

Modélisation robuste de données non-négatives

Sommaire

9.1 Distributions Positives α-stables	147
9.1.1 Lois stables	147
9.1.2 Variables à support non-négatif	147
9.1.3 Distribution de Lévy	148
9.2 Modèle de Lévy NMF	150
9.2.1 Modèle de mélange	150
9.2.2 Estimation du modèle	150
9.2.3 Algorithmes de Lévy NMF	159
9.2.4 Interprétation des mises à jour	159
9.3 Estimateur des sources	160
9.3.1 Somme de 2 variables	160
9.3.2 Somme de K variables	164
9.4 Expériences	164
9.4.1 Algorithmes de Lévy NMF	165
9.4.2 Représentation de bruits impulsionnels	166
9.4.3 Applications	167
9.5 Conclusion	172

Dans la deuxième partie de ce manuscrit, nous nous sommes intéressés à des modèles de signaux qui permettent d'obtenir des contraintes de phase dans un but de reconstruction de celle-ci. Au chapitre précédent, nous avons proposé un modèle probabiliste qui prend en compte les à priori que l'on peut obtenir sur la phase grâce à de tels modèles. Néanmoins, nous avons jusqu'à présent supposé que les amplitudes des sources étaient connues (cas Oracle) ou bien estimées par avance. En tout cas, elles étaient déterministes. Dans le but de mettre au point un modèle de sources complet pour une application de séparation de sources réaliste, nous allons chercher à modéliser les amplitudes des sources. En outre, nous avons constaté expérimentalement (aux chapitres 5 et 8 en particulier) que la performance des techniques de reconstruction de phase dépend de la qualité des amplitudes estimées préalablement.

C'est pourquoi, dans ce chapitre, nous nous intéressons à la modélisation de données non-négatives. En particulier, nous mettons en évidence le fait que les lois usuelles (telles que la loi de Poisson [VIRTANEN et al. \(2008\)](#); [CEMGIL \(2009\)](#)) employées pour modéliser des données non-négatives, notamment en traitement du signal audio, ne sont pas à *queue lourde* : cela implique que de telles distributions représentent mal des valeurs éloignées de leurs modes et qui sont considérées comme "abbérantes", alors que de telles valeurs sont pourtant fréquentes en audio. Les lois α -stables [NOLAN \(2015\)](#), en revanche, présentent cette propriété de robustesse aux valeurs abbérantes tout en conservant la propriété d'additivité des lois de Poisson. Ces distributions ont été utilisées avec succès dans des applications de traitement du signal robuste, notamment en audio [SAMORADNITSKY et TAQQU \(1994\)](#); [GODSILL et KURUOGLU \(1999\)](#); [BASSIOU et al. \(2013\)](#). Nous introduisons un sous-ensemble des lois stables : les distributions Positives α -stables (P α S), qui sont à support non-négatif, et peuvent donc modéliser des données non-négatives telles que des amplitudes de TFCT de signaux audio.

Après avoir introduit cette famille de lois, nous en étudions un cas particulier : la distribution de Lévy. Il s'agit du seul cas de loi P α S pour lequel nous pouvons écrire analytiquement la densité de probabilité. Nous proposons un modèle de mélange de données non-négatives appelé *Lévy NMF*, qui est appliqué à la séparation de sources non-négatives. Ceci ne porte que sur un sous-problème particulier de notre objectif final, qui consiste à séparer des sources complexes (en utilisant notamment un modèle de données non-négatives), mais nous avons trouvé intéressant d'étudier plus en détail ce modèle, car il s'agit d'une thématique de recherche active. La séparation de sources non-négatives est en effet un problème fondamental dans de nombreux domaines tels que l'imagerie par résonance magnétique [SAJDA et al. \(2004\)](#), la reconnaissance de visage [GUILLAMET et VITRIA \(2002\)](#) ou la fouille de données textuelles [PAUCA et al. \(2004\)](#). Nous proposons des applications en débruitage de spectrogrammes audio et en séparation de sources en spectroscopie de fluorescence [LIU et al. \(2013\)](#).

Un des résultats forts de ce chapitre, la généralisation du filtrage de Wiener aux variables P α S, a fait l'objet d'une publication dans un rapport technique déposé dans la base de données de Télécom ParisTech [MAGRON et al. \(2016c\)](#). Le modèle de Lévy NMF ainsi que la partie expérimentale ont fait l'objet d'un article soumis dans la revue *IEEE Signal Processing Letters* [MAGRON et al. \(2017c\)](#). Ce travail ayant été mené conjointement avec Antoine Liutkus de l'INRIA Nancy, nous profitons de ce préambule pour le remercier.

Dans la section 9.1, nous introduisons les distributions P α S. Dans la section 9.2, nous présentons un modèle de mélange de données non-négatives, dont nous estimons les paramètres par différentes méthodes. Dans la section 9.3, nous fournissons une justification théorique à l'utilisation d'un filtrage de Wiener généralisé pour estimer les sources de mélanges P α S. Une validation expérimentale est proposée dans la section 9.4 avec des exemples d'application en audio et en spectroscopie de fluorescence. Enfin, nous concluons dans la section 9.5.

9.1 Distributions Positives α -stables

9.1.1 Lois stables

Les distributions α -stables [NOLAN \(2015\)](#), notées $\mathcal{S}(\alpha, \sigma, \mu, \beta)$, sont définies comme l'ensemble des distributions sur la variable aléatoire X à valeurs dans \mathbb{R} dont la fonction caractéristique est :

$$\varphi_X(t) = \mathbb{E}(e^{itX}) = \begin{cases} e^{it\mu - \sigma^\alpha |t|^\alpha (1 - i\beta \Phi \operatorname{sg}(t))} & \text{si } \alpha \neq 1, \\ e^{it\mu - \sigma |t| (1 + i\beta \frac{2}{\pi} \log(|t|) \operatorname{sg}(t))} & \text{si } \alpha = 1, \end{cases} \quad (9.1)$$

où $\operatorname{sg}(t)$ désigne le signe de $t \in \mathbb{R}$, et $\Phi = \tan(\frac{\pi\alpha}{2})$. Elles dépendent de 4 paramètres :

- Un exposant caractéristique ou paramètre de forme $\alpha \in]0, 2]$ qui détermine la forme de la queue de la distribution (pour des valeurs faibles de α , la distribution est dite à queue lourde) ;
- Un paramètre d'échelle $\sigma \in]0, +\infty[$ qui mesure la dispersion de la distribution autour de son mode ;
- Un paramètre de localisation $\mu \in \mathbb{R}$;
- Un paramètre d'asymétrie $\beta \in [-1, 1]$.

La propriété de stabilité se traduit par le fait que si K variables X_k suivent une loi α -stable et sont indépendantes, alors la somme de ces variables suit également une loi α -stable. En particulier, si $X_k \sim \mathcal{S}(\alpha, \sigma_k, \mu_k, \beta_k)$, alors :

$$X = \sum_k X_k \sim \mathcal{S}(\alpha, \sigma, \mu, \beta) \text{ avec } \begin{cases} \sigma^\alpha = \sum_k \sigma_k^\alpha, \\ \mu = \sum_k \mu_k, \\ \beta = \frac{\sum_k \beta_k \sigma_k^\alpha}{\sum_k \sigma_k^\alpha}. \end{cases} \quad (9.2)$$

Cette propriété est utilisée pour manipuler les grandeurs X et obtenir un certain nombre de résultats. Elle donne des informations sur la densité du mélange, ce qui permet l'application de méthodes de type ML ou MAP pour l'estimation des paramètres.

Les distributions dont le paramètre d'asymétrie β est nul sont dites Symétriques α -stables (SaS) [LIUTKUS et BADEAU \(2015\)](#), notée $\mathcal{S}\alpha\mathcal{S}(\alpha, \sigma, \mu) = \mathcal{S}(\alpha, \sigma, \mu, 0)$. Ces distributions sont à support dans \mathbb{R} et peuvent être étendues à \mathbb{C} en raisonnant sur les parties réelles et imaginaires des variables complexes considérées (lois SaS isotropes).

9.1.2 Variables à support non-négatif

Les distributions SaS ne sont pas adaptées à la modélisation de données non-négatives puisqu'elles sont à support dans \mathbb{R} (ou \mathbb{C}). Nous proposons d'introduire un autre cas particulier de distributions α -stables qui soient à support non-négatif : les distributions Positives α -stables (PaS).

Lorsque $\beta = 1$ et $\alpha < 1$, on peut montrer [NOLAN \(2015\)](#) que la distribution a pour support $[\mu, +\infty[$. Pour $\mu = 0$, on obtient un ensemble de distributions qui sont à support non-négatif. Une distribution PaS est donc telle que $\mathcal{P}\alpha\mathcal{S}(\sigma) = \mathcal{S}(\alpha, \sigma, 0, 1)$ avec $\alpha < 1$.

Soit $X_k \sim \mathcal{P}\alpha\mathcal{S}(\sigma_k)$ pour $k \in \llbracket 1, K \rrbracket$, les X_k étant indépendants. On a, par stabilité :

$$X = \sum_k X_k \sim \mathcal{P}\alpha\mathcal{S}(\sigma) \text{ avec } \sigma^\alpha = \sum_k \sigma_k^\alpha. \quad (9.3)$$

En règle générale, nous ne pouvons pas écrire analytiquement la densité de probabilité des distributions α -stables. Ainsi, pour estimer les paramètres de modèles de mélanges comme (9.3), il faut mettre en oeuvre des méthodes de type MCMC. C'est ce qui a été effectué dans [SIMSEKLI et al. \(2015\)](#) pour les distributions $\mathcal{S}\alpha\mathcal{S}$, et pourrait par la suite être étendu au cas de distributions $\mathcal{P}\alpha\mathcal{S}$.

9.1.3 Distribution de Lévy

Il existe trois cas de figure pour lesquels on a une expression analytique de la densité pour des distributions α -stables :

- $\alpha = 2$ et $\beta = 0$: C'est la loi normale, très populaire notamment en audio [FÉVOTTE et al. \(2009\)](#) ;
- $\alpha = 1$ et $\beta = 0$: C'est la loi de Cauchy, utilisée dans [LIUTKUS et al. \(2015\)](#) ;
- $\alpha = 1/2$ et $\beta = 1$: C'est la loi de Lévy.

Les deux premiers cas correspondent à des distributions $\mathcal{S}\alpha\mathcal{S}$ et ne nous intéressent donc pas ici puisqu'on cherche à modéliser des données non-négatives. Par contre, la loi de Lévy présente l'avantage d'être une distribution $\mathcal{P}\alpha\mathcal{S}$ puisque $\beta = 1$ et $\alpha < 1$. Son paramètre de localisation peut être fixé arbitrairement, et comme on cherche à modéliser des données non-négatives, on choisit $\mu = 0$. Par abus de langage, on désignera par loi de Lévy cette loi dans le cas particulier où le paramètre de localisation est nul. Sa densité de probabilité, illustrée sur la figure 9.1 est donnée par :

$$p(x|\sigma) = \begin{cases} \sqrt{\frac{\sigma}{2\pi}} \frac{1}{x^{3/2}} e^{-\frac{\sigma}{2x}} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases} \quad (9.4)$$

Outre le fait d'avoir une densité de probabilité que l'on peut écrire analytiquement, la distribution de Lévy possède une queue lourde, comme illustré sur la figure 9.2. Cela signifie que des valeurs éloignées du mode de la distribution restent tout de même probables. Cette propriété se traduit par la robustesse d'un modèle de Lévy par rapport aux valeurs aberrantes : en effet, celles-ci modifient moins l'estimation des paramètres de la loi que pour des distributions à queue non lourde.

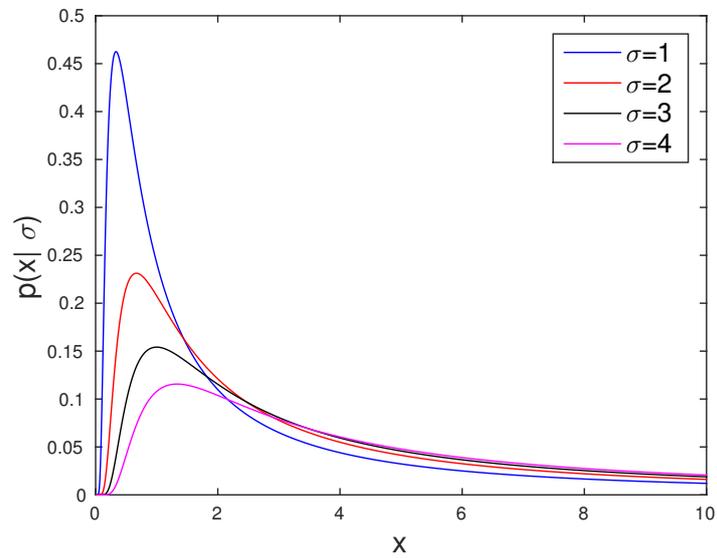


FIGURE 9.1 – Densité de probabilité d'une loi de Lévy de paramètre de localisation nul.

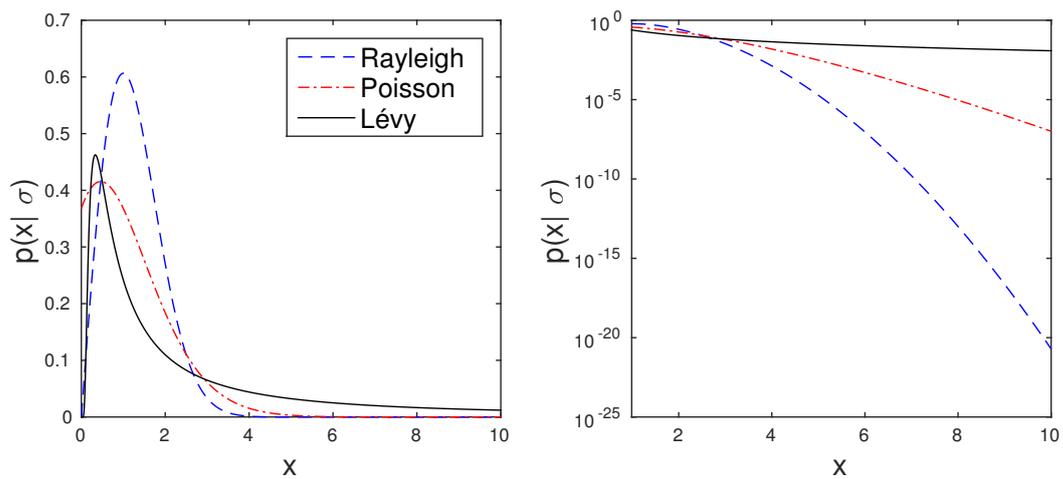


FIGURE 9.2 – Densité de probabilité pour plusieurs lois avec le même paramètre $\sigma = 1$. On observe que la distribution de Lévy possède la queue la plus lourde.

9.2 Modèle de Lévy NMF

Dans cette partie, nous proposons un modèle de mélange de sources non-négatives basé sur la distribution de Lévy.

9.2.1 Modèle de mélange

Considérons une matrice de données non-négatives $X \in \mathbb{R}_+^{F \times T}$ qui est le mélange de K sources non-négatives $X_k \in \mathbb{R}_+^{F \times T}$. On modélise les sources X_k comme des variables aléatoires matricielles suivant une loi de Lévy de paramètre $\sigma_k \in \mathbb{R}_{+*}^{F \times T}$, notée $\mathcal{L}(\sigma_k)$, $\forall k \in \llbracket 1, K \rrbracket$. On étend en effet les définitions aux variables matricielles, étant donné que l'on traite tous les points TF indépendamment. Ainsi, on a, $\forall (f, t)$, $X_k \sim \mathcal{L}(\sigma_k(f, t))$.

Comme $X = \sum_k X_k$, on a, par additivité des distributions α -stables :

$$X \sim \mathcal{L}(\sigma), \text{ avec } \sigma^{\odot 1/2} = \sum_k \sigma_k^{\odot 1/2}. \quad (9.5)$$

Comme les observations et paramètres ci-dessus sont matriciels, on propose de structurer les paramètres de dispersion par un modèle NMF :

$$\sqrt{\sigma_k(f, t)} = W(f, k)H(k, t). \quad (9.6)$$

Ainsi, la dispersion du mélange est :

$$\sigma^{\odot 1/2} = WH. \quad (9.7)$$

Ce modèle de mélange est appelé *Lévy NMF*. Un tel modèle permet donc, similairement au modèle de ISNMF, de conserver une structuration des données (qui peuvent être par exemple des spectres de notes de musique), tout en autorisant en outre aux observations de s'éloigner localement du modèle. Cela suggère donc que l'estimation des paramètres du modèle sera moins affectée par des valeurs aberrantes que dans un modèle moins robuste (comme ISNMF, qui n'est pas basé sur une distribution à queue lourde).

9.2.2 Estimation du modèle

Afin d'estimer les paramètres W et H du modèle précédemment établi, nous maximisons la vraisemblance des données (ou de façon équivalente mais plus pratique, sa log-vraisemblance). La log-vraisemblance est :

$$\begin{aligned} L(W, H) &= \sum_f \sum_t \log(p(X(f, t) | \sigma(f, t))) \\ &= \sum_f \sum_t \frac{1}{2} \log(\sigma(f, t)) - \frac{1}{2} \log(2\pi) - \frac{3}{2} \log(X(f, t)) - \frac{\sigma(f, t)}{2X(f, t)} \\ &\stackrel{c}{=} \frac{1}{2} \sum_f \sum_t \log([WH](f, t)^2) - \frac{[WH](f, t)^2}{X(f, t)}. \end{aligned}$$

On remarque alors que :

$$L(W, H) \stackrel{c}{=} -\frac{1}{2} D_{IS}([WH]^{\odot 2}, X), \quad (9.8)$$

où D_{IS} désigne la divergence d'Itakura-Saito. Ainsi, maximiser la vraisemblance des données dans le modèle Lévy NMF revient à minimiser la divergence IS entre $[WH]^{\odot 2}$ et X . En retirant les termes constants, cela revient à minimiser la fonction de coût suivante :

$$\mathcal{C}(W, H) = \sum_{f,t} \frac{[WH](f,t)^2}{X(f,t)} - 2 \log([WH](f,t)). \quad (9.9)$$

Approche naïve

L'optimisation de la fonction \mathcal{C} donnée par (9.9) peut-être effectuée par une approche heuristique similaire à LEE et SEUNG (1999) et rappelée au chapitre 2, section 2.2.4. La dérivée partielle de cette fonction par rapport à une variable θ est écrite sous forme d'une différence entre deux termes positifs :

$$\frac{\partial \mathcal{C}}{\partial \theta} = \nabla_{\theta}^+ - \nabla_{\theta}^-. \quad (9.10)$$

On considère alors la règle de mise à jour multiplicative suivante :

$$\theta \leftarrow \theta \times \frac{\nabla_{\theta}^-}{\nabla_{\theta}^+}. \quad (9.11)$$

Une telle mise à jour permet de s'assurer que le signe de θ ne change pas au cours des itérations (ce qui garantit la positivité des variables tant que l'initialisation est faite avec des valeurs positives), et que θ varie dans le sens de la décroissance (locale) de \mathcal{C} (ce qui, rappelons-le, n'est pas une garantie de décroissance de la fonction de coût).

On calcule donc la dérivée partielle de \mathcal{C} par rapport à $W(f, k)$:

$$\frac{\partial \mathcal{C}}{\partial W(f, k)} = \underbrace{\sum_t 2H(k, t) \frac{[WH](f, t)}{X(f, t)}}_{\nabla_{W(f, k)}^+} - \underbrace{\sum_t 2 \frac{H(k, t)}{[WH](f, t)}}_{\nabla_{W(f, k)}^-}. \quad (9.12)$$

La mise à jour de $W(f, k)$ est donnée par :

$$W(f, k) \leftarrow W(f, k) \frac{\sum_t \frac{H(k, t)}{[WH](f, t)}}{\sum_t H(k, t) \frac{[WH](f, t)}{X(f, t)}}. \quad (9.13)$$

De façon tout à fait similaire, on peut calculer la mise à jour de $H(k, t)$:

$$H(k, t) \leftarrow H(k, t) \frac{\sum_f \frac{W(f, k)}{[WH](f, t)}}{\sum_f W(f, k) \frac{[WH](f, t)}{X(f, t)}}. \quad (9.14)$$

Sous forme matricielle, on obtient donc :

$$W \leftarrow W \odot \frac{[WH]^{\odot -1} H^T}{([WH] \odot X^{\odot -1}) H^T}, \quad (9.15)$$

et

$$H \leftarrow H \odot \frac{W^T [WH]^{\odot -1}}{W^T ([WH] \odot X^{\odot -1})}. \quad (9.16)$$

Les équations (9.15) et (9.16) fournissent des règles de mise à jour pour les variables W et H qui assurent que les variables sont actualisées dans le sens de la décroissance locale de la fonction de coût. Nous avons cependant observé expérimentalement que ces règles ne faisaient pas décroître la fonction de coût (9.9), et que celle-ci ne convergeait pas, même après un grand nombre d'itérations. D'après [BADEAU et al. \(2010\)](#), il peut être souhaitable d'utiliser les règles de mise à jour suivantes :

$$W \leftarrow W \odot \left(\frac{[WH]^{\odot -1} H^T}{([WH] \odot X^{\odot -1}) H^T} \right)^{\odot \eta}, \quad (9.17)$$

et

$$H \leftarrow H \odot \left(\frac{W^T [WH]^{\odot -1}}{W^T ([WH] \odot X^{\odot -1})} \right)^{\odot \eta}, \quad (9.18)$$

où η est un exposant qui, choisi judicieusement, augmente significativement la vitesse de convergence. Nous proposons par la suite d'utiliser une nouvelle technique d'optimisation, qui permet de se ramener à ce type de règles.

Approche Majoration-Minimisation

L'algorithme Majoration-Minimisation (MM) [HUNTER et LANGE \(2004\)](#) fournit un cadre théorique qui permet de minimiser la fonction de coût $\mathcal{C}(\theta)$. Le principe de l'algorithme MM (*cf.* chapitre 2, section 2.2.4) est de majorer la fonction de coût $\mathcal{C}(\theta)$ par une fonction auxiliaire $G(\theta, \bar{\theta})$, telle que $G(\bar{\theta}, \bar{\theta}) = \mathcal{C}(\bar{\theta})$. À $\bar{\theta}$ fixé, la minimisation de G par rapport à θ produit une mise à jour de ce paramètre qui conduit, par construction de G , à la décroissance de \mathcal{C} .

Pour construire la fonction auxiliaire G , on s'inspire de la démarche conduite dans [FÉVOTTE et IDIER \(2011\)](#); [FÉVOTTE \(2011\)](#), qui consiste à écrire \mathcal{C} comme somme de fonctions convexes et concaves. Les termes convexes sont majorés via l'inégalité de Jensen, et les termes concaves sont majorés par leurs tangentes.

Nous détaillons ici la construction de la fonction auxiliaire en un point $\bar{W}(f, k)$ (cette notation ne désigne pas ici la conjugaison complexe, puisqu'on travaille sur des données non-négatives). Un calcul semblable permet d'obtenir la fonction auxiliaire en un point $\bar{H}(k, t)$ pour obtenir la règle de mise à jour sur H .

On s'intéresse tout d'abord au premier terme dans l'expression de \mathcal{C} donnée par l'équation (9.9) :

$$\mathcal{C}_1(W, H) = \sum_{f,t} \frac{[WH](f, t)^2}{X(f, t)}. \quad (9.19)$$

On note $\bar{V}(f, t) = [\bar{W}H](f, t)$ et $\rho_k(f, t) = \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)}$. On peut alors introduire ces termes dans la somme ci-dessous :

$$\begin{aligned} [WH](f, t)^2 &= \left(\sum_k W(f, k)H(k, t) \right)^2 \\ &= \left(\sum_k \rho_k(f, t) \frac{W(f, k)H(k, t)}{\rho_k(f, t)} \right)^2. \end{aligned}$$

Les termes ρ_k jouent le rôle de poids et ont pour propriété de sommer à l'unité (par définition de \bar{V}). Comme la fonction carré est convexe et que les poids somment à l'unité, on peut appliquer l'inégalité de Jensen :

$$\begin{aligned} [WH](f, t)^2 &\leq \sum_k \rho_k(f, t) \left(\frac{W(f, k)H(k, t)}{\rho_k(f, t)} \right)^2 \\ &\leq \sum_k \frac{W(f, k)^2 H(k, t)^2}{\bar{W}(f, k)H(k, t)} \bar{V}(f, t) \\ &\leq \bar{V}(f, t) \sum_k \frac{W(f, k)^2 H(k, t)}{\bar{W}(f, k)}. \end{aligned}$$

Ainsi, on obtient finalement la majoration suivante du premier terme dans la fonction de coût :

$$\mathcal{C}_1(W, H) \leq \sum_{f, t} \frac{\bar{V}(f, t)}{X(f, t)} \sum_k \frac{W(f, k)^2 H(k, t)}{\bar{W}(f, k)}. \quad (9.20)$$

On procède à présent au même type de majoration sur le deuxième terme de la fonction de coût :

$$-2 \log([WH](f, t)) = -2 \log \left(\sum_k \rho_k(f, t) \frac{W(f, k)H(k, t)}{\rho_k(f, t)} \right). \quad (9.21)$$

La fonction $-2 \log$ étant convexe (le logarithme étant concave), on applique de nouveau l'inégalité de convexité et on obtient :

$$\begin{aligned} -2 \log([WH](f, t)) &\leq \sum_k \rho_k(f, t) \left(-2 \log \frac{W(f, k)H(k, t)}{\rho_k(f, t)} \right) \\ &\leq -2 \sum_k \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)} \log \left(W(f, k) \frac{\bar{V}(f, t)}{\bar{W}(f, k)} \right). \end{aligned}$$

Cela nous mène à une majoration du deuxième terme de la fonction de coût :

$$\mathcal{C}_2(W, H) \leq -2 \sum_{f, t} \sum_k \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)} \log \left(W(f, k) \frac{\bar{V}(f, t)}{\bar{W}(f, k)} \right). \quad (9.22)$$

En combinant (9.20) et (9.22), on obtient finalement une majorante de la fonction de coût : $\mathcal{C}(W, H) \leq G(W, H, \bar{W})$ avec :

$$G(W, H, \bar{W}) = \sum_{f, t, k} \frac{\bar{V}(f, t)H(k, t)}{X(f, t)\bar{W}(f, k)} W(f, k)^2 - 2 \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)} \log \left(W(f, k) \frac{\bar{V}(f, t)}{\bar{W}(f, k)} \right). \quad (9.23)$$

Il est par ailleurs aisé de constater que $\mathcal{C}(W, H) = G(W, H, W)$ (c'est obtenu par construction). G est donc une fonction auxiliaire à \mathcal{C} .

On applique alors la méthode MM : on minimise la fonction auxiliaire G par rapport à la variable $W(f, k)$ à H et \bar{W} fixés. La dérivée partielle est :

$$\frac{\partial G}{\partial W(f, k)}(W, H, \bar{W}) = \sum_t 2 \frac{\bar{V}(f, t)H(k, t)}{X(f, t)\bar{W}(f, k)} W(f, k) - 2 \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)} \frac{1}{W(f, k)}, \quad (9.24)$$

et l'annulation de cette dérivée conduit à :

$$\frac{W(f, k)}{\overline{W}(f, k)} \sum_t \frac{\overline{V}(f, t) H(k, t)}{X(f, t)} = \frac{\overline{W}(f, k)}{W(f, k)} \sum_t \frac{H(k, t)}{\overline{V}(f, t)}, \quad (9.25)$$

soit, en remplaçant \overline{V} par sa définition $[\overline{W}H]$:

$$W(f, k)^2 = \overline{W}(f, k)^2 \frac{\sum_t \frac{H(k, t)}{[\overline{W}H](f, t)}}{\sum_t H(k, t) \frac{[\overline{W}H](f, t)}{X(f, t)}}, \quad (9.26)$$

et donc (en ne gardant que la solution positive, la seule qui nous intéresse) :

$$W(f, k) = \overline{W}(f, k) \left(\frac{\sum_t \frac{H(k, t)}{[\overline{W}H](f, t)}}{\sum_t H(k, t) \frac{[\overline{W}H](f, t)}{X(f, t)}} \right)^{1/2}. \quad (9.27)$$

On retrouve une mise à jour similaire à celle de l'approche naïve, exceptée une puissance 1/2 supplémentaire. Comme on le faisait remarquer dans la section précédente, nous avons trouvé un exposant η qui permette d'assurer la décroissance de la fonction de coût. En tenant compte du fait que la variable \overline{W} est ensuite actualisée en prenant la nouvelle valeur de W , on peut synthétiser la mise à jour sur W sous forme matricielle :

$$W \leftarrow W \odot \left(\frac{[WH]^{\odot -1} H^T}{([WH] \odot X^{\odot -1}) H^T} \right)^{\odot 1/2}. \quad (9.28)$$

Par un calcul similaire (on introduit la variable auxiliaire \overline{H} et on utilise les mêmes inégalités de convexité), on construit une fonction auxiliaire à \mathcal{C} en un point \overline{H} , et on peut obtenir la mise à jour suivante :

$$H \leftarrow H \odot \left(\frac{W^T [WH]^{\odot -1}}{W^T ([WH] \odot X^{\odot -1})} \right)^{\odot 1/2}. \quad (9.29)$$

Les équations (9.28) et (9.29) fournissent donc les règles de mise à jour pour l'estimation des paramètres W et H du modèle de Lévy NMF par l'approche MM. Par construction de ces règles, nous avons la garantie de la monotonie de la fonction de coût, contrairement au cas de l'approche naïve.

Approche Majoration-Égalisation

L'approche MM présentée ci-dessus est un moyen d'obtenir des règles de mise à jour pour lesquelles on a une garantie de décroissance de la fonction de coût. Néanmoins, il y a d'autres façons d'obtenir de telles mises à jour qu'en minimisant la fonction majorante : en effet, toute mise à jour vérifiant $\mathcal{C}(\theta^{(it+1)}) \leq \mathcal{C}(\theta^{(it)})$ convient. L'approche Majoration-Égalisation (ME) consiste à choisir $\theta^{(it+1)}$ tel que $G(\theta^{(it+1)}, \theta^{(it)}) = G(\theta^{(it)}, \theta^{(it)})$. Cette approche est illustrée sur la figure 9.3.

Par construction de la fonction auxiliaire :

$$\mathcal{C}(\theta^{(it+1)}) \leq G(\theta^{(it+1)}, \theta^{(it)}), \quad (9.30)$$

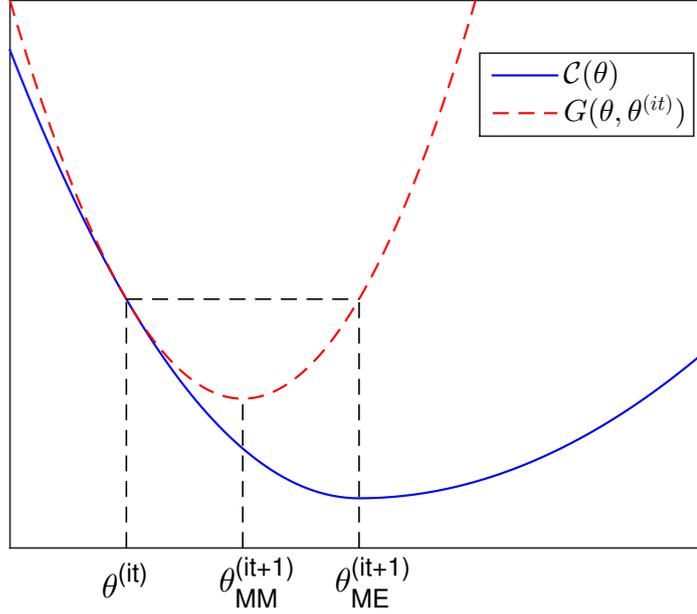


FIGURE 9.3 – Illustration des approches MM et ME : la fonction de coût $\mathcal{C}(\theta)$ est majorée à l'itération (it) par la fonction auxiliaire $G(\theta, \theta^{(it)})$. À partir de celle-ci, on peut obtenir les nouveaux estimateurs de θ , par minimisation (MM) ou par égalisation (ME).

et donc, en appliquant l'approche ME, on a :

$$\mathcal{C}(\theta^{(it+1)}) \leq G(\theta^{(it+1)}, \theta^{(it)}) = G(\theta^{(it)}, \theta^{(it)}) = \mathcal{C}(\theta^{(it)}), \quad (9.31)$$

ce qui prouve que la fonction objectif \mathcal{C} est décroissante sous cette mise à jour. Nous proposons donc d'appliquer cette méthode pour obtenir une nouvelle procédure d'estimation des paramètres. En effet, cette approche peut être intéressante car il est montré dans **FÉVOTTE et IDIER (2011)** que les mises à jour ME convergent plus rapidement que les mises à jour MM dans le cas de la NMF avec β -divergence. Il est possible que pour notre cas d'application également, de telles mises à jour accélèrent la convergence de l'algorithme. Pour l'estimation de W (encore une fois, celle de H est équivalente, nous ne la détaillons donc pas), on est amené à résoudre l'équation suivante :

$$G(W, H, \bar{W}) = G(\bar{W}, H, \bar{W}). \quad (9.32)$$

En utilisant la séparabilité des variables dans l'expression de G , c'est équivalent à résoudre, pour tout (f, k) :

$$\begin{aligned} \sum_t \frac{\bar{V}(f, t)H(k, t)}{X(f, t)\bar{W}(f, k)} W(f, k)^2 - 2 \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)} \log \left(W(f, k) \frac{\bar{V}(f, t)}{\bar{W}(f, k)} \right) \\ = \sum_t \frac{\bar{V}(f, t)H(k, t)}{X(f, t)\bar{W}(f, k)} \bar{W}(f, k)^2 - 2 \frac{\bar{W}(f, k)H(k, t)}{\bar{V}(f, t)} \log \left(\bar{W}(f, k) \frac{\bar{V}(f, t)}{\bar{W}(f, k)} \right), \end{aligned} \quad (9.33)$$

que l'on peut simplifier et réécrire :

$$\left(\frac{W(f, k)^2}{\bar{W}(f, k)} - \bar{W}(f, k) \right) \sum_t \frac{\bar{V}(f, t)H(k, t)}{X(f, t)} = 2\bar{W}(f, k) \log \left(\frac{W(f, k)}{\bar{W}(f, k)} \right) \sum_t \frac{H(k, t)}{\bar{V}(f, t)}. \quad (9.34)$$

En notant :

$$a_W(f, k) = \frac{\sum_t \frac{H(k, t)}{\bar{V}(f, t)}}{\sum_t \frac{\bar{V}(f, t)H(k, t)}{X(f, t)}}, \quad (9.35)$$

on a alors :

$$\frac{W(f, k)^2}{\bar{W}(f, k)^2} - 1 = 2a_W(f, k) \log \left(\frac{W(f, k)}{\bar{W}(f, k)} \right). \quad (9.36)$$

Dans le but d'obtenir des règles de mise à jour multiplicatives, on cherche, sans perte de généralité, les solutions sous la forme $W = \bar{W}a_W^{\odot \eta_W}$ où $\eta_W \geq 0$ est à déterminer. On peut alors réécrire l'équation ci-dessus :

$$a_W(f, k)^{2\eta_W(f, k)} = 1 + 2\eta_W(f, k)a_W(f, k) \log a_W(f, k). \quad (9.37)$$

On obtient exactement la même équation en raisonnant sur H plutôt que sur W : on utilise alors le paramètre

$$a_H(k, t) = \frac{\sum_f \frac{W(f, k)}{[W\bar{H}](f, t)}}{\sum_f W(f, k) \frac{[W\bar{H}](f, t)}{X(f, t)}}, \quad (9.38)$$

et un exposant η_H . On peut donc écrire, sous forme matricielle, l'équation définissant l'exposant η_θ (avec $\theta = W$ ou H) qui mène à l'estimation ME :

$$a_\theta^{2\eta_\theta} = 1 + 2\eta_\theta \odot a_\theta \odot \log a_\theta. \quad (9.39)$$

Comme pour les deux matrices W et H , et pour toutes les entrées de ces matrices, on est amené à résoudre l'équation ci-dessus, on choisit, par souci de clarté, de considérer une variable (W ou H) et une entrée de cette variable, et on s'affranchit des notations d'indices et de matrices. On cherche donc à résoudre sur \mathbb{R}_+ l'équation suivante :

$$a^{2\eta} = 1 + 2\eta a \log a, \quad (9.40)$$

ce qui autrement dit, revient à trouver un (ou plusieurs) zéro de la fonction définie sur \mathbb{R}_+ par :

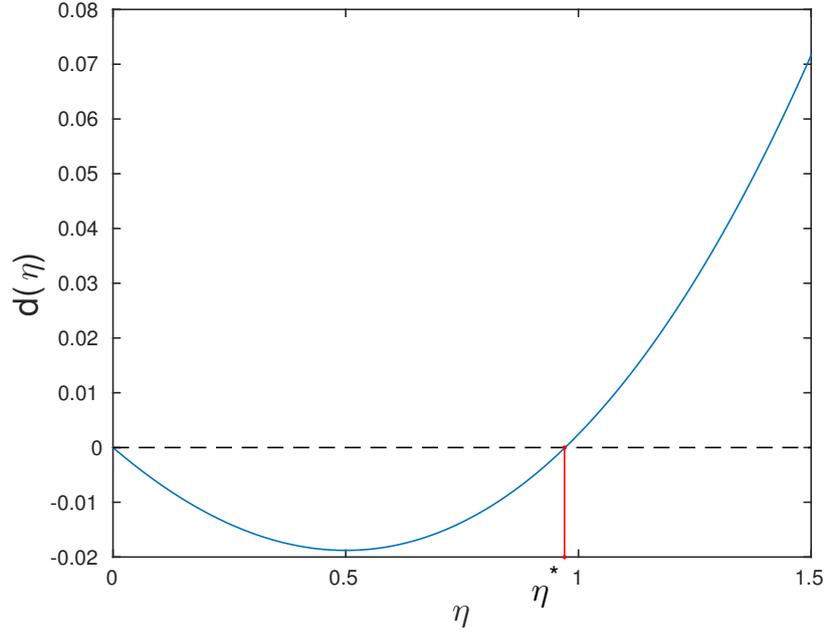
$$d(\eta) = a^{2\eta} - 1 - 2\eta a \log a. \quad (9.41)$$

La fonction d est illustrée sur la figure 9.4.

On considère le cas non-trivial où $a \neq 1$. Il est aisé de constater que $d(0) = 0$ et que $\lim_{\eta \rightarrow +\infty} d(\eta) = +\infty$. Un calcul élémentaire de dérivée montre que cette fonction possède un minimum atteint en $\eta = 1/2$, que cette valeur minimale est négative, et que d est strictement décroissante sur $[0, 1/2]$ et croissante sur $[1/2, +\infty[$. Ainsi, lorsque $a \neq 1$, le théorème des valeurs intermédiaires garantit qu'il existe exactement 2 solutions à l'équation (9.40). Une de ces solutions est triviale et correspond à $\eta = 0$, ce qui signifie que l'on a alors $\theta = \bar{\theta}$: la mise à jour ne modifie pas la valeur des paramètres, ce qui ne nous intéresse pas ici. Nous cherchons donc à obtenir l'autre solution de l'équation. Les solutions de cette équation peuvent être écrites de façon exacte par le biais des fonctions de Lambert [CORLESS et al. \(1996\)](#), mais cela ne nous fournit pas d'expression analytique simple de la solution.

Nous proposons donc d'approcher cette solution en utilisant un développement limité. Au voisinage de $a = 1$, on a :

$$a^{2\eta} = (1 + (a - 1))^{2\eta} \approx 1 + 2\eta(a - 1) + \frac{2\eta(2\eta - 1)}{2}(a - 1)^2. \quad (9.42)$$


 FIGURE 9.4 – Fonction d définie par (9.41) pour $a = 1.2$.

En réinjectant cette expression dans (9.40), on obtient :

$$1 + 2\eta(a - 1) + \frac{2\eta(2\eta - 1)}{2}(a - 1)^2 = 1 + 2\eta a \log a, \quad (9.43)$$

et, en simplifiant par 2η (que l'on suppose non-nul pour s'affranchir de cette solution triviale) :

$$\begin{aligned} (a - 1) + \frac{(2\eta - 1)}{2}(a - 1)^2 &= a \log a \\ (2\eta - 1)(a - 1)^2 &= 2(a \log a - (a - 1)) \\ 2\eta - 1 &= 2 \frac{a \log a - (a - 1)}{(a - 1)^2}, \end{aligned}$$

soit finalement l'approximation de l'exposant optimale suivante :

$$\eta^* = \frac{1}{2} + \frac{a \log a - (a - 1)}{(a - 1)^2}. \quad (9.44)$$

Cette expression permet donc d'obtenir une approximation de l'exposant à utiliser dans les règles de mise à jour de Lévy NMF par l'approche ME.

Nous souhaitons examiner la validité de cette approximation. Nous calculons, pour différentes valeurs de a , la valeur de η^* . Nous calculons également une solution approchée $\tilde{\eta}$ de l'équation (9.40) par la fonction Matlab `fzero`. Cette valeur est considérée comme une référence (à la précision de Matlab près, elle annule l'équation (9.40)). Sur la figure 9.5, nous traçons les exposants approché et de référence, ainsi que l'erreur relative entre les deux.

Nous constatons tout d'abord que l'erreur d'approximation est relativement faible (de l'ordre de 3 %) pour des valeurs de a proches de 1. Cela signifie que plus on est proche de la convergence (ce qui correspond à $a = 1$), plus l'approximation que nous avons proposée est de bonne qualité. Par ailleurs, nous constatons que l'approximation η^* semble être inférieure à la valeur de référence $\tilde{\eta}$. Ce constat est intéressant car il traduit la propriété suivante : la mise

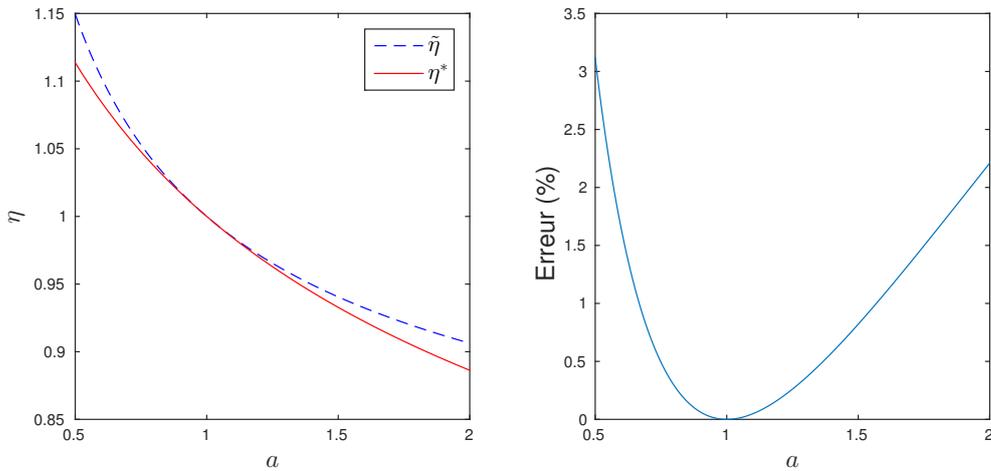


FIGURE 9.5 – Estimations de l'exposant optimal dans l'approche ME (gauche), et erreur relative (en %) entre référence et approximation (droite).

à jour obtenue avec l'exposant η^* est "moins" éloignée de la valeur courante du paramètre qu'avec l'exposant $\tilde{\eta}$. Autrement dit, la variation du paramètre θ est moins grande lors d'une mise à jour utilisant cette approximation, ce qui conduit à une garantie sur la décroissance de la fonction de coût. Considérons en effet le cas $a > 1$. On a alors :

$$\theta^{(it)} \leq \theta^{(it)} a^{\eta^*} \leq \theta^{(it)} a^{\tilde{\eta}}, \quad (9.45)$$

soit

$$\theta^{(it)} \leq \theta_{ME}^{(it+1)*} \leq \theta_{ME}^{(it+1)}, \quad (9.46)$$

et par ailleurs, $G(\theta^{(it)}, \theta^{(it)}) = G(\theta_{ME}^{(it+1)*}, \theta^{(it)})$. Donc, en utilisant la convexité de la fonction G , on a :

$$G(\theta_{ME}^{(it+1)*}, \theta^{(it)}) \leq G(\theta^{(it)}, \theta^{(it)}) = \mathcal{C}(\theta^{(it)}). \quad (9.47)$$

Enfin, on a par construction de la fonction auxiliaire $\mathcal{C}(\theta_{ME}^{(it+1)*}) \leq G(\theta_{ME}^{(it+1)*}, \theta^{(it)})$, ce qui conduit à :

$$\mathcal{C}(\theta_{ME}^{(it+1)*}) \leq \mathcal{C}(\theta^{(it)}), \quad (9.48)$$

ce qui montre la décroissance de la fonction de coût avec cette mise à jour. La démonstration est similaire lorsque $a < 1$. En d'autres termes, tant que l'on utilise un exposant η^* qui est inférieur à l'exposant $\tilde{\eta}$, on a toujours une garantie de décroissance de la fonction de coût. Il est donc intéressant de noter, à partir de la figure 9.5, que l'approximation que nous avons proposée respecte cette condition. Néanmoins, il s'agit d'une observation empirique sur une échelle réduite, et nous n'avons aucune garantie que cette propriété reste vraie pour toute valeur de a .

On s'attend donc à constater la décroissance de la fonction de coût à l'approche de la convergence (lorsque a se rapproche de 1), mais celle-ci n'est pas globalement garantie, notamment pour les premières itérations. On pourrait alors imaginer un algorithme hybride qui utilise les règles de mises à jour MM pour les premières itérations, puis les règles ME ensuite.

Méthode	Naïve	MM	ME
η_θ	1	$\frac{1}{2}$	$\frac{1}{2} + \frac{a_\theta \odot \log a_\theta - a_\theta + 1}{(a_\theta - 1)^{\odot 2}}$

TABLEAU 9.1 – Exposant dans les règles de mise à jour pour l’estimation du modèle de Lévy NMF selon la méthode choisie.

Algorithme 11 Lévy NMF (Approche Majoration-Minimisation)

Entrées :

Données $X \in \mathbb{R}_+^{F \times T}$,

Matrices initiales $W \in \mathbb{R}_+^{F \times K}$ et $H \in \mathbb{R}_+^{K \times T}$,

Nombre d’itérations N_{it} .

pour $j = 1$ à N_{it} **faire**

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

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

Normalisation de W et H .

fin pour

Sortie :

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

9.2.3 Algorithmes de Lévy NMF

Dans les trois cas que nous avons étudiés, nous avons abouti à des règles de mise à jour des paramètres W et H qui sont multiplicatives, et qui possèdent une forme similaire :

$$\theta \leftarrow \theta \odot a_\theta^{\odot \eta_\theta}, \quad (9.49)$$

les paramètres a_θ étant donnés par :

$$a_W = \frac{[WH]^{\odot -1} H^T}{([WH] \odot X^{\odot -1}) H^T} \text{ et } a_H = \frac{W^T [WH]^{\odot -1}}{W^T ([WH] \odot X^{\odot -1})}. \quad (9.50)$$

L’exposant η_θ est donné, selon la méthode choisie, par le tableau 9.1. On présente dans l’algorithme 11 la procédure itérative détaillée d’estimation des paramètres du modèle de Lévy NMF par l’approche MM. La forme des autres algorithmes est donc la même, il suffit de modifier l’exposant dans les règles de mise à jour conformément aux valeurs données dans le tableau 9.1.

9.2.4 Interprétation des mises à jour

Nous proposons une interprétation des règles de mise à jour obtenues dans le cadre de l’approche MM. Plaçons-nous dans le cas $K = 1$ et supposons que $W(f) = 1$ pour tout f . La mise à jour sur H dans l’algorithme de Lévy NMF (par approche MM) s’écrit alors :

$$H(t) \leftarrow \sqrt{\frac{F}{\sum_f \frac{1}{X(f,t)}}}, \quad (9.51)$$

et, à titre de comparaison, nous donnons également la mise à jour dans le cas de l'algorithme de ISNMF [FÉVOTTE et al. \(2009\)](#) :

$$H(t) \leftarrow \frac{1}{F} \sum_f X(f, t). \quad (9.52)$$

On constate donc que dans l'algorithme de ISNMF, la mise à jour sur H revient à effectuer une moyenne arithmétique sur les lignes de X , alors que dans l'algorithme de Lévy NMF, on effectue une opération de moyenne harmonique (élevée à la puissance $1/2$).

Considérons que X est de dimensions 10×10 et ne contient que des 1, à l'exception d'une entrée (à la t -ième ligne) qui vaut 10^8 . On cherche à factoriser X grâce à un modèle NMF de rang 1 (c'est-à-dire tel que tous les coefficients de W et H valent 1, à la normalisation près) : la valeur 10^8 est donc considérée comme aberrante dans un tel modèle. On souhaite mettre au point une technique d'estimation du modèle robuste à cette valeur. Supposons qu'à une itération donnée, on a $W(f) = 1$ pour tout f , et on cherche à estimer $H(t)$. La mise à jour dans le modèle de Lévy serait, d'après (9.51), $H(t) \leftarrow 1.05$, alors que dans le modèle de ISNMF, d'après (9.52), elle serait $H(t) \leftarrow 10^7$.

Cet exemple illustre la propriété de robustesse aux valeurs aberrantes de l'algorithme de Lévy NMF (construit sur une distribution à queue lourde) comparativement à l'algorithme de ISNMF (construit sur un modèle gaussien, donc à queue non-lourde).

9.3 Estimateur des sources

L'algorithme de Lévy NMF permet d'estimer les paramètres des sources. On cherche à présent à obtenir un estimateur des sources construit à partir de ces paramètres. Ce qui suit est valable pour toute loi positive α -stable, et pas uniquement la distribution de Lévy.

Afin de clarifier la démonstration qui suit, on considère dans un premier temps un mélange de 2 sources uniquement ($K = 2$) et une seule entrée matricielle. On note alors, pour un jeu d'indices (f, t) , $x = X(f, t)$, $s_1 = X_1(f, t)$ et $s_2 = X_2(f, t)$, de sorte à ce que $x = s_1 + s_2$. Le résultat sera par la suite étendu à K sources et à des variables matricielles.

9.3.1 Somme de 2 variables

Soit $\alpha \in]0, 1[$. Considérons deux variables s_1 et s_2 indépendantes et distribuées selon une loi PaS de paramètres de dispersion σ_1 et σ_2 respectivement. Nous allons nous inspirer de la démonstration qui est conduite dans [BADEAU et LIUTKUS \(2014\)](#) (pour des variables aléatoires SaS) pour obtenir les estimateurs des sources donnés par l'espérance à postériori des sources sachant les observations, que nous notons \hat{s}_1 et \hat{s}_2 . L'estimateur de s_1 est donc :

$$\hat{s}_1 = \mathbb{E}_{s_1|x}(s_1), \quad (9.53)$$

et par ailleurs, cette espérance est définie si et seulement si la fonction caractéristique de $s_1|x$, $\varphi_{s_1|x}(t_1) = \mathbb{E}_{s_1|x}(e^{it_1 s_1})$, est différentiable en $t_1 = 0$, auquel cas on a (par exemple d'après [BADEAU et LIUTKUS \(2014\)](#)) :

$$\mathbb{E}_{s_1|x}(s_1) = \frac{1}{i} \frac{d\varphi_{s_1|x}}{dt_1}(0). \quad (9.54)$$

Ainsi, en déterminant la fonction caractéristique de $s_1|x$ puis la dérivée de celle-ci, on pourra en déduire l'estimateur de s_1 . L'intérêt de cette méthode est d'utiliser les fonctions caractéristiques de lois stables, alors que l'on ne sait pas exprimer (dans le cas général) leurs densités de probabilités.

Étape 1 : Fonction caractéristique de $s_1|x$. Tout d'abord, par stabilité des processus P α S, x suit une loi P α S de paramètre σ tel que $\sigma^\alpha = \sigma_1^\alpha + \sigma_2^\alpha$. La fonction caractéristique d'une telle variable est donnée par :

$$\varphi_x(t_x) = e^{-\sigma^\alpha |t_x|^\alpha + i\Phi\sigma^\alpha |t_x|^\alpha \text{sg}(t_x)}, \quad (9.55)$$

où $\Phi > 0$ comme rappelé dans la section 9.1. La fonction caractéristique du vecteur (s_1, x) est :

$$\begin{aligned} \varphi_{s_1, x}(t_1, t_x) &= \mathbb{E}(e^{i(t_1 s_1 + t_x x)}) \\ &= \mathbb{E}(e^{i(t_1 s_1 + t_x (s_1 + s_2))}) \\ &= \mathbb{E}(e^{i((t_1 + t_x)s_1 + t_x s_2)}) \\ &= \varphi_{s_1, s_2}(t_1 + t_x, t_x). \end{aligned}$$

Or, les variables s_1 et s_2 étant indépendantes, $\varphi_{s_1, s_2}(t_1 + t_x, t_x) = \varphi_{s_1}(t_1 + t_x)\varphi_{s_2}(t_x)$. On obtient donc :

$$\varphi_{s_1, x}(t_1, t_x) = e^{-\sigma_1^\alpha |t_1 + t_x|^\alpha - \sigma_2^\alpha |t_x|^\alpha + i\Phi(\sigma_1^\alpha |t_1 + t_x|^\alpha \text{sg}(t_1 + t_x) + \sigma_2^\alpha |t_x|^\alpha \text{sg}(t_x))}. \quad (9.56)$$

Afin d'obtenir la fonction caractéristique de $s_1|x$ à partir de celle du vecteur (s_1, x) , nous allons utiliser un résultat issu de SAMORADNITSKY et TAQQU (1994) (eq. (5.1.7) p. 226) :

$$\varphi_{s_1|x}(t_1) = \frac{\int_{\mathbb{R}} \varphi_{s_1, x}(t_1, t_x) e^{-it_x x} dt_x}{\int_{\mathbb{R}} \varphi_{s_1, x}(0, t_x) e^{-it_x x} dt_x}. \quad (9.57)$$

Étape 2 : Dérivation de la fonction caractéristique. La dérivée de la fonction caractéristique de $s_1|x$ est donnée par :

$$\frac{d\varphi_{s_1|x}}{dt_1}(t_1) = \frac{\int_{\mathbb{R}} \frac{\partial \varphi_{s_1, x}}{\partial t_1}(t_1, t_x) e^{-it_x x} dt_x}{\int_{\mathbb{R}} \varphi_{s_1, x}(0, t_x) e^{-it_x x} dt_x}, \quad (9.58)$$

que l'on applique en $t_1 = 0$:

$$\frac{d\varphi_{s_1|x}}{dt_1}(0) = \frac{\int_{\mathbb{R}} \frac{\partial \varphi_{s_1, x}}{\partial t_1}(0, t_x) e^{-it_x x} dt_x}{\int_{\mathbb{R}} \varphi_{s_1, x}(0, t_x) e^{-it_x x} dt_x}. \quad (9.59)$$

Notons que la relation (9.58) est valable à condition qu'il soit possible de dériver sous le signe \int , ce que nous montrons ci-après. On distingue deux cas à partir de l'équation (9.56) :

— Si $t_1 > -t_x$, on a :

$$\varphi_{s_1, x}(t_1, t_x) = e^{-\sigma_1^\alpha (t_1 + t_x)^\alpha - \sigma_2^\alpha |t_x|^\alpha + i\Phi(\sigma_1^\alpha (t_1 + t_x)^\alpha + \sigma_2^\alpha |t_x|^\alpha \text{sg}(t_x))}, \quad (9.60)$$

et donc :

$$\frac{\partial \varphi_{s_1, x}}{\partial t_1}(t_1, t_x) = [-\alpha\sigma_1^\alpha (t_1 + t_x)^{\alpha-1} + i\alpha\Phi\sigma_1^\alpha (t_1 + t_x)^{\alpha-1}] \varphi_{s_1, x}(t_1, t_x) \quad (9.61)$$

— Si $t_1 < -t_x$, on a :

$$\varphi_{s_1,x}(t_1, t_x) = e^{-\sigma_1^\alpha(-t_1-t_x)^\alpha - \sigma_2^\alpha|t_x|^\alpha + i\Phi(-\sigma_1^\alpha(-t_1-t_x)^\alpha + \sigma_2^\alpha|t_x|^\alpha)sg(t_x)}, \quad (9.62)$$

et donc :

$$\frac{\partial \varphi_{s_1,x}}{\partial t_1}(t_1, t_x) = [\alpha\sigma_1^\alpha(-t_1-t_x)^{\alpha-1} + i\alpha\Phi\sigma_1^\alpha(-t_1-t_x)^{\alpha-1}]\varphi_{s_1,x}(t_1, t_x). \quad (9.63)$$

On obtient dans les deux cas la même expression de la dérivée :

$$\frac{\partial \varphi_{s_1,x}}{\partial t_1}(t_1, t_x) = \alpha\sigma_1^\alpha[-(t_1+t_x)|t_1+t_x|^{\alpha-2} + i\Phi|t_1+t_x|^{\alpha-1}]\varphi_{s_1,x}(t_1, t_x), \quad (9.64)$$

qui, appliquée en $t_1 = 0$, conduit à :

$$\frac{\partial \varphi_{s_1,x}}{\partial t_1}(0, t_x) = \alpha\sigma_1^\alpha[-t_x|t_x|^{\alpha-2} + i\Phi|t_x|^{\alpha-1}]\varphi_{s_1,x}(0, t_x), \quad (9.65)$$

avec :

$$\varphi_{s_1,x}(0, t_x) = e^{-\sigma_1^\alpha|t_x|^\alpha - \sigma_2^\alpha|t_x|^\alpha + i\Phi(\sigma_1^\alpha|t_x|^\alpha)sg(t_x) + \sigma_2^\alpha|t_x|^\alpha sg(t_x)} \quad (9.66)$$

$$= e^{-(\sigma_1^\alpha + \sigma_2^\alpha)|t_x|^\alpha + i\Phi(\sigma_1^\alpha + \sigma_2^\alpha)|t_x|^\alpha)sg(t_x)}. \quad (9.67)$$

Revenons à présent sur la dérivation sous l'intégrale (9.59). Pour montrer qu'il est possible d'écrire cette dérivée, nous allons prouver que :

$$\frac{\partial \int_{\mathbb{R}_+} \varphi_{s_1,x}(t_1, t_x) e^{-it_x x} dt_x}{\partial t_1}(t_1 = 0) = \int_{\mathbb{R}_+} \frac{\partial \varphi_{s_1,x}}{\partial t_1}(0, t_x) e^{-it_x x} dt_x, \quad (9.68)$$

$$\frac{\partial \int_{\mathbb{R}_-} \varphi_{s_1,x}(t_1, t_x) e^{-it_x x} dt_x}{\partial t_1}(t_1 = 0) = \int_{\mathbb{R}_-} \frac{\partial \varphi_{s_1,x}}{\partial t_1}(0, t_x) e^{-it_x x} dt_x. \quad (9.69)$$

Montrons l'équation (9.68) (la preuve est similaire pour (9.69)). Il nous faut trouver une borne supérieure de $\left| \frac{\partial \varphi_{s_1,x}}{\partial t_1}(t_1, t_x) e^{-it_x x} \right|$. Comme $\alpha \in]0, 1[$, on montre avec les équations (9.56) et (9.64) que :

$$\forall t_1 \in \mathbb{R}_+, \forall t_x \in \mathbb{R}_+ \setminus \{0\}, \left| \frac{\partial \varphi_{s_1,x}}{\partial t_1}(t_1, t_x) e^{-it_x x} \right| \leq g(t_1 + t_x) h(t_x) \leq \|g\|_\infty h(t_x) \quad (9.70)$$

avec $g(t) = \alpha\sigma_1^\alpha \sqrt{1 + \Phi^2} e^{-\sigma_1^\alpha|t|^\alpha} \in L^\infty(\mathbb{R}_+)$ et $h(t) = \frac{1}{|t|^{1-\alpha}} e^{-\sigma_2^\alpha|t|^\alpha} \in L^1(\mathbb{R}_+)$. Comme la borne (terme de droite dans (9.70)) est indépendante de t_1 et intégrable au sens de Lebesgue sur $\mathbb{R}_+ \setminus \{0\}$, et comme $\{0\}$ est un ensemble négligeable, on peut conclure que :

$$\forall t_1 \in \mathbb{R}_+, \frac{\partial \int_{\mathbb{R}_+} \varphi_{s_1,x}(t_1, t_x) e^{-it_x x} dt_x}{\partial t_1} = \int_{\mathbb{R}_+} \frac{\partial \varphi_{s_1,x}}{\partial t_1}(t_1, t_x) e^{-it_x x} dt_x, \quad (9.71)$$

ce qui prouve l'équation (9.68). La preuve de (9.69) étant similaire, on montre finalement (9.59).

Étape 3 : Intégration du numérateur dans (9.59). Calculons la dérivée du terme $\tilde{\varphi}(t_x) = \varphi_{s_1,x}(0, t_x)$ par rapport à t_x . On peut appliquer la même technique que précédemment (découper \mathbb{R} en deux parties pour éliminer la valeur absolue) dans l'équation (9.67), et on a alors :

$$\frac{d\tilde{\varphi}}{dt_x}(t_x) = \alpha(\sigma_1^\alpha + \sigma_2^\alpha)[-t_x|t_x|^{\alpha-2} + i\Phi|t_x|^{\alpha-1}]\varphi_{s_1,x}(0, t_x). \quad (9.72)$$

Ainsi, en combinant les équations (9.65) et (9.72), on obtient la relation :

$$\frac{\partial \varphi_{s_1,x}}{\partial t_1}(0, t_x) = \frac{\sigma_1^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} \frac{d\tilde{\varphi}}{dt_x}(t_x). \quad (9.73)$$

L'équation (9.73) est utile car elle nous permet de calculer l'intégrale au numérateur dans l'équation (9.59). En effet :

$$\int_{\mathbb{R}} \frac{\partial \varphi_{s_1,x}}{\partial t_1}(0, t_x) e^{-it_x x} dt_x = \frac{\sigma_1^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} \int_{\mathbb{R}} \frac{d\tilde{\varphi}}{dt_x}(t_x) e^{-it_x x} dt_x, \quad (9.74)$$

et une intégration par parties permet d'écrire¹ :

$$\int_{\mathbb{R}} \frac{d\tilde{\varphi}}{dt_x}(t_x) e^{-it_x x} dt_x = - \int_{\mathbb{R}} \tilde{\varphi}(t_x) \frac{d(e^{-it_x x})}{dt_x} dt_x = ix \int_{\mathbb{R}} \tilde{\varphi}(t_x) e^{-it_x x} dt_x. \quad (9.75)$$

En combinant les équations (9.74) et (9.75), on obtient :

$$\int_{\mathbb{R}} \frac{\partial \varphi_{s_1,x}}{\partial t_1}(0, t_x) e^{-it_x x} dt_x = i \frac{\sigma_1^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} x \int_{\mathbb{R}} \varphi_{s_1,x}(0, t_x) e^{-it_x x} dt_x, \quad (9.76)$$

et on ré-injecte cette expression dans (9.59), ce qui conduit à :

$$\frac{d\varphi_{s_1|x}}{dt_1}(0) = i \frac{\sigma_1^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} x. \quad (9.77)$$

Étape 4 : Obtention de l'estimateur. Finalement, en utilisant (9.77) avec (9.53) et (9.54), on obtient l'estimateur de la source s_1 :

$$\hat{s}_1 = \frac{\sigma_1^\alpha}{\sigma_1^\alpha + \sigma_2^\alpha} x. \quad (9.78)$$

De façon complètement analogue, on a naturellement l'estimateur de la source s_2 . Ainsi, pour $K = 2$, on a le résultat suivant :

$$\forall k \in \llbracket 1, K \rrbracket, \hat{s}_k = \frac{\sigma_k^\alpha}{\sum_{l=1}^K \sigma_l^\alpha} x. \quad (9.79)$$

Remarque : En fait, ce résultat reste valable pour n'importe-quelle valeur de β (nous avons ici supposé qu'il valait 1 pour se placer dans le cas des distributions P α S). En effet, si $\alpha \in]0, 1[$, la fonction caractéristique (9.55) devient :

$$\forall t_x \in \mathbb{R}, \varphi_x(t_x) = \mathbb{E}_x(e^{it_x x}) = e^{-\sigma^\alpha |t_x|^\alpha + i\beta\Phi\sigma^\alpha |t_x|^\alpha \text{sg}(t_x)}, \quad (9.80)$$

Ainsi, si on remplace Φ par $\beta\Phi$ dans notre démonstration, le résultat reste valable. La nouvelle constante $\beta\Phi$ peut cependant s'annuler (pour $\beta = 0$), ce qui perturbe la démonstration, mais la distribution est alors symétrique α -stable (S α S) : le résultat (9.79) est toujours valable, comme démontré dans LIUTKUS et BADEAU (2015).

1. En toute rigueur, cette intégration par partie fait intervenir le produit des termes non dérivés évalué en $+\infty$ et $-\infty$. Or ce produit de termes a pour module $e^{-(\sigma_1^\alpha + \sigma_2^\alpha)|t_x|^\alpha}$, il est donc nul en $\pm\infty$.

9.3.2 Somme de K variables

Nous montrons à présent que (9.79) est valable quelque soit le nombre de sources K . Considérons une somme de $K \geq 2$ variables s_k qui suivent une loi P α S de paramètre σ_k . On définit alors, $\forall k \in \llbracket 1, K \rrbracket$, $\tilde{s}_k = \sum_{l \neq k} s_l$. Ainsi :

$$s_k + \tilde{s}_k = x, \quad (9.81)$$

et par propriété de stabilité des lois P α S, \tilde{s}_k est aussi une variable P α S de paramètre $\tilde{\sigma}_k$ telle que $\tilde{\sigma}_k^\alpha = \sum_{l \neq k} \sigma_l^\alpha$. On utilise alors (9.79) avec les deux sources s_k and \tilde{s}_k :

$$\hat{s}_k = \frac{\sigma_k^\alpha}{\sigma_k^\alpha + \tilde{\sigma}_k^\alpha} x = \frac{\sigma_k^\alpha}{\sigma_k^\alpha + \sum_{l \neq k} \sigma_l^\alpha} x = \frac{\sigma_k^\alpha}{\sum_l \sigma_l^\alpha} x. \quad (9.82)$$

Comme ce résultat est valable quel que soit $k \in \llbracket 1, K \rrbracket$, l'équation (9.79) est donc valable pour tout K . Enfin, on peut écrire ce résultat matriciellement en considérant que tous les entrées de X sont indépendantes :

$$\hat{X}_k = \frac{\sigma_k^{\odot \alpha}}{\sum_l \sigma_l^{\odot \alpha}} \odot X. \quad (9.83)$$

En conclusion, nous avons démontré que le *filtrage de Wiener généralisé* (9.83) déjà justifié par LIUTKUS et BADEAU (2015) dans le cas des distributions S α S, restait valable pour les distributions P α S. Nous avons à présent un outil pour estimer des sources à partir de mélange de données non-négatives.

Dans le cas du modèle de Lévy NMF, on a donc, $\forall k \in \llbracket 1, K \rrbracket$,

$$\hat{X}_k = \frac{\sigma_k^{\odot 1/2}}{\sum_l \sigma_l^{\odot 1/2}} \odot X = \frac{W_k H_k}{\sum_l W_l H_l} \odot X, \quad (9.84)$$

où W_k (resp. H_k) désigne la k -ième colonne (resp. ligne) de la matrice W (resp. H).

Remarque : Cet estimateur des sources est généralement optimal au sens des moindres carrés (estimateur MMSE). Néanmoins, pour montrer que c'est le cas, il faut prouver que la variance de $s_k|x$ est bien définie (c'est le cas pour des distributions gaussiennes).

9.4 Expériences

Nous menons une série d'expériences pour montrer l'intérêt des distributions P α S et de l'algorithme de Lévy NMF pour diverses applications. Les données utilisées sont des mélanges de notes de piano issues de la base MAPS EMIYA et al. (2010), des morceaux de guitare issus de la base IDMT-SMT-GUITAR KEHLING et al. (2014), et des extraits de morceaux de musique polyphoniques issus de la base DSD100 ONO et al. (2015). Les signaux sont échantillonnés à 44100 Hz. La TFCT est calculée avec une fenêtre de Hann de longueur 93 ms et 75 % de recouvrement. Plusieurs méthodes sont comparées :

- ISNMF : NMF avec divergence d'Itakura-Saito, correspondant au modèle gaussien FÉVOTTE et al. (2009) ;
- KLNMF : NMF avec divergence de Kullback-Leibler, correspondant au modèle de Poisson VIRTANEN et al. (2008) ;

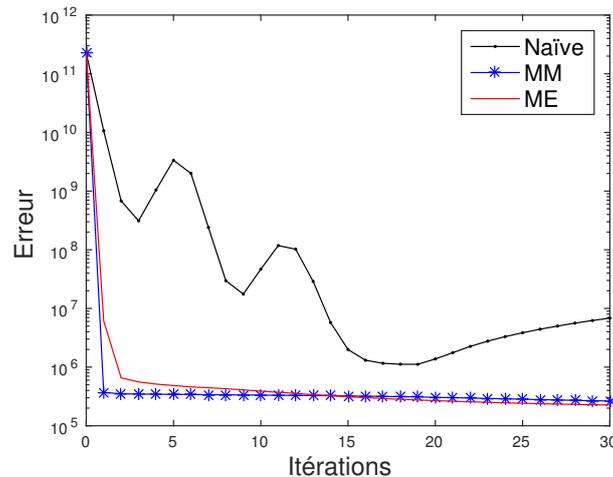


FIGURE 9.6 – Évolution de la fonction de coût au cours des itérations pour les différentes approches retenues pour l'algorithme de Lévy NMF.

- Cauchy NMF : modèle de Cauchy [LIUTKUS et al. \(2015\)](#) ;
- Lévy NMF : modèle de Lévy, présenté dans ce chapitre ;
- RPCA : Analyse en Composantes Principales Robuste [CANDÈS et al. \(2011\)](#). Nous avons utilisé l'implémentation [HUANG et al. \(2012\)](#). Il est en effet intéressant de comparer les approches NMF sus-citées à la PCA robuste, car celle-ci est précisément utilisée pour sa robustesse aux bruits, ce qui, on le verra, est l'un des atouts de la Lévy NMF.

Les algorithmes utilisent 200 itérations (sauf explicitement mentionné), valeur pour laquelle la convergence est observée et au-delà de laquelle les résultats ne sont pas améliorés.

9.4.1 Algorithmes de Lévy NMF

Ces premières expériences étudient certaines propriétés des algorithmes de Lévy NMF présentés dans la section 9.2.

Convergence des algorithmes

Nous testons les différents algorithmes sur un extrait de la base DSD100 [ONO et al. \(2015\)](#), qui est un morceau de musique polyphonique. Nous effectuons une Lévy NMF par les différentes approches avec un ordre de factorisation $K = 40$ et 30 itérations. Nous traçons sur la figure 9.6 la fonction de coût (qui est, rappelons-le, la divergence IS entre le modèle $[WH]^{\odot 2}$ et les données X , ici le spectrogramme d'amplitude du signal) pour les différentes approches : naïve, MM et ME.

Nous constatons tout d'abord que l'utilisation de l'approche naïve produit une fonction de coût non monotone. En outre, la convergence n'est pas observée (même en augmentant significativement le nombre d'itérations). Les approches MM et ME, quant à elles, conduisent à une fonction de coût décroissante. Conformément à notre résultat théorique, l'approche ME fait décroître la fonction de coût plus lentement que l'approche MM durant les premières itérations, mais à proximité du minimum local, c'est-à-dire pour les dernières itérations, cette approche permet une décroissance de la fonction de coût légèrement plus importante.

Nous utiliserons néanmoins l'approche MM dans la suite des expériences, préférant en effet avoir une garantie de décroissance de la fonction de coût. En outre, les résultats obtenus par les deux approches ne diffèrent pas significativement.

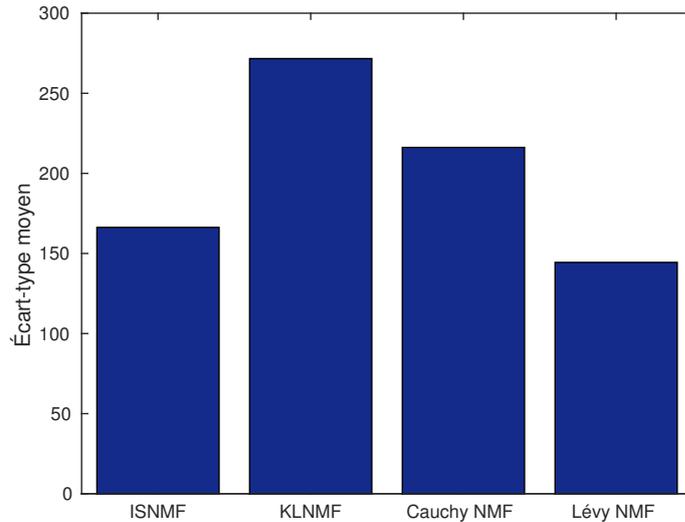


FIGURE 9.7 – Robustesse des algorithmes de NMF à l’initialisation (écart-type moyen de la distribution d’erreur pour plusieurs initialisation aléatoires).

Robustesse à l’initialisation

Nous proposons de tester la robustesse à l’initialisation, c’est-à-dire la capacité d’un algorithme à converger vers la même solution, avec des valeurs initiales différentes. Pour ce faire, nous considérons des signaux de mélanges de notes de piano issus de la base MAPS. Afin de mesurer la robustesse à l’initialisation, nous reconstruisons, par application du filtrage de Wiener généralisé (9.83), les spectrogrammes d’amplitude des différentes sources \hat{X}_k , ce qui permet de calculer l’erreur moyenne de reconstruction :

$$\frac{1}{K} \sum_k \|X_k - \hat{X}_k\|_2. \quad (9.85)$$

Cette erreur est calculée pour 30 initialisations aléatoires et pour chaque méthode. On mesure la robustesse à l’initialisation par l’écart-type de la distribution de l’erreur (9.85) : plus celui-ci est faible, plus l’erreur est concentrée autour de sa moyenne, plus l’algorithme est alors dit robuste. Enfin, cet écart-type est moyenné sur 30 signaux différents, l’écart-type moyen étant représenté sur la figure 9.7. On constate que l’algorithme de Lévy NMF est le plus robuste parmi les 4 considérés. C’est un avantage pour des applications pratiques puisque une variation de l’initialisation conduit à faire moins varier le résultat de la factorisation qu’avec les autres approches.

9.4.2 Représentation de bruits impulsionnels

Nous testons à présent la capacité du modèle de Lévy NMF à représenter des bruits de nature impulsive. Pour ce faire, nous considérons des données synthétiques qui sont créées en générant $K = 5$ paires de composantes W et H , obtenues en prenant la puissance quatrième d’un bruit blanc gaussien (ce qui permet d’obtenir des composantes parcimonieuses). Le produit obtenu WH est alors de dimensions $F \times T = 50 \times 50$. On utilise celui-ci comme paramètre d’échelle de distributions $P\alpha S$ ($\sigma^{\odot\alpha} = WH$) pour diverses valeurs de α , notamment de faibles valeurs (comprises entre 0.01 et 0.5) : cela permet de représenter des signaux très impulsionnels. Pour générer ces signaux tests, on utilise la boîte à outils [WERON \(2010\)](#).

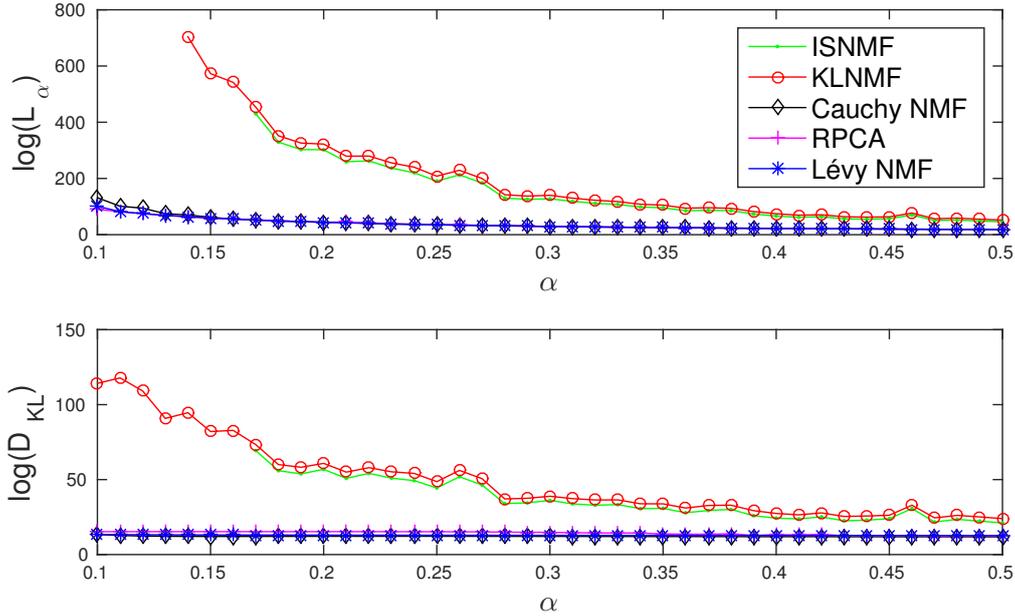


FIGURE 9.8 – Mesures de la qualité d'estimation des paramètres d'échelle.

On applique alors les différentes méthodes d'estimation des paramètres d'échelle $\hat{\sigma}$ sur ces données. Le rang de la factorisation est choisi à 5. Afin de quantifier la qualité d'estimation des paramètres d'échelle, on utilise deux mesures, similairement à ce qui est fait dans [LIUTKUS et al. \(2015\)](#) :

- La α -dispersion $L_\alpha = \sum_{f,t} |\sigma(f,t) - \hat{\sigma}(f,t)|^{1/\alpha}$;
- La divergence de Kullback-Leibler $D_{KL} = \sum_{f,t} \sigma(f,t) \log \frac{\sigma(f,t)}{\hat{\sigma}(f,t)} - \hat{\sigma}(f,t) + \sigma(f,t)$.

On calcule ces deux indicateurs (moyennés sur 100 signaux synthétiques) pour chaque méthode et diverses valeurs de α . Les résultats sont présentés sur la figure 9.8.

Le modèle de Lévy NMF fournit des résultats comparables à la RPCA et à la Cauchy NMF. En particulier, pour de petites valeurs de α (bruits très impulsionnels), la Lévy NMF fournit des résultats légèrement meilleurs que les autres. Cela montre le potentiel de ce modèle pour représenter des signaux de nature variable, qui peuvent aller jusqu'à des bruits très impulsionnels.

9.4.3 Applications

Nous proposons enfin de tester la Lévy NMF dans le cadre de plusieurs applications sur des données réalistes.

Restauration de signaux de musique corrompus synthétiquement

Nous reprenons l'idée de l'expérience précédente mais cette fois-ci en considérant des morceaux de musique réels, qui sont corrompus synthétiquement avec des bruits impulsionnels dans le domaine TF. Bien que ces bruits ne soient pas très réalistes, cela illustre le potentiel

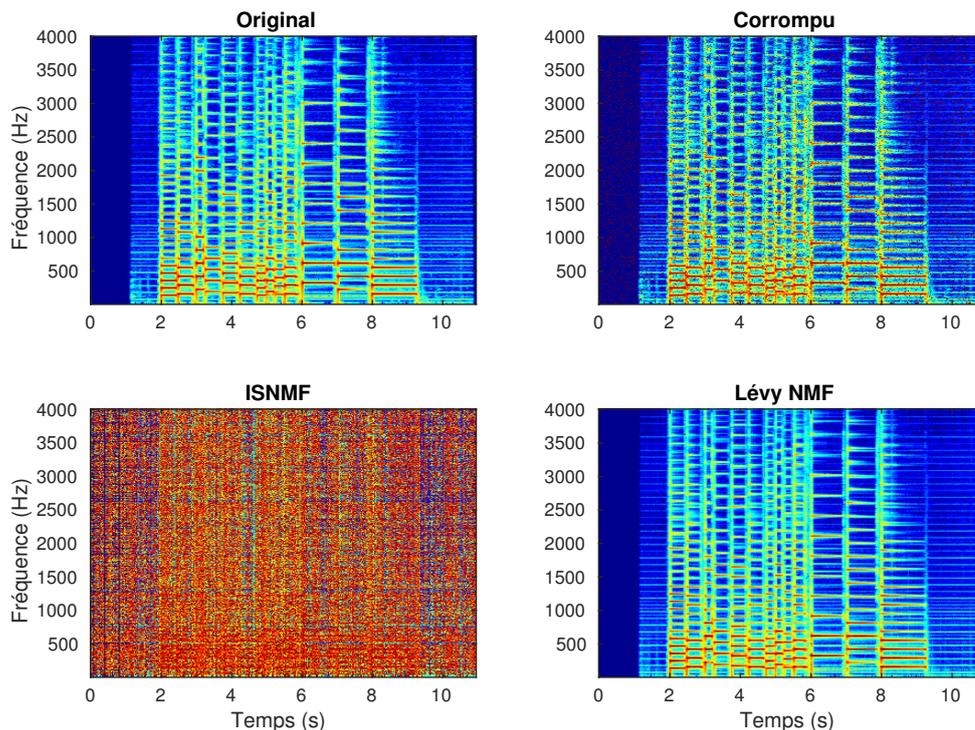


FIGURE 9.9 – Reconstruction de spectrogrammes corrompus grâce à diverses méthodes.

de notre technique en matière de robustesse par rapport à des bruits impulsionnels pour une tâche de restauration de signaux audio.

Les signaux considérés sont issus de la base IDMT-SMT-GUITAR [KEHLING et al. \(2014\)](#) et sont de courts extraits de morceaux de guitare. Nous traçons sur la figure 9.9 divers spectrogrammes que l'on obtient après débruitage par simple application des méthodes de NMF et de RPCA (KLNMF et ISNMF conduisent à des résultats similaires, de même que RPCA, Cauchy NMF et Lévy NMF). On constate que la Lévy NMF semble très robuste aux bruits et que l'estimation des paramètres d'échelle n'est pas perturbée par ceux-ci. Aucun modèle de bruit n'est injecté dans le modèle, ce qui traduit la robustesse naturelle du modèle de Lévy NMF aux bruits impulsionnels.

Ce constat est quantifiable par la divergence KL mesurée entre les spectrogrammes originaux et restaurés. On présente celle-ci sur la figure 9.10. On constate que la Lévy NMF conduit aux meilleurs résultats (c'est-à-dire à la plus basse divergence KL), ce qui confirme l'intuition préalablement faite à partir des spectrogrammes.

On reconstruit ensuite les signaux temporels en appliquant la phase du signal original non bruité (ce qui est un cas Oracle, mais n'a aucune justification théorique) aux spectrogrammes estimés. On mesure la qualité de reconstruction par le SDR [VINCENT et al. \(2006\)](#) et on présente les résultats sous forme de boîtes à moustaches sur la figure 9.10. On constate que la Lévy NMF donne les meilleurs résultats, confirmant le potentiel de celle-ci comme un outil d'estimation de paramètres très robuste aux bruits impulsionnels dans le domaine TF, notamment dans le cas de signaux audio.

Nous avons enfin mené une expérience complémentaire, qui consiste à informer le modèle ISNMF par la localisation des bruits : on applique alors une ISNMF pondérée [LIMEM et al. \(2013\)](#) par un masque qui vaut 0 pour les points TF corrompus, et 1 ailleurs. Les résultats s'en

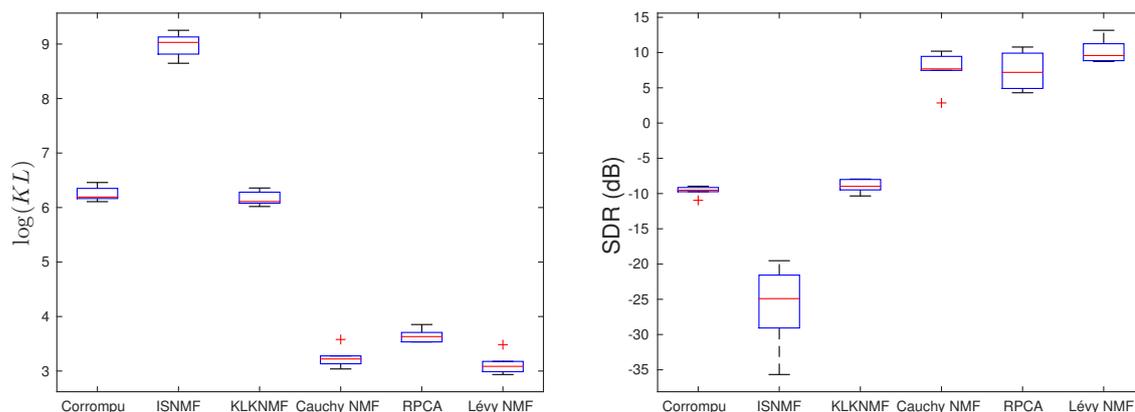


FIGURE 9.10 – Qualité de restauration des spectrogrammes mesurée par la divergence KL (gauche) et de reconstruction de signal mesurée par le SDR (droite).

voient améliorés (on passe en moyenne de 9 u. environ² pour la ISNMF non informée à 3.8 u. pour la ISNMF pondérée). Néanmoins, les résultats restent moins bons que ceux obtenus avec la Lévy NMF (environ 3.2 u.) sachant que celle-ci n'est pas informée. Cela traduit la supériorité de ce modèle, qui mène de façon aveugle à des performances meilleures que certains modèles informés (ISNMF pondérée) ou construits sur des propriétés de robustesse (RPCA).

Rehaussement de l'accompagnement musical

Nous considérons 30 signaux tirés de la base DSD100, composés d'une piste d'accompagnement et d'une piste comportant la voix chantée. En appliquant une Lévy NMF (sans aucun à priori ou modèle sur la voix), on s'attend à ce que la voix, traitée comme du bruit, soit partiellement supprimée au profit de l'accompagnement. On choisit un rang de factorisation égal à 30 et 200 itérations de NMF. L'accompagnement est alors estimé par filtrage de Wiener généralisé pour les différentes méthodes. Dans le cas de la Lévy NMF (et de la RPCA), on applique la phase du mélange aux amplitudes estimées (ce qui pourrait être amélioré par une technique plus avancée de reconstruction de phase). On calcule enfin le SDR entre l'accompagnement musical original et sa version estimée. Les résultats sont présentés sur la figure 9.11.

On constate que les méthodes Cauchy NMF, RPCA et Lévy NMF donnent des résultats similaires et supérieures aux techniques plus traditionnelles de ISNMF et KLKMF (avec un très léger avantage pour la Lévy NMF). Perceptivement, ce sont les seules méthodes sur lesquelles on entend effectivement un rehaussement de l'accompagnement musical et une quasi-suppression de la partie voix. Cette expérience montre donc le potentiel de Lévy NMF pour de telles applications (ainsi que celui de modèles comme Cauchy NMF).

Application à la spectroscopie de fluorescence

L'intérêt du modèle de distribution PaS est de traiter directement des données non-négatives. Les expériences précédentes ont montré le potentiel de la Lévy NMF, mais dans le cadre de données audio, qui sont intrinsèquement complexes. L'avantage par rapport à la méthode Cauchy NMF n'apparaît donc pas clairement. Cette dernière possède en effet des pro-

2. On désigne par u. l'unité du logarithme de la divergence de Kullback-Leibler, par commodité d'écriture.

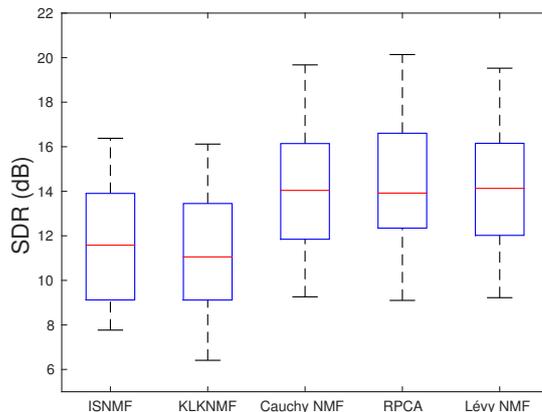


FIGURE 9.11 – Qualité du rehaussement de l’accompagnement (SDR en dB).

priétés similaires (stabilité, robustesse) et pourrait paraître plus adapté pour des applications audio, puisque la loi de Cauchy modélise directement des données complexes.

Le principal intérêt de la Lévy NMF est la modélisation et l’estimation de données non négatives, aussi nous proposons une application différente de l’audio, où son potentiel peut être encore davantage révélé. Dans un tel contexte, les modèles de type Cauchy NMF ou même ISNMF (gaussiennes complexes) ne sont pas appropriés. Nous avons choisi de considérer l’application de séparation de sources pour des données issues de la chimie, et plus spécifiquement de la spectroscopie de fluorescence [LIU et al. \(2013\)](#).

Pour cette application, les données considérées sont des spectres d’émission de certaines espèces chimiques : une fois une espèce excitée par une lumière incidente à une certaine longueur d’onde (365 nm pour nos données), elle émet avec une intensité variable selon la longueur d’onde d’émission considérée. En séparant le spectre d’émission d’un mélange d’espèces données, on peut alors estimer les espèces chimiques pures qui composent le mélange (spectres W), et dans quelles proportions (matrice de concentration H).

Le jeu de données dont nous disposons est celui employé dans [GOBINET et al. \(2004\)](#)³. Ce jeu de données est composé de spectres d’émission d’un grain de blé, dans la bande de longueurs d’onde allant de 350 à 670 nm (avec une précision d’environ 2.5 nm, soit $F = 128$ canaux fréquentiels). Ces spectres sont obtenus en différents points d’une section de grain (sur une grille de 20×20 points), ce qui conduit à $T = 400$ spectres en tout.

Les spectres d’émission de grains de blé sont composés principalement de 3 composantes, qui sont les spectres de l’acide férulique, de l’acide férulique libre et de l’acide paracoumarique. Nous avons à notre disposition les spectres de ces composés purs. Ne connaissant pas les concentrations exactes des espèces dans nos mélanges, on commence par appliquer une NMF dans un cas "Oracle", en supposant les spectres des composantes isolées connues et en n’estimant que les concentrations (c’est-à-dire la matrice H). Cela nous donne un point de référence qui servira par la suite à mesurer l’erreur de séparation de sources. La figure 9.12 illustre les cartes de concentrations obtenues pour les différents composants dans le cas Oracle (c’est donc une carte de 20×20 coordonnées spatiales).

Ensuite, nous effectuons sur les mélanges diverses NMF aveugles (en apprenant donc les spectres et les concentrations) utilisées dans la littérature dédiée à la spectroscopie de fluorescence : NMF avec distance Euclidienne (Euc) [GOBINET et al. \(2004\)](#); [MONTCUQUET](#)

3. Les mesures ont été effectuées à l’INRA Montpellier et nous ont été fournies par Cyril Gobinet, l’auteur de cet article. Nous en profitons donc pour le remercier.

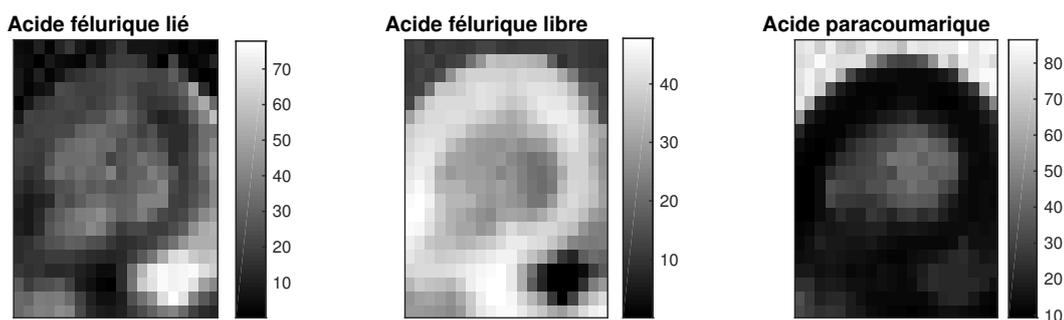


FIGURE 9.12 – Cartographie des concentrations des diverses espèces obtenues dans le cas Oracle.

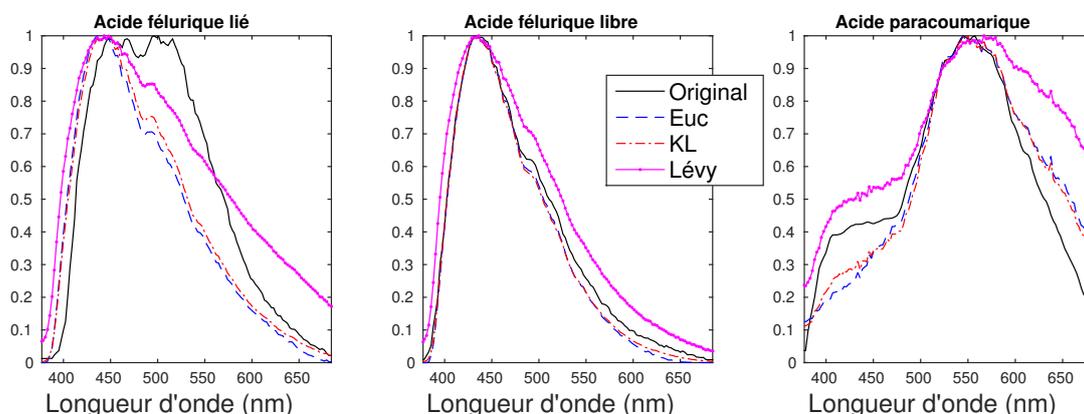


FIGURE 9.13 – Spectres d'émission (normalisés) réels et estimés par différentes méthodes pour les 3 composantes.

et al. (2009) et NMF avec divergence KL [GOBINET et al. \(2005\)](#). Les spectres obtenus sont représentés sur la figure 9.13. On constate le potentiel du modèle de Lévy NMF pour apprendre avec une certaine précision des spectres à partir de mélanges, avec des résultats similaires aux autres méthodes de NMF. On remarquera notamment que la Lévy NMF a tendance à approcher les spectres de référence par valeurs supérieures.

On peut alors estimer les sources par filtrage de Wiener généralisé. Pour chaque source et pour chaque méthode, on évalue la qualité de l'estimation par la corrélation entre les sources Oracle et estimées. Le résultat est présenté sur la figure 9.14. On constate que la reconstruction des sources par Lévy NMF conduit à un résultat plus proche de l'oracle que les autres approches. Cela confirme donc le potentiel de la Lévy NMF pour des applications de séparation de sources non-négatives.

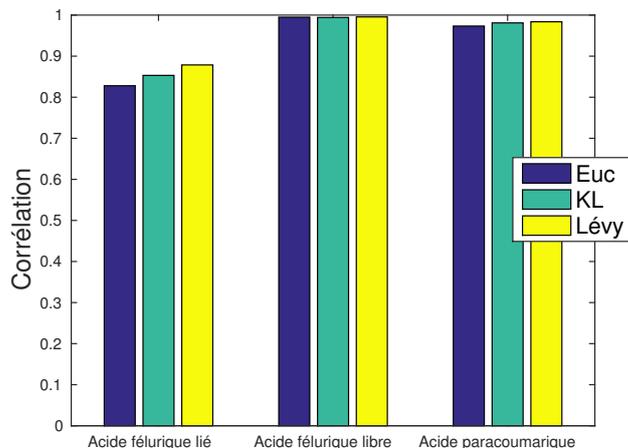


FIGURE 9.14 – Qualité de la reconstruction des sources pour différentes méthodes NMF.

9.5 Conclusion

Nous avons introduit dans ce chapitre les distributions positives α -stables pour représenter des données non-négatives. En prenant le cas particulier de la distribution de Lévy, nous avons mis au point une procédure d'estimation des paramètres des sources dans un mélange, par un modèle NMF estimé de différentes façons. Un estimateur des sources généralisant le filtrage de Wiener a également été obtenu. Expérimentalement, nous avons constaté le potentiel de cet outil en ce qui concerne sa robustesse au bruit impulsionnel dans le domaine TF, et vu qu'il était compétitif avec d'autres méthodes de l'état de l'art pour diverses tâches.

De nombreuses disciplines s'intéressent à la modélisation et à la séparation de données non-négatives par approches NMF, comme la physique appliquée [SAJDA et al. \(2004\)](#) ou la vision par ordinateur [LEE et SEUNG \(1999\)](#). Par ailleurs, la distribution de Lévy est utilisée dans divers domaines tels que l'optique [ROGERS \(2008\)](#) ou l'analyse du champ magnétique terrestre [CARBONE et al. \(2006\)](#).

Outre les nombreuses applications potentielles, la méthode présentée dans ce chapitre peut servir de fondation à des techniques plus avancées. En estimant les paramètres au sens MAP, il est possible d'intégrer certaines informations sur les matrices W (comme l'harmonicité) et H (parcimonie, continuité temporelle), et ainsi obtenir des résultats plus satisfaisants. Dans le cas général des distributions $P\alpha S$, il serait intéressant de mettre en place une technique d'estimation (basée sur les MCMC par exemple) afin de pleinement exploiter le potentiel de ces distributions pour la modélisation robuste de données non négatives. Alternativement, on pourrait s'intéresser aux distributions inverse-gamma [KOUNADES-BASTIAN et al. \(2016\)](#), dont la loi de Lévy est un cas particulier, et dont la densité est exprimable analytiquement. Néanmoins, cette famille de lois n'est pas additive dans le cas général, et il n'est pas évident d'obtenir la loi d'une somme de variables inverse-gamma [WITKOVSKY \(2001\)](#), ce qui requiert la mise en place de méthodes variationnelles.

Nous avons considéré, dans nos expériences en audio, que les spectrogrammes d'amplitude suivaient une loi de Lévy : il est possible que de meilleurs résultats soient obtenus en modélisant non plus les spectrogrammes d'amplitude, mais des spectrogrammes fractionnaires, comme le suggèrent certains résultats récents sur le sujet [LIUTKUS et BADEAU \(2015\)](#); [VORAN \(2015\)](#). Enfin, ces modèles pourraient être combinés à un modèle de phase non uniforme (*cf.* chapitre 8) pour la séparation de sources audio.