

Soutenance d'habilitation à diriger des recherches

Contributions à la factorisation en matrices non-négatives

Cédric Févotte

Laboratoire Lagrange, Nice



Observatoire
de la CÔTE d'AZUR



13 octobre 2014

Plan

Curriculum vitæ

Contributions à la factorisation en matrices non-négatives

- Introduction à la NMF

- Algorithmes pour la NMF fondée sur la β -divergence

- Cas particulier de la NMF avec la divergence d'Itakura-Saito

- Estimation par maximum de vraisemblance marginalisée

- Remarques finales

Parcours

Carrière

2007–	chargé de recherche CNRS (CR1, section 7)
depuis 2013	Laboratoire Lagrange (CNRS, OCA & UNS), Nice
2007–2012	LTCI (CNRS & Télécom ParisTech), Paris
2007	chercheur contractuel à Télécom ParisTech (8 mois)
2006–2007	ingénieur de recherche à Mist-Technologies, Paris (10 mois)
2003–2006	chercheur post-doctorant à l'Université de Cambridge, UK

Formation

2000–2003	thèse de doctorat à l'IRCCyN, Nantes (avec C. Doncarli)
1999–2000	DEA auto/info à l'École Centrale de Nantes
1997–2000	études d'ingénieur généraliste à l'École Centrale de Nantes

Thématiques : traitement statistique du signal & apprentissage, pour le débruitage, la séparation de sources, les problèmes inverses, notamment en audio.

Contributions reconnues

Modèles probabilistes temps-fréquence pour la décomposition du signal

► modèles NMF (depuis mon entrée au CNRS)

- C. Févotte, N. Bertin, and J.-L. Durrieu. [Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis.](#) *Neural Computation*, 2009. **~400 citations**
- C. Févotte and J. Idier. [Algorithms for nonnegative matrix factorization with the beta-divergence.](#) *Neural Computation*, 2011. **~90 citations**
- A. Ozerov and C. Févotte. [Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation.](#) *IEEE TASLP*, 2010. **~150 citations**

► modèles parcimonieux (thèse, post-doc)

- C. Févotte and S. J. Godsill. [A Bayesian approach to blind separation of sparse sources.](#) *IEEE TASLP*, 2006. **~150 citations**
- C. Févotte and C. Doncarli. [Two contributions to blind source separation using time-frequency distributions.](#) *IEEE SPL*, 2004. **~90 citations**

Évaluation des méthodes de séparation de sources audio (thèse, post-doc)

Co-auteur de la boîte à outils Matlab [BSS-Eval](#)

E. Vincent, R. Gribonval, and C. Févotte. [Performance measurement in blind audio source separation.](#) *IEEE TASLP*, 2006. **~700 citations**

Environnements internationaux

IEEE Signal Processing Society

- ▶ publications (et reviews) majoritairement IEEE SPS (14 revues, 23 conférences).
- ▶ senior member depuis septembre 2014.
- ▶ membre élu du comité technique IEEE Machine Learning for Signal Processing depuis 2010.
- ▶ éditeur associé pour les IEEE Trans. Signal Processing depuis 2014.
- ▶ chairman de sessions à SSP 2011, ICASSP 2013 & 2014.
- ▶ ICASSP 2014 Best student paper award (attribué à D. Sun).

Communauté “machine learning”

- ▶ 2 publications dans Neural Computation (2009, 2011).
- ▶ 2 publications dans NIPS (2011, 2014).

Un cadre national : le GdR ISIS

Pendant la thèse

- ▶ projet jeunes chercheurs “Ressources pour la séparation de sources audio” avec E. Vincent & R. Gribonval (2001–2003).
- ▶ séjour doctoral d'un mois à l'université de Maribor (2002).

Depuis mon entrée au CNRS

- ▶ **membre du comité de direction** depuis 2009, en charge du réseau des doctorants.
- ▶ co-organisateur de la journée “Carrières en Signal, Image & Vision” (120 participants, 2013).
- ▶ co-organisateur de deux journées scientifiques (2011 & 2012).
- ▶ directeur scientifique de l'école d'été de Peyresq en 2010.
- ▶ membre du jury du prix de thèse Grets/ISIS/EEA depuis 2013.

Mobilité(s)

Nationale

IRCCyN [Nantes](#), LTCI [Paris](#), Lagrange [Nice](#).

+ courts séjours à l'IRIT [Toulouse](#) et au LATP [Marseille](#).

Internationale

- ▶ Université de [Cambridge UK](#) (post-doc de 2 ans et demi).
- ▶ Mitsubishi Electric Research Laboratories (MERL) [Cambridge MA](#) (3 mois en disponibilité du CNRS).

+ courts séjours à l'Université de [Maribor](#) et à l'Université du Bosphore [Istanbul](#).

Industrielle

- ▶ 10 mois en [start-up](#) (Mist-Technologies).
- ▶ 3 mois dans un [grand département de R&D](#) (MERL).

+ activité de [consultance](#) pour CGG Veritas.

2 brevets américains (déposés).

Divers

Encadrement

- ▶ 2 thèses soutenues, 1 en cours (+ 1 interrompue)
- ▶ 2 post-doctorants (18 et 24 mois)
- ▶ 2 séjours de doctorants américains (3 mois)
- ▶ 2 stages de master

Financements

- ▶ 2 projets ANR, dont 1 ANR Jeunes Chercheurs
- ▶ 1 bourse de thèse Institut Télécom
- ▶ divers “petits” financements (CNRS PEPS, GdR ISIS, BQR Lagrange, etc.)

Expertise

- ▶ 4 participations à des jurys de thèse à l'étranger (Stockholm, Cambridge, Istanbul)
- ▶ 4 participations à des comités de recrutement MCF

Bonnes œuvres

- ▶ participation à un rapport IGAENR sur la simplification administrative dans les unités de recherche

Plan

Curriculum vitæ

Contributions à la factorisation en matrices non-négatives

- Introduction à la NMF

- Algorithmes pour la NMF fondée sur la β -divergence

- Cas particulier de la NMF avec la divergence d'Itakura-Saito

- Estimation par maximum de vraisemblance marginalisée

- Remarques finales

Factorisation en matrices non-négatives (NMF)

Étant donnée une matrice non-négative \mathbf{V} de dimensions $F \times N$, la NMF consiste à produire l'approximation

$$\mathbf{V} \approx \mathbf{WH},$$

où \mathbf{W} et \mathbf{H} sont des matrices non-négatives de dimensions $F \times K$ et $K \times N$, respectivement.

Factorisation en matrices non-négatives (NMF)

Étant donnée une matrice non-négative \mathbf{V} de dimensions $F \times N$, la NMF consiste à produire l'approximation

$$\mathbf{V} \approx \mathbf{WH},$$

où \mathbf{W} et \mathbf{H} sont des matrices non-négatives de dimensions $F \times K$ et $K \times N$, respectivement.

Cadre courant : $K < F < N$ (approximation de rang faible)

$$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \mathbf{V} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \approx \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \mathbf{W} & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \mathbf{H} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Représentation adaptative par parties

Comme la PCA ou l'ICA, la NMF produit une **représentation linéaire non-supervisée** des données :

$$\begin{array}{ccccc}
 \mathbf{v}_n & \approx & \mathbf{W} & & \mathbf{h}_n \\
 \text{donnée} & & \begin{array}{l} \text{"variables explicatives"} \\ \text{"dictionnaire"} \\ \text{"motifs"} \end{array} & & \begin{array}{l} \text{"régresseurs"} \\ \text{"coefficients de décomposition"} \\ \text{"activations"} \end{array}
 \end{array}$$

\mathbf{W} est *appris* grâce à l'ensemble des données $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_N]$.

Représentation adaptative par parties

Comme la PCA ou l'ICA, la NMF produit une **représentation linéaire non-supervisée** des données :

$$\mathbf{v}_n \approx \mathbf{W} \mathbf{h}_n$$

donnée “variables explicatives” “régresseurs”
 “dictionnaire” “coefficients de décomposition”
 “motifs” “activations”

\mathbf{W} est *appris* grâce à l'ensemble des données $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_N]$.

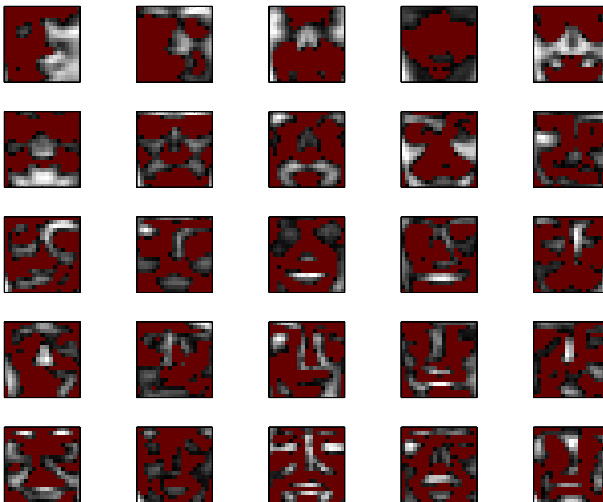
- ▶ **non-négativité de \mathbf{W}** assure l'*interprétabilité* du dictionnaire (motifs \mathbf{w}_k et données \mathbf{v}_n appartiennent au même espace).
- ▶ **non-négativité de \mathbf{H}** tend à produire une représentation *par parties* car le modèle interdit les combinaisons soustractives.

Travaux préliminaires de Paatero and Tapper (1994), papier phare dans *Nature* par Lee and Seung (1999).

49 images parmi 2429 des données CBCL



Dictionnaire PCA avec $K = 25$



(les pixels rouges indiquent des valeurs négatives)

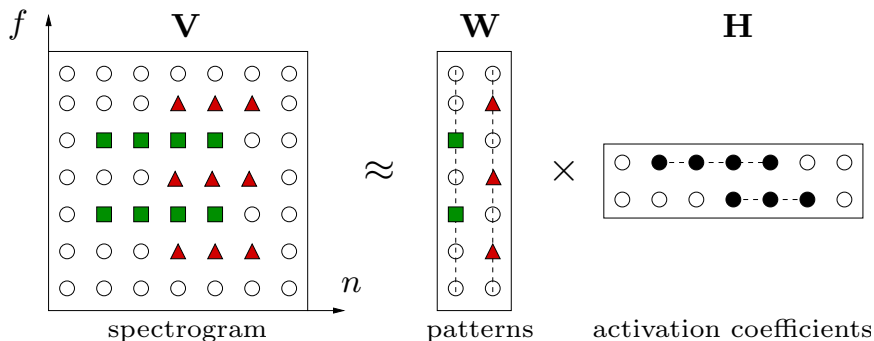
Dictionnaire NMF avec $K = 25$



(expérience reproduite de (Lee and Seung, 1999))

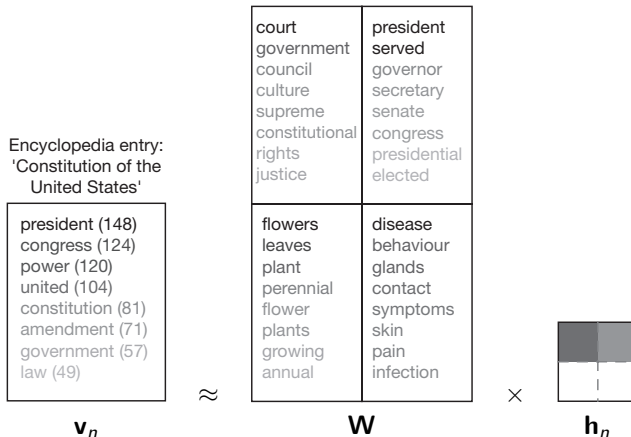
Traitement du signal audio

(Smaragdis and Brown, 2003)



Analyse sémantique de données textuelles

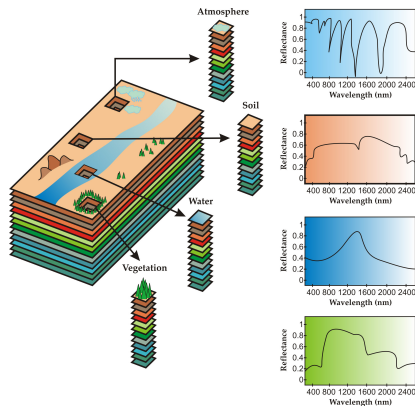
(Lee and Seung, 1999; Hofmann, 1999; Blei, Ng, and Jordan, 2003)



(figure reproduite de (Lee and Seung, 1999))

Imagerie hyperspectrale

(Berry, Browne, Langville, Pauca, and Plemmons, 2007)



(figure reproduite de Bioucas-Dias et al. (2012))

Plan

Curriculum vitæ

Contributions à la factorisation en matrices non-négatives

Introduction à la NMF

Algorithmes pour la NMF fondée sur la β -divergence

Cas particulier de la NMF avec la divergence d'Itakura-Saito

Estimation par maximum de vraisemblance marginalisée

Remarques finales

Un problème d'optimisation contraint

Minimiser une mesure de dissemblance entre \mathbf{V} et \mathbf{WH} :

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D(\mathbf{V}|\mathbf{WH}) = \sum_{fn} d([\mathbf{V}]_{fn} | [\mathbf{WH}]_{fn}),$$

où $d(x|y)$ est une fonction coût scalaire.

Termes de régularisation souvent ajoutés à $D(\mathbf{V}|\mathbf{WH})$ afin de structurer \mathbf{W} et/ou \mathbf{H} (e.g., parcimonie, lissage).

β -divergence

(Basu, Harris, Hjort, and Jones, 1998; Cichocki and Amari, 2010)

Fonction coût courante en NMF.

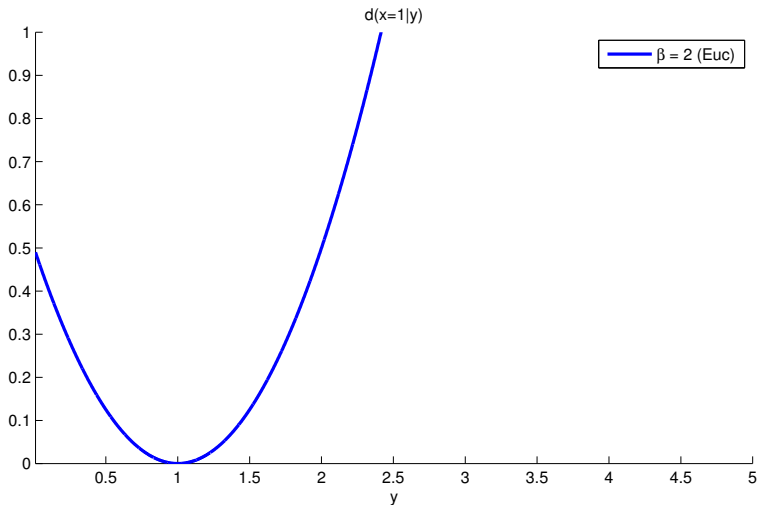
$$d_{\beta}(x|y) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{\beta(\beta-1)} (x^{\beta} + (\beta-1)y^{\beta} - \beta x y^{\beta-1}) & \beta \in \mathbb{R} \setminus \{0, 1\} \\ x \log \frac{x}{y} + (y - x) & \beta = 1 \\ \frac{x}{y} - \log \frac{x}{y} - 1 & \beta = 0 \end{cases}$$

Cas particuliers :

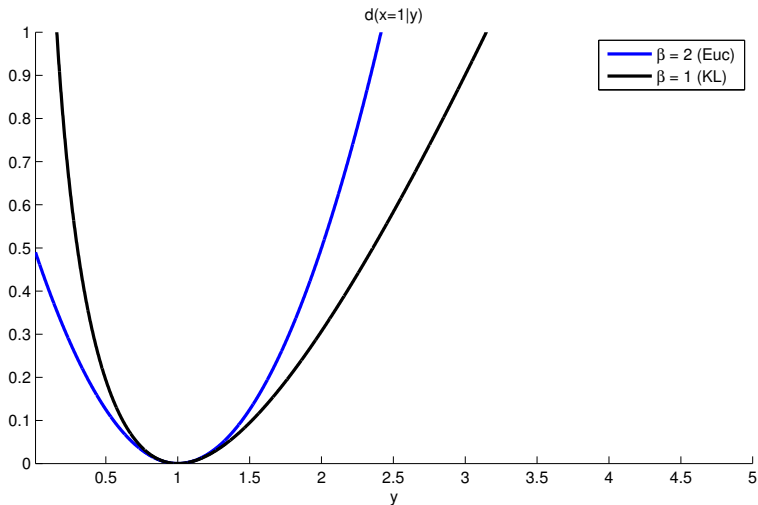
- ▶ erreur quadratique ($\beta = 2$)
- ▶ divergence de Kullback-Leibler (KL) ($\beta = 1$)
- ▶ divergence d'Itakura-Saito (IS) ($\beta = 0$)

Possibilité d'ajuster β à un problème donné par apprentissage, validation croisée, induction, etc.

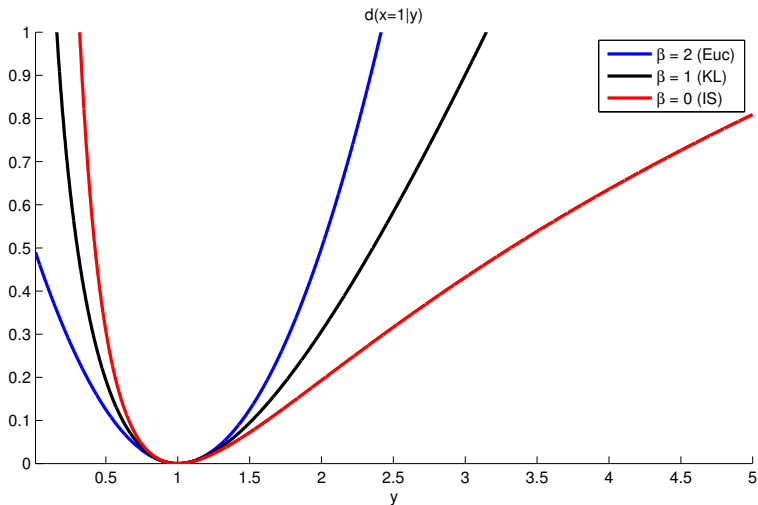
β -divergence



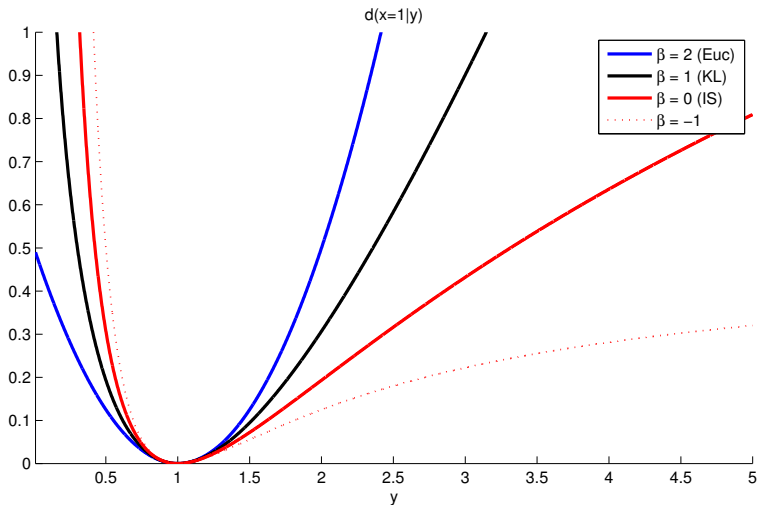
β -divergence



β -divergence



β -divergence



β -divergence

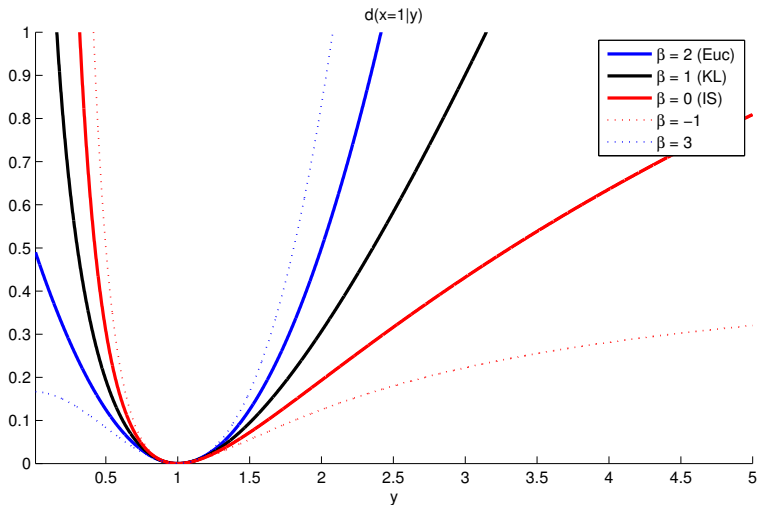


Schéma de descente par bloc

- mise à jour de \mathbf{H} sachant $\mathbf{W}^{(i-1)}$ et de \mathbf{W} sachant $\mathbf{H}^{(i)}$:

$$\min_{\mathbf{H} \geq 0} D(\mathbf{V} | \mathbf{W}^{(i-1)} \mathbf{H})$$

$$\min_{\mathbf{W} \geq 0} D(\mathbf{V} | \mathbf{W} \mathbf{H}^{(i)})$$

(seule garantie de convergence locale)

Schéma de descente par bloc

- mise à jour de \mathbf{H} sachant $\mathbf{W}^{(i-1)}$ et de \mathbf{W} sachant $\mathbf{H}^{(i)}$:

$$\min_{\mathbf{H} \geq 0} D(\mathbf{V} | \mathbf{W}^{(i-1)} \mathbf{H})$$

$$\min_{\mathbf{W} \geq 0} D(\mathbf{V} | \mathbf{W} \mathbf{H}^{(i)})$$

(seule garantie de convergence locale)

- mises à jour de \mathbf{W} et \mathbf{H} équivalentes par transposition :

$$\mathbf{V} \approx \mathbf{W} \mathbf{H} \iff \mathbf{V}^T \approx \mathbf{H}^T \mathbf{W}^T$$

Schéma de descente par bloc

- mise à jour de \mathbf{H} sachant $\mathbf{W}^{(i-1)}$ et de \mathbf{W} sachant $\mathbf{H}^{(i)}$:

$$\min_{\mathbf{H} \geq 0} D(\mathbf{V} | \mathbf{W}^{(i-1)} \mathbf{H})$$

$$\min_{\mathbf{W} \geq 0} D(\mathbf{V} | \mathbf{W} \mathbf{H}^{(i)})$$

(seule garantie de convergence locale)

- mises à jour de \mathbf{W} et \mathbf{H} équivalentes par transposition :

$$\mathbf{V} \approx \mathbf{W} \mathbf{H} \iff \mathbf{V}^T \approx \mathbf{H}^T \mathbf{W}^T$$

- fonction objectif séparable selon les colonnes de \mathbf{H} et les lignes de \mathbf{W} :

$$D(\mathbf{V} | \mathbf{W} \mathbf{H}) = \sum_n D(\mathbf{v}_n | \mathbf{W} \mathbf{h}_n)$$

Schéma de descente par bloc

Au final, un problème de *régression linéaire* sous contrainte de *non-négativité* :

$$\min_{\mathbf{h} \geq 0} C(\mathbf{h}) \stackrel{\text{def}}{=} D(\mathbf{v} | \mathbf{W}\mathbf{h})$$

Schéma de descente par bloc

Au final, un problème de *régression linéaire* sous contrainte de *non-négativité* :

$$\min_{\mathbf{h} \geq 0} C(\mathbf{h}) \stackrel{\text{def}}{=} D(\mathbf{v} | \mathbf{W}\mathbf{h})$$

- ▶ quelques cas connus en déconvolution d'image, traités par majoration-minimisation :
 - ▶ coût quadratique : (Daube-Witherspoon and Muehllehner, 1986; De Pierro, 1993)
 - ▶ divergence KL : (Richardson, 1972; Lucy, 1974)
 - ▶ divergence IS : (Cao, Eggermont, and Terebey, 1999)

Schéma de descente par bloc

Au final, un problème de *régression linéaire* sous contrainte de *non-négativité* :

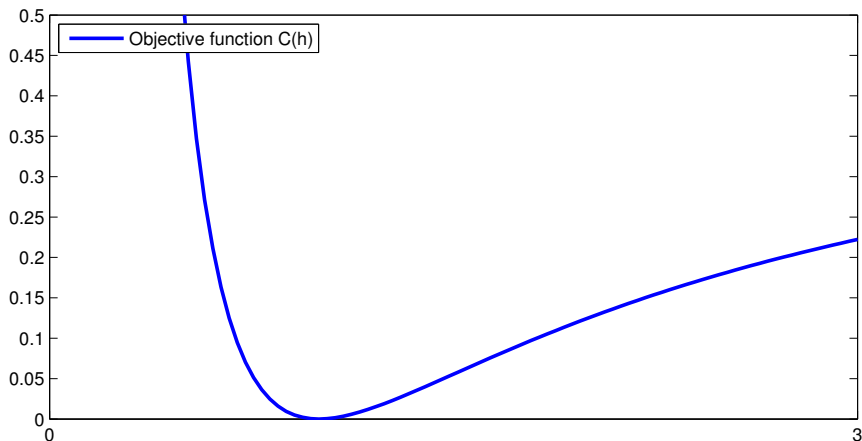
$$\min_{\mathbf{h} \geq 0} C(\mathbf{h}) \stackrel{\text{def}}{=} D(\mathbf{v} | \mathbf{W}\mathbf{h})$$

- ▶ quelques cas connus en déconvolution d'image, traités par majoration-minimisation :
 - ▶ coût quadratique : (Daube-Witherspoon and Muehllehner, 1986; De Pierro, 1993)
 - ▶ divergence KL : (Richardson, 1972; Lucy, 1974)
 - ▶ divergence IS : (Cao, Eggermont, and Terebey, 1999)
- ▶ avant nos travaux, contributions éparées pour les autres valeurs de β (Cichocki et al., 2006; Kompass, 2007), avec notamment des approches heuristiques.

Majoration-minimisation (MM)

(Févotte and Idier, 2011)

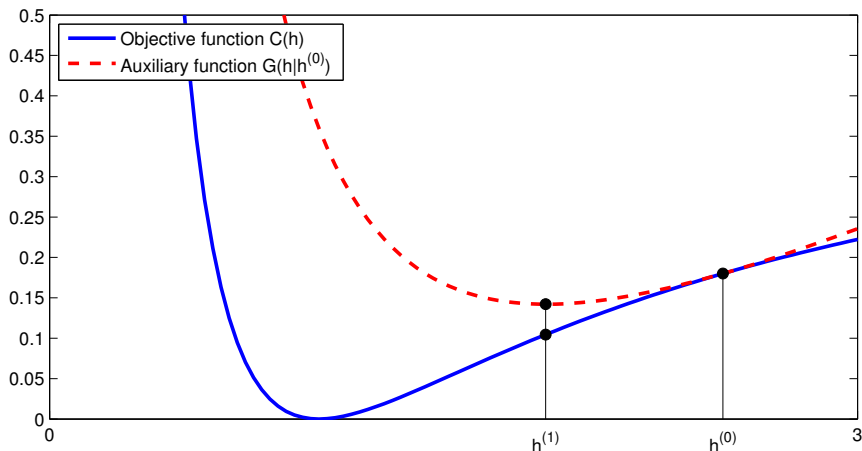
Construire $G(\mathbf{h}|\tilde{\mathbf{h}})$ tel que $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ et $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Minimiser (iterativement) $G(\mathbf{h}|\tilde{\mathbf{h}})$ au lieu de $C(\mathbf{h})$.



Majoration-minimisation (MM)

(Févotte and Idier, 2011)

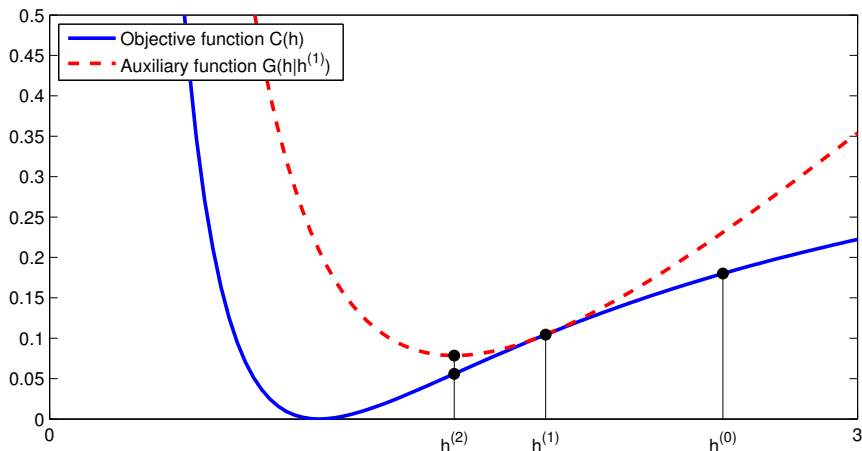
Construire $G(\mathbf{h}|\tilde{\mathbf{h}})$ tel que $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ et $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Minimiser (iterativement) $G(\mathbf{h}|\tilde{\mathbf{h}})$ au lieu de $C(\mathbf{h})$.



Majoration-minimisation (MM)

(Févotte and Idier, 2011)

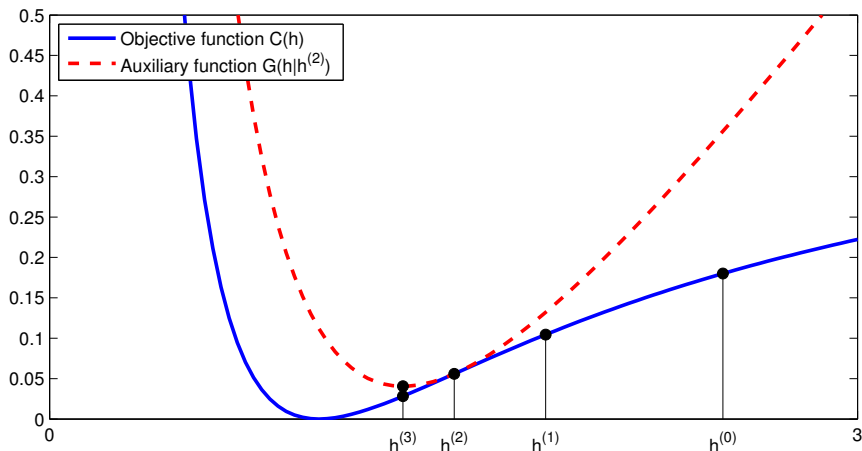
Construire $G(\mathbf{h}|\tilde{\mathbf{h}})$ tel que $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ et $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Minimiser (iterativement) $G(\mathbf{h}|\tilde{\mathbf{h}})$ au lieu de $C(\mathbf{h})$.



Majoration-minimisation (MM)

(Févotte and Idier, 2011)

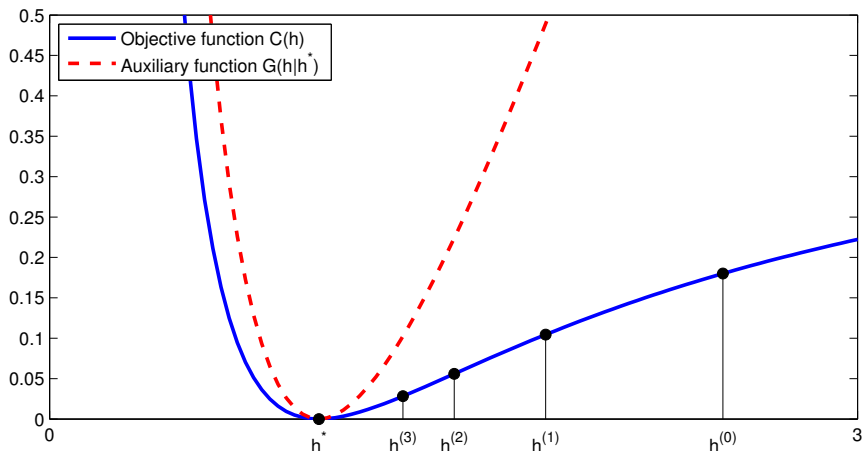
Construire $G(\mathbf{h}|\tilde{\mathbf{h}})$ tel que $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ et $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Minimiser (iterativement) $G(\mathbf{h}|\tilde{\mathbf{h}})$ au lieu de $C(\mathbf{h})$.



Majoration-minimisation (MM)

(Févotte and Idier, 2011)

Construire $G(\mathbf{h}|\tilde{\mathbf{h}})$ tel que $G(\mathbf{h}|\tilde{\mathbf{h}}) \geq C(\mathbf{h})$ et $G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}}) = C(\tilde{\mathbf{h}})$.
Minimiser (iterativement) $G(\mathbf{h}|\tilde{\mathbf{h}})$ au lieu de $C(\mathbf{h})$.



Majoration-minimisation (MM)

(Févotte and Idier, 2011)

Majorer les parties convexes et concaves de la fonction objectif séparément :

$$C(\mathbf{h}) = \underbrace{\widetilde{C}(\mathbf{h})}_{\text{Maj. par inég. de Jensen}} + \underbrace{\widehat{C}(\mathbf{h})}_{\text{Maj. par tangente}}$$

Majoration-minimisation (MM)

(Févotte and Idier, 2011)

Majorer les parties convexes et concaves de la fonction objectif séparément :

$$C(\mathbf{h}) = \underbrace{\tilde{C}(\mathbf{h})}_{\text{Maj. par inég. de Jensen}} + \underbrace{\widehat{C}(\mathbf{h})}_{\text{Maj. par tangente}}$$

Conduit à des mises à jour multiplicatives :

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \left[\frac{\mathbf{W}^T [(\mathbf{W}\mathbf{H})^{(\beta-2)} \cdot \mathbf{V}]}{\mathbf{W}^T [\mathbf{W}\mathbf{H}]^{(\beta-1)}} \right]^{\gamma(\beta)}$$

où $\gamma(\beta)$ est un exposant scalaire ≤ 1 . Mise à jour équivalente pour \mathbf{W} .

Majoration-minimisation (MM)

(Févotte and Idier, 2011)

Majorer les parties convexes et concaves de la fonction objectif séparément :

$$C(\mathbf{h}) = \underbrace{\widetilde{C}(\mathbf{h})}_{\text{Maj. par inég. de Jensen}} + \underbrace{\widehat{C}(\mathbf{h})}_{\text{Maj. par tangente}}$$

Conduit à des mises à jour multiplicatives :

$$\mathbf{H} \leftarrow \mathbf{H} \cdot \left[\frac{\mathbf{W}^T [(\mathbf{W}\mathbf{H})^{(\beta-2)} \cdot \mathbf{V}]}{\mathbf{W}^T [\mathbf{W}\mathbf{H}]^{(\beta-1)}} \right]^{\gamma(\beta)}$$

où $\gamma(\beta)$ est un exposant scalaire ≤ 1 . Mise à jour équivalente pour \mathbf{W} .

- ▶ forme multiplicative préserve la non-négativité.
- ▶ implémentation facile, complexité $\mathcal{O}(FKN)$ par itération.

Autres contributions

- ▶ **Sur-relaxations de l'algorithme MM** (Févotte and Idier, 2011)
 - ▶ Recherche de valeurs plus "agressives" pour $\gamma(\beta)$.
 - ▶ Majoration-égalisation : résoudre $G(\mathbf{h}|\tilde{\mathbf{h}}) = G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}})$.

Autres contributions

- ▶ **Sur-relaxations de l'algorithme MM** (Févotte and Idier, 2011)
 - ▶ Recherche de valeurs plus "agressives" pour $\gamma(\beta)$.
 - ▶ Majoration-égalisation : résoudre $G(\mathbf{h}|\tilde{\mathbf{h}}) = G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}})$.
- ▶ **Approche pénalisée pour la détermination de l'ordre** (Tan and Févotte, 2013)

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D_{\beta}(\mathbf{V}|\mathbf{WH}) + \lambda \sum_{k=1}^K \log(\|\mathbf{w}_k\| + \|\underline{h}_k\| + b)$$

Parcimonie de groupe au niveau des colonnes de \mathbf{W} & lignes de \mathbf{H} .

Autres contributions

- ▶ **Sur-relaxations de l'algorithme MM** (Févotte and Idier, 2011)
 - ▶ Recherche de valeurs plus "agressives" pour $\gamma(\beta)$.
 - ▶ Majoration-égalisation : résoudre $G(\mathbf{h}|\tilde{\mathbf{h}}) = G(\tilde{\mathbf{h}}|\tilde{\mathbf{h}})$.
- ▶ **Approche pénalisée pour la détermination de l'ordre** (Tan and Févotte, 2013)

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D_{\beta}(\mathbf{V}|\mathbf{WH}) + \lambda \sum_{k=1}^K \log(\|\mathbf{w}_k\| + \|\mathbf{h}_k\| + b)$$

Parcimonie de groupe au niveau des colonnes de \mathbf{W} & lignes de \mathbf{H} .

- ▶ **Co-factorisation douce en matrices non-négatives** (Seichepine, Essid, Févotte, and Cappé, 2014)

$$\min_{\mathbf{W}_1, \mathbf{W}_2, \mathbf{H}_1, \mathbf{H}_2} D_{\beta_1}(\mathbf{V}_1|\mathbf{W}_1\mathbf{H}_1) + \alpha D_{\beta_2}(\mathbf{V}_2|\mathbf{W}_2\mathbf{H}_2) + \lambda D(\mathbf{H}_1|\mathbf{H}_2)$$

Plan

Curriculum vitæ

Contributions à la factorisation en matrices non-négatives

Introduction à la NMF

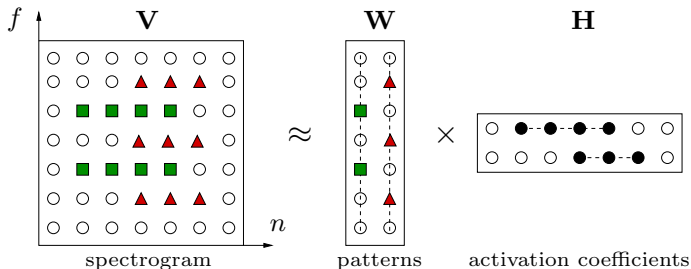
Algorithmes pour la NMF fondée sur la β -divergence

Cas particulier de la NMF avec la divergence d'Itakura-Saito

Estimation par maximum de vraisemblance marginalisée

Remarques finales

Décomposition spectrale (audio)



- ▶ quelle mesure de dissemblance utiliser ?
- ▶ quel type de spectrogramme ?
- ▶ la NMF approche V par une somme de spectrogrammes de rang 1. Comment les inverser ? Comment traiter la phase ?

Itakura-Saito NMF : une approche générative

(Févotte, Bertin, and Durrieu, 2009)

- ▶ soit $\mathbf{X} = \{x_{fn}\}$ la TFCT du signal (à valeurs complexes).

Itakura-Saito NMF : une approche générative

(Févotte, Bertin, and Durrieu, 2009)

- ▶ soit $\mathbf{X} = \{x_{fn}\}$ la TFCT du signal (à valeurs complexes).
- ▶ soit le modèle composite

$$x_{fn} = \sum_{k=1}^K c_{k,fn}$$
$$c_{k,fn} \sim N_c(0, w_{fk} h_{kn})$$

avec les composantes $c_{1,fn}, \dots, c_{K,fn}$ indépendantes sachant \mathbf{W} et \mathbf{H} .

Itakura-Saito NMF : une approche générative

(Févotte, Bertin, and Durrieu, 2009)

- ▶ soit $\mathbf{X} = \{x_{fn}\}$ la TFCT du signal (à valeurs complexes).
- ▶ soit le modèle composite

$$x_{fn} = \sum_{k=1}^K c_{k,fn}$$

$$c_{k,fn} \sim \mathcal{N}_c(0, w_{fk} h_{kn})$$

avec les composantes $c_{1,fn}, \dots, c_{K,fn}$ indépendantes sachant \mathbf{W} et \mathbf{H} .

- ▶ alors

$$-\log p(\mathbf{X}|\mathbf{W}, \mathbf{H}) = D_{\text{IS}}(|\mathbf{X}|^2|\mathbf{WH}) + \text{cst}$$

Additivité des composantes dans le domaine complexe.

Phase conservée dans le modèle, bien que de manière non-informative.

Itakura-Saito NMF : une approche générative

(Févotte, Bertin, and Durrieu, 2009)

- ▶ composantes latentes peuvent être reconstruites de façon naturelle :

$$\hat{c}_{k,fn} = \mathbb{E}\{c_{k,fn} | \mathbf{X}, \mathbf{W}, \mathbf{H}\} = \frac{w_{fk} h_{kn}}{\underbrace{\sum_j w_{fj} h_{jn}}_{\text{masque temps-freq.}}} x_{fn}$$

Décomposition sans perte : $x_{fn} = \sum_k \hat{c}_{k,fn}$.

Itakura-Saito NMF : une approche générative

(Févotte, Bertin, and Durrieu, 2009)

- composantes latentes peuvent être reconstruites de façon naturelle :

$$\hat{c}_{k,fn} = \mathbb{E}\{c_{k,fn} | \mathbf{X}, \mathbf{W}, \mathbf{H}\} = \frac{w_{fk} h_{kn}}{\underbrace{\sum_j w_{fj} h_{jn}}_{\text{masque temps-freq.}}} x_{fn}$$

Décomposition sans perte : $x_{fn} = \sum_k \hat{c}_{k,fn}$.

- modèle s'écrit aussi plus simplement :

$$p(\mathbf{x}_n | \mathbf{W}\mathbf{H}_n) = \prod_f N_c(x_{fn} | 0, [\mathbf{W}\mathbf{H}]_{fn})$$

(indépendance conditionnelle des coefficients de Fourier similaire à l'approximation de Whittle pour les processus stationnaires)

Itakura-Saito NMF : une approche générative

(Févotte, Bertin, and Durrieu, 2009)

Autres propriétés

- ▶ IS-NMF est une forme de NMF avec du bruit multiplicatif :

$$v_{fn} = [\mathbf{WH}]_{fn} \cdot \epsilon_{fn}$$

où ϵ_{fn} a une distribution Gamma d'espérance 1.

Itakura-Saito NMF : une approche générative

(Févotte, Bertin, and Durrieu, 2009)

Autres propriétés

- ▶ IS-NMF est une forme de NMF avec du bruit multiplicatif :

$$v_{fn} = [\mathbf{WH}]_{fn} \cdot \epsilon_{fn}$$

où ϵ_{fn} a une distribution Gamma d'espérance 1.

- ▶ divergence IS invariante à l'échelle : $d_{\text{IS}}(\lambda x | \lambda y) = d_{\text{IS}}(x | y)$
 - ▶ induit une représentation *fine* de données à échelle de valeurs très contrastée (tels que spectres audio).
 - ▶ en comparaison, $d_{\text{EUC}}(\lambda x | \lambda y) = \lambda^2 d_{\text{EUC}}(x | y)$, $d_{\text{KL}}(\lambda x | \lambda y) = \lambda d_{\text{KL}}(x | y)$.

Exemple

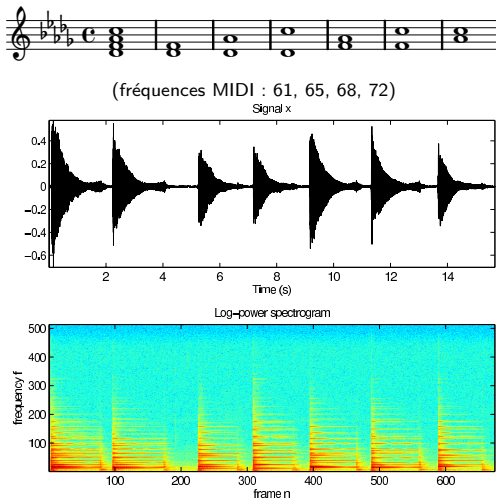
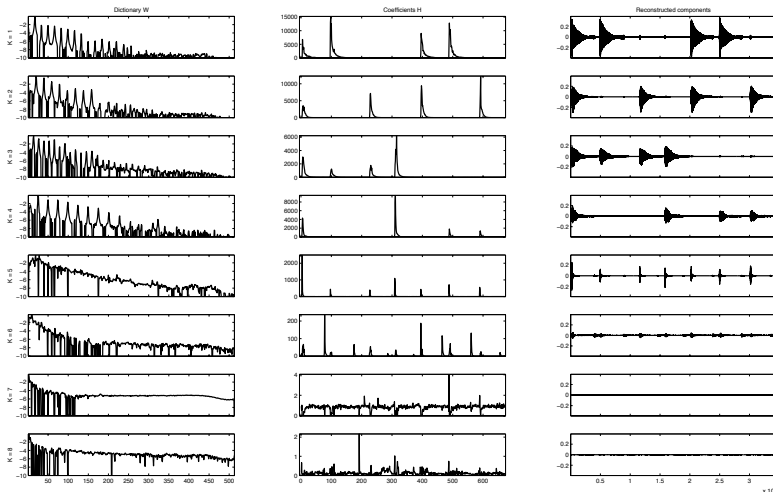


FIGURE: Trois représentations des données.

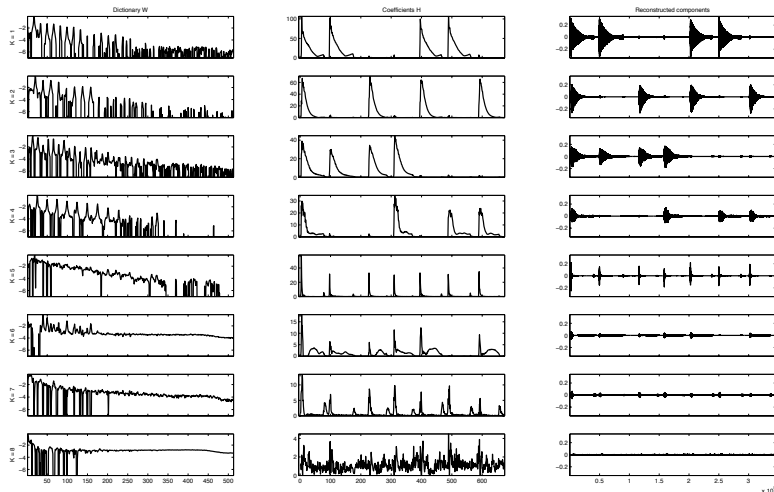
IS-NMF appliquée au spectrogramme de puissance



fréquences MIDI estimées : 65.0 68.0 61.0 72.0 0 0 0 0

(Vraies valeurs : 61, 65, 68, 72)

KL-NMF appliquée au spectrogramme de magnitude



fréquences MIDI estimées : 65.2 68.2 61.0 72.2 0 56.2 0 0
 (Vraies valeurs : 61, 65, 68, 72)

Autres contributions

- ▶ **validation exhaustive** du modèle en séparation de sources (King, Févotte, and Smaragdis, 2012).

Autres contributions

- ▶ **validation exhaustive** du modèle en séparation de sources (King, Févotte, and Smaragdis, 2012).
- ▶ **IS-NMF dynamique** (Févotte, 2011; Févotte, Le Roux, and Hershey, 2013)

$$\min_{\mathbf{W}, \mathbf{H}, \mathbf{A} \geq 0} D_{\text{IS}}(\mathbf{V} | \mathbf{W}\mathbf{H}) + \lambda \sum_n D(\mathbf{h}_n | \mathbf{A}\mathbf{h}_{n-1})$$

Autres contributions

- ▶ **validation exhaustive** du modèle en séparation de sources (King, Févotte, and Smaragdis, 2012).
- ▶ **IS-NMF dynamique** (Févotte, 2011; Févotte, Le Roux, and Hershey, 2013)

$$\min_{\mathbf{W}, \mathbf{H}, \mathbf{A} \geq 0} D_{\text{IS}}(\mathbf{V} | \mathbf{WH}) + \lambda \sum_n D(\mathbf{h}_n | \mathbf{A} \mathbf{h}_{n-1})$$

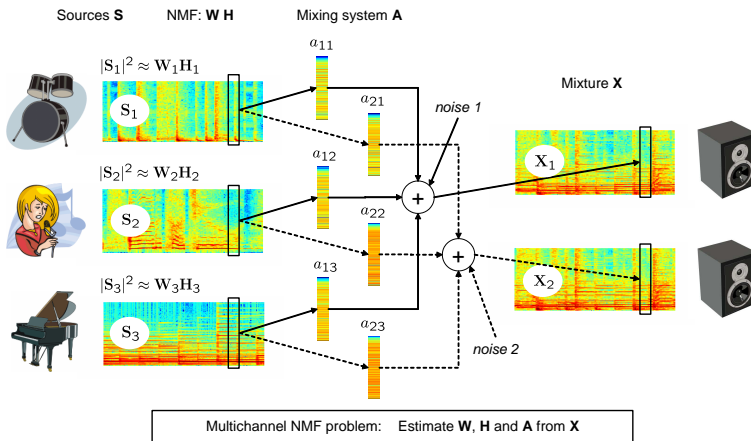
- ▶ **IS-NMF et parcimonie de groupe** (Lefèvre, Bach, and Févotte, 2011a)

$$\min_{\mathbf{W}, \mathbf{H} \geq 0} D_{\text{IS}}(\mathbf{V} | \mathbf{WH}) + \lambda \sum_n \sum_g \log(\|\mathbf{h}_{gn}\| + b)$$

- ▶ **IS-NMF incrémental** (Lefèvre, Bach, and Févotte, 2011b) : mise à jour séquentielle de \mathbf{W} .

IS-NMF multicanal

(Ozerov and Févotte, 2010)

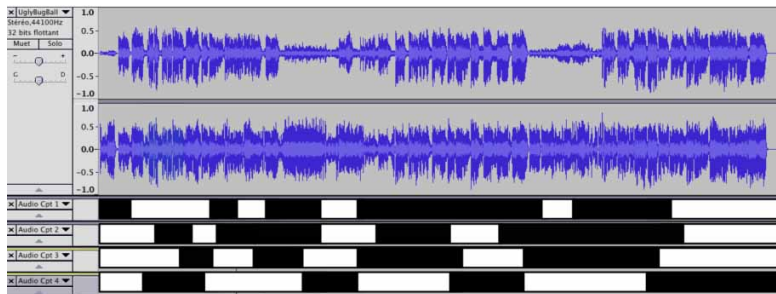


Meilleurs scores parmi 14 méthodes pour la tâche *underdetermined speech & music separation* à la campagne d'évaluation internationale (SiSEC) 2008.

IS-NMF multicanal semi-supervisée

(Ozerov, Févotte, Blouet, and Durrieu, 2011)

- ▶ décomposition “guidée” par l’opérateur : les plages d’activation de chaque source sont données au système.
- ▶ information traduite par des coefficients nuls fixes dans la matrice \mathbf{H} .



Brevet américain déposé en 2011.

Plan

Curriculum vitæ

Contributions à la factorisation en matrices non-négatives

Introduction à la NMF

Algorithmes pour la NMF fondée sur la β -divergence

Cas particulier de la NMF avec la divergence d'Itakura-Saito

Estimation par maximum de vraisemblance marginalisée

Remarques finales

NMF et maximum de vraisemblance

Mesures de dissemblance souvent assimilables à une log-vraisemblance :

$$\operatorname{argmin}_{\mathbf{W}, \mathbf{H}} D(\mathbf{V} | \mathbf{WH}) = \operatorname{argmin}_{\mathbf{W}, \mathbf{H}} -\log p(\mathbf{V} | \mathbf{WH}).$$

Dissemblance	Bruit
erreur quadratique	additif gaussien
divergence KL	Poisson
divergence IS	multiplicatif Gamma
β -divergence	Tweedie

NMF \approx estimateur de *maximum de vraisemblance jointe* de \mathbf{W} et \mathbf{H} .

⚠ Estimation mal-posée car le nombre de paramètres croît avec le nombre de données (un \mathbf{h}_n pour chaque \mathbf{v}_n).

Estimateur de maximum de vraisemblance marginale

(Dikmen and Févotte, 2011, 2012)

- ▶ traiter \mathbf{H} comme une *variable latente aléatoire* de prior $p(\mathbf{H})$.
- ▶ traiter \mathbf{W} comme une *variable déterministe* (donc sans prior).
- ▶ optimiser la vraisemblance *marginale* de \mathbf{V} et \mathbf{W} :

$$\min_{\mathbf{W} \geq 0} C(\mathbf{W}) \stackrel{\text{def}}{=} -\log p(\mathbf{V}|\mathbf{W}) = -\log \int_{\mathbf{H}} p(\mathbf{V}|\mathbf{W}\mathbf{H})p(\mathbf{H})d\mathbf{H}.$$

Nombre de paramètres fixe (estimation bien posée).

Inspiré par ICA, Latent Dirichlet Analysis (LDA), estimation statistique en présence de paramètres de nuisance.

Estimateur de maximum de vraisemblance marginale

(Dikmen and Févotte, 2011, 2012)

Algorithme EM : compléter les données \mathbf{V} avec \mathbf{H} and optimiser

$$Q(\mathbf{W}|\tilde{\mathbf{W}}) = - \int_{\mathbf{H}} \log p(\mathbf{V}, \mathbf{H}|\mathbf{W}) p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}}) d\mathbf{H}.$$

Pas de forme analytique pour $p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}})$ en général.

Estimateur de maximum de vraisemblance marginale

(Dikmen and Févotte, 2011, 2012)

Algorithme EM : compléter les données \mathbf{V} avec \mathbf{H} and optimiser

$$Q(\mathbf{W}|\tilde{\mathbf{W}}) = - \int_{\mathbf{H}} \log p(\mathbf{V}, \mathbf{H}|\mathbf{W}) p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}}) d\mathbf{H}.$$

Pas de forme analytique pour $p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}})$ en général.

Recours à

► **EM variationnel** : $q(\mathbf{H}) \approx p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}})$

$$Q^{\text{VB}}(\mathbf{W}|\tilde{\mathbf{W}}) = - \int_{\mathbf{H}} \log p(\mathbf{V}, \mathbf{H}|\mathbf{W}) q(\mathbf{H}) d\mathbf{H}$$

Estimateur de maximum de vraisemblance marginale

(Dikmen and Févotte, 2011, 2012)

Algorithme EM : compléter les données \mathbf{V} avec \mathbf{H} and optimiser

$$Q(\mathbf{W}|\tilde{\mathbf{W}}) = - \int_{\mathbf{H}} \log p(\mathbf{V}, \mathbf{H}|\mathbf{W}) p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}}) d\mathbf{H}.$$

Pas de forme analytique pour $p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}})$ en général.

Recours à

► **EM variationnel** : $q(\mathbf{H}) \approx p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}})$

$$Q^{\text{VB}}(\mathbf{W}|\tilde{\mathbf{W}}) = - \int_{\mathbf{H}} \log p(\mathbf{V}, \mathbf{H}|\mathbf{W}) q(\mathbf{H}) d\mathbf{H}$$

► **Monte-Carlo EM** : $\mathbf{H}^{(i)} \sim p(\mathbf{H}|\mathbf{V}, \tilde{\mathbf{W}})$

$$Q^{\text{MC}}(\mathbf{W}|\tilde{\mathbf{W}}) = - \sum_i \log p(\mathbf{V}, \mathbf{H}^{(i)}|\mathbf{W})$$

Expériences

Modèles considérés

	Gamma-Poisson	Gamma-Exponentiel
$p(\mathbf{V} \mathbf{WH})$	$\prod_{fn} \text{Pois}(v_{fn} [\mathbf{WH}]_{fn})$ (bruit Poisson)	$\prod_{fn} \text{Exp}(v_{fn} [\mathbf{WH}]_{fn})$ (bruit multiplicatif exponentiel)
$p(\mathbf{H} \beta)$	$\prod_{kn} G(h_{kn} \alpha_k, \beta_k)$	

Expériences

Modèles considérés

	Gamma-Poisson	Gamma-Exponentiel
$p(\mathbf{V} \mathbf{WH})$	$\prod_{fn} \text{Pois}(v_{fn} [\mathbf{WH}]_{fn})$ (bruit Poisson)	$\prod_{fn} \text{Exp}(v_{fn} [\mathbf{WH}]_{fn})$ (bruit multiplicatif exponentiel)
$p(\mathbf{H} \beta)$	$\prod_{kn} G(h_{kn} \alpha_k, \beta_k)$	

Estimateurs comparés

- ▶ Estimateur de maximum de vraisemblance **marginale** (MMLE)

$$C_{\text{ML}}(\mathbf{W}, \beta) = -\log p(\mathbf{V}|\mathbf{W}, \beta)$$

Optimisation avec EM variationnel ou MC-EM.

- ▶ Estimateur de maximum de vraisemblance **jointe** (MJLE)

$$C_{\text{JL}}(\mathbf{W}, \mathbf{H}, \beta) = -\log p(\mathbf{V}, \mathbf{H}|\mathbf{W}, \beta)$$

Équivalent à NMF pénalisée. Optimisation MM.

Données swimmer

Modèle Gamma-Exponentiel

Images synthétiques corrompues par un bruit multiplicatif.

(a) échantillons purs

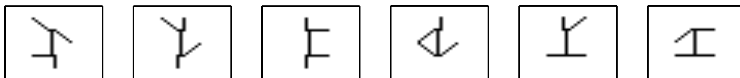


Données swimmer

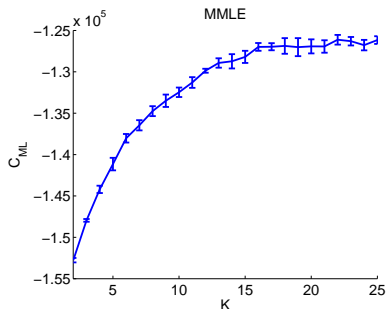
Modèle Gamma-Exponentiel

Images synthétiques corrompues par un bruit multiplicatif.

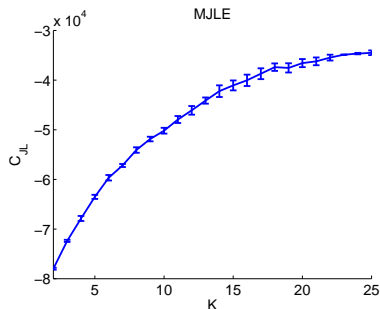
(a) échantillons purs



(b) log-vraisemblances marginales

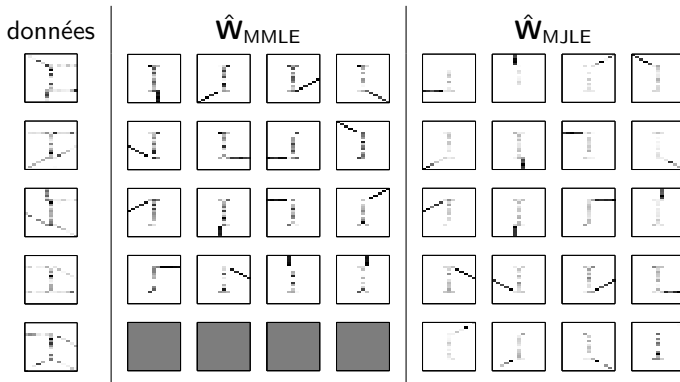


(c) log-vraisemblances jointes



Données swimmer

Modèle Gamma-Exponentiel



MMLE retourne 4 colonnes nulles dans \hat{W} (auto-régularisation de l'ordre).
MJLE sur-apprend.

Données musixmatch

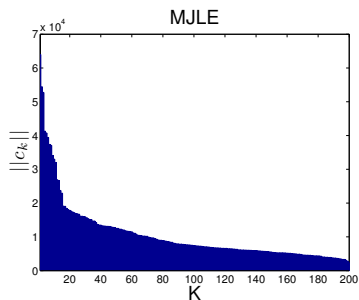
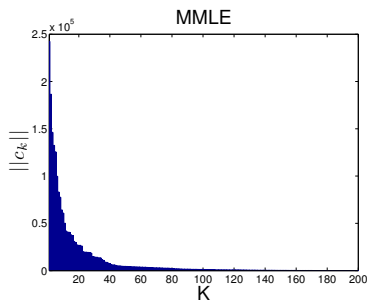
Modèle Gamma-Poisson

- ▶ Paroles de 10.000 chansons (texte).
- ▶ Représentation “sac de mots” de chaque chanson (vocabulaire constitué des 5,000 mots plus fréquents).
- ▶ Analyse sémantique avec MMLE et MJLE.
- ▶ $K = 200$.
- ▶ Reconstruction des composantes sémantiques, tel que $\mathbf{V} = \sum_k \hat{\mathbf{C}}_k$.

Données musixmatch

Modèle Gamma-Poisson

Normes $\|\hat{\mathbf{C}}_k\|$ des composantes reconstruites avec les deux estimateurs.



MMLE retourne 50 composantes nulles (auto-régularisation de l'ordre).

Données musixmatch

Modèle Gamma-Poisson

4 motifs sémantiques extraits par MMLE et leurs 5 meilleurs représentants.

(a) hip-hop

$(k = 2)$ get nigga the ya shit like fuck em got hit bitch up off yall ass they that cmon money and	
UGK (Underground Kingz) - Murder	i the to nigga my a you got murder and it is am from we so with they yo cuz
Big Punisher - Nigga Shit	shit that nigga the i and my what to out am in on for love me with gettin you do
E-40 - Turf Drop [Clean]	gasolin the my i hey to a it on you some fuck spit of what one ride nigga sick gold
Cam'Ron - Sports Drugs & Entertainment	a the you i got yo stop shot is caus or street jump short wick either to on but in
Foxy Brown - Chyna Whyte	the nigga and you shit i not yall to a on with bitch no fuck uh it money white huh

(b) métal

$(k = 8)$ god of blood soul death die fear pain hell power within shall earth blind human bleed scream evil holi peac	
Demolition Hammer - Epidemic Of Violence	of pain death reign violenc and a kill rage vicious the to in down blue dead cold
Disgorge - Parallels Of Infinite Torture	of the tortur by their within upon flow throne infinit are no they see life eye befor
Tacere - Beyond Silence	silenc beyond a dark beauti i the you to and me it not in my is of your that do
Cannibal Corpse - Perverse Suffering	to my pain of i me for agoni in by and from way etern lust tortur crave the not be
Showbread - Sampsas Meets Kafka	to of no one die death loneli starv i the you and a me it not in my is your

(c) filles

$(k = 26)$ she her girl beauti woman & queen sex sexi cloth herself doll shes pink gypsi bodi midnight callin dress hair	
Headhunter - Sex & Drugs & Rock'N Roll	& sex drug rock roll n is good veri inde and not my are all need dead bodi brain i
Holy Barbarians - She	she of kind girl my is the a littl woman like world and gone destroy tiger me on an
X - Devil Doll	devil doll her she and a the in is of eye bone & shoe rag batter you to on no
Kittie - Paperdoll	her she you i now soul pain to is down want eat fit size and not in all dead bodi
Ottawan - D.I.S.C.O.	is she oh disco i o s d c super incred a crazi such desir sexi complice special candi

(d) français

$(k = 13)$ je et les le pas dan pour des cest qui de tout mon moi au comm ne sur jai	
Veronique Sanson - Féminin	cest comm le car de bien se les mai a fait devant heur du et une quon quelqu etre
Nevrotic Explosion - Heritage	quon faut mieux pour nous qui nos ceux de la un plus tous honor parent ami oui
Kells - Sans teint	de la se le san des est loin peur reve pour sa sang corp lumier larm
Stille Volk - Corps Magicien	de les ell dan la se le du pass est sa par mond leur corp vivr lair voyag feu
Florent Pagny - Tue-Moi	si plus que un tu mon mes jour souvenir parc

Plan

Curriculum vitæ

Contributions à la factorisation en matrices non-négatives

Introduction à la NMF

Algorithmes pour la NMF fondée sur la β -divergence

Cas particulier de la NMF avec la divergence d'Itakura-Saito

Estimation par maximum de vraisemblance marginalisée

Remarques finales

Problèmes ouverts

- ▶ **vers des algorithmes à convergence globale**
 - ▶ convexifications
 - ▶ mises à jour conjointes des facteurs
 - ▶ mariage de l'optimisation déterministe et de la simulation stochastique
- ▶ **algorithmes séquentiels et distribués**
 - ▶ big data compatible
- ▶ **identifiabilité de la NMF**
 - ▶ résultats existants basés sur des hypothèses "pure pixel"
 - ▶ pas d'études dans le cas pénalisé
- ▶ **théorie de l'estimation par maximum de vraisemblance marginalisée**
 - ▶ analyse de l'effet d'auto-régularisation de l'ordre
 - ▶ analyse des propriétés de $C_{ML}(\mathbf{W})$
- ▶ **vers de nouvelles représentations**
 - ▶ low-rank time-frequency synthesis (Févotte and Kowalski, 2014)
 - ▶ low-rank inducing transforms (travail en cours avec P. Smaragdis)

References I

- A. Basu, I. R. Harris, N. L. Hjort, and M. C. Jones. Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3) :549–559, Sep. 1998.
- M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons. Algorithms and applications for approximate nonnegative matrix factorization. *Computational Statistics & Data Analysis*, 52(1) :155–173, Sep. 2007.
- J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2) :354–379, 2012.
- D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3 :993–1022, Jan. 2003.
- Y. Cao, P. P. B. Eggermont, and S. Terebey. Cross Burg entropy maximization and its application to ringing suppression in image reconstruction. *IEEE Transactions on Image Processing*, 8(2) :286–292, Feb. 1999. doi : 10.1109/83.743861.
- A. Cichocki and S. Amari. Families of Alpha- Beta- and Gamma- divergences : Flexible and robust measures of similarities. *Entropy*, 12(6) :1532–1568, June 2010.
- A. Cichocki, R. Zdunek, and S. Amari. Csiszar's divergences for non-negative matrix factorization : Family of new algorithms. In *Proc. International Conference on Independent Component Analysis and Blind Signal Separation (ICA)*, pages 32–39, Charleston SC, USA, Mar. 2006.
- M. Daube-Witherspoon and G. Muehllehner. An iterative image space reconstruction algorithm suitable for volume ECT. *IEEE Transactions on Medical Imaging*, 5(5) :61 – 66, 1986. doi : 10.1109/TMI.1986.4307748.

References II

- A. R. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Trans. Medical Imaging*, 12(2) :328–333, 1993. doi : 10.1109/42.232263.
- O. Dikmen and C. Févotte. Nonnegative dictionary learning in the exponential noise model for adaptive music signal representation. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2267–2275, Granada, Spain, Dec. 2011. URL <http://www.unice.fr/cfevotte/publications/proceedings/nips11.pdf>.
- O. Dikmen and C. Févotte. Maximum marginal likelihood estimation for nonnegative dictionary learning in the Gamma-Poisson model. *IEEE Transactions on Signal Processing*, 60(10) :5163–5175, Oct. 2012. doi : <http://dx.doi.org/10.1109/TSP.2012.2207117>. URL http://www.unice.fr/cfevotte/publications/journals/ieee_sp_mmle.pdf.
- C. Févotte. Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011. URL <http://www.unice.fr/cfevotte/publications/proceedings/icassp11a.pdf>.
- C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the beta-divergence. *Neural Computation*, 23(9) :2421–2456, Sep. 2011. doi : 10.1162/NECO_a.00168. URL <http://www.unice.fr/cfevotte/publications/journals/neco11.pdf>.
- C. Févotte and M. Kowalski. Low-rank time-frequency synthesis. In *Advances in Neural Information Processing Systems (NIPS)*, 2014.
- C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis. *Neural Computation*, 21(3) :793–830, Mar. 2009. doi : 10.1162/neco.2008.04-08-771. URL http://www.unice.fr/cfevotte/publications/journals/neco09_is-nmf.pdf.

References III

- C. Févotte, J. Le Roux, and J. R. Hershey. Non-negative dynamical system with application to speech and audio. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013. URL <http://www.unice.fr/cfevotte/publications/proceedings/icassp13a.pdf>.
- T. Hofmann. Probabilistic latent semantic indexing. In *Proc. 22nd International Conference on Research and Development in Information Retrieval (SIGIR)*, 1999. URL <http://www.cs.brown.edu/~th/papers/Hofmann-SIGIR99.pdf>.
- B. King, C. Févotte, and P. Smaragdis. Optimal cost function and magnitude power for NMF-based speech separation and music interpolation. In *Proc. IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Santander, Spain, Sep. 2012. URL <http://www.unice.fr/cfevotte/publications/proceedings/mlsp12.pdf>.
- R. Kompass. A generalized divergence measure for nonnegative matrix factorization. *Neural Computation*, 19(3) :780–791, 2007.
- D. D. Lee and H. S. Seung. Learning the parts of objects with nonnegative matrix factorization. *Nature*, 401 :788–791, 1999.
- A. Lefèvre, F. Bach, and C. Févotte. Itakura-Saito nonnegative matrix factorization with group sparsity. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011a. URL <http://www.unice.fr/cfevotte/publications/proceedings/icassp11c.pdf>.
- A. Lefèvre, F. Bach, and C. Févotte. Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Mohonk, NY, Oct. 2011b. URL <http://www.unice.fr/cfevotte/publications/proceedings/waspaa11.pdf>.

References IV

- L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astronomical Journal*, 79 :745–754, 1974. doi : 10.1086/111605.
- A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Transactions on Audio, Speech and Language Processing*, 18(3) : 550–563, Mar. 2010. doi : 10.1109/TASL.2009.2031510. URL http://www.unice.fr/cfevotte/publications/journals/ieee_asl_multinmf.pdf.
- A. Ozerov, C. Févotte, R. Blouet, and J.-L. Durrieu. Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011. URL <http://www.unice.fr/cfevotte/publications/proceedings/icassp11d.pdf>.
- P. Paatero and U. Tapper. Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5 :111–126, 1994.
- W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62 :55–59, 1972.
- N. Seichepine, S. Essid, C. Févotte, and O. Cappé. Soft nonnegative matrix co-factorization. *IEEE Transactions on Signal Processing*, 2014.
- P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03)*, Oct. 2003.
- V. Y. F. Tan and C. Févotte. Automatic relevance determination in nonnegative matrix factorization with the beta-divergence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7) : 1592 – 1605, July 2013. URL http://www.unice.fr/cfevotte/publications/journals/pami13_ardnmf.pdf.