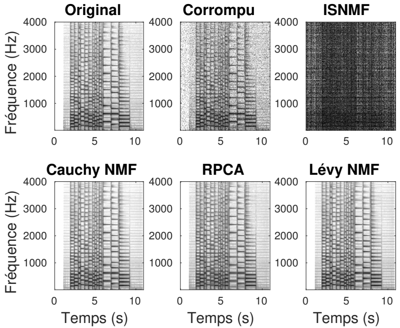
$\rightarrow$ **Filtrage de Wiener généralisé**

- Pour Lévy NMF :

$$\hat{X}_k = \frac{W_k H_k}{\sum_l W_l H_l} \odot X \tag{2}$$

Restauration de spectrogrammes musicaux

- Données : 6 morceaux de guitare ;
- Spectrogrammes corrompus par des bruits impulsionnels.
- Modèles appris directement sur les données corrompues ;
- La localisation du bruit est inconnue.



Restauration de spectrogrammes musicaux

- Résultats (log-divergence de Kullback-Leibler entre spectrogrammes originaux et estimés) :

ISNMF	KLNMF	Cauchy NMF	RPCA	Lévy NMF
9.0	6.2	3.4	3.6	**3.2**

Orig	Corrompu	IS	RPCA	Lévy

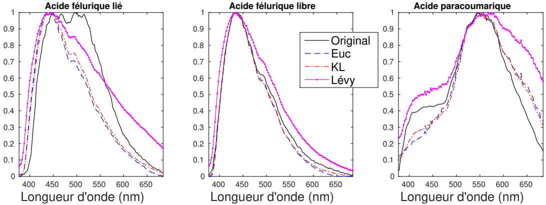
- NMFs "classiques" (IS et KL) : mauvais résultats ;
- Lévy NMF est comparable à d'autres méthodes robustes.

Spectroscopie de fluorescence

Spectres d'excitation-émission d'un mélange de composants = somme des spectres purs W pondérée par leurs concentrations H.

$\rightarrow$ identifier les espèces et leurs concentrations.

- Données : $T = 400$ spectres, $F = 128$ fréquences, $K = 3$ composantes.



Spectroscopie de fluorescence

Corrélation entre Oracle et sources estimées (%) :

	Euc	KL	Lévy
Acide férulique lié	82.8	85.3	**87.8**
Acide férulique libre	99.5	99.4	**99.6**
Acide paracoumarique	97.3	98.1	**98.4**

Potentiel de Lévy NMF pour décomposer des données non-négatives.

Conclusion

Un modèle robuste de données non-négatives

- Nombreux domaines d'application : fouille de données, physique appliquée, classification d'images...
- Extension du filtre de Wiener aux données non-négatives.

Recherches futures

- Estimation MAP : à prioris sur les paramètres (parcimonie, continuité temporelle...) ;
- Généralisation aux modèles PαS ou inverse-Gamma.