

Lévy NMF : un modèle robuste de séparation de sources non-négatives

Paul MAGRON^{1,2}, Roland BADEAU², Antoine LIUTKUS³ *

¹Signal Processing Laboratory, Tampere University of Technology (TUT), Finland

²LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

³Inria, Nancy Grand-Est, Multispeech team, LORIA UMR 7503, France

paul.magron@tut.fi, roland.badeau@telecom-paristech.fr, antoine.liutkus@inria.fr

Résumé – Dans cet article, nous nous intéressons à la séparation robuste de sources non-négatives. Nous introduisons les distributions Positives α -stables (P α S), un sous-ensemble de la famille des lois stables, qui modélisent les variables latentes non-négatives. Comme ces distributions sont à queue lourde, elles possèdent naturellement une propriété de robustesse aux valeurs aberrantes. En étudiant plus particulièrement la loi de Lévy, la seule loi P α S dont la densité s’exprime sous forme analytique, nous mettons au point un modèle de mélange dans lequel nous structurons les paramètres de dispersion des variables de Lévy par un modèle de factorisations en matrices non-négatives (NMF). Les paramètres de ce modèle, appelé Lévy NMF, sont estimés au sens du maximum de vraisemblance. Nous obtenons également un estimateur des sources qui généralise le filtrage de Wiener aux distributions P α S. Des expériences conduites sur des spectrogrammes musicaux et des spectres de fluorescence démontrent le potentiel de ce modèle pour décomposer des données non-négatives.

Abstract – In this paper, we address the problem of robust source separation of nonnegative data. We introduce the P α S distributions, which are a subclass of the stable distributions family, to model the nonnegative latent sources. Since those distributions are heavy-tailed, they are expected to be robust to outliers. Considering the Lévy distribution, the only P α S distribution whose density admits a closed-form expression, we propose a mixture model called Lévy Nonnegative Matrix Factorization (Lévy NMF). The parameters of the model are estimated in a maximum-likelihood sense. We also derive an estimator of the sources which extends the validity of the generalized Wiener filtering to the P α S case. Experiments on musical spectrograms and fluorescence spectra highlight the potential of the Lévy NMF model for decomposing nonnegative data.

1 Introduction

La séparation de sources consiste à extraire automatiquement les composantes appelées *sources* qui s’ajoutent pour former un signal observable appelé *mélange*. Cette tâche trouve des applications dans de nombreux domaines tels que la vision par ordinateur [1] ou le traitement de signaux audio [2].

Une idée novatrice présentée dans [1] part du constat que les observations étant souvent non-négatives¹, elles devraient être décomposées comme la somme de termes non-négatifs. La factorisation en matrices non-négatives (NMF) est une méthode de réduction de rang qui approche une matrice de données non-négatives X par le produit de deux matrices non-négatives W et H , de rang inférieur. Cette factorisation peut être obtenue par l’optimisation d’une fonction de coût qui mesure l’erreur entre X et WH , comme la distance euclidienne ou les divergences de Kullback-Leibler (KL [1]) et d’Itakura-Saito (IS [2]). Cela peut être formalisé dans un cadre probabiliste, dans lequel la factorisation n’est plus faite sur les observations, mais sur les

paramètres de dispersion des variables latentes (par exemple gaussiennes [2] ou de Poisson [3]) : la fonction de coût correspondante peut alors être vue comme l’opposé de la log-vraisemblance des données.

Cependant, ces modèles ne mènent pas à des résultats satisfaisants lorsque les données sont impulsives ou contiennent des valeurs aberrantes. Cela provient de la queue non-lourde des distributions sous-jacentes, incapables de prendre en compte des observations très peu probables. Les lois *stables* [4], une famille de distributions à queue lourde, ont donc été utilisées dans ce but, avec des applications notamment en audio [5, 6]. Néanmoins, ces travaux s’appuient sur des lois à support réel ou complexe, qui ne sont donc pas adaptées à la modélisation de données non-négatives.

Dans cet article, nous proposons un modèle pour la séparation de sources non-négatives qui bénéficie de cette propriété de robustesse. Nous considérons un sous-ensemble des lois stables à support non-négatif : les lois Positives α -stables (P α S). Ces distributions sont à queue lourde, et ont donc le potentiel pour fournir des estimées robustes des sources. Comme leur densité n’a en général pas d’expression analytique simple, nous considérons plus particulièrement la loi de Lévy, la seule distribution P α S possédant une forme analytique. Nous introduisons le modèle de *Lévy NMF*, dans lequel les paramètres de dispersion sont structurés par un modèle NMF et dont les réali-

* Ce travail a été en partie supporté par l’Agence Nationale de la Recherche dans le cadre du projet KAMouloX (ANR-15-CE38-0003-01). Les auteurs souhaitent remercier Cyril Gobinet de l’Université de Reims Champagne-Ardenne pour leur avoir fourni les données utilisées en spectroscopie de fluorescence.

1. Par souci de clarté, nous employons l’anglicisme "non-négatif" pour traduire "non-negative", qui se dit normalement "positif" en français.

sations sont nécessairement non-négatives. Les paramètres sont estimés au sens du Maximum de Vraisemblance (MV). Nous fournissons également un estimateur des sources qui généralise le filtrage de Wiener au cas PaS. Des expériences conduites sur des données musicales et de chimie montrent le potentiel de cette méthode pour décomposer des données non-négatives.

Cet article est organisé comme suit. La section 2 introduit le modèle de Lévy NMF. La section 3 détaille l'estimation des paramètres et des sources. Dans la section 4, nous montrons expérimentalement le potentiel de cette méthode, et nous présentons nos conclusions dans la section 5.

2 Modèle de Lévy NMF

2.1 Distributions Positives α -stables

Les lois stables scalaires, notées $\mathcal{S}(\mu, \sigma^\alpha, \beta)$, sont des distributions à queue lourde qui dépendent de 4 paramètres : un paramètre de forme $\alpha \in]0; 2]$ qui détermine la lourdeur de la queue de la distribution (plus α est petit, plus la queue est lourde), un paramètre de localisation $\mu \in \mathbb{R}$, un paramètre d'échelle $\sigma \in]0; +\infty[$ qui mesure la dispersion de la distribution autour de son mode, et un paramètre d'asymétrie $\beta \in [-1; 1]$. Les lois symétriques α -stables (SaS), telles que $\beta = 0$, forment un sous-ensemble important de ces distributions, utilisées notamment en audio [6, 7]. Des exemples notables sont la loi normale ($\alpha = 2$) [2] et la loi de Cauchy ($\alpha = 1$) [5]. Ces distributions sont dites stables en raison de leur propriété d'additivité : la somme X de K variables stables indépendantes $X_k \sim \mathcal{S}(\mu_k, \sigma_k^\alpha, \beta)$ est aussi stable : $X \sim \mathcal{S}(\mu, \sigma^\alpha, \beta)$, avec $\mu = \sum_k \mu_k$ et $\sigma^\alpha = \sum_k \sigma_k^\alpha$.

En règle générale, ces distributions ne sont pas à support non-négatif, mais on peut montrer [4] que si $\beta = 1$ et $\alpha < 1$, leur support est $[\mu; +\infty[$. Dans cet article, nous supposons que $\mu = 0$, ainsi le support devient \mathbb{R}_+ , ce qui définit les distributions Positives α -stables : $\mathcal{PS}(\sigma^\alpha) = \mathcal{S}(0, \sigma^\alpha, 1)$ avec $\alpha < 1$.

2.2 Modèle de mélange : Lévy NMF

La seule valeur de α pour laquelle on peut écrire la densité d'une loi PaS sous forme analytique simple est $\alpha = 1/2$, ce qui correspond à la loi de Lévy $\mathcal{L}(\sigma) = \mathcal{PS}(\sigma^{1/2})$ de densité :

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

Nous modélisons chaque entrée (f, t) d'une matrice $X \in \mathbb{R}_+^{F \times T}$ (f et t désignent par exemple les canaux fréquentiels et les trames temporelles d'une représentation temps-fréquence) comme la somme de K variables aléatoires indépendantes de Lévy de paramètre $\sigma_k(f, t)$. Supposant indépendantes ces entrées, nous pouvons étendre nos notations aux matrices : ainsi, par propriété de stabilité, on a $X \sim \mathcal{L}(\sigma)$ avec $\sigma^{\odot 1/2} = \sum_k \sigma_k^{\odot 1/2}$, où \odot désigne la puissance matricielle terme à terme.

Les paramètres d'échelle sont alors structurés par un modèle NMF [1] qui préserve la propriété d'additivité : $\sigma^{\odot 1/2} = WH$, avec $W \in \mathbb{R}_+^{F \times K}$ et $H \in \mathbb{R}_+^{K \times T}$. Nous appelons donc ce modèle Lévy NMF.

3 Estimation du modèle

3.1 Maximum de Vraisemblance

On peut montrer que la log-vraisemblance des données s'écrit sous la forme $L(W, H) \stackrel{c}{=} -\frac{1}{2}\mathcal{C}(W, H)$ avec :

$$\mathcal{C}(W, H) = \sum_{f,t} \frac{[WH](f, t)^2}{X(f, t)} - 2 \log([WH](f, t)), \quad (2)$$

et où $\stackrel{c}{=}$ désigne l'égalité à une constante additive près. L'estimation MV revient donc à minimiser \mathcal{C} . Une approche communément adoptée [1] consiste à écrire le gradient de \mathcal{C} par rapport au paramètre θ (W ou H) comme la différence entre deux termes non-négatifs : $\nabla_\theta \mathcal{C} = \nabla_\theta^+ - \nabla_\theta^-$, puis à considérer les mises à jour (MàJ) suivantes : $\theta \leftarrow \theta \odot \frac{\nabla_\theta^-}{\nabla_\theta^+}$, où \odot (resp. la barre de fraction) désigne la multiplication (resp. la division) matricielle terme à terme.

Ces MàJ garantissent que W et H restent non-négatives au cours des itérations à condition que l'initialisation le soit aussi. Cependant, elles ne garantissent pas la décroissance de la fonction de coût : bien que, pour de nombreux modèles, la décroissance soit observée en pratique, ce n'est pas le cas pour le modèle de Lévy NMF, ce qui motive le recours à une autre méthode d'optimisation.

3.2 Approche Majoration-Minimisation

Alternativement, on peut obtenir des règles de MàJ des paramètres par l'approche Majoration-Minimisation (MM) [8] : on cherche une fonction auxiliaire G qui majore \mathcal{C} :

$$\forall(\theta, \bar{\theta}), \mathcal{C}(\theta) \leq G(\theta, \bar{\theta}), \text{ et } \mathcal{C}(\bar{\theta}) = G(\bar{\theta}, \bar{\theta}). \quad (3)$$

La minimisation de $G(\theta, \bar{\theta})$ à $\bar{\theta}$ fixé fournit alors une MàJ sur θ . Cette approche garantit la décroissance de \mathcal{C} au cours des itérations. Cette fonction est obtenue d'une façon similaire à [9]. On commence par introduire les paramètres auxiliaires :

$$[WH](f, t)^2 = \left(\sum_k \rho_k(f, t) \frac{W(f, k)H(k, t)}{\rho_k(f, t)} \right)^2, \quad (4)$$

avec :

$$\rho_k(f, t) = \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)}, \quad \bar{V} = \bar{W}H, \quad (5)$$

et \bar{W} est un paramètre auxiliaire. Comme $\sum_k \rho_k(f, t) = 1$, on applique l'inégalité de Jensen à la fonction convexe $x \mapsto x^2$:

$$[WH](f, t)^2 \leq \sum_k \rho_k(f, t) \left(\frac{W(f, k)H(k, t)}{\rho_k(f, t)} \right)^2, \quad (6)$$

d'où la majoration suivante du premier terme dans (2) :

$$\sum_{f,t} \frac{[WH](f,t)^2}{X(f,t)} \leq \sum_{f,t} \frac{\bar{V}(f,t)}{X(f,t)} \sum_k \frac{W(f,k)^2 H(k,t)}{\bar{W}(f,k)}. \quad (7)$$

Similairement, nous obtenons une majoration du second terme dans (2) ($x \mapsto -\log(x)$ est aussi une fonction convexe), ce qui conduit à la fonction auxiliaire G suivante :

$$G(W, H, \bar{W}) = \sum_{f,t,k} \frac{\bar{V}(f,t)H(k,t)}{X(f,t)\bar{W}(f,k)} W(f,k)^2 - 2 \frac{\bar{W}(f,k)H(k,t)}{\bar{V}(f,t)} \log \left(W(f,k) \frac{\bar{V}(f,t)}{\bar{W}(f,k)} \right). \quad (8)$$

Annuler la dérivée de G par rapport à W conduit alors à une M&J sur W . Celle sur H est obtenue exactement de la même manière. Nous constatons que ces M&J peuvent s'écrire sous la forme $\theta \leftarrow \theta \odot a_\theta^{\odot 1/2}$, avec :

$$a_W = \frac{[WH]^{\odot -1} H^T}{([WH] \odot X^{\odot -1}) H^T}, \quad (9)$$

et

$$a_H = \frac{W^T [WH]^{\odot -1}}{W^T ([WH] \odot X^{\odot -1})}. \quad (10)$$

Nous fournissons une implémentation MATLAB de l'algorithme correspondant sur notre site web².

3.3 Estimation des composantes

Une fois les paramètres estimés, il est nécessaire d'obtenir un estimateur \hat{X}_k des composantes X_k pour achever la séparation. Dans un cadre probabiliste, un estimateur naturel est l'espérance à postériori d'une composante conditionnellement à l'observation du mélange $\mathbb{E}(X_k|X)$. Cet estimateur est donné par un filtrage de Wiener généralisé dans le cas de distributions $S\alpha S$ [7]. Dans le rapport technique annexe [10], nous prouvons que ce résultat est toujours valable pour la famille $P\alpha S$. Ainsi, pour des variables aléatoires de Lévy ($\alpha = 1/2$), on a :

$$\hat{X}_k = \mathbb{E}(X_k|X) = \frac{\sigma_k^{\odot 1/2}}{\sum_l \sigma_l^{\odot 1/2}} \odot X = \frac{W_k H_k}{\sum_l W_l H_l} \odot X. \quad (11)$$

4 Évaluation expérimentale

4.1 Restauration de spectrogrammes musicaux

Nous testons dans cette section la capacité du modèle de Lévy NMF à restaurer des spectrogrammes musicaux corrompus par des bruits impulsionnels. Nous considérons 6 morceaux de guitare issus de la base de données IDMT-SMT-GUITAR³ échantillonnés à 8000 Hz. La matrice X est le spectrogramme

2. <http://www.cs.tut.fi/~magron/ressources/levyNMF.m>

3. https://www.idmt.fraunhofer.de/en/business_units/m2d/smt/guitar.html

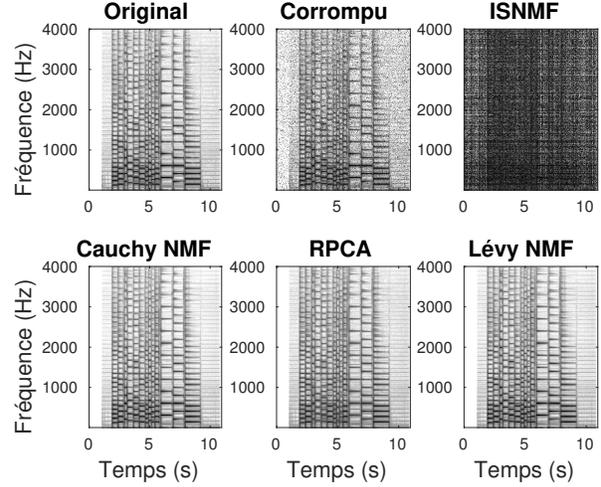


FIG. 1 – Restauration de spectrogrammes musicaux.

TAB. 1 – Log-divergence KL entre spectrogrammes originaux et débruités.

IS	KL	Cauchy	Lévy	RPCA
9.0	6.2	3.4	3.2	3.6

d'amplitude (module de la transformée de Fourier à court-terme) des signaux, calculé avec une fenêtre d'analyse de Hann de longueur 125 ms et 75 % de recouvrement. Les spectrogrammes sont corrompus par des bruits impulsionnels qui représentent 10 % des données. Ils sont ensuite estimés par 200 itérations des algorithmes de Lévy, IS [2], KL [1] et Cauchy NMF [5], ainsi que l'algorithme d'Analyse en Composantes Principales Robuste (RPCA) [11], avec un rang de factorisation de $K = 30$.

Les spectrogrammes obtenus sont présentés sur la figure 1 (les résultats obtenus avec KLNMF et ISNMF sont similaires). Les techniques de NMF usuelles ne semblent pas capables de débruiter les données : l'estimation des paramètres est fortement détériorée par la présence de bruits impulsionnels. À l'inverse, le bruit est entièrement supprimé en appliquant Lévy NMF. Cette performance est confirmée par le tableau 1 qui présente la divergence KL entre spectrogrammes originaux et estimés moyennée sur les 6 morceaux. Les meilleurs résultats sont obtenus par les modèles de Cauchy et Lévy NMF, ce qui est d'autant plus remarquable qu'aucune information sur la localisation du bruit n'est injectée dans ces algorithmes.

4.2 Séparation de spectres de fluorescence

La spectroscopie de fluorescence [12] consiste à mesurer le spectre d'excitation-émission d'un mélange de composants, qui est modélisé comme la somme des spectres des composants purs W pondérée par leurs concentrations H . La séparation de sources de tels spectres permet donc d'identifier les espèces présentes dans un mélange ainsi que leurs concentrations. Les

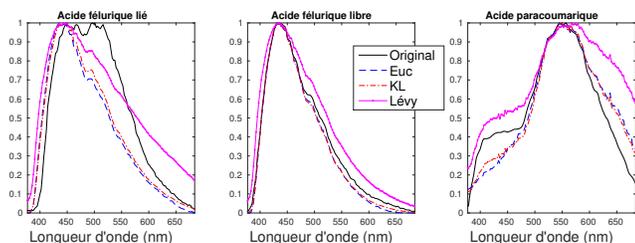


FIG. 2 – Spectres des composants purs.

TAB. 2 – Corrélation (%) entre sources Oracles et 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

modèles NMF ont donc leur place dans ce cadre [12] puisque les données sont non-négatives et qu'ils sont conformes à l'hypothèse d'additivité des spectres.

Nous comparons donc le modèle de Lévy NMF aux modèles précédemment employés dans la littérature : la NMF avec distance euclidienne (EucNMF) [13] et avec divergence KL [12]. Chaque algorithme utilise 50 itérations. Les données (détaillées dans [12]) consistent en une collection de $T = 400$ spectres d'émission (avec $F = 128$ canaux fréquentiels) qui sont la somme de spectres (connus) de $K = 3$ composants de concentrations inconnues. Nous estimons tout d'abord W et H à partir des mélanges. Comme l'indique la figure 2, les trois méthodes permettent d'apprendre les spectres des composants purs, les spectres de Lévy ayant la particularité d'approcher les spectres de référence par valeur supérieure.

Enfin, nous estimons les sources isolées par filtrage de Wiener ((11) pour Lévy NMF). Nous apprenons également la matrice H en supposant W connue : ce cas Oracle sert alors de référence de comparaison. La similarité entre les sources Oracles et estimées est mesurée par la corrélation, dont les valeurs sont présentées dans le tableau 2. Lévy NMF conduit à la meilleure performance pour toutes les sources, ce qui confirme son potentiel pour une tâche de séparation de données non-négatives.

5 Conclusion

Le modèle de Lévy NMF introduit dans cet article apparaît prometteur pour une tâche de décomposition robuste de données non-négatives. Outre l'audio et la spectroscopie, il peut être utile dans de nombreux domaines où l'on rencontre de telles données, la vision par ordinateur [1] ou l'optique, où la loi de Lévy apparaît naturellement [14].

Enfin, nous envisageons de généraliser cette approche à toute distribution PaS : ces modèles pourront être estimés par des méthodes MCMC [6]. Alternativement, on pourra s'intéresser à la famille inverse-Gamma (dont la loi de Lévy est un cas particulier) qui possède une densité analytiquement simple.

Références

- [1] Daniel D. LEE et H. Sebastian SEUNG. « Learning the parts of objects by non-negative matrix factorization ». *Nature*, 401(6755) :788–791, 1999.
- [2] Cédric FÉVOTTE, Nancy BERTIN et Jean-Louis DURRIEU. « Nonnegative matrix factorization with the Itakura-Saito divergence : With application to music analysis ». *Neural computation*, 21(3) :793–830, mars 2009.
- [3] Tuomas VIRTANEN. « Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria ». *IEEE Transactions on Audio, Speech, and Language Processing*, 15(3) :1066–1074, mars 2007.
- [4] John P. NOLAN. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston, 2015.
- [5] Antoine LIUTKUS, Derry FITZGERALD et Roland BADEAU. « Cauchy Nonnegative Matrix Factorization ». Dans *Proc. IEEE WASPAA*, New Paltz, NY, USA, octobre 2015.
- [6] Umut SIMSEKLI, Antoine LIUTKUS et Ali Taylan CEMGIL. « Alpha-Stable Matrix Factorization ». *IEEE Signal Processing Letters*, 22(12) :2289–2293, décembre 2015.
- [7] Antoine LIUTKUS et Roland BADEAU. « Generalized Wiener filtering with fractional power spectrograms ». Dans *Proc. IEEE ICASSP*, Brisbane, Australie, avril 2015.
- [8] David R. HUNTER et Kenneth LANGE. « A tutorial on MM algorithms ». *The American Statistician*, 58(1) :30–37, 2004.
- [9] Cédric FÉVOTTE et Jérôme IDIER. « Algorithms for non-negative matrix factorization with the beta-divergence ». *Neural Computation*, 23(9) :2421–2456, septembre 2011.
- [10] Paul MAGRON, Roland BADEAU et Antoine LIUTKUS. « Generalized Wiener filtering for positive alpha-stable random variables ». Rapport Technique D004, Télécom ParisTech, 2016.
- [11] Emmanuel J. CANDÈS, Xiaodong LI, Yi MA et John WRIGHT. « Robust principal component analysis? ». *Journal of the ACM*, 58(3), 2011.
- [12] Cyril GOBINET, Eric PERRIN et Régis HUEZ. « Application of non-negative matrix factorization to fluorescence spectroscopy ». Dans *Proc. EUSIPCO*, Vienne, Autriche, septembre 2004.
- [13] Anne-Sophie MONTCUQUET, Lionel HERVE, Laurent GUYON, Jean-Marc DINTEN et Jérôme I. MARS. « Non-negative Matrix Factorization : A blind sources separation method to unmix fluorescence spectra ». Dans *Proc. WHISPERS*, Grenoble, France, août 2009.
- [14] Geoffrey L. ROGERS. « Multiple path analysis of reflectance from turbid media ». *Journal of the Optical Society of America*, 25(11) :2879–2883, 2008.