

NMF complexe à phase contrainte

Sommaire

7.1	Modèle de NMF complexe à phase contrainte	112
7.1.1	Approche intuitive	112
7.1.2	Modélisation probabiliste	114
7.2	Estimation du modèle	117
7.3	Résultats expérimentaux	121
7.3.1	Données et protocole	121
7.3.2	Mélanges simples	121
7.3.3	Morceaux de musique polyphoniques	123
7.4	Conclusion	125

Les modèles de phase étudiés jusqu'à présent ont été utilisés dans un cadre de séparation de sources, mais les amplitudes étaient supposées connues (ou estimées préalablement à la reconstruction de phase). Nous nous intéressons dans ce chapitre à un modèle inspiré de la NMF, de type NMF complexe (CNMF), qui factorise les spectrogrammes d'amplitude tout en modélisant les phases. Comme nous l'avons vu dans le chapitre 3, les représentations CNMF sont prometteuses pour la séparation de sources, mais requièrent que la phase soit contrainte pour conduire à une séparation de qualité. Nous proposons donc des contraintes de phase issues des modèles étudiés dans ce manuscrit : le déroulé linéaire introduit au chapitre 4, et un modèle de phase dans les trames d'attaque (*cf.* chapitre 6).

Les résultats de ce travail ont fait l'objet d'une publication à la conférence ICASSP 2016 [MAGRON et al. \(2016b\)](#).

La section 7.1 introduit le modèle de CNMF à phase contrainte. La section 7.2 présente les algorithmes par lesquels le modèle est estimé. La section 7.3 consiste en une validation expérimentale de cette technique appliquée à la séparation de sources, et nous livrons nos principales conclusions dans la section 7.4.

7.1 Modèle de NMF complexe à phase contrainte

7.1.1 Approche intuitive

Le modèle de NMF complexe que nous avons présenté dans le chapitre 2 consiste à approcher une matrice de données complexes X (une TFCT en général) par le modèle \hat{X} tel que pour tout canal fréquentiel f et trame temporelle t :

$$\hat{X}(f, t) = \sum_{k=1}^K W(f, k) H(k, t) e^{i\phi_k(f, t)}, \quad (7.1)$$

avec $W \in \mathbb{R}_+^{F \times K}$ et $H \in \mathbb{R}_+^{K \times T}$. L'estimation de ce modèle [KAMEOKA et al. \(2009\)](#) conduit à minimiser une fonction de coût qui s'exprime comme la somme de deux termes : la distance euclidienne entre X et \hat{X} , et un terme de parcimonie $\mathcal{C}_s(H) = 2 \sum_{k,t} H(k, t)^p$ ($p \in]0, 2[$) multiplié par un poids σ_s .

Intuitivement, on peut considérer le problème d'estimation d'un modèle de NMF complexe à phase contrainte comme un problème de minimisation d'une fonction de coût \mathcal{C} qui serait la somme de divers termes, qui traduisent les contraintes intégrées au modèle :

- Un terme de distance euclidienne D entre le modèle et les données, qui évalue la précision de la reconstruction (terme d'attache aux données) ;
- Un terme de parcimonie \mathcal{C}_s qui force cette propriété ;
- Un terme \mathcal{C}_u qui tient compte du déroulé linéaire de phases ;
- Un terme \mathcal{C}_r qui tient compte d'un modèle de phases d'attaque.

On reprend, pour D et \mathcal{C}_s , les expressions obtenues dans le modèle originel [KAMEOKA et al. \(2009\)](#), et on intègre les nouvelles contraintes décrites ci-après.

Déroulé linéaire Le déroulé linéaire de phase introduit dans le chapitre 4 modélise la phase ϕ_k d'une source de la façon suivante :

$$\phi_k(f, t) = \phi_k(f, t - 1) + 2\pi S \nu_k(f), \quad (7.2)$$

où S est le décalage temporel entre deux trames successives (en échantillons) et $\nu_k(f)$ est la fréquence réduite de la source k dans le canal f .

Afin d'injecter cette propriété de déroulé linéaire sous forme de contrainte dans notre problème d'estimation du modèle (7.1), on considère la fonction de coût suivante :

$$\mathcal{C}_u(\phi) = \sum_{f,k} \sum_{t \notin \Omega_k} |X(f,t)|^2 |e^{i\phi_k(f,t)} - e^{i\phi_k(f,t-1)} e^{2i\pi S\nu_k(f)}|^2, \quad (7.3)$$

où Ω_k désigne l'ensemble des trames d'attaque de la source k . En effet, le modèle de déroulé linéaire n'est valable qu'en dehors des trames d'attaque. Ce type d'approche a déjà été utilisé dans la littérature, notamment dans BRONSON et DEPALLE (2014) et plus récemment dans RODRIGUEZ-SERRANO et al. (2016). Contrairement à ces approches, la notre est adaptée à la séparation aveugle car elle ne suppose pas connues les nombres d'harmoniques et les fréquences fondamentales, et ne requiert pas d'information externe (partition).

En posant $\forall(k, f, t)$, $u_k(f, t) = |e^{i\phi_k(f,t)} - e^{i\phi_k(f,t-1)} e^{2i\pi S\nu_k(f)}|^2$, (7.3) se réécrit :

$$\mathcal{C}_u(\phi) = \sum_{f,k} \sum_{t \notin \Omega_k} |X(f,t)|^2 u_k(f, t). \quad (7.4)$$

Nous proposons d'estimer les fréquences réduites par QIFFT sur les colonnes de W , et reprenons l'expression de la région d'influence utilisée dans le chapitre 4 pour associer à chaque canal fréquentiel f la fréquence réduite correspondante $\nu_k(f)$.

Remarque : Dans nos précédents travaux, la fréquence instantanée dépendait de la trame temporelle considérée t , afin de prendre en compte les variations de celle-ci, et donc les signaux non-stationnaires. Ici, le modèle NMF utilisé suppose que les fréquences sont fixes au cours du temps puisque le dictionnaire W d'atomes spectraux ne dépend pas du temps (les variations de fréquence fondamentale ne sont pas prises en compte dans ce modèle). Ainsi, nous estimerons les fréquences ν_k à partir de la matrice W , ces valeurs étant ensuite utilisées pour la totalité du déroulé. Notons cependant qu'un modèle plus fin prenant en compte les variations de fréquence pourra être ultérieurement envisagé, basé sur les travaux de HENNEQUIN et al. (2011a).

Phases dans les trames d'attaque Afin de tenir compte du caractère répétitif des atomes temporels, nous proposons d'utiliser le modèle introduit dans le chapitre 6. La phase d'une source ϕ_k dans une trame d'attaque $t \in \Omega_k$ est modélisée par une phase de référence ψ_k à laquelle est ajoutée un délai linéaire en fréquence de pente λ_k :

$$\phi_k(f, t) = \psi_k(f) + \lambda_k(t)f. \quad (7.5)$$

La notation $t \in \Omega_k$ est préférée à la notation t_m du chapitre 6 pour alléger les notations. On a donc le critère suivant :

$$\mathcal{C}_r(\phi, \psi, \lambda) = \sum_{f,k} \sum_{t \in \Omega_k} |X(f,t)|^2 |e^{i\phi_k(f,t)} - e^{i\psi_k(f)} e^{i\lambda_k(t)f}|^2. \quad (7.6)$$

soit, en posant $r_k(f, t) = |e^{i\phi_k(f,t)} - e^{i\psi_k(f)} e^{i\lambda_k(t)f}|^2$,

$$\mathcal{C}_r(\phi, \psi, \lambda) = \sum_{f,k} \sum_{t \in \Omega_k} |X(f,t)|^2 r_k(f, t). \quad (7.7)$$

Fonction de coût globale En ajoutant les termes (7.4) et (7.7) à la distance euclidienne entre les données et le modèle ainsi qu'à la contrainte de parcimonie, on obtient alors la fonction de coût globale suivante :

$$\mathcal{C}(\theta) = D(X, \hat{X}) + \sigma_u \mathcal{C}_u(\phi) + \sigma_r \mathcal{C}_r(\phi, \psi, \lambda) + \sigma_s \mathcal{C}_s(H). \quad (7.8)$$

Dans l'article [MAGRON et al. \(2016b\)](#), nous avons formulé la fonction de coût de cette façon, mais on peut également utiliser un cadre probabiliste pour y parvenir. C'est ce que nous détaillons dans la section suivante.

7.1.2 Modélisation probabiliste

Estimateurs ML et MAP

On considère que les données X sont égales au modèle \hat{X} défini par (7.1) auquel est ajouté un terme d'erreur, modélisé par un bruit blanc gaussien : $\forall(f, t)$,

$$X(f, t) = \hat{X}(f, t) + \epsilon(f, t), \quad (7.9)$$

où les $\epsilon(f, t)$ sont indépendants et de même loi $\mathcal{N}(0, \tilde{\sigma}^2)$. On a alors $X(f, t) \sim \mathcal{N}(\hat{X}(f, t), \tilde{\sigma}^2)$. En notant $\theta = \{W, H, \phi\}$ l'ensemble des paramètres du modèle, la log-vraisemblance est donc donnée par :

$$\begin{aligned} L(\theta) &= \log(p_{X|\theta}(X)) = \sum_{f,t} \log(p_{X(f,t)|\theta}(X(f,t))) \\ &= \sum_{f,t} -\log(\pi\tilde{\sigma}^2) - \frac{|X(f,t) - \sum_k \hat{X}_k(f,t)|^2}{\tilde{\sigma}^2} \\ &\stackrel{c}{=} - \sum_{f,t} \frac{|X(f,t) - \sum_k W(f,k)H(k,t)e^{i\phi_k(f,t)}|^2}{\tilde{\sigma}^2} \\ &\stackrel{c}{=} - \frac{1}{\tilde{\sigma}^2} \sum_{f,t} |X(f,t) - \sum_k W(f,k)H(k,t)e^{i\phi_k(f,t)}|^2 \\ &\stackrel{c}{=} - \frac{1}{\tilde{\sigma}^2} D(X, \hat{X}). \end{aligned}$$

On constate donc que maximiser la log-vraisemblance des données (approche ML) revient à minimiser D . Néanmoins, on ne s'intéresse pas ici à la méthode ML puisque celle-ci ne permet pas d'injecter des à priori sur les paramètres. Pour ce faire, on adopte plutôt une approche MAP. Comme nous l'avons déjà rappelé dans le chapitre 2, section 2.2.4, l'approche MAP consiste à maximiser la loi à postérieure, ce qui revient à maximiser :

$$\mathcal{C}_{MAP}(\theta) = L(\theta) + \log(p_\theta), \quad (7.10)$$

où p_θ désigne la loi à priori sur les variables θ . Nous allons donc incorporer la parcimonie via un à priori sur H et les contraintes de phase via un à priori sur ϕ .

Parcimonie

Pour introduire une contrainte de parcimonie, on modélise chaque $H(k, t)$ comme une variable aléatoire suivant une loi normale généralisée [KAMEOKA et al. \(2009\)](#) :

$$p_{H(k,t)}(H(k,t)) = \frac{1}{2\Gamma(1 + \frac{1}{p})b} e^{-\frac{|H(k,t)|^p}{b^p}}, \quad (7.11)$$

où p et b sont deux paramètres qui déterminent la forme de la distribution, et Γ désigne la fonction Gamma d'Euler [ARTIN \(2015\)](#). Ainsi, en supposant toutes les sources et trames

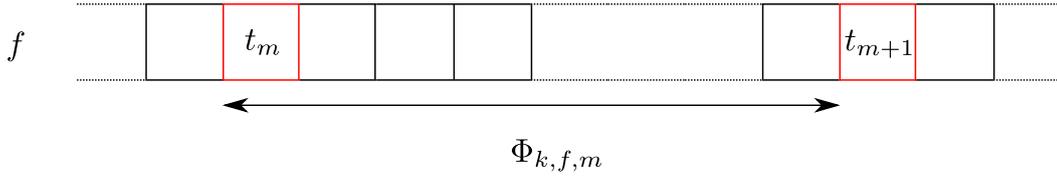


FIGURE 7.1 – Découpage de l'ensemble des trames selon les trames d'attaque.

temporelles indépendantes, on a :

$$\begin{aligned} \log(p_H) &= \sum_{k,t} \log(p_{H(k,t)}(H(k,t))) \\ &\stackrel{c}{=} -\frac{1}{b^p} \sum_{k,t} |H(k,t)|^p. \end{aligned}$$

En posant $\sigma_s = \frac{\tilde{\sigma}^2}{2b^p}$ et $\mathcal{C}_s(H) = 2 \sum_{k,t} |H(k,t)|^p$, on a :

$$\log(p_H) = -\frac{\sigma_s}{\tilde{\sigma}^2} \mathcal{C}_s(H), \quad (7.12)$$

Remarque : Sans à priori sur la phase, maximiser la distribution à postériori revient à minimiser la fonction de coût $D(X, \hat{X}) + \sigma_s \mathcal{C}_s(H)$: on retrouve exactement le modèle de CNMF complexe originel [KAMEOKA et al. \(2009\)](#).

Contraintes de phase

Nous enrichissons le modèle précédent avec un terme d'à priori sur la phase ϕ . On commence par décomposer ces termes, dans chaque bande de fréquences et pour chaque source, sous hypothèse d'indépendance des sources et canaux fréquentiels :

$$p_\Phi = \prod_{f,k} p_{\Phi_{k,f}}. \quad (7.13)$$

On introduit les trames d'attaque t_m (nous revenons temporairement à cette notation, similaire à celle du chapitre 6, par souci de clarté) pour la source k , avec $m \in \llbracket 0, M-1 \rrbracket$ où M désigne le nombre de trames d'attaque pour la source k ¹. On découpe les composantes $\Phi_{k,f}$ selon les différentes attaques, comme c'est illustré sur la figure 7.1 :

$$p_{\Phi_{k,f}} = \prod_{m=0}^{M-1} p_{\Phi_{k,f,m}}. \quad (7.14)$$

On utilise une structure d'à priori en chaînes de Markov pour tenir compte de la dépendance des phases entre trames successives, comme cela est proposé dans [BERTIN et al. \(2010\)](#) pour modéliser la continuité temporelle des activations dans un modèle NMF :

$$p_{\Phi_{k,f,m}} = p(\phi_k(f, t_m)) \prod_{t=t_m+1}^{t_{m+1}-1} p(\phi_k(f, t) | \phi_k(f, t-1)). \quad (7.15)$$

1. En toute rigueur, M et t_m devraient dépendre de k , mais on retire ces indices pour plus de lisibilité. Nous revenons temporairement à cette notation car elle rend le découpage des trames plus simple.

On considère que $\phi_k(f, t) | \phi_k(f, t-1)$ suit une loi de Von Mises [MARDIA et ZEMROCH \(1975\)](#) de mode $\phi_k(f, t-1) + 2\pi S\nu_k(f)$ et de paramètre de concentration $\kappa_k(f, t)$. Nous détaillons cette loi et ses paramètres dans le chapitre 8, section 8.1.1. Cette distribution modélise des variables 2π -périodiques, ce qui est adapté à la phase. Sa densité de probabilité est :

$$p(\phi_k(f, t) | \phi_k(f, t-1)) = \frac{e^{\kappa_k(f, t) \cos(\phi_k(f, t) - \phi_k(f, t-1) - 2\pi S\nu_k(f))}}{2\pi I_0(\kappa_k(f, t))}, \quad (7.16)$$

où I_0 désigne la fonction de Bessel modifiée de première espèce d'ordre 0. Ainsi :

$$\log(p_\Phi) \stackrel{c}{=} \sum_{k, f} \sum_m \log(p(\phi_k(f, t_m))) + \sum_{k, f} \sum_m \sum_{t=t_m+1}^{t_{m+1}-1} \kappa_k(f, t) \cos(\phi_k(f, t) - \phi_k(f, t-1) - 2\pi S\nu_k(f)), \quad (7.17)$$

ce que l'on peut réécrire en utilisant la notation Ω_k :

$$\log(p_\Phi) \stackrel{c}{=} \sum_{k, f} \sum_{t \in \Omega_k} \log(p(\phi_k(f, t))) + \sum_{k, f} \sum_{t \notin \Omega_k} \kappa_k(f, t) \cos(\phi_k(f, t) - \phi_k(f, t-1) - 2\pi S\nu_k(f)). \quad (7.18)$$

Par ailleurs, on considère que les phases dans les trames d'attaque sont distribuées selon une loi de Von Mises de localisation $\psi_k(f) + \lambda_k(t)f$, afin d'introduire le modèle de phase dans les trames d'attaque :

$$\forall t \in \Omega_k, p(\phi_k(f, t)) = \frac{e^{\kappa_k(f, t) \cos(\phi_k(f, t) - \psi_k(f) - \lambda_k(t)f)}}{2\pi I_0(\kappa_k(f, t))}. \quad (7.19)$$

Ainsi :

$$\begin{aligned} \log(p_\Phi) \stackrel{c}{=} & \sum_{k, f} \sum_{t \in \Omega_k} \kappa_k(f, t) \cos(\phi_k(f, t) - \psi_k(f) - \lambda_k(t)f) \\ & + \sum_{k, f} \sum_{t \notin \Omega_k} \kappa_k(f, t) \cos(\phi_k(f, t) - \phi_k(f, t-1) - 2\pi S\nu_k(f)). \end{aligned} \quad (7.20)$$

Remarque : On aurait pu aboutir au même résultat en structurant les phases dans les trames d'attaque avec un modèle de chaîne de Markov. Les termes ψ et λ auraient alors eu une signification différente : ψ désignerait la phase dans la première trame d'attaque, et λ le délai entre deux trames d'attaque consécutives. Dans le modèle tel que présenté ici, les λ représentent le délai par rapport à une phase de référence ψ . Ces deux modèles sont équivalents à condition de fixer $\lambda = 0$ pour la première trame d'attaque.

Afin de réécrire cette expression, on utilise le résultat suivant : $\forall (a, b) \in \mathbb{R}^2$,

$$|e^{ia} - e^{ib}|^2 = |e^{ia}|^2 + |e^{ib}|^2 - 2\Re(e^{ia}e^{-ib}) = 2 - 2\cos(a - b). \quad (7.21)$$

Ainsi :

$$\cos(\phi_k(f, t) - \psi_k(f) - \lambda_k(t)f) \stackrel{c}{=} -\frac{1}{2} |e^{i\phi_k(f, t)} - e^{i\psi_k(f)} e^{i\lambda_k(t)f}|^2, \quad (7.22)$$

et :

$$\cos(\phi_k(f, t) - \phi_k(f, t-1) - 2\pi S\nu_k(f)) \stackrel{c}{=} -\frac{1}{2} |e^{i\phi_k(f, t)} - e^{i\phi_k(f, t-1)} e^{2i\pi S\nu_k(f)}|^2. \quad (7.23)$$

On pose $\kappa_k(f, t) = 2 \frac{\sigma_r |X(f, t)|^2}{\tilde{\sigma}^2}$ dans les trames d'attaque et $\kappa_k(f, t) = 2 \frac{\sigma_u |X(f, t)|^2}{\tilde{\sigma}^2}$ dans les autres trames², avec $\sigma_u > 0$ et $\sigma_r > 0$. Ainsi,

$$\begin{aligned} \log(p_\Phi) \stackrel{c}{=} & \sum_{k,f} \sum_{t \in \Omega_k} \frac{\sigma_r}{\tilde{\sigma}^2} |X(f, t)|^2 |e^{i\phi_k(f,t)} - e^{i\psi_k(f)} e^{i\lambda_k(t)f}|^2 \\ & + \sum_{k,f} \sum_{t \notin \Omega_k} \frac{\sigma_u}{\tilde{\sigma}^2} |X(f, t)|^2 |e^{i\phi_k(f,t)} - e^{i\phi_k(f,t-1)} e^{2i\pi S\nu_k(f)}|^2, \end{aligned} \quad (7.24)$$

soit, en utilisant les expressions (7.3) et (7.6),

$$\log(p_\Phi) \stackrel{c}{=} -\frac{1}{\tilde{\sigma}^2} (\sigma_r \mathcal{C}_r(\phi, \psi, \lambda) + \sigma_u \mathcal{C}_u(\phi)). \quad (7.25)$$

Modèle complet

La maximisation de la distribution à postériori revient donc à maximiser :

$$\begin{aligned} \mathcal{C}_{MAP}(\theta) &= L(\theta) + \log(p_\Phi) + \log(p_H) \\ &\stackrel{c}{=} -\frac{1}{\tilde{\sigma}^2} \left(D(X, \hat{X}) + \sigma_u \mathcal{C}_u(\phi), +\sigma_r \mathcal{C}_r(\phi, \psi, \lambda) + \sigma_s \mathcal{C}_s(H) \right), \end{aligned}$$

ce qui revient à minimiser la fonction de coût complète suivante :

$$\mathcal{C}(\theta) = D(X, \hat{X}) + \sigma_u \mathcal{C}_u(\phi), +\sigma_r \mathcal{C}_r(\phi, \psi, \lambda) + \sigma_s \mathcal{C}_s(H). \quad (7.26)$$

On retrouve donc la fonction de coût introduite de façon intuitive dans la section 7.1.1.

7.2 Estimation du modèle

Nous présentons dans cette section les algorithmes d'estimation du modèle de NMF complexe contrainte. Nous introduisons certaines notations afin de simplifier l'écriture des règles de mise à jour. On commence par définir les matrices suivantes, $\forall k \in \llbracket 1, K \rrbracket$:

$$\begin{aligned} \mu_k &\in \mathbb{C}^{F \times 1}, \mu_k(f) = e^{2i\pi S\nu_k(f)}, \\ \Lambda_k &\in \mathbb{C}^{F \times T}, \Lambda_k(f, t) = \mathbb{1}_k(t) e^{if\lambda_k(t)}, \\ \Psi_k &\in \mathbb{C}^{F \times 1}, \Psi_k(f) = e^{i\psi_k(f)}, \\ \Phi_k &\in \mathbb{C}^{F \times T}, \Phi_k(f, t) = e^{i\phi_k(f,t)}. \end{aligned}$$

On désigne par H_k la k -ième ligne de H et par W_k la k -ième colonne de W . On utilise également les notations suivantes :

- M_\downarrow (respectivement M_\uparrow) est la matrice obtenue en retirant la dernière ligne (respectivement la première ligne) de M .
- M_\rightarrow (respectivement M_\leftarrow) désigne la matrice obtenue en retirant la dernière (respectivement la première) colonne de M et en insérant une colonne de 0 en tant que première (respectivement en temps que dernière) colonne.

2. Les hyper-paramètres dépendent donc des observations $|X(f, t)|^2$, ce qui peut sembler étrange, mais donne expérimentalement de bons résultats. Ce type d'approches est connu sous le nom d'approches *Bayésiennes empiriques* [EFRON \(2012\)](#).

-
- $\text{diag}_v(v)$ désigne la matrice diagonale dont les éléments diagonaux sont les termes du vecteur v , et $\text{diag}_m(M)$ désigne le vecteur colonne obtenu par extraction des éléments diagonaux de la matrice M .
 - $\text{vand}(v)$ désigne la matrice de Vandermonde obtenue à partir du vecteur v : si $v \in \mathbb{C}^{1 \times T}$, alors $M = \text{vand}(v) \in \mathbb{C}^{F \times T}$, avec $\forall(f, t), M(f, t) = v(t)^f$.

La minimisation de la fonction de coût (7.8) peut être effectuée par minimisation successive par rapport à chacune des variables (annulation des dérivées partielles). Cela permet d'aboutir à une procédure itérative. On n'a pas de garantie de convergence dans le cas général. Nous donnons le détail mathématique de cette technique d'optimisation appliquée à la fonction de coût (7.8) dans l'Annexe C (section C.1) de ce manuscrit afin de ne pas surcharger ce chapitre. La procédure complète est décrite dans l'Algorithme 8. Notons que l'ordre dans lequel sont effectuées les mises à jour des paramètres est arbitraire, et il pourrait être intéressant de considérer un ordre différent pour en évaluer l'impact sur les résultats.

Alternativement, on peut appliquer la méthode de la fonction auxiliaire, qui fournit un cadre théorique pour obtenir des règles de mise à jour. Les détails de calcul sont donnés dans l'Annexe C (section C.2). Essentiellement, la différence avec la première méthode se trouve dans la présence d'un terme G_k qui est un gain (ici choisi comme étant celui de Wiener) qui permet de "redistribuer" l'erreur d'approximation sur les diverses composantes. Ces gains proviennent des inégalités de convexité utilisées pour obtenir une fonction auxiliaire, et pourraient tout à fait prendre des valeurs différentes que celles proposées ici. Les mises à jour obtenues sont synthétisées dans l'Algorithme 9.

Pour l'initialisation, nous proposons d'appliquer une première NMF à $|X|$ afin d'obtenir une première approximation de W et de H . Alternativement, celles-ci peuvent être gardées aléatoires et l'algorithme de NMF Complexe peut être appliqué directement sur ces valeurs. Nous initialisons les phases en leur donnant celle du mélange : $\forall k \in \llbracket 1, K \rrbracket, \Phi_k = \frac{X}{|X|}$, et les paramètres de phases d'attaque Ψ_k et Λ_k par des valeurs aléatoires de module 1. Il faut également déterminer $\mathbf{1}_k$, qui est l'indicatrice de Ω_k . Il est possible d'estimer directement celle-ci à partir de H_k en utilisant par exemple PAULUS et VIRTANEN (2005). Nous souhaitons avoir une estimation précise des trames d'attaque (afin de se concentrer essentiellement sur l'influence de la reconstruction de phase), aussi nous avons utilisé la boîte à outils MATLAB Tempogram Toolbox GROSCHE et MÜLLER (2011) appliquée sur chaque spectrogramme de source isolément.

Algorithme 8 Estimation du modèle CNMF à phase contrainte (méthode de relaxation)

Entrées :

$X, K, \sigma_r, \sigma_u, \sigma_s.$

Initialisation $\forall k \in \llbracket 1, K \rrbracket :$

$W_k, H_k, \Phi_k,$

$\Lambda_k, \Psi_k, \mathbb{1}_k, \bar{\mathbb{1}}_k,$

$\hat{X}_k = (W_k H_k) \odot \Phi_k,$

$B_k = X - \sum_{l \neq k} \hat{X}_l.$

tant que critère non atteint **faire**

pour $k = 1$ à K **faire**

Calculer μ_k

 QIFFT sur W_k et découpage en régions d'influence.

Calculer Ψ_k

$$\Psi_k = \frac{\text{diag}_m((\Phi_k \odot X^{\odot 2})(\Lambda_k)^H)}{|\text{diag}_m((\Phi_k \odot X^{\odot 2})(\Lambda_k)^H)|}.$$

Calculer Λ_k

$$\Lambda_k = \text{vand} \left(\frac{(\bar{\Psi}_{k,\downarrow} \odot \Psi_{k,\uparrow})^H (|X|_{\downarrow} \odot |X|_{\uparrow} \odot \bar{\Phi}_{k,\downarrow} \odot \Phi_{k,\uparrow})}{|(\bar{\Psi}_{k,\downarrow} \odot \Psi_{k,\uparrow})^H (|X|_{\downarrow} \odot |X|_{\uparrow} \odot \bar{\Phi}_{k,\downarrow} \odot \Phi_{k,\uparrow})|} \right) \text{diag}_v(\mathbb{1}_k).$$

Calculer ρ_k

$$\rho_k = \sigma_r(\Psi_k \mathbb{1}_k) \odot \Lambda_k \odot |X|^{\odot 2} + \sigma_u(\mu_k \bar{\mathbb{1}}_k) \odot \Phi_{k,\rightarrow} \odot |X|^{\odot 2} + \sigma_u(\bar{\mu}_k \bar{\mathbb{1}}_{k,\leftarrow}) \odot \Phi_{k,\leftarrow} \odot |X_{\leftarrow}|^{\odot 2}.$$

Calculer Φ_k

$$\Phi_k = \frac{B_k \odot (W_k H_k) + \rho_k}{|B_k \odot (W_k H_k) + \rho_k|}.$$

Actualiser \hat{X}_k et B_k

Calculer β_k

$$\beta_k = \Re(B_k \odot \bar{\Phi}_k).$$

Calculer W

$$W_k = \frac{\beta_k (H_k)^T}{\alpha ((H_k)^{\odot 2})^T}.$$

Calculer H

$$H_k = \frac{(W_k)^T \beta_k}{p \sigma_s (H_k)^{\odot p-2} + ((W_k)^{\odot 2})^T \alpha}.$$

Projection de W et H **sur l'orthant positif**

Normaliser W et H

Actualiser \hat{X}^k et B_k

fin pour

fin tant que

Sorties :

$\hat{X}_k, W_k, H_k, \Phi_k,$

$\Lambda_k, \Psi_k, \mu_k.$

Algorithme 9 Estimation du modèle CNMF à phase contrainte (méthode de la fonction auxiliaire)

Entrées :

$X, K, \sigma_r, \sigma_u, \sigma_s.$

Initialisation $\forall k \in \llbracket 1, K \rrbracket :$

$W_k, H_k, \Phi_k,$

$\Lambda_k, \Psi_k, \mathbf{1}_k, \bar{\mathbf{1}}_k,$

$\hat{X}_k = (W_k H_k) \odot \Phi_k,$

$G_k = \sum_l \frac{W_k H_k}{W_l H_l},$

$B_k = \hat{X}_k + G_k \odot (X - \hat{X}).$

tant que critère non atteint **faire**

pour $k = 1$ à K **faire**

Calculer μ_k

 QIFFT sur W_k et découpage en régions d'influence.

Calculer Ψ_k

$$\Psi_k = \frac{\text{diag}_m((\Phi_k \odot X^{\odot 2})(\Lambda_k)^H)}{|\text{diag}_m((\Phi_k \odot X^{\odot 2})(\Lambda_k)^H)|}.$$

Calculer Λ_k

$$\Lambda_k = \text{vand} \left(\frac{(\bar{\Psi}_{k,\downarrow} \odot \Psi_{k,\uparrow})^H (|X|_{\downarrow} \odot |X|_{\uparrow} \odot \bar{\Phi}_{k,\downarrow} \odot \Phi_{k,\uparrow})}{|(\bar{\Psi}_{k,\downarrow} \odot \Psi_{k,\uparrow})^H (|X|_{\downarrow} \odot |X|_{\uparrow} \odot \bar{\Phi}_{k,\downarrow} \odot \Phi_{k,\uparrow})|} \right) \text{diag}_v(\mathbf{1}_k).$$

Calculer ρ_k

$$\rho_k = \sigma_r (\Psi_k \mathbf{1}_k) \odot \Lambda_k \odot |X|^{\odot 2} + \sigma_u (\mu_k \bar{\mathbf{1}}_k) \odot \Phi_{k,\rightarrow} \odot |X|^{\odot 2} + \sigma_u (\bar{\mu}_k \bar{\mathbf{1}}_{k,\leftarrow}) \odot \Phi_{k,\leftarrow} \odot |X_{\leftarrow}|^{\odot 2}.$$

Calculer Φ_k

$$\Phi_k = \frac{B_k \odot (W_k H_k) / G_k + \rho_k}{|B_k \odot (W_k H_k) / G_k + \rho_k|}.$$

Actualiser \hat{X}_k, B_k et G_k

Calculer β_k

$$\beta_k = \Re(B_k \odot \bar{\Phi}_k).$$

Calculer W

$$W_k = \frac{\frac{\beta_k}{G_k} (H_k)^T}{\frac{1}{G_k} ((H_k)^{\odot 2})^T}.$$

Calculer H

$$H_k = \frac{(W_k)^T \frac{\beta_k}{G_k}}{p \sigma_s (\bar{H}_k)^{\odot p-2} + ((W_k)^{\odot 2})^T \frac{1}{G_k}}.$$

Projection de W et H sur l'orthant positif

Normaliser W et H

Actualiser \hat{X}_k, B_k et G_k

fin pour

fin tant que

Sorties :

$\hat{X}_k, W_k, H_k, \Phi_k,$

$\Lambda_k, \Psi_k, \mu_k.$

7.3 Résultats expérimentaux

Nous avons conduit un ensemble d'expériences afin d'évaluer la performance de notre modèle de NMF complexe. Tout d'abord, nous avons cherché à étudier l'influence des paramètres σ_r et σ_u sur la performance de l'algorithme, mesurée grâce au SDR, SIR et SAR [VINCENT et al. \(2006\)](#). Puis, nous avons comparé cette approche dans le cadre de la séparation de sources à d'autres méthodes : la NMF complexe non contrainte, et une NMF classique avec filtrage de Wiener pour reconstruire la phase.

7.3.1 Données et protocole

Les jeux de données sont :

- A : 30 mélanges de deux sources, composées de sinusoides amorties synthétiques. Les sources se recouvrent dans le domaine TF.
- B : 30 mélanges de deux notes de piano tirées de la base de données MAPS [EMIYA et al. \(2010\)](#). Les sources se recouvrent également dans le domaine TF.
- C : 100 mélanges de morceaux de musique polyphonique tirés de la base DSD100 [ONO et al. \(2015\)](#), que nous avons eu l'occasion de présenter dans le chapitre 4.

Pour les données A et B, chaque source est observée seule, puis les deux sont activées simultanément. Les signaux sont échantillonnés à 11025 Hz sur ces jeux de données, et à 44100 Hz sur la base C. La TFCT est calculée avec une fenêtre de Hann de longueur 46 ms (512 échantillons pour les données A et B et 2048 échantillons pour les données C). Le taux de recouvrement est de 75 %. Les paramètres de parcimonie sont quant à eux fixés à des valeurs régulièrement utilisées dans la littérature (*cf.* par exemple [KAMEOKA et al. \(2009\)](#)) : $p = 1$ et $\sigma_s = \|X\|_2^2 K^{-(1-p/2)} 10^{-5}$.

7.3.2 Mélanges simples

Influence des paramètres

Dans cette expérience, nous évaluons, pour les jeux de données A et B, l'influence des paramètres σ_r et σ_u sur la qualité de la séparation effectuée avec 10 itérations de l'algorithme 9, initialisé avec 30 itérations de KLNMF.

Les résultats présentés dans l'article ICASSP [MAGRON et al. \(2016b\)](#) étaient obtenus par l'algorithme 8. Nous présentons ici les résultats obtenus avec la méthode de la fonction auxiliaire, qui sont globalement meilleurs. Néanmoins, en utilisant l'algorithme 9, nous constatons que le paramètre σ_r a une influence négligeable sur les résultats (ce qui n'est pas le cas avec l'algorithme 8). Pour les expériences sur les données A et B, on considère donc $\sigma_r = 0$. Les résultats de l'impact du paramètre σ_u sur la séparation de sources sont donnés sur la figure 7.2.

Globalement, des valeurs non-nulles de σ_u conduisent à améliorer les résultats par rapport à une valeur nulle, ce qui montre l'intérêt de notre méthode par rapport à une CNMF non contrainte. Un compromis entre SDR, SIR et SAR semble obtenu pour $\sigma_u = 1$ (données synthétiques) et $\sigma_u = 0.1$ (notes de piano). Des valeurs plus faibles sont insuffisantes pour contraindre efficacement le problème (on se rapproche alors du modèle de CNMF non contrainte), et des valeurs plus importantes sont au contraire trop contraignantes car les données ne respectent pas "suffisamment" le modèle (et l'influence du terme d'attache aux données dans la fonction de coût devient négligeable).

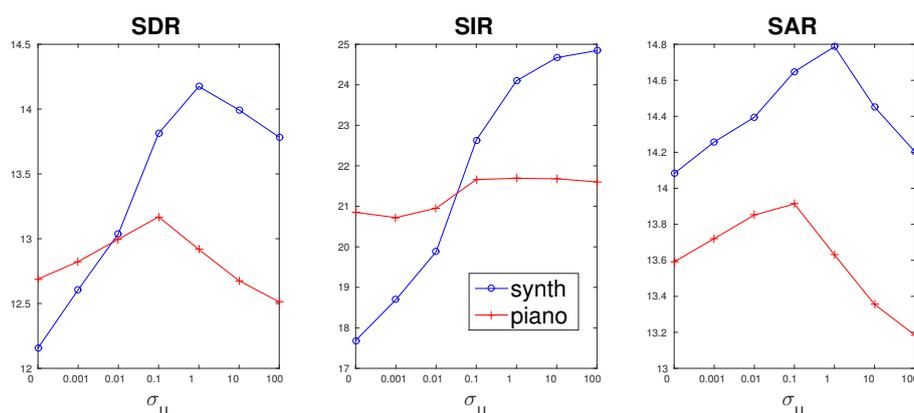


FIGURE 7.2 – Influence du paramètre σ_u sur la qualité de séparation de sources.

Données	Méthode	SDR	SIR	SAR
A	NMF-W	11.7	16.8	13.5
	CNMF	9.6	16.7	10.7
	CNMF- ϕ	12.3	23.2	12.8
B	NMF-W	14.7	18.5	17.4
	CNMF	13.3	19.6	14.9
	CNMF- ϕ	14.6	22.1	15.7

TABLEAU 7.1 – Performance de la séparation de source (SDR, SIR et SAR en dB) pour divers jeux de données et méthodes.

Séparation de sources

Nous effectuons une tâche de séparation de sources avec différentes méthodes :

- **NMF-W** : Une KLNMF³ suivie d'un filtrage de Wiener pour la reconstruction de phase ;
- **CNMF** : La NMF complexe sans contraintes de phase ;
- **CNMF- ϕ** : Notre algorithme de NMF complexe à phase contrainte, avec $\sigma_u = 1$ pour les données A et $\sigma_u = 0.1$ pour les données B.

La KLNMF est effectuée avec 30 itérations de mises à jour multiplicatives. Celle-ci sert d'initialisation aux NMF complexes, qui utilisent ensuite 10 itérations de l'algorithme 9 (en effet, l'algorithme **CNMF** décrit dans [KAMEOKA et al. \(2009\)](#) revient à utiliser l'algorithme 9 avec $\sigma_r = \sigma_u = 0$). Par ailleurs, à partir de la KLNMF initiale, on effectue 10 itérations supplémentaires de KLNMF afin que les différentes méthodes utilisent le même nombre d'itérations en tout. Les résultats de la qualité de la séparation de sources (SDR, SIR et SAR) sont présentés dans le tableau 7.1.

On constate un résultat similaire à celui observé dans le chapitre précédent, à savoir que le gain en performance dû à cet algorithme est à nuancer selon les indicateurs. Pour des signaux synthétiques (données A), on note une augmentation modérée de SDR et un gain plus important de SIR, au détriment d'une légère dégradation de SAR par rapport au filtrage de Wiener. L'amélioration reste significative par rapport à une NMF complexe non contrainte.

Sur les signaux de piano, le filtrage de Wiener conduit aux meilleurs résultats (en SDR et SAR), bien que notre algorithme permette une nette augmentation de SIR. Il est possible que

3. Par souci d'équité, la KLNMF est également assortie d'une contrainte de parcimonie, pour ne pas avantager les autres méthodes.

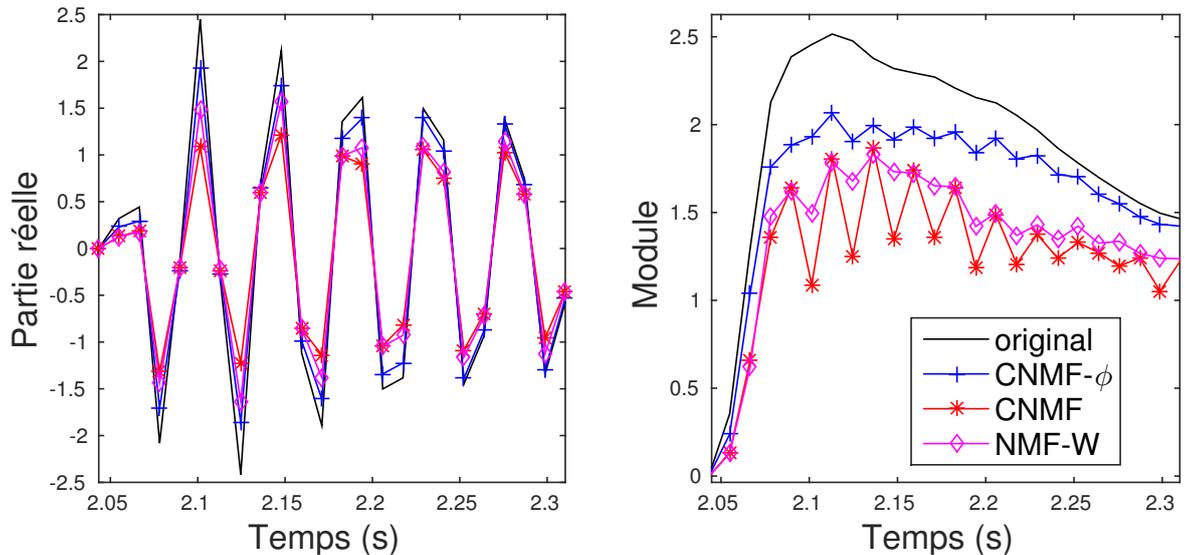


FIGURE 7.3 – Reconstruction d’un partiel de la note B2 à partir d’un mélange (E2 et B2), dans le canal fréquentiel à 495 Hz où les deux notes se recouvrent. Partie réelle (gauche) et amplitude (droite).

la NMF avec filtrage de Wiener initial soit déjà un résultat de bonne qualité, aussi poursuivre la séparation avec une CNMF ne semble pas forcément judicieux. Dans l’expérience suivante, nous verrons si tel est toujours le cas lorsque l’on traite des signaux plus compliqués. Pour nuancer ce résultat, notons que sur les notes de piano, la dégradation du SDR entre **NMF-W** et **CNMF- ϕ** est faible (0.1 dB) alors que le gain en SIR est plus important (3.6 dB). La diminution du SAR est également modérée.

Par ailleurs, on représente les signaux reconstruits (partie réelle et amplitude) dans une bande de fréquences et sur une durée où il y a recouvrement afin d’illustrer ces résultats. On choisit pour cela un mélange de deux notes de piano E2 et B2 et on représente le résultat sur la figure 7.3. On constate que les signaux reconstruits par les méthodes **NMF-W** et **CNMF** sont assez éloignés du signal original, dans le sens où le phénomène de battement dû au recouvrement des sources impacte les signaux séparés. Le fait de contraindre la phase permet de s’en affranchir partiellement, ce qui se traduit également par une amélioration perceptive de la qualité des signaux.

Enfin, en analysant plus précisément les résultats, on constate que pour des notes de piano graves, la CNMF à phase contrainte donne de meilleurs résultats que **NMF-W**, alors que cette tendance s’inverse à mesure que l’on monte dans les aigus. Ce phénomène peut s’expliquer par le fait que dans les basses fréquences, les recouvrements et les battements qui en découlent sont plus marqués que dans les hautes fréquences, d’où l’intérêt particulier d’une contrainte de phase pour cette gamme de sons.

7.3.3 Morceaux de musique polyphoniques

Nous reprenons les expériences précédentes, appliquées ici au jeu de données C. Plus précisément, cette base de données comprend 50 morceaux d’apprentissage et 50 morceaux de test. Nous étudions l’influence des paramètres sur les morceaux d’apprentissage, afin d’en déduire les valeurs optimales pour effectuer la séparation de sources sur la base de test. L’algorithme 9 est utilisé. Les sources sont reconstruites par clustering Oracle **BARKER et al. (2013)** à partir des atomes NMF, dont le principe est rappelé dans la section 2.2.6.

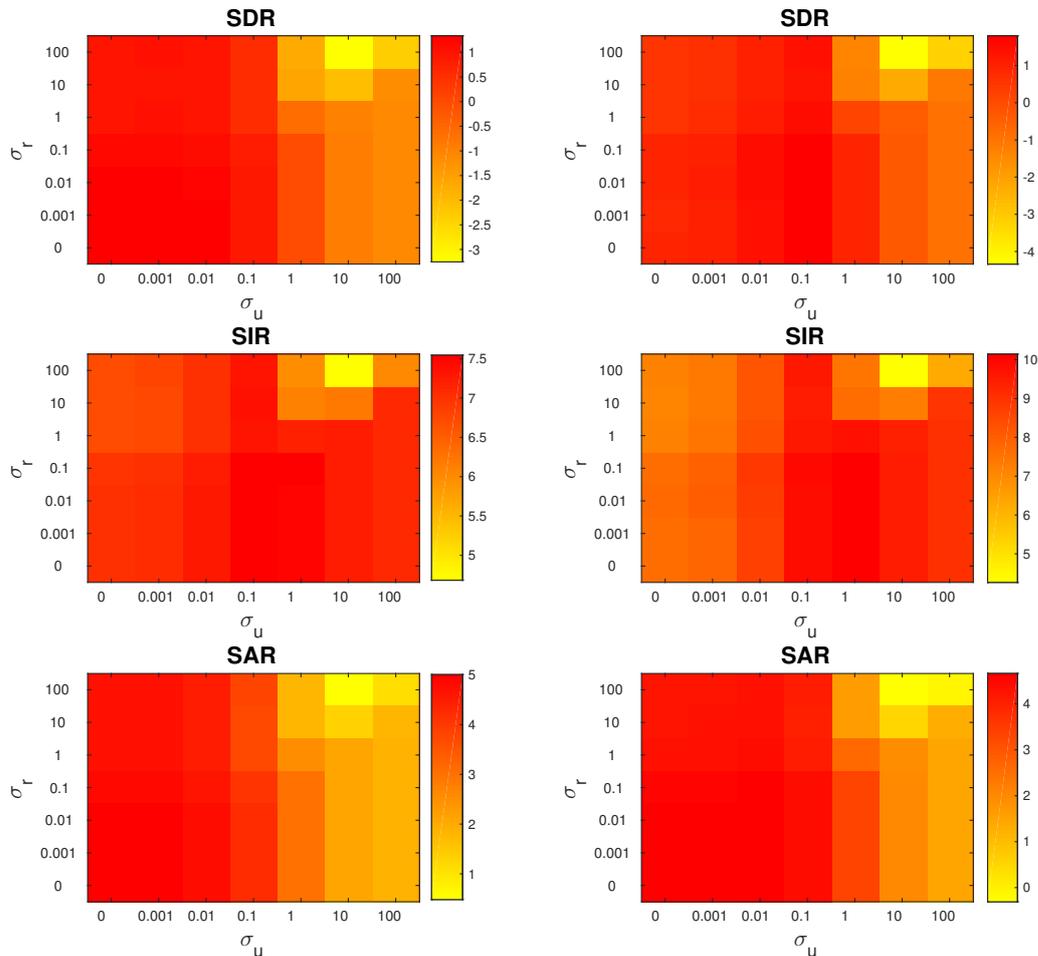


FIGURE 7.4 – Influence des paramètres σ_u et σ_r sur la qualité de séparation de sources (données DSD100), pour une initialisation par NMF (à gauche) et aléatoire (à droite).

Influence des paramètres

Nous testons deux initialisations :

- Une initialisation aléatoire, auquel cas l’algorithme de CNMF utilise 30 itérations ;
- Une initialisation par 30 itérations d’une KLNMF préalable, auquel cas l’algorithme de CNMF utilise 10 itérations.

Les résultats obtenus sont illustrés sur la figure 7.4. Nous constatons que le jeu de paramètres qui permet de maximiser la qualité de la séparation de sources (au niveau des trois indicateurs) semble être localisé autour du couple $(\sigma_r, \sigma_u) \approx (0.1, 0.1)$. Des valeurs plus faibles sont insuffisantes pour contraindre efficacement le problème (on se rapproche alors du modèle de CNMF non contrainte), et des valeurs plus importantes semblent au contraire trop contraignantes car les données ne respectent pas suffisamment le modèle. La valeur du paramètre σ_r est peu influente : les résultats sont principalement dépendants de la contrainte de déroulé linéaire, et non de la contrainte de répétition dans les trames d’attaque.

L’initialisation par KLNMF semble conduire à des résultats plus robustes (moins de variabilité notamment en SIR) mais en revanche, les valeurs de SIR en particulier sont moindres.

Méthode	SDR	SIR	SAR
NMF-W	1.9	10.2	3.7
CNMF	1.4	10.9	2.9
CNMF- ϕ	1.7	12.2	2.9

TABLEAU 7.2 – Performance de la séparation de source (SDR, SIR et SAR en dB) pour diverses méthodes sur la base DSD100.

En d'autres termes, le choix de l'initialisation n'impacte que peu les SDR et SAR, mais avec une initialisation aléatoire, on peut s'attendre à un gain d'environ 3 dB en SIR. Ce résultat, bien que suprenant, peut s'expliquer par le fait que la KLNMF initiale conduit à un minimum local de la fonction de coût, dont il est ensuite difficile de s'éloigner par application de la CNMF (similairement aux résultats de HRNMF dans le cas aveugle dans l'étude comparative du chapitre 3). En outre, cette approche est moins coûteuse en nombre d'itérations. Ce point est néanmoins à relativiser puisqu'on n'utilise alors aucune itération de KLNMF, mais davantage d'itérations de CNMF (qui sont un peu plus lourdes en temps de calcul).

Pour la dernière expérience, nous choisissons donc une initialisation aléatoire et on fixe les valeurs des paramètres comme suit : $(\sigma_r, \sigma_u) \approx (0.1, 0.1)$.

Séparation de sources

Nous testons la performance de notre méthode pour la séparation de sources sur la base de test DSD100. Chacune des méthodes utilise 30 itérations des algorithmes impliqués, et est initialisée avec des valeurs aléatoires pour W et H . Les résultats sont présentés dans le tableau 7.2.

La CNMF à phase contrainte donne de meilleurs résultats en SDR et SIR que la CNMF non contrainte, et une valeur similaire de SAR. Cela confirme l'intérêt de contraindre la NMF complexe pour en améliorer la performance. Par rapport au filtrage de Wiener, elle conduit à une baisse de 0.2 dB et 0.8 dB en SDR et SAR respectivement, et à un gain de 2 dB en SIR. Aussi, le choix d'une technique plutôt qu'une autre pourrait être motivée par la recherche d'un compromis entre ces différents indicateurs. On observe en particulier que l'utilisation d'une contrainte de déroulé linéaire de phase a tendance à réduire les interférences entre sources, conclusion déjà observée au chapitre 5.

7.4 Conclusion

Le modèle de NMF complexe à phase contrainte introduit dans ce chapitre est un outil prometteur pour la séparation de sources, notamment pour la réduction des interférences qui surviennent lorsqu'une méthode plus traditionnelle de reconstruction de phase est appliquée à des mélanges dans lesquels les sources se recouvrent dans le domaine TF. Ces contraintes de phase, basées sur les modèles de signaux que nous avons développés dans ce manuscrit, améliorent les résultats par rapport à une approche non contrainte. Nous avons observé expérimentalement que la contrainte au niveau des trames d'attaque n'avait pas un impact significatif sur les résultats. On pourrait donc envisager, à l'avenir, de travailler principalement à la reconstruction des phases des partiels dans le cadre d'une NMF complexe.

Nous avons supposé que le mélange était égal au modèle auquel est ajouté un bruit. Il pourrait être plus judicieux de directement modéliser les sources comme des variables latentes, ce qui permettrait de nouvelles techniques d'estimation. En outre, le modèle n'est pas conservatif, c'est-à-dire que la somme des sources n'est (en général) pas égale au mélange observé.

Le cadre probabiliste est donc plus adapté à ce type de problèmes, et c'est la direction que nous envisageons de suivre dans la troisième partie de ce manuscrit.