



**HAL**  
open science

# Séparation de sources non-négatives : application au traitement des signaux de spectroscopie

Saïd Moussaoui

► **To cite this version:**

Saïd Moussaoui. Séparation de sources non-négatives : application au traitement des signaux de spectroscopie. Autre. Université Henri Poincaré - Nancy 1, 2005. Français. NNT : 2005NAN10164 . tel-01748096

**HAL Id: tel-01748096**

**<https://hal.univ-lorraine.fr/tel-01748096v1>**

Submitted on 29 Mar 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

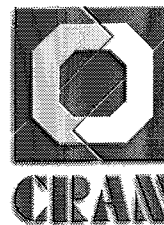
## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



UFR Sciences et Techniques Mathématiques Informatique Automatique  
École Doctorale IAEM Lorraine  
DFD Automatique et Production Automatisée

## THÈSE

présentée pour l'obtention du

Doctorat de l'Université Henri Poincaré, Nancy 1  
(Spécialité Automatique, Traitement du Signal et Génie Informatique)

par

Saïd MOUSSAOUI

Séparation de sources non-négatives.  
Application au traitement des signaux de spectroscopie

Soutenue publiquement le 07 décembre 2005

### Composition du jury

<i>Rapporteurs :</i>	P.-O. AMBLARD	Chargé de Recherche au CNRS, HdR, I.N.P. de Grenoble
	J.-Y. TOURNERET	Professeur à l'Institut National Polytechnique de Toulouse
<i>Examineurs :</i>	C. JUTTEN	Professeur à l'Université Joseph Fourier, Grenoble
	A. MOHAMMAD-DJAFARI	Directeur de Recherche au CNRS, Université Paris-Sud, Orsay
	D. BRIE	Professeur à l'Université Henri Poincaré, CRAN, Nancy
	C. CARTERET	Maître de Conférences à l'Université Henri Poincaré, LCPME, Nancy
<i>Invité :</i>	B. HUMBERT	Professeur à l'Université Henri Poincaré, LCPME, Nancy

*A mes chers parents, Zohra et Ali*

*A ma soeur, Ghalia*

*A mes frères, Djamel, Farid, Mustapha, Hamid et Kaci.*

*A Nassira.*

# Remerciements

Je tiens tout d'abord à exprimer ma gratitude à tous les membres du jury qui m'on fait l'honneur de participer à l'examen de ce travail.

Je remercie M. **Jean-Yves Tournet**, Professeur à l'Institut Polytechnique de Toulouse, et M. **Pierre-Olivier Amblard**, Chargé de Recherche au CNRS, Institut National Polytechnique de Grenoble, pour avoir accepté de rapporter sur ce travail et d'avoir consacré une partie de leur temps pour examiner minutieusement ce manuscrit. Je les remercie vivement pour tous leurs commentaires et suggestions.

Je remercie M. **Christian Jutten**, Professeur à l'Université Joseph Fourier, Grenoble, pour avoir accepté d'apporter son regard critique sur ce travail.

Je remercie M. **Ali Mohammad-Djafari**, Directeur de recherche au CNRS, Université Paris-Sud, d'avoir participé à ce jury et apporté son regard bayésien sur ce travail. Je tiens également à lui exprimer ma reconnaissance de m'avoir accueilli au sein de son équipe, au Groupe Problèmes Inverses du Laboratoire des Signaux et des Systèmes, durant le mois de mai 2003. J'ai particulièrement apprécié nos échanges et notamment avec ses anciens doctorants : M. **Hisham Snoussi** et M. **Mahieddine Ichir**.

Je tiens à exprimer ma gratitude à M. **Bernard Humbert**, Professeur à l'Université Henri Poincaré, Nancy 1, au sein du Laboratoire de Chimie, Physique et Microbiologie pour l'Environnement, et M. **David Brie**, Professeur à l'Université Henri Poincaré au sein du Centre de Recherche en Automatique de Nancy, d'avoir permis à cette thèse d'exister. Il faut dire que leur rencontre au rayon "Cocktails" du supermarché MATCH a fortement inspiré le lancement d'une thèse sur l'analyse de mélanges.

Tout au long de cette thèse j'ai eu la chance de bénéficier d'un excellent co-encadrement de la part de M. **David Brie** et M. **Cédric Carteret**, Maître de Conférences à l'Université Henri Poincaré au sein du Laboratoire de Physique, Chimie et Microbiologie pour l'Environnement. Merci à **Cédric** d'avoir pris le temps et la patience nécessaires pour m'expliquer des choses qui ne sont pas évidentes pour un non spécialiste de la spectroscopie. Sa contribution à l'analyse des signaux réels, à la mise au point des simulations et à la rédaction du dernier chapitre me fût si précieuse. Merci à **David** pour son suivi constant, ses encouragements, sa rigueur scientifique et sa présence incontestable dans tous les moments, aussi bien joyeux que difficiles, et enfin, mais surtout, pour son amitié.

Je tiens à remercier vivement M. **Olivier Caspary**, Maître de Conférences à l'Université Henri Poincaré au sein de l'I.U.T de Saint-Dié, d'avoir commencé l'aventure de la thèse avec moi et qui a mis à ma disposition toute la documentation nécessaire.

Je tiens à exprimer ma gratitude M. **Jérôme Idier**, Chargé de Recherches au CNRS, IRCCyN Nantes, pour ses suggestions et longues discussions. Au cours de nos échanges, j'ai été particulièrement marqué par son intégrité scientifique et son esprit d'analyse critique.

Je tiens à remercier M. **Alain Richard**, Professeur à l'Université Henri Poincaré, Nancy 1, pour m'avoir offert la possibilité d'effectuer mon DEA et ma thèse au sein du CRAN et de m'avoir constamment encouragé à poursuivre mes travaux de recherche.

Merci à **El-Hadi Djermoune** et **Magalie Thomassin** d'avoir partagé mon bureau pendant trois ans et pour la bonne humeur quotidienne. Merci, non seulement pour toutes les discussions scientifiques, et autres, que nous avons eues, mais également pour les longues soirées de jeux de cartes.

Merci à **Frédéric Hamelin**, **Cédric Join**, **Hisham Jamouli** et **Eric Huselstein** pour tous les instants de délire et l'ambiance quotidienne remarquable dans la salle informatique. Merci également à **Olivier Cervelin** (dit Zitoune), notre ingénieur informaticien, pour sa disponibilité infaillible. Merci à **Sabine Huraux**, notre secrétaire dévouée, pour son assistance administrative irréprochable.

Merci à **Vincent** pour tous les moments de travail que nous avons partagés et de m'avoir soutenu durant les moments difficiles. Vince, merci de m'avoir ramené chez moi après toutes les longues soirées, même si mes conseils de conduite ne sont pas toujours à ton goût. Merci à **Mylène** pour sa gentillesse.

Merci à **Ludovic** pour sa générosité, ses éclats de rire remarquables et aussi pour ses coups de gueule époustouflants. Ludo, tu es invité à une soirée spéciale pavé de rumsteck.

Merci à **Samir** et **Mickaël** pour leur amitié et pour tous les moments de joie que nous avons partagés, notamment, à l'ENP d'Alger avec Sam et sur le terrain de foot avec Mike. Merci également à **Adeline** et à **Emilie** pour leur aimable sagesse.

Merci à **Philippe** pour sa gentillesse et ses histoires délirantes. Philo, j'aurais bientôt besoin de tes conseils en gestion de budget et construction de maison. Merci à **Fabienne** pour sa sympathie et pour les soirées de belote.

Merci à nos syndicalistes dévoués **Yann** et **Stéphane**. Que serait la situation des doctorants au CRAN sans votre engagement ? Je vous souhaite du courage pour cette fin de thèse. Merci également à **Jean-Pierre** pour ses conseils de séduction et à **Cyril** pour ses mails réjouissants.

Merci aux autres membres du laboratoire, notamment **Jean-Marc**, le stagiaire, et les doctorants **Fateh**, **Sinuhé**, **Bryan**, **Sébastien**, **Bryan** et **Kamel**, courage les gars !, et les permanents **Marion**, **Hugues**, **Thierry**, **Jean-Marie**, **Marc**, **Patrick**, **Taha**, **Didier**, **Christophe**, **Jean-Christophe**, **Dominique**, **Charles** et **Joseph** pour les pauses café et les moments de travail partagés dans la bonne humeur.

Enfin, un grand merci à toutes celles et ceux, qui de près ou de loin, ont contribué au bon déroulement de ces trois années de thèse.

# Résumé

Lors de l'analyse physico-chimique d'une substance multi-composantes par des techniques spectroscopiques, les signaux mesurés sont des mélanges linéaires des spectres purs de différentes composantes de cette substance. L'objectif de l'analyse est d'identifier la composition de la substance via l'estimation des spectres purs et la détermination de leurs concentrations. C'est un problème de séparation de sources dans lequel les spectres purs sont les signaux sources et les coefficients de mélange permettent de déduire les proportions des différentes composantes dans la substance. Dans le cas de données spectroscopiques, la contrainte principale est la non-négativité des signaux sources et des coefficients de mélange. Il s'agit donc d'un problème de séparation de sources non-négatives.

La séparation de sources est un problème fondamental en traitement du signal dont une hypothèse forte est celle de l'indépendance statistique des signaux sources. Compte tenu du recouvrement entre les spectres purs, leur corrélation mutuelle devient parfois importante. Dans une telle situation, l'application d'une méthode fondée sur l'hypothèse d'orthogonalité s'avère infructueuse. Par ailleurs, une analyse des solutions admissibles sous la contrainte de non-négativité montre que cette contrainte toute seule ne permet d'obtenir une solution unique que dans certains cas particuliers. Ces deux constats motivent le développement de deux méthodes qui considèrent conjointement l'hypothèse d'indépendance et l'information de non-négativité. La première méthode est issue de l'approche de séparation par maximum de vraisemblance et la deuxième se fonde sur une inférence bayésienne. Une évaluation numérique des performances des méthodes développées à l'aide de données synthétiques et le traitement de signaux expérimentaux permettent, d'une part, de mettre en évidence les avantages de ces méthodes par rapport aux approches usuelles et, d'autre part, de déceler leurs limitations. Des applications au traitement de signaux réels issus de trois types de spectroscopies (Infrarouge, Raman et Ultraviolet-Visible) illustrent l'apport de la séparation de sources non-négatives à l'analyse physico-chimique.





# Table des matières

Résumé	v
Notations	xi
Abréviations	xiii
Introduction générale	1
<b>I Position des problèmes de la séparation de sources non-négatives</b>	<b>5</b>
<b>1 Séparation de sources spectrales : position des problèmes</b>	<b>7</b>
1.1 Introduction à la séparation de sources . . . . .	7
1.1.1 Modèle du mélange . . . . .	8
1.1.2 Principe de la séparation de sources . . . . .	9
1.2 Séparation par analyse en composantes indépendantes . . . . .	10
1.2.1 Mesures de l'indépendance statistique . . . . .	10
1.2.2 Présentation de quelques méthodes de séparation . . . . .	13
1.2.3 Séparation de sources spectrales par analyse en composantes indépendantes . . . . .	17
1.2.4 Quelques approches proposées pour améliorer les performances . . . . .	18
1.3 Séparation par prise en compte de la non-négativité . . . . .	22
1.3.1 Approche algébrique . . . . .	22
1.3.2 Séparation par moindres carrés sous contrainte de non-négativité . . . . .	23
1.3.3 Séparation par moindres carrés pénalisés . . . . .	24
1.4 Conclusions . . . . .	26
<b>2 Solutions admissibles et conditions d'unicité de la solution</b>	<b>27</b>
2.1 Introduction . . . . .	27
2.2 Formulation du problème . . . . .	28
2.2.1 Indétermination d'échelle . . . . .	29
2.2.2 Indétermination d'ordre . . . . .	29
2.2.3 Définitions . . . . .	29
2.3 Solutions admissibles et conditions d'unicité dans le cas de deux sources . . . . .	29

2.3.1	Solutions admissibles . . . . .	30
2.3.2	Conditions d'unicité de la solution . . . . .	31
2.3.3	Illustration des solutions admissibles . . . . .	32
2.4	Solutions admissibles et conditions d'unicité dans la cas de plus de deux sources . . . . .	32
2.4.1	Conditions nécessaires pour l'unicité . . . . .	32
2.4.2	Conditions suffisantes pour l'unicité . . . . .	35
2.4.3	Formulation géométrique . . . . .	36
2.4.4	Détermination des solutions admissibles . . . . .	38
2.5	Séparation de sources non-négatives : quel algorithme ? . . . . .	40
2.5.1	Coefficients non-négatifs et sources non-négatives et indépendantes . . . . .	41
2.5.2	Coefficients non-négatifs et sources indépendantes et strictement positives . . . . .	41
2.5.3	Coefficients non-négatifs et sources non-négatives et corrélées . . . . .	42
2.5.4	Coefficients strictement positifs et sources strictement positives et corrélées . . . . .	42
2.6	Conclusion . . . . .	43
<b>II Développement d'algorithmes pour la séparation de sources spectrales</b>		<b>45</b>
<b>3 Séparation de sources spectrales par maximum de vraisemblance</b>		<b>47</b>
3.1	Introduction . . . . .	47
3.2	Séparation de sources par maximum de vraisemblance . . . . .	48
3.2.1	Formulation de la vraisemblance . . . . .	48
3.2.2	Optimisation en utilisant l'algorithme du gradient . . . . .	50
3.2.3	Optimisation en utilisant un algorithme EM . . . . .	51
3.2.4	Choix de la densité de probabilité des sources . . . . .	54
3.3	Séparation de sources spectrales par l'approche du maximum de vraisemblance . . . . .	56
3.3.1	Modélisation de la distribution des sources . . . . .	56
3.3.2	Séparation en utilisant l'algorithme du gradient . . . . .	57
3.3.3	Séparation en utilisant l'algorithme MCEM . . . . .	60
3.4	Illustration des algorithmes sur des données synthétiques . . . . .	62
3.4.1	Séparation en utilisant l'algorithme du gradient relatif . . . . .	64
3.4.2	Séparation en utilisant l'algorithme MCEM . . . . .	65
3.5	Conclusion . . . . .	66
<b>4 Séparation de sources spectrales par approche bayésienne</b>		<b>67</b>
4.1	Introduction . . . . .	67
4.2	Approche bayésienne pour la séparation de sources . . . . .	68
4.2.1	Principe de la séparation bayésienne . . . . .	68
4.2.2	Techniques d'estimation bayésienne pour la séparation de sources . . . . .	69
4.3	Séparation de sources non-négatives par approche bayésienne . . . . .	70
4.3.1	Modélisation probabiliste . . . . .	70

4.3.2	Interprétation du critère et liens avec les méthodes existantes . . . . .	72
4.3.3	Estimation par simulation MCMC . . . . .	74
4.3.4	Densités conditionnelles <i>a posteriori</i> . . . . .	77
4.4	Illustration de l'algorithme . . . . .	80
4.4.1	Analyse des chaînes de Markov . . . . .	81
4.4.2	Temps de calcul . . . . .	83
4.5	Conclusions . . . . .	84
<b>5</b>	<b>Analyse des performances et comparaisons</b>	<b>85</b>
5.1	Mise au point des simulations . . . . .	85
5.1.1	Simulation des spectres purs . . . . .	85
5.1.2	Simulation des coefficients de mélange . . . . .	86
5.1.3	Synthèse des observations . . . . .	87
5.1.4	Indices de performances . . . . .	87
5.2	Comparaisons avec des méthodes usuelles . . . . .	88
5.2.1	Cas d'un faible niveau de bruit . . . . .	88
5.2.2	Cas d'un niveau de bruit non-négligeable . . . . .	93
5.3	Analyse des performances . . . . .	93
5.3.1	Influence de la corrélation mutuelle des sources . . . . .	93
5.3.2	Influence du niveau de bruit . . . . .	94
5.3.3	Influence de la présence d'un arrière-plan . . . . .	95
5.4	Conclusions . . . . .	96
<b>III</b>	<b>Application à des données expérimentales</b>	<b>99</b>
<b>6</b>	<b>Application au traitement de signaux spectroscopiques</b>	<b>101</b>
6.1	Introduction à la spectroscopie . . . . .	101
6.1.1	Rayonnement électromagnétique et types de spectroscopie . . . . .	101
6.1.2	Interprétation qualitative des spectres . . . . .	102
6.1.3	Interprétation quantitative des spectres . . . . .	103
6.1.4	Application de la séparation de sources à l'analyse de mélanges . . . . .	105
6.2	Application à des données expérimentales . . . . .	105
6.2.1	Signaux de spectroscopie Raman . . . . .	105
6.2.2	Signaux de spectroscopie Infrarouge . . . . .	107
6.3	Application à des problèmes d'analyse physico-chimique . . . . .	109
6.3.1	Méthodologie de traitement . . . . .	110
6.3.2	Analyse de l'hydratation de matériaux poreux et/ou lamellaires . . . . .	111
6.3.3	Analyse des polymorphes du carbonate de calcium . . . . .	114
6.3.4	Analyse des complexes cations métalliques - biomolécules . . . . .	118
6.4	Conclusion . . . . .	120

Conclusion générale	121
IV Annexes	125
A Calcul de la densité conditionnelle <i>a posteriori</i> des signaux sources	127
B Simulation d'une densité normale-gamma	129
C Simulation des paramètres d'une densité gamma	131
Bibliographie	146

# Notations

## Notations générales :

- $\mathbb{R}$  : corps des réels.
- $\mathbb{C}$  : corps des complexes.
- $s$  : les scalaires sont notés par des minuscules.
- $\mathbf{v}$  : les vecteurs sont notés par des minuscules en gras.
- $\mathbf{v}^T$  : transposé du vecteur  $\mathbf{v}$ .
- $\mathbf{M}$  : les matrices sont notées par des majuscules en gras.
- $\mathbf{M}^T$  : transposée de la matrice  $\mathbf{M}$ .
- $\mathbf{M}^*$  : transposée conjuguée de la matrice  $\mathbf{M}$ .
- $\mathbf{M}^{-1}$  : matrice inverse de la matrice carrée  $\mathbf{M}$ .
- $\mathbf{M}^\#$  : inverse généralisée de la matrice  $\mathbf{M}$ .
- $\det \mathbf{M}$  : déterminant de la matrice  $\mathbf{M}$ .
- $\text{Tr}(\mathbf{M})$  : trace de la matrice  $\mathbf{M}$ .
- $\text{Diag}(\mathbf{v})$  : matrice diagonale dont la diagonale est le vecteur  $\mathbf{v}$ .
- $\mathbf{m}_k$  :  $k^{\text{ème}}$  vecteur colonne de la matrice  $\mathbf{M}$ .
- $m_{ij}$  :  $(ij)^{\text{ème}}$  élément de la matrice  $\mathbf{M}$ .
- $s_{(i,k)}$  :  $k^{\text{ème}}$  échantillon du  $i^{\text{ème}}$  signal  $s_i$ .
- $\mathbf{I}_n$  : matrice identité de dimension  $(n \times n)$ .
- $f(x) \propto g(x)$  :  $f(x)$  est proportionnelle à  $g(x)$ .
- $f'(x)$  : dérivée première de  $f(x)$  par rapport à  $x$ .
- $\nabla_\theta f(\theta)$  : gradient de la fonction  $f(\theta)$  par rapport à  $\theta$ .
- $L_\alpha$  : norme  $\alpha$ .
- $p(\cdot)$  : densité de probabilité.
- $x \sim p(x)$  :  $x$  suit une loi de probabilité  $p(x)$ .
- $x|y$  :  $x$  conditionnelement à  $y$ .
- $\mathbb{E}[\cdot]$  : espérance mathématique.
- $\text{Cum}[\cdot, \dots, \cdot]$  : cumulants.
- $\Gamma(\cdot)$  : fonction gamma.
- $\gamma(\cdot, \cdot)$  : fonction gamma incomplète.
- $\psi(\cdot)$  : fonction digamma.
- $\mathbb{I}_{\mathbb{K}}(x)$  : fonction indicatrice, qui vaut 1 pour  $x \in \mathbb{K}$ .

**Densités de probabilité :**

- $\mathcal{U}(a, b)$  : loi uniforme sur l'intervalle  $[a, b]$ .  
 $\mathcal{N}(\mu, \sigma^2)$  : loi normale de moyenne  $\mu$  et variance  $\sigma^2$ .  
 $\mathcal{N}_+(\mu, \sigma^2)$  : loi normale à support positif, de moyenne  $\mu$  et variance  $\sigma^2$ .  
 $\mathcal{G}(\alpha, \beta)$  : loi gamma de paramètre de forme  $\alpha$  et d'échelle  $\beta$ .

**Modèles de mélange et de séparation :**

- $m$  : nombre de mesures.  
 $n$  : nombre d'échantillons.  
 $p$  : nombre de sources.  
 $\mathbf{x}_k$  : vecteur de taille  $(m \times 1)$  contenant les  $m$  échantillons des observations à l'instant  $k$ .  
 $\mathbf{s}_k$  : vecteur de taille  $(p \times 1)$  contenant les  $p$  échantillons des sources à l'instant  $k$ .  
 $\mathbf{X}$  : matrice des observations, de dimension  $(m \times n)$ .  
 $\mathbf{A}$  : matrice de mélange, de dimension  $(m \times p)$ .  
 $\mathbf{S}$  : matrice des signaux sources, de dimension  $(p \times n)$ .  
 $\mathbf{E}$  : matrice de bruit sur les observations, de dimension  $(m \times n)$ .  
 $\mathbf{R}_x$  : matrice d'autocovariance des observations, de dimension  $(m \times m)$ .  
 $\mathbf{R}_s$  : matrice d'autocovariance des sources, de dimension  $(p \times p)$ .  
 $\mathbf{R}_e$  : matrice d'autocovariance du bruit, de dimension  $(m \times m)$ .  
 $\mathbf{W}$  : matrice de blanchiment, de dimension  $(p \times m)$ .  
 $\mathbf{B}$  : matrice de séparation, de dimension  $(p \times m)$ .  
 $\mathbf{T}$  : matrice de transformation, de dimension  $(p \times p)$ .  
 $\mathbf{U}$  : matrice de transformation unitaire, de dimension  $(p \times p)$ .

# Abréviations

<b>ALS</b>	: Alternating Least Squares [TIRC93].
<b>BPSS</b>	: Bayesian positive source separation [MMDBC04].
<b>MLPSS</b>	: Maximum Likelihood Positive Source Separation [MBC05a].
<b>OPA</b>	: Orthogonal Projection Approach [STvdBM96].
<b>SIMPLISMA</b>	: SIMPLe to use Interactive Self-modeling Mixture Analysis [WG91].
<b>NMF</b>	: Non-negative Matrix Factorization [LS99].
<b>PMF</b>	: Positive Matrix factorization [PT94, Van97].
<b>NNSC</b>	: Non-Negative Sparse Coding [Hoy02].
<b>NNICA</b>	: Non-Negative Independent Component Analysis [Plu02].
<b>ICA</b>	: Independent Component Analysis [Com94].
<b>PCA</b>	: Principal Component Analysis [Hot33].
<b>FastICA</b>	: Fast Independent Component Analysis [HO97].
<b>AMUSE</b>	: Algorithm for Multiple Unknown Signals Extraction [TLSH91].
<b>SOBI</b>	: Second Order Blind Identification [BMC97].
<b>FOBI</b>	: Fourth Order Blind Identification [Car89].
<b>JADE</b>	: Joint Approximate Diagonalization of Eigen-Matrices [CS93].
<b>i.i.d</b>	: indépendants et identiquement distribués.
<b>MAP</b>	: Maximum A Posteriori.
<b>JMAP</b>	: Joint Maximum A Posteriori.
<b>MMAP</b>	: Marginal Maximum A Posteriori.
<b>EM</b>	: Expectation Maximization [DLR77].
<b>SEM</b>	: Stochastic Expectation Maximization [CD85].
<b>SAEM</b>	: Stochastic Approximation Expectation Maximization [DLM99].
<b>MCEM</b>	: Monte Carlo Expectation Maximization [WT90]
<b>MCMC</b>	: Markov Chain Monte Carlo [GRS99, Rob99, CSI00].
<b>M-H</b>	: Metropolis-Hastings [MRR <sup>+</sup> 53, Has70].
<b>RSB</b>	: Rapport Signal à Bruit. Défini par l'équation (5.6)
<b>RSD</b>	: Rapport Signal à Distortion. Défini par l'équation (5.8)
<b>IP</b>	: Indice de Performance. Défini par l'équation (5.7)





# Table des figures

1.1	Modèle de mélange et principe de séparation par analyse en composantes indépendantes	10
1.2	Sources spectrales simulées. . . . .	17
1.3	Diagramme de dispersion ( <i>scatter plot</i> ) ou nuage de points de sources non-orthogonales.	19
1.4	Diagramme de dispersion des composantes indépendantes. . . . .	19
1.5	Avantage de la séparation à partir des dérivées secondes des observations. . . . .	21
1.6	Approche de Lawton-Sylvestre pour la séparation de sources spectrales . . . . .	22
2.1	Signaux sources et profils d'évolution des coefficients de mélange . . . . .	32
2.2	Quelques solutions admissibles des sources et des coefficients de mélange . . . . .	33
2.3	Intersection de $\text{Im } \mathbf{S}$ avec $\mathbb{R}_+^n$ , dans le cas $p = 2$ et $n = 3$ . . . . .	37
2.4	Différents cas d'unicité de la factorisation non-négative. Les cercles représentent les vecteurs de $\mathbf{X}$ nécessaire pour assurer l'unicité de la factorisation. . . . .	38
2.5	Signaux sources et coefficients de mélange simulés. . . . .	39
2.6	Valeurs des paramètres des matrices de transformation permettant d'obtenir des solutions admissibles. . . . .	40
2.7	Illustration de quelques solutions admissibles en termes de signaux sources (a-c) et coefficients de mélange (d-f). La solution correcte est tracée en pointillés. . . . .	40
2.8	Cas 1 : Coefficients de mélange non-négatifs et sources mutuellement indépendantes et non-négatives . . . . .	41
2.9	Cas 2 : Coefficients de mélange non-négatifs et sources strictement positives et mutuellement indépendantes . . . . .	41
2.10	Cas 3 : Coefficients non-négatifs et sources non-négatives mais mutuellement dépendantes . . . . .	42
2.11	Cas 4 : coefficients de mélange strictement positifs et sources corrélées et strictement positives . . . . .	42
3.1	Illustration de la forme de la densité de probabilité Gamma. . . . .	57
3.2	Illustration de distribution gamma modifiée et de la fonction non-linéaire résultante. . . . .	58
3.3	Illustration de l'algorithme de simulation de la densité conditionnelle <i>a posteriori</i> des sources. . . . .	62
3.4	Organigramme simplifié de la méthode de séparation par MCEM . . . . .	63
3.5	Signaux sources simulés. . . . .	63

3.6	Séparation par maximum de vraisemblance et gradient relatif. . . . .	64
3.7	Séparation par maximum de vraisemblance et algorithme SAEM. . . . .	65
3.8	Évolution des estimations des coefficients de mélange au cours des itérations de l'algorithme MCEM. . . . .	65
4.1	Illustration de la densité <i>a priori</i> associée à la méthode PMF pour (a) $\alpha = 0, \beta = 1$ , (b) $\alpha = 0.1, \beta = 1$ et (c) $\alpha = 2, \beta = 1$ . . . . .	74
4.2	Organigramme simplifié de la méthode de séparation par simulation MCMC . . . . .	75
4.3	Signaux sources et coefficients de mélange simulés. . . . .	80
4.4	Chaînes de Markov des variances du bruit et des coefficients de mélange . . . . .	81
4.5	Histogrammes des paramètres des lois <i>a priori</i> des signaux sources et des coefficients de mélange. . . . .	82
4.6	Histogrammes des coefficients de mélange simulés. . . . .	83
4.7	Signaux sources estimés par approche bayésienne. . . . .	83
4.8	Temps de calcul par itération. . . . .	84
5.1	Allure des signaux sources simulés et leurs histogrammes. Les paramètres de synthèse sont $K = 15, \gamma_{max} = 10, \gamma_{min} = 2n$ . Ces sources présentent les coefficients de corrélation mutuelle suivants $\text{corr}(s_1, s_2) = -0.16, \text{corr}(s_1, s_3) = -0.24$ , et $\text{corr}(s_2, s_3) = +0.27$ . . . . .	86
5.2	Coefficients de mélange simulés à partir d'une cinétique chimique. . . . .	87
5.3	Coefficients de mélange et observations simulées pour $m = 10, p = 3$ et $n = 1000$ . . . . .	87
5.4	Résultat de la séparation en utilisant JADE : (a,b,c) sources originales (discontinu) et sources estimées (continu). Les flèches indiquent les endroits où apparaissent des composantes inexistantes ou aberrantes. (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles). . . . .	89
5.5	Résultat de la séparation en utilisant NNICA : (a,b,c) sources originales (discontinu) et sources estimées (continu). (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles). . . . .	90
5.6	Résultat de la séparation en utilisant NMF : (a,b,c) sources originales (discontinu) et sources estimées (continu). (d) coefficients de mélange originaux (discontinu) et coefficients identifiés (continu). . . . .	91
5.7	Résultat de la séparation en utilisant NNSC : (a,b,c) sources originales (discontinu) et sources estimées (continu). (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles). . . . .	91
5.8	Résultat de la séparation en utilisant l'approche MLPSS-gamma : (a,b,c) signaux sources originales (discontinu) et sources estimées (continu). Le résidu de l'estimation est aussi tracé pour évaluer l'erreur de reconstruction. (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles). . . . .	92
5.9	Influence de la corrélation mutuelle des sources sur les performances de la séparation. . . . .	94
5.10	Influence du niveau de bruit sur les performances de la séparation. . . . .	95

5.11 (a,b,c) Allure des signaux sources pour un arrière-plan d'amplitude égal à l'unité et (d) évolution des performances en fonction de l'amplitude de l'arrière-plan pour un RSB de 20 dB. . . . .	96
6.1 Régions du spectre électromagnétique présentant un intérêt pour l'analyse spectroscopique et type de mouvements associés. . . . .	102
6.2 Spectres Raman des quatre composés organiques ( $p=4$ et $n = 3400$ ). . . . .	106
6.3 Profils d'évolution des coefficients de mélange ( $m=10$ ). . . . .	106
6.4 Spectres infrarouge des trois alcanes ( $p = 3$ et $n = 5200$ ). . . . .	108
6.5 Spectres infrarouges mesurés dans la gamme spectrale ( $3000-6000 \text{ cm}^{-1}$ ). . . . .	108
6.6 Comparaison visuelle entre les sources reconstruites (continu) et les spectres purs (discontinu) sur deux régions spectrales). . . . .	110
6.7 Comparaisons entre les coefficients de mélange estimés (discontinu) et les proportions des trois alcanes dans les mélanges. . . . .	110
6.8 Observations enregistrées pour plusieurs valeurs du $P/P_0$ . ( $m = 33$ et $n = 4000$ ). . . . .	113
6.9 Résultats de la séparation en termes de spectres purs et proportions des différentes sources dans les mélanges. . . . .	114
6.10 Spectres mesurés à plusieurs températures. . . . .	116
6.11 Spectres purs estimés par BPSS. . . . .	117
6.12 Analyse de corrélation entre sources estimées et spectres purs. . . . .	117
6.13 Proportions des trois espèces dans les mélanges : (a) suivi la transformation de phase à $40^\circ\text{C}$ et (b) polymorphes après précipitation. . . . .	118
6.14 (a) Observations enregistrées pour plusieurs valeurs du pH et (b) valeurs propres de la matrice de covariance empirique des observations. . . . .	119
6.15 Résultats de la séparation en termes de spectres purs et proportions des différentes sources dans les mélanges. . . . .	119
B.1 Quelques formes typiques de la densité normale-gamma. . . . .	130
B.2 Histogrammes de 5000 échantillons simulés en utilisant l'algorithme M-H indépendant. . . . .	130
C.1 (a) Distribution conjointe <i>a posteriori</i> des deux paramètres de la distribution gamma, (b) distribution des paramètres simulés par Gibbs, (c) chaînes de Markov associées aux deux paramètres et (d) taux d'acceptation de l'algorithme M-H lors de la simulation du paramètre de forme de la densité gamma. . . . .	133



# Liste des tableaux

1.1 Performances de la séparation en utilisant quelques algorithmes d'analyse en composantes indépendantes. . . . .	18
3.1 Quelques densités de probabilité et fonctions non-linéaires correspondantes. . . . .	54
3.2 Quelques modèles de densité de probabilité et leurs fonctions de séparation associées .	55
4.1 Valeurs des paramètres des lois <i>a priori</i> associées aux variances du bruit, aux signaux sources et aux coefficients de mélange. . . . .	81
5.1 Comparaison des performances de la séparation dans le cas d'une un RSB de 50 dB. .	93
5.2 Comparaison des performances de la séparation dans le cas d'une un RSB de 10 dB. .	93
6.1 Comparaison des performances de quelques méthodes de séparation pour un RSB de 30 dB. . . . .	107
6.2 Comparaison des performances de quelques algorithmes de séparation. . . . .	109



# Introduction générale

Ce travail a été mené au Centre de Recherche en Automatique de Nancy (CNRS UMR 7039) au sein du projet "Signal et Information" du groupe thématique IRIS (Identification, Restauration, Images et Signaux) dans le cadre de la résolution de problèmes inverses appliquée à la spectroscopie. Ce travail rentre dans le cadre d'une collaboration avec le Laboratoire de Chimie-Physique et Microbiologie pour l'Environnement (LCPME, CNRS UMR 7564). L'objectif de cette collaboration est le développement de méthodes de traitement de signaux spectroscopiques. Un premier aspect concerne l'amélioration de la lisibilité d'un spectre via la soustraction de la ligne de base et l'estimation du spectre de raies ; il est traité dans la thèse de mon ami Vincent Mazet [Maz05]. Le deuxième aspect, qui constitue le sujet de ce travail, consiste à retrouver, à partir d'une collection de spectres enregistrés dans des conditions physico-chimiques différentes, les spectres des composantes pures présentes au sein des mélanges ainsi que leurs concentrations.

## Problématique

Dans le domaine de la chimie que ce soit académique ou industrielle, l'utilisation des spectromètres occupe une place très importante. Ces instruments permettent d'accéder à des informations sur la composition physico-chimique de nombreux matériaux. Ainsi, il est nécessaire de développer des méthodes d'analyse des spectres fournis par ces instruments en vue de l'extraction de ces informations. Durant ces trois dernières décennies, l'intérêt croissant porté au traitement de ces données chimiques a donné naissance à la chimiométrie qui est une discipline qui se fonde sur les mathématiques appliquées, les statistiques et les techniques du traitement de signal.

Afin d'illustrer la problématique qui nous intéresse, imaginons une situation simple dans laquelle une réaction chimique est provoquée par le mélange de plusieurs espèces. Un spectromètre est employé pour enregistrer la réponse du mélange à des fréquences différentes. Des dizaines de spectres peuvent être accumulées tout au long de la réaction qui peut durer de quelques minutes à quelques heures. Les données ainsi enregistrées contiennent des informations sur les espèces qui apparaissent ou disparaissent au cours de la réaction ainsi que sur l'évolution de leurs proportions. Afin de décrire finement la réaction en utilisant toutes ces mesures, des méthodes statistiques bien adaptées doivent être développées. Dans la plupart des applications réelles, les réactions chimiques sont très complexes et les espèces ne sont connues que théoriquement, leurs signatures spectrales ne sont pas disponibles individuellement ;

parfois même l'évolution de la réaction est imprévisible. L'analyse dans ce cas devient délicate car il s'agit d'un problème d'identification qui, en reprenant la terminologie utilisée dans le domaine de l'identification de systèmes dynamiques, peut être qualifié de *boite grise* ou *boite noire*. Dans le cas d'un modèle boite grise certaines informations sur les espèces initiales, intermédiaires et finales ou sur l'évolution de la réaction sont disponibles, alors que dans un modèle boite noire, mis à part les données spectrales, aucune connaissance *a priori* sur les espèces chimiques et sur leurs concentrations n'est disponible. Ce problème d'identification et de restauration est un problème inverse dont la résolution doit permettre d'atteindre les objectifs suivants :

1. Déterminer le nombre d'espèces qui interviennent dans le processus en utilisant seulement les données spectrales mesurées par un spectromètre.
2. Obtenir les spectres purs des différentes espèces en utilisant les données fournies par le spectromètre.
3. Déterminer les concentrations des espèces et différentes caractéristiques physico-chimiques du système en combinant les résultats de l'analyse avec les connaissances théoriques.

La résolution de ce problème inverse n'est pas aisée car elle nécessite une modélisation réaliste du processus d'observation avant de passer aux développements méthodologiques. Dans le cadre de cette thèse, nous considérons que les signatures spécifiques (spectres purs) des différentes espèces restent inchangées durant les observations par le spectromètre et nous supposons également une contribution linéaire de ces spectres purs dans le spectre global mesuré par l'instrument. Ce modèle d'observation nous mène au problème inverse de la séparation de sources dans le cas d'un mélange linéaire invariant. De plus, étant donné la nature spectrale des données et le caractère positif des abondances des espèces au cours de la réaction, ce problème de séparation de sources sera soumis aux contraintes de non-négativité des coefficients de mélange et des signaux sources. Il s'agit donc d'un problème de séparation de sources non-négatives.

**Nous envisageons à travers cette thèse de résoudre un problème de séparation de sources non-négatives en vue d'une application au traitement de données issues d'une famille aussi large que possible de spectromètres.**

## Organisation du document

Une première partie de ce document présente une étude des méthodes existantes afin de mieux comprendre leur fonctionnement et déceler leur limitations. Une première classe de méthodes consiste à effectuer une analyse en composantes indépendantes. L'influence de la validité de l'hypothèse d'indépendance statistique sur les résultats de la séparation est analysée de façon à discuter le cas où cette approche permet de réaliser la séparation et quelques solutions qui permettent d'améliorer les performances dans le cas contraire. Une seconde classe de méthodes consiste à effectuer une factorisation de la matrice de données en un produit de matrices non-négatives. Dans ce cadre, nous étudions théoriquement les conditions de l'unicité d'une telle factorisation ainsi que, lorsque



l'unicité n'est pas garantie, le domaine des solutions admissibles. Les résultats de cette première partie permettent de comprendre l'échec de ce type de méthodes car, dans la majorité des cas, les conditions d'unicité ne sont pas satisfaites par les signaux de spectroscopie. Les résultats et les discussions de cette première partie sont présentés dans les chapitres 1 et 2 de ce document.

La deuxième partie de ce document a pour objectif le développement de méthodes de séparation de sources spectrales. Une première méthode repose sur le principe du maximum de vraisemblance. Dans ce cadre, nous montrons que l'approche du maximum de vraisemblance peut être appliquée à la séparation de sources non-négatives en considérant une loi gamma comme modèle probabiliste qui permet de prendre en compte la non-négativité et la parcimonie des sources. Ce résultat est très intéressant car il montre que les méthodes d'analyse en composantes indépendantes peuvent être appliquées au traitement de données spectrales à condition de ne pas imposer de façon stricte l'orthogonalité des sources et de spécifier *a priori* leurs distributions. Tous ces développements sont présentés dans le chapitre 3. La principale limitation de cette approche est de ne pas garantir la positivité des coefficients de mélange. Une méthode originale qui permet de palier cette limitation est développée en posant le problème de séparation dans un cadre bayésien. Les signaux sources et les coefficients de mélange sont supposés distribués selon des lois gamma et le problème d'estimation conjointe des sources, des coefficients de mélange et des hyperparamètres est résolu en utilisant les méthodes Monte Carlo par chaînes de Markov (MCMC). Le chapitre 4 développe les aspects théoriques et méthodologiques liés à cette approche. Finalement, les performances des approches proposées sont évaluées et comparées à celles des approches classiques sur des signaux synthétiques de formes similaires aux données spectroscopiques. Tous les résultats de ces comparaisons et analyses de performances sont rapportés dans le chapitre 5.

La troisième partie du document présente une application des méthodes développées à des signaux expérimentaux fournis par le LCPME. Les résultats de cette partie sont présentés dans le chapitre 6. Une première expérience est réalisée de façon à tester la méthode sur des données infrarouge dans lesquelles les spectres purs et leur proportions dans les mélanges sont connus au préalable. Un deuxième type d'expériences illustre l'application de l'approche bayésienne dans des conditions réelles, en traitant des données issues de trois problèmes d'analyse physico-chimique (l'hydratation d'une silice, la caractérisation des polymorphes du carbonate de calcium et l'étude de l'interaction des cations métalliques avec des biomolécules).

Les travaux menés durant cette thèse ont fait l'objet de communications en conférences [MMDBC04, MBCMD04, MBI05b, MBC05a, MBC05b] et d'articles acceptés pour publication en revues [MBMDC05, MCBMD05].



## Première partie

# Position des problèmes de la séparation de sources non-négatives



# Chapitre 1

## Séparation de sources spectrales : position des problèmes

Ce chapitre introduit le problème de séparation de sources d'une façon générale et fournit une discussion sur le cas de sources spectroscopiques. Après une présentation de quelques approches de séparation fondées sur l'hypothèse d'indépendance statistique, nous analysons les difficultés posées par ce type de sources. La contrainte principale étant celle de la non-négativité des signaux sources et des coefficients de mélange, les approches qui prennent en compte cette contrainte sont classées en fonction de la manière dont celle-ci est introduite dans les algorithmes de séparation. Tout au long de cette présentation, nous discutons les limitations de ces approches et nous formulons quelques orientations possibles pour le développement de méthodes de séparation bien adaptées.

### 1.1 Introduction à la séparation de sources

Le concept de séparation de sources peut être décrit ainsi : retrouver des signaux qui ont été mélangés et bruités éventuellement, en tenant compte de toute information disponible sur les signaux d'intérêt et sur le processus de mélange. Les signaux d'intérêt sont appelés *sources* alors que les signaux mesurés sont appelés *observations*. Outre le domaine du traitement de la parole et des signaux audio (problème dit *cocktail party*), ce problème se rencontre dans plusieurs domaines : en astronomie lorsqu'un radio-télescope observe simultanément plusieurs étoiles dont on veut séparer les contributions pour les analyser séparément, en communications numériques ainsi qu'en radar et sonar lorsque les signaux provenant de plusieurs émetteurs ou réflecteurs interfèrent au niveau des antennes de réception, en prospection sismique, en imagerie médicale ainsi qu'en chimie analytique lorsque les spectres d'absorption d'une substance hétérogène résultent de la superposition des spectres spécifiques des différentes composantes pures de la substance.

Avant de se focaliser sur le problème inverse de la séparation, un point tout aussi important est celui de l'établissement du modèle d'observation (problème direct) dont les méthodes de séparation à envisager dépendent fortement.

### 1.1.1 Modèle du mélange

Le modèle de mélange décrit la transformation liant les sources aux observations et peut être linéaire ou non linéaire, convolutif ou instantané, variant ou invariant dans le temps. Le modèle linéaire instantané invariant dans le temps est le plus utilisé car il s'agit d'un modèle relativement simple, mais dont les applications sont nombreuses. Il peut être vu comme une approximation au premier ordre d'une transformation non-linéaire à horizon d'observation suffisamment petit.

#### 1.1.1.1 Mélange linéaire instantané invariant dans le temps

Ce modèle de mélange suppose qu'à chaque instant (point d'observation)  $k$ , les  $m$  observations  $\{x_{(i,k)}\}_{i=1}^m$ , sont des mélanges linéaires instantanés des  $p$  sources  $\{s_{(j,k)}\}_{j=1}^p$  :

$$x_{(i,k)} = \sum_{j=1}^p a_{ij} s_{(j,k)} + e_{(i,k)}, \text{ pour } i = 1, \dots, m \quad (1.1)$$

où  $a_{ij} \in \mathbb{R}$  pour  $i \in \{1, \dots, m\}$  et  $j \in \{1, \dots, p\}$  sont les *coefficients de mélange*. Ce modèle de mélange (1.1) s'exprime aussi sous la forme :

$$\mathbf{x}_k = \mathbf{A} \mathbf{s}_k + \mathbf{e}_k, \text{ pour } k = 1, \dots, n \quad (1.2)$$

où les  $m$  composantes du vecteur  $\mathbf{x}_k$  représentent les observations, les  $p$  sources sont contenues dans le vecteur  $\mathbf{s}_k$ , tandis que les coefficients de mélange forment une matrice notée  $\mathbf{A}$ , de dimension  $(m \times p)$ , appelée *matrice de mélange*. On note  $\mathbf{e}_k$  le vecteur dont les composantes correspondent à un possible *bruit additif* d'autocovariance  $\mathbf{R}_e$  et supposé indépendant des signaux sources. Ayant  $n$  échantillons pour  $k = 1, \dots, n$  et en utilisant une notation matricielle, le modèle de mélange s'écrit

$$\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{E}, \quad (1.3)$$

où  $\mathbf{X} \in \mathbb{R}^{(m \times n)}$  est la matrice des observations,  $\mathbf{A} \in \mathbb{R}^{(m \times p)}$  est la matrice de mélange,  $\mathbf{S} \in \mathbb{R}^{(p \times n)}$  est la matrice des signaux sources et  $\mathbf{E} \in \mathbb{R}^{(m \times n)}$  est la matrice du bruit additif.

Ce modèle de mélange est utilisé en séparation de sources spectrales où on introduit la contrainte supplémentaire de non-négativité des sources et des coefficients de mélange. Cependant, une attention particulière doit être accordée aux non-linéarités qui peuvent survenir lors de la mesure des spectres caractéristiques de substances chimiques multi-composantes. Néanmoins, la validité du modèle est justifiée par la loi de Beer [BD00, Hol98] et l'utilisation d'un modèle de mélange avec un bruit additif permet de prendre en compte les incertitudes liées aux erreurs de modélisation. Dans ce qui suit nous considérons donc un modèle de mélange linéaire invariant dans le temps.

#### 1.1.1.2 Indéterminations

Dans le cas où aucune information *a priori* sur le mélange n'est disponible, une identification complète de la matrice de mélange est impossible. En effet, l'échange d'un facteur quelconque entre un signal

source et le vecteur colonne de la matrice de mélange correspondant n'affecte en rien les observations. Ceci est montré par la relation suivante

$$x_{(i,k)} = \sum_{j=1}^m a_{ij} s_{(j,k)} + e_{(i,k)}, \quad (1.4)$$

$$= \sum_{j=1}^p \frac{a_{ij}}{\lambda_j} \lambda_j s_{(j,k)} + e_{(i,k)}. \quad (1.5)$$

C'est l'*indétermination d'échelle*. Afin de pallier cette indétermination, plusieurs stratégies peuvent être utilisées : (i) fixer la norme  $L_\alpha$  de chaque source à l'unité, (ii) fixer la norme  $L_\alpha$  de chaque vecteur colonne de la matrice de mélange à l'unité, (iii) fixer les éléments diagonaux de la matrice de mélange à 1. Or, ces solutions ne sont que des conventions de représentation et n'ont aucune incidence sur la forme d'onde des sources reconstruites. Par ailleurs, l'indice associé à chaque source est arbitraire car les sources ne seront connues qu'à une permutation près des vraies sources. C'est l'*indétermination d'ordre*.

### 1.1.2 Principe de la séparation de sources

La séparation de sources est l'opération qui, à partir des observations, permet d'obtenir un ensemble de signaux proportionnels aux sources et d'identifier la contribution de chacune des sources à l'intérieur du mélange observé. Ainsi on distingue deux sous-problèmes : (i) l'*identification* du mélange et (ii) la *reconstruction* des sources.

Ce problème inverse est mal-posé car sans aucune information sur les sources et sur le mélange une infinité de solutions seraient admissibles. Il est alors nécessaire de formuler des hypothèses supplémentaires et prendre en compte des informations supplémentaires sur le mélange et sur les sources. On peut aborder le problème de séparation de sources de deux points de vue. Le premier est celui de la décomposition des observations sur une base de signaux élémentaires permettant d'éliminer la redondance d'information entre les différentes observations. Il est alors nécessaire de définir une mesure de cette redondance qui mène à la spécification de contraintes fortes sur ces composantes. Par exemple, la spécification de l'orthogonalité avec un minimum d'énergie de la décomposition aboutit à l'analyse en composantes principales (PCA). Une deuxième approche, plus récente, est celle de l'analyse en composantes indépendantes (ICA) qui se base sur l'hypothèse d'indépendance statistique des composantes recherchées. Dans le cas du mélange linéaire, l'objectif est de chercher une transformation linéaire dont l'application aux observations fournit un ensemble de signaux qui respectent aux mieux les hypothèses formulées au préalable (décorrélation mutuelle, indépendance statistique, etc.). Le deuxième point de vue est celui de la reconstruction où on cherche à appliquer une transformation linéaire ou non-linéaire aux observations afin de reconstituer le plus fidèlement possible les sources et d'identifier au mieux le processus de mélange. Mais, ces points de vue de décomposition et de reconstruction se rejoignent dans le cas où les sources à restaurer possèdent les mêmes propriétés que les composantes élémentaires recherchées. Cette remarque établit le point de connection entre l'analyse en composantes indépendantes et la séparation de sources en utilisant l'hypothèse d'indépendance

mutuelle. Cependant, dans certaines applications, comme nous le verrons par la suite, ces deux points de vue (décomposition et reconstruction) peuvent ne pas être concordants. Dans ce cas, il est alors nécessaire de définir des critères de séparation qui, au lieu de se baser sur une mesure d'indépendance statistique, évaluent l'adéquation des sources reconstruites et des coefficients de mélange estimés aux propriétés préalablement spécifiées. Il peut s'agir d'hypothèses sur leurs distributions, sur leurs structure temporelle ainsi que d'autres contraintes telles que la parcimonie et la positivité. Néanmoins, dans la plupart des modèles de séparation de sources, l'hypothèse d'indépendance statistique est toujours prise en compte mais avec un poids relativement plus faible que dans le cas de l'analyse en composantes indépendantes.

## 1.2 Séparation par analyse en composantes indépendantes

Le principe général consiste à appliquer des transformations aux observations pour obtenir des signaux statistiquement indépendants. Cette définition fait apparaître une hypothèse très importante qu'est l'indépendance statistique des signaux sources. Néanmoins, l'utilisation de l'indépendance statistique comme hypothèse de séparation ne garantit l'unicité de la séparation que dans le cas sur-déterminé ( $m \geq p$ ) et qu'au plus une source est de distribution gaussienne [TIL93, Com94]. Nous présentons dans cette section quelques mesures, directes ou indirectes, de l'indépendance statistique ainsi que quelques algorithmes résultants de l'utilisation de ces mesures pour la synthèse de critères de séparation.

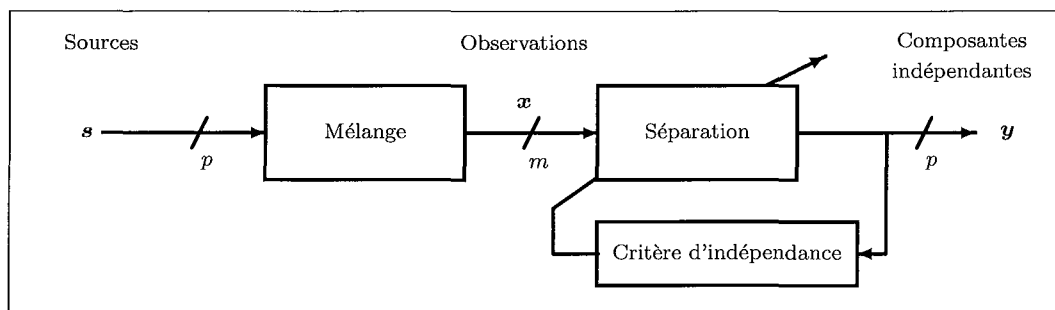


FIG. 1.1 – Modèle de mélange et principe de séparation par analyse en composantes indépendantes

### 1.2.1 Mesures de l'indépendance statistique

Dans un contexte de séparation par analyse en composantes indépendantes, il est nécessaire de pouvoir évaluer la dépendance statistique de variables aléatoires avant de définir des critères de séparation. Nous donnons d'abord quelques rappels sur la définition et les mesures de l'indépendance statistique de variables aléatoires indépendantes et identiquement distribuées (i.i.d), avant de présenter succinctement certains critères et algorithmes de séparation.



### 1.2.1.1 Définition

Des variables aléatoires  $\{x_j\}_{j=1}^p$  sont dites *statistiquement mutuellement indépendantes* si et seulement si :

$$p(x_1, x_2, \dots, x_p) = \prod_{j=1}^p p(x_j). \quad (1.6)$$

Par conséquent,

- si deux variables aléatoires  $x_1$  et  $x_2$  sont statistiquement indépendantes, alors :

$$\mathbb{E}[f(x_1)g(x_2)] = \mathbb{E}[f(x_1)] \mathbb{E}[g(x_2)], \quad (1.7)$$

pour toute fonction  $f$  et  $g$ .

- l'indépendance statistique implique la décorrélation, mais l'inverse n'est pas toujours vrai, sauf pour le cas de vecteurs aléatoires gaussiens.

### 1.2.1.2 Divergence de Kullback-Leibler

La mesure de la distance entre les densités de probabilité de variables aléatoires permet de caractériser leur indépendance mutuelle. La Divergence de Kullback-Leibler [Kul59] permet d'effectuer cette mesure. Dans le cas d'un vecteur de variables aléatoires  $\mathbf{y} = [y_1, \dots, y_p]^T$ ,

$$\mathcal{KL} \left( p(\mathbf{y}), \prod_{j=1}^p p(y_j) \right) \triangleq \int_{\mathbb{R}^p} p(\mathbf{y}) \log \left( \frac{p(\mathbf{y})}{\prod_{j=1}^p p_j(y_j)} \right) d\mathbf{y}. \quad (1.8)$$

La divergence  $\mathcal{KL}$  possède les propriétés suivantes : (i) la divergence  $\mathcal{KL}$  est non-négative et n'est nulle que lorsque les variables  $\{y_j\}_{j=1}^p$  sont statistiquement indépendantes. (ii) la divergence  $\mathcal{KL}$  est invariante par permutation ou par changement d'échelle. Au vu de ces deux propriétés, la divergence  $\mathcal{KL}$  est une mesure d'indépendance qui, par son annulation ou minimisation, permet d'obtenir une séparation des signaux sources.

### 1.2.1.3 Information mutuelle

L'indépendance de variables aléatoires peut être aussi quantifiée en utilisant l'information mutuelle qui s'exprime par :

$$\mathcal{I}(\mathbf{y}) = \sum_{j=1}^p \mathcal{H}(y_j) - \mathcal{H}(\mathbf{y}), \quad (1.9)$$

où :

$$\begin{aligned} \mathcal{H}(\mathbf{y}) &= - \int_{\mathbb{R}^p} p(\mathbf{y}) \log p(\mathbf{y}) d\mathbf{y} = - \mathbb{E}[\log p(\mathbf{y})] \\ \mathcal{H}(y_j) &= - \int_{\mathbb{R}} p_j(y_j) \log p_j(y_j) dy = - \mathbb{E}[\log p_j(y_j)] \end{aligned}$$

sont, respectivement, les entropies différentielles conjointes et marginales de  $\mathbf{y}$  et  $y_j$ . Il est à noter que l'information mutuelle se déduit à partir de la divergence de Kullback-Leibler. Considérons que

les variables  $\{y_j\}_{j=1}^p$  sont les estimées des sources obtenues par l'application d'une matrice séparante, notée  $\mathbf{B}$ . L'information mutuelle de ces variables sera exprimée par :

$$\mathbf{y} = \mathbf{B}\mathbf{x} \Rightarrow \mathcal{I}(\mathbf{y}) = \sum_{j=1}^p \mathcal{H}(y_j) - \mathcal{H}(\mathbf{x}) - \log \det \mathbf{B}, \quad (1.10)$$

car  $\mathcal{H}(\mathbf{y}) = \mathcal{H}(\mathbf{B}\mathbf{x}) = \mathcal{H}(\mathbf{x}) + \log |\det \mathbf{B}|$ .

#### 1.2.1.4 Néguentropie

Si l'on considère des variables aléatoires  $\mathbf{v} = \{v_1, \dots, v_p\}$  de densité de probabilité gaussienne, alors :

$$\mathcal{H}(v_i) = \frac{1}{2}(\log \sigma_i^2 + \log 2\pi + 1) \quad (1.11)$$

$$\mathcal{H}(\mathbf{v}) = \frac{1}{2}(\log \det \mathbf{\Sigma} + p(1 + \log 2\pi)) \quad (1.12)$$

où  $\sigma_j^2$  et  $\mathbf{\Sigma}$  représentent, respectivement, la variance et la covariance de  $v_j$  et  $\mathbf{v}$ . La néguentropie est définie comme étant une mesure de l'éloignement entre la distribution d'une variable aléatoire et la densité gaussienne. Cette mesure est définie par

$$\mathcal{J}(y_j) = \mathcal{H}(y_j) - \mathcal{H}(v_j), \quad (1.13)$$

où les variables  $y_j$  et  $v_j$  possèdent la même variance. Cette quantité est toujours positive et n'est nulle que lorsque la variable  $y_j$  est gaussienne. Sa maximisation correspond donc à la recherche de composantes non-gaussiennes dans un mélange, ce qui d'après la théorème central-limite tend à rechercher les composantes indépendantes [Don81, Hyv99, HKO01]. En appliquant les définitions (1.13) et (1.11) à l'équation (1.9), on obtient :

$$\mathcal{I}(\mathbf{y}) = \sum_{j=1}^p \mathcal{J}(y_j) - \mathcal{J}(\mathbf{y}) + \frac{1}{2} \log \left[ \frac{\sigma_j^2}{\det \mathbf{\Sigma}} \right], \quad (1.14)$$

En réalisant un blanchiment spatial des observations, le deuxième terme sera réduit à une constante. En seconde étape, les deux derniers termes seront invariants par transformation orthonormale. Il suffit alors de rechercher une matrice orthonormale qui minimise la somme des néguentropies marginales, ce qui nécessite également l'approximation des densités de probabilité des sources. Nous pouvons consulter à ce propos [Com94, HO97, Hyv99].

#### 1.2.1.5 Statistiques d'ordre supérieur

Les statistiques d'ordres supérieur sont des outils pratiques pour évaluer l'indépendance statistique de variables aléatoires non-gaussiennes. Ces derniers présentent des propriétés très intéressantes car ils sont d'une part multilinéaires et, d'autre part, additifs dans la superposition des variables aléatoires indépendantes. Si un sous-ensemble d'une famille de variables aléatoires est indépendant d'un autre sous-ensemble complémentaire, alors les cumulants croisés s'annulent. Enfin les cumulants d'ordre supérieur à deux d'une variable gaussienne sont tous nuls. Les cumulants sont utilisés soit d'une façon directe pour la construction de mesures d'indépendance [MTM99, PM01, Mor01], ou comme des outils d'approximation d'autres mesures d'indépendance tel que la néguentropie [Com94, Car99a]. Particulièrement le Kurtosis est utilisé pour une approximation polynomiale de la néguentropie [HO97].

## 1.2.2 Présentation de quelques méthodes de séparation

D'une façon générale, toute méthode d'analyse en composantes indépendantes est caractérisée par : (i) une fonction de contraste qui définit le critère de séparation. (ii) un algorithme d'optimisation de ce critère.

### 1.2.2.1 Méthodes algébriques

Ces méthodes se basent sur les étapes suivantes : (i) blanchir la partie signal des observations par l'application d'une matrice dite de *blanchiment*, notée  $\mathbf{W}$  et de dimension  $(p \times m)$ , afin d'obtenir un processus  $\mathbf{z} = \mathbf{W} \mathbf{x}$  *décorrélé spatialement* (processus dont la matrice d'autocovariance est une matrice identité). L'intérêt de cette étape est de réduire le problème de détermination de la matrice de mélange de dimension  $(m \times p)$  à celui de la recherche d'une matrice unitaire de dimension  $(p \times p)$ . (ii) estimer une matrice de rotation  $\hat{\mathbf{U}}$  qui permet d'obtenir des signaux  $\mathbf{y} = \hat{\mathbf{U}} \mathbf{z}$  mutuellement indépendants (au sens du critère de séparation ou de mesure d'indépendance définis). Finalement, la matrice de mélange se déduit alors directement selon :

$$\hat{\mathbf{A}} = \mathbf{W}^\# \hat{\mathbf{U}}, \quad (1.15)$$

où  $\mathbf{W}^\#$  représente la pseudo-inverse de la matrice  $\mathbf{W}$ , et les signaux sources sont estimés par :

$$\hat{\mathbf{S}} = \hat{\mathbf{U}} \mathbf{W} \mathbf{X}. \quad (1.16)$$

Cette approche est dite algébrique car les étapes (i) et (ii) sont réalisées par de simples manipulations de matrices.

**Méthodes utilisant les statistiques de second ordre :** dans le cas où les signaux sources possèdent une corrélation temporelle, la séparation est réalisée en utilisant les matrices d'intercovariance des signaux observés [TLSH91, BAM93, MS94]. Les sources sont supposés stationnaires au second ordre, de moyenne nulle, de puissance finie et mutuellement décorrélés. C'est principalement la procédure avec laquelle cette matrice de rotation est recherchée qui diffère entre les deux algorithmes de séparation de sources au second ordre : AMUSE et SOBI. L'algorithme AMUSE, pour *Algorithm for Multiple Unknown Signals Extraction*, [TLSH91, TIL93] se fonde sur la diagonalisation exacte d'une seule matrice d'intercovariance décalée d'un retard  $\tau$ , qui est un paramètre de synthèse de la méthode, alors que l'algorithme SOBI, pour *Second Order Blind Identification* [BAM93, BMCM97], se fonde sur la diagonalisation conjointe approchée de  $L$  matrices d'intercovariance décalées

$$\mathbf{R}_\ell = \mathbb{E} [\mathbf{z}_k^T \mathbf{z}_{(k+\tau_\ell)}], \text{ pour } \ell = 1, \dots, L. \quad (1.17)$$

La diagonalisation conjointe approchée de ces matrices correspond à la minimisation du critère défini par :

$$\mathcal{D}_{SOBI}(\mathbf{U}) = \sum_{\ell=1}^L \sum_{i=1}^p \sum_{j=1, j \neq i}^p [\mathbf{U}^T \mathbf{R}_\ell \mathbf{U}]_{ij}^2. \quad (1.18)$$

**Méthodes utilisant les statistiques d'ordre supérieur :** dans le cas où les signaux sources ne présentent pas de cohérence temporelle, il n'est plus possible de séparer les sources en utilisant uniquement les statistiques d'ordre deux. En plus des matrices de covariance des observations, les statistiques d'ordres supérieurs sont utilisées. Cardoso a défini la notion de matrice cumulante  $\mathbf{N} = Q_x(\mathbf{M})$  associée à une matrice  $\mathbf{M}$  de dimension  $(p \times p \text{ par})$  :

$$\mathbf{N} = Q_x(\mathbf{M}) \stackrel{\Delta}{\iff} \left\{ N_{ij} = \sum_{k=1}^p \sum_{\ell=1}^p Q_{ij}^{k\ell} M_{\ell k} \mid 1 \leq i, j \leq p \right\}. \quad (1.19)$$

où :

$$Q_x \stackrel{\Delta}{=} \left\{ Q_{ij}^{k\ell} = \text{Cum}[x_i, x_j^*, x_k, x_\ell^*] \mid 1 \leq i, j, k, \ell \leq n \right\}. \quad (1.20)$$

L'algorithme FOBI pour *Fourth Order Blind Identification* [Car89, Car92] se fonde sur la diagonalisation d'une seule matrice cumulante  $Q_z(\mathbf{M})$  pour obtenir la matrice de rotation permettant d'assurer l'indépendance des sources estimées  $\hat{\mathbf{y}} = \hat{\mathbf{U}}\mathbf{z}$ . La forme la plus simple de cette méthode s'obtient en choisissant  $\mathbf{M} = \mathbf{I}_p$ . Par ailleurs, le choix de la matrice  $\mathbf{M}$  n'est pas toujours aussi simple, car par exemple lorsque  $\mathbf{M} = \mathbf{I}_p$  il se peut qu'il y'ait des sources ayant le même kurtosis et par conséquent on ne peut pas diagonaliser la matrice cumulante  $Q_z(\mathbf{I}_p)$  à cause de la dégénérescence de ses valeurs propres [CS93] et on ne peut pas choisir une autre matrice  $\mathbf{M}$  qui permet d'assurer *a priori* la séparation des sources. Pour cela, l'algorithme JADE (*Joint Approximate Diagonalization of Eigen-matrices*) a été proposé par [CS93], et dont le principe repose sur la diagonalisation conjointe approchée de plusieurs matrices cumulantes. Considérons un ensemble de  $K$  matrices de dimension  $p \times p$ ,  $\mathbb{M} = \{\mathbf{M}_k \mid k = 1, K\}$ . Une première solution consiste à considérer  $K = p^2$  matrices creuses (ne possédant qu'un seul élément non nul pris égal à 1), de façon à former une base orthonormée de matrices et dans ce cas l'identifiabilité des sources est assurée. Néanmoins, cette solution se heurte à une complexité arithmétique accrue. Une autre alternative se base sur le choix de  $K = p$  matrices propres du tenseur cumulant. La diagonalisation conjointe approchée de ces matrices correspond à la minimisation du critère défini par :

$$\mathcal{D}_{JADE}(\mathbf{U}) = \sum_{\ell=1}^L \sum_{i=1}^p \sum_{j=1, j \neq i}^p [U^T Q_z(\mathbf{M}_\ell) \mathbf{U}]_{ij}^2. \quad (1.21)$$

### 1.2.2.2 Méthodes fondées sur la minimisation d'une fonction de contraste

**Fonction de contraste :** en estimation statistique, la notion de fonction de contraste est introduite pour la construction d'estimateurs dits de minimum de contraste [Pfa73, Egu83, Egu85]. En traitement du signal, la notion de fonction de contraste est initialement définie dans [Don81] pour le problème de déconvolution myope et [Com94] pour l'analyse en composantes indépendantes. La définition donnée par Comon est la suivante : une fonction  $\Psi$  est dite fonction de contraste pour les vecteurs aléatoires  $\mathbf{z}$ , si celle-ci vérifie les conditions :

(C1)  $\Psi(\mathbf{P}\mathbf{z}) = \Psi(\mathbf{z})$ ,  $\forall \mathbf{P}$ , matrice de permutation,

(C2)  $\Psi(\mathbf{D}\mathbf{z}) = \Psi(\mathbf{z})$ ,  $\forall \mathbf{D}$ , matrice diagonale,

(C3) si les composantes de  $\mathbf{z}$  sont indépendantes alors :

- $\Psi(\mathbf{M}\mathbf{z}) \leq \Psi(\mathbf{z}), \forall \mathbf{M}$  matrice inversible,
- $\Psi(\mathbf{M}\mathbf{z}) = \Psi(\mathbf{z}) \Leftrightarrow \mathbf{M} = \mathbf{DP}$ .

Les conditions (C1) et (C2) signifient que l'on ne discrimine aucune solution parmi l'ensemble des solutions admissibles et que l'on impose donc un contraste indépendant des indéterminations d'ordre et d'échelle. La dernière condition indique qu'il s'agit de maximiser la fonction de contraste afin d'obtenir la matrice de rotation adéquate.

**Contrastes fondés sur l'information mutuelle :** dans la mesure définie dans (1.10), l'entropie différentielle des observations  $\mathbf{x}$  étant constante, alors la variation de  $\mathcal{I}(\mathbf{y})$  dépend uniquement de la matrice de séparation  $\mathbf{B}$  et de la densité de probabilité des sources recherchées. La séparation est réalisée par minimisation de cette information mutuelle par rapport à la matrice de séparation. Se pose alors le problème d'évaluation de l'information mutuelle qui nécessite l'estimation de la densité de probabilité des sources. Nous pouvons consulter à ce propos [Pha02, Pha04] où une approche semi-paramétrique est employée.

Dans le cas de variables aléatoires de distribution uniforme, la minimisation de l'information mutuelle correspond à la maximisation de l'entropie conjointe. Cette propriété est mise à profit en utilisant une fonction non linéaire  $\mathcal{G}(\cdot)$ , qui, appliquée aux signaux sources estimés  $\mathbf{y}_k$ , permet d'obtenir des signaux  $\tilde{\mathbf{y}}_k = \mathcal{G}(\mathbf{y}_k)$  de distribution uniforme. Autrement dit, une fonction non-linéaire correspondant à la fonction de répartition des sources. Pour séparer les sources il suffit alors de maximiser l'entropie conjointe de  $\mathbf{z}$  par rapport à la matrice séparante  $\mathbf{B}$ , d'où la dénomination *InfoMax* [BS95]. Le lien avec la méthode du maximum de vraisemblance est présenté dans [Mac96, Car97, OD98].

Sous la contrainte de blancheur spatiale des sources estimées, i.e.  $\mathbb{E}[\mathbf{y}_k \mathbf{y}_k^*] = \mathbf{I}_p$ , l'entropie conjointe de  $\mathbf{y}$  est invariante par transformation orthonormale. Alors, la minimisation de l'information mutuelle revient à la minimisation de l'entropie marginale des sources estimées. Cette approche nécessite également l'estimation de l'entropie marginale des sources. Nous pouvons consulter à ce propos [BV00, LMF03].

**Contrastes exploitant la décorrélation non-linéaire :** une autre définition équivalente de l'indépendance entre deux variables  $y_1$  et  $y_2$  est la suivante :

$$\mathbb{E}[f(y_1)g(y_2)] = \mathbb{E}[f(y_1)] \mathbb{E}[g(y_2)], \quad (1.22)$$

pour toutes fonctions continues  $f$  et  $g$ . C'est donc une généralisation de la définition de la décorrélation obtenue en prenant les deux fonctions  $f$  et  $g$  égales à l'identité. Le travail pionnier en analyse en composantes indépendantes de Jutten, Héroult et Ans [HA84, HJ85, JH91, Jut87], repris d'une manière plus générale dans le cadre des "fonctions d'estimation" par Amari et Cardoso [ACC97], s'appuie sur l'équation (1.22) comme condition d'équilibre d'un algorithme de séparation. Autrement dit, on construit en général un algorithme de type gradient proportionnel à la différence des deux termes de

l'équation (1.22) et on étudie *a posteriori* la convergence et la stabilité. La décorrélation non-linéaire peut être reliée à la définition des contrastes via les cumulants, si l'on considère des approximations polynômiales des fonctions non-linéaires utilisées dans ces algorithmes. D'autre part, Bach et al. [BJ02] ont présenté une famille d'algorithmes fondée sur la définition d'une mesure de dépendance utilisant la corrélation non-linéaire.

**Contrastes utilisant les statistiques d'ordre supérieur :** ces critères sont construits en utilisant les cumulants d'ordre supérieur ou égal à trois. La forme générale d'un contraste fondé sur les cumulants est exprimée par :

$$\psi_r^f(\mathbf{z}) = \sum_{i=1}^p f(C_r[z_i]), \quad \text{pour } r \geq 3 \quad (1.23)$$

où  $C_r[z_i] = \text{Cum}[\underbrace{z_i, \dots, z_i}_{r \text{ termes}}]$  et  $f$  est une fonction convexe strictement croissante. Pour  $f(u) = u^2$ , on retrouve le contraste  $\psi_r^2$  pour lequel il peut être montré que la fonction

$$\Phi_r(\mathbf{z}) = \sum_{i_1, i_2, \dots, i_r=1}^p (\text{Cum}[z_{i_1}, z_{i_2}, \dots, z_{i_r}])^2 \quad (1.24)$$

est invariante par transformation unitaire de  $\mathbf{z}$  [GL90b, Com94]. En d'autres termes  $\Phi_r(\mathbf{U}\mathbf{z}) = \Phi_r(\mathbf{z})$ , si  $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ . Ainsi la maximisation du contraste  $\psi_r^2$  est équivalente à la minimisation de la somme des carrés des cumulants croisés d'ordre  $r$ . Cette remarque permet de justifier, d'une part, le recours à l'optimisation de la fonction de contraste en utilisant la méthode de Jacobi pour la diagonalisation conjointe approchée de plusieurs matrices et, d'autre part, de définir une mesure directe de l'indépendance statistique des variables à l'ordre  $r$ . Une discussion plus détaillée sur l'utilisation explicite des cumulants pour la synthèse de fonctions de contraste est fournie dans [Car99a, MTM99, Mor01].

### 1.2.2.3 Séparation de sources par maximum de vraisemblance

L'approche du maximum de vraisemblance considère le problème d'identification de la matrice de mélange ou de la matrice de séparation, en vue de la reconstruction des sources [GL90b, PGJ92, Pha96]. L'estimation de la matrice de mélange au sens du maximum de vraisemblance est solution de

$$\hat{\mathbf{A}} = \arg \max_{\mathbf{A}} p(\mathbf{X}|\mathbf{A}). \quad (1.25)$$

En considérant, le modèle de mélange dans le cas carré et non bruité et des sources indépendantes et identiquement distribuées, l'approche du maximum de vraisemblance rejoint l'approche InfoMax [Car97, OD98]. Néanmoins, au lieu de fixer la forme de la non-linéarité, comme dans le cas de l'InfoMax, l'approche du maximum de vraisemblance offre l'avantage de pouvoir choisir *a priori* la forme des distributions des composantes indépendantes. Un autre avantage de cette approche provient de la possibilité de considérer le cas de mélanges bruités ou sur-déterminés [Bel95, Ber00]. De plus, l'hypothèse d'orthogonalité des sources peut être relâchée pour exploiter une éventuelle structure temporelle ou non-stationarité des sources ou toute autre information sur les sources [CP01].

### 1.2.2.4 Séparation de sources par approche bayésienne

Les premiers travaux concernant l'application de l'approche bayésienne en séparation de sources sont [Rob98, Knu98, MD99]. Dans le cadre de la séparation par maximum de vraisemblance aucune information spécifique n'est introduite sur la matrice de mélange ou sur la matrice de séparation. Cependant, en adoptant une telle approche le problème de séparation est implicitement traité dans un cadre bayésien. D'une part, le choix de la fonction non-linéaire ou du modèle de distribution des sources revient à spécifier la loi *a priori*  $p(\mathbf{S})$ . D'autre part, la vraisemblance  $p(\mathbf{X}|\mathbf{A})$  est obtenue par marginalisation par rapport aux sources. Concernant, les coefficients de mélange ou les coefficients de la matrice de séparation, l'approche du maximum de vraisemblance suppose une loi *a priori* uniforme. Par contre, pour tenir compte des informations supplémentaires éventuelles, on peut introduire une autre loi *a priori*  $p(\mathbf{A})$  sur la distribution des coefficients de mélange et former la densité *a posteriori* conjointe

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}) = p(\mathbf{X}|\mathbf{S}, \mathbf{A}) \times p(\mathbf{S}) \times p(\mathbf{A}). \quad (1.26)$$

L'estimation à partir de cette loi *a posteriori* peut être envisagée en utilisant plusieurs stratégies et différentes techniques de calcul bayésien.

### 1.2.3 Séparation de sources spectrales par analyse en composantes indépendantes

Afin d'illustrer la possibilité de séparer des sources spectrales en utilisant les méthodes d'analyse en composantes indépendantes, considérons deux sources de formes similaires aux signaux de spectroscopie et montrées dans la figure 1.2. Ces sources sont mélangées avec la matrice suivante

$$\mathbf{A} = \begin{bmatrix} 0.60 & 0.40 \\ 0.40 & 0.60 \end{bmatrix}. \quad (1.27)$$

Le coefficient de corrélation entre les deux sources est de l'ordre de  $-0.005$ , ce qui indique que ces sources peuvent être considérées orthogonales.

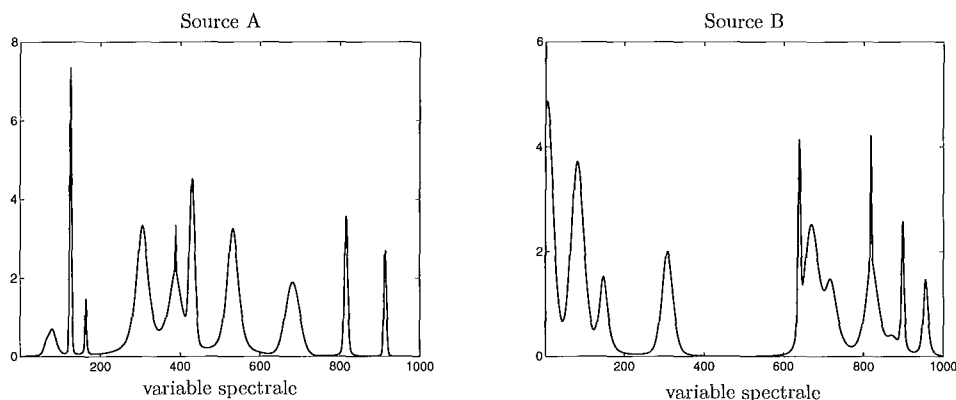


FIG. 1.2 – Sources spectrales simulées.

Le tableau 1.1 résume les performances de la séparation en utilisant quelques méthodes classiques. Les indices de performances utilisés sont définis dans le chapitre 5. D'autres méthodes peuvent aussi

bien être appliquées, mais seulement les résultats de cinq algorithmes sont présentés. Nous pouvons noter que toutes les méthodes appliquées atteignent un niveau de performance très satisfaisant, ce qui indique que la seule hypothèse d'indépendance suffit pour réaliser la séparation de telles sources. De plus, le rapport signal à distorsion sur chacune des sources estimées est très grand, ce qui indique une bonne estimation des signaux sources. Ainsi, la positivité est implicitement assurée durant la séparation. Ce premier exemple montre que la non-négativité est assurée implicitement dans le cas où les signaux sources sont indépendants, c'est à dire que les composantes indépendantes du mélange coïncident avec les sources recherchées. Or, dans le cas des signaux spectroscopiques cette hypothèse peut ne pas être vérifiée, notamment dans le cas de mélange d'espèces qui présentent des spectres purs très similaires.

TAB. 1.1 – Performances de la séparation en utilisant quelques algorithmes d'analyse en composantes indépendantes.

Indice	SOBI	JADE	ICA-Comon	FastICA	InfoMax
$RSD_{S_1}$	40.93	41.42	42.30	33.94	38.35
$RSD_{S_2}$	43.65	38.19	38.90	36.61	39.93
$IP$	-43.22	-38.16	-39.35	-35.67	-37.31

Afin d'analyser l'effet de la corrélation mutuelle des sources sur la séparation, considérons des mélanges avec des sources spectrales simulées avec une corrélation mutuelle significative. Le premier cas est celui de sources avec un coefficient de corrélation positif et le deuxième cas est celui avec un coefficient de corrélation négatif. En analysant les sources comme des observations pour lesquelles une méthode d'analyse en composantes indépendantes est appliquée, on obtient des composantes qui sont orthogonales. Les résultats de cette analyse sont montrés dans la figure 1.4. L'analyse des sources déduites montre que celles-ci peuvent être négatives lorsque le coefficient de corrélation mutuelle des vraies sources est positif. Ceci s'explique par le fait que la matrice appliquée pour la séparation transforme les axes principaux pour passer d'un angle aigu à un angle droit, basculant ainsi certains échantillons vers les parties négatives. De façon similaire, dans le cas de sources à coefficient de corrélation négatif, la matrice séparante préservera leur positivité mais l'application de sa transformation inverse aux coefficients de mélange tendra à rendre ces coefficients négatifs. Ainsi l'application d'une méthode d'analyse en composantes indépendantes à des observations obtenues à partir de sources positives mutuellement corrélées tendra à produire des sources négatives et/ou des coefficients de mélange négatifs.

## 1.2.4 Quelques approches proposées pour améliorer les performances

### 1.2.4.1 Séparation à partir d'une décomposition en sous bandes

Cette approche présentée par [CG03] suppose que les signaux sources sont à large bande et présentent des corrélations mutuelles importantes, mais ces sources sont mutuellement indépendants sur certaines sous bandes. Les signaux sources sont alors décomposés en sous bandes et la séparation est réalisée



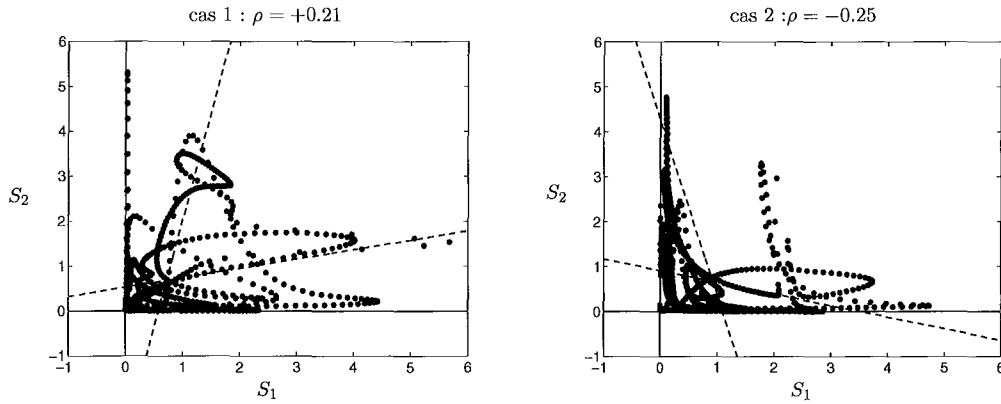
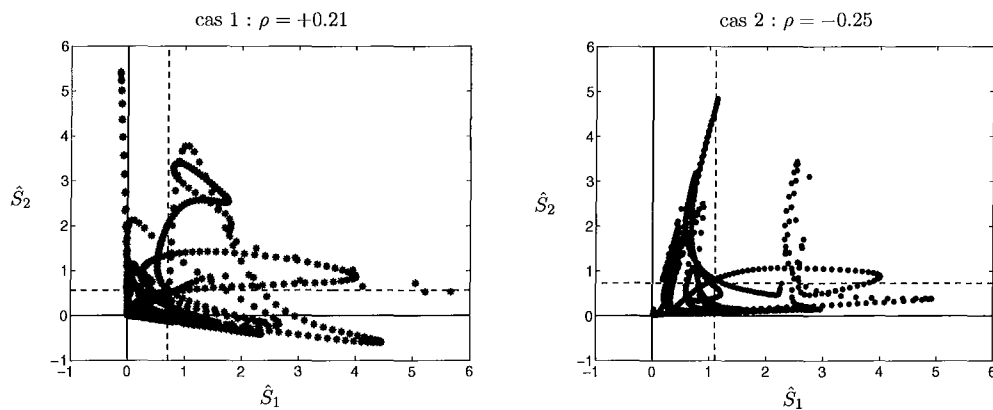
FIG. 1.3 – Diagramme de dispersion (*scatter plot*) ou nuage de points de sources non-orthogonales.

FIG. 1.4 – Diagramme de dispersion des composantes indépendantes.

en utilisant un algorithme d'analyse en composantes indépendantes classique sur une sous bande où les sources sont indépendantes. Néanmoins, cette approche requiert la connaissance au préalable ou la sélection manuelle de cette sous bande. La solution consiste alors à effectuer la recherche de la matrice de séparation en filtrant les observations par un banc de filtres passe-bandes. Un exemple simple est celui où chaque source peut être décomposée en une composante basses fréquences  $s_{(j,k)}^b$  et une deuxième composante hautes fréquences  $s_{(j,k)}^h$

$$s_{(j,k)} = s_{(j,k)}^b + s_{(j,k)}^h \quad \text{pour } j = 1, \dots, p, \quad (1.28)$$

et les composantes hautes fréquence ont tendance à être plus indépendantes que les sources. Ainsi, pour effectuer la séparation il suffit d'appliquer un filtre passe haut aux observations et ensuite appliquer une méthode classique d'analyse en composantes indépendantes afin d'obtenir la matrice de mélange. Cette approche est illustrée sur un exemple synthétique dans [CG03]. Afin de surmonter la difficulté du choix de la sous bande adéquate, on peut imaginer d'exploiter la contrainte de non-négativité en choisissant une structure de filtre passe-bande linéaire et rechercher les paramètres optimaux de ce filtre (bande passante et fréquence centrale) afin d'obtenir des sources positives. Ceci fournira une méthode de sélection automatique de la sous bande adéquate.

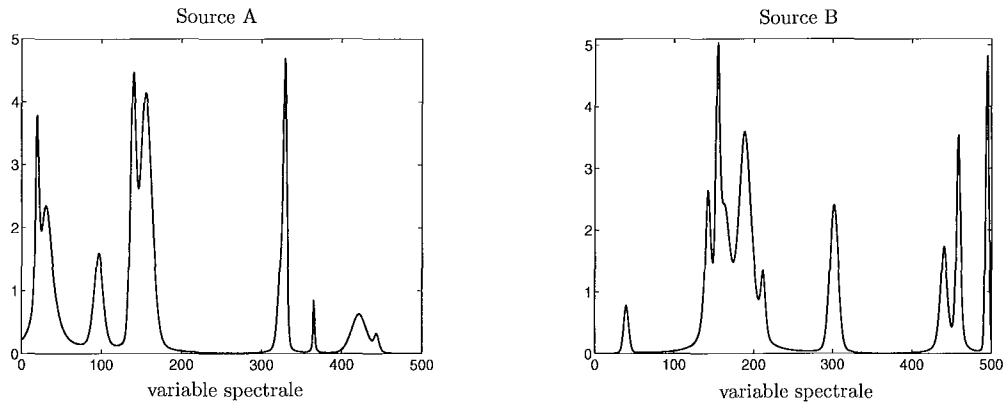
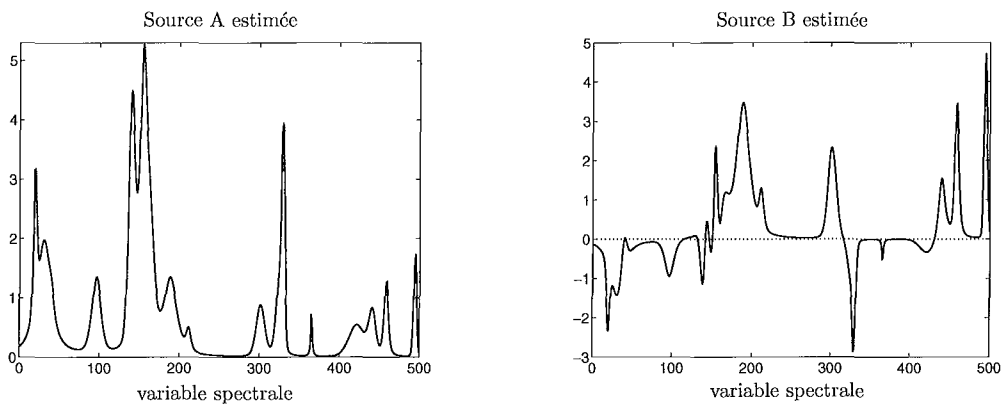
### 1.2.4.2 Séparation à partir des dérivées des observations

Cette approche consiste à réaliser la séparation à partir de la dérivée première ou seconde des observations. Une méthode présentée par Hyvarinen [Hyv98] et dont l'objectif est de réduire la corrélation temporelle des sources et de réaliser la séparation à partir de leurs processus d'innovations. La justification de cette approche vient du fait que les processus d'innovations sont plus indépendants et plus non-gaussiens que les signaux sources [Kai68, Hay96], ce qui conduit donc à une meilleure estimation de la matrice de mélange. Cette approche a été utilisée pour la séparation de signaux spectroscopiques dans [CW01, BN05]. Le recours à l'hypothèse d'indépendance des dérivées premières ou secondes des sources revient à supposer des sources ayant une structure markovienne et à formuler l'hypothèse d'indépendance mutuelle à partir de leurs densités conditionnelles. Cette méthode peut donc être vue comme un cas particulier de l'approche proposée dans [PG97, HJP03], car la même structure "temporelle" est considérée pour toutes les sources. Afin d'illustrer cette approche, considérons un exemple simple avec deux sources et deux mélanges obtenus en utilisant une matrice de mélange positive choisie arbitrairement. La séparation est réalisée en appliquant l'algorithme FastICA sur les observations et sur leurs dérivées secondes. La matrice de séparation estimée est ensuite utilisée pour estimer les sources. D'après la figure 1.5 on constate l'amélioration de la qualité de reconstruction des sources en réalisant l'estimation de la matrice de séparation à partir des dérivées secondes des observations. Cependant, cet exemple reste académique car en pratique il faut déterminer non seulement l'ordre du filtre de dérivation mais aussi choisir une fréquence de coupure du filtre de régularisation pour éviter l'amplification du bruit. Dans ce cas, un lien entre cette approche avec la méthode de décomposition en sous bandes peut être établi. En effet, le filtre dérivateur régularisé est un filtre passe bande dont la fréquence centrale et la bande passante dépendent de l'ordre de la dérivée et de la fréquence de coupure du filtre de régularisation. Ainsi l'utilisation de cette approche rejoint la méthode de décomposition en sous-bandes proposée par Cichocki et al [CG03].

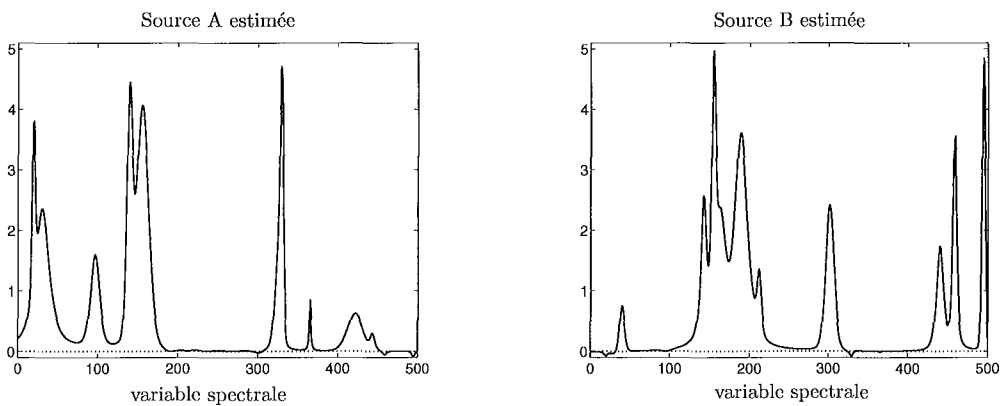
### 1.2.4.3 Analyse en composantes indépendantes non-négatives

Cette approche a pour point de départ la prise en compte simultanée de la contrainte de non-négativité et de l'hypothèse d'indépendance statistique des sources. Pour garantir la séparabilité en utilisant une approche d'analyse en composantes indépendantes non-négatives, deux conditions sont nécessaires [Plu02] : (i) Les sources doivent être orthogonales. Cette condition est une conséquence directe de l'hypothèse d'indépendance statistique. (ii) Les sources doivent avoir des distributions non nulles au voisinage positif du point zéro (sources dites *well grounded*). Comme la plupart des méthodes d'analyse en composantes indépendantes, cette approche se fonde d'abord sur une étape de blanchiment spatial préalable des observations avant de rechercher une matrice de rotation qui optimise un critère permettant de prendre en compte explicitement la contrainte de non-négativité des composantes indépendantes. Le critère minimisé dans le cadre de la méthode NNICA (pour *Non-negative independent Component Analysis*), est donné par la formule suivante [Plu03]

$$C_{nnica}(\mathbf{B}) = \frac{1}{n} \sum_{k=1}^n (\mathbf{z}_k - \mathbf{B}^T \mathbf{y}_k^+)^2, \quad (1.29)$$

(a) Signaux sources avec un coefficient de corrélation  $\rho=+0.28$ .

(b) Composantes indépendantes estimées à partir des mélanges bruts



(c) Sources estimées en utilisant une analyse en composantes indépendantes sur les dérivées secondes

FIG. 1.5 – Avantage de la séparation à partir des dérivées secondes des observations.

où les éléments du vecteur  $\mathbf{y}_k^+$  sont donnés par la  $\{y_{(j,k)}^+ = \max\{0, y_{(j,k)}\}; \forall j = 1, \dots, p\}$ . La méthode consiste donc en la recherche d'une matrice  $\mathbf{B}$  orthonormale qui minimise ce critère. Nous pouvons noter que le critère (1.29) devient nul pour  $\{y_k \geq 0, \forall k = 1, \dots, n\}$ . Récemment, ce critère a été formulé à partir d'une méthode d'analyse en composantes principales non-linéaire [PO04], qui, pour certains types de fonctions non-linéaires, est reliée également aux méthodes d'analyse en composantes indépendantes. Nous pouvons consulter pour cela [KPO98]. Cependant cette approche requiert la

contrainte d'orthogonalité des sources, ce qui est la difficulté principale posée par les sources spectroscopiques, donc la méthode NNICA n'est pas adaptée dans le cas de sources non-orthogonales.

### 1.3 Séparation par prise en compte de la non-négativité

En fonction de la manière avec laquelle cette contrainte est prise en compte, les méthodes de séparation de sources spectrales peuvent être classées en deux catégories. La première est algébrique alors que la deuxième se fonde sur une estimation par des moindres carrés sous la contrainte de non-négativité.

#### 1.3.1 Approche algébrique

Cette approche est issue des travaux de Lawton et Sylvestre [LS71], et, comme illustré par la figure 1.6, réalise la séparation en deux étapes. La première est une étape d'analyse en composantes principales des observations et la deuxième consiste à appliquer une transformation linéaire aux composantes principales afin d'obtenir des signaux sources et des coefficients de mélange non-négatifs. Lawton et Sylvestre ont développé une analyse complète du cas de deux sources et plus tard une analyse du cas de trois sources a été présentée par Ohta [Oht73] qui a utilisé une méthode de Monte Carlo pour la recherche des matrices de transformations qui mènent à des solutions non-négatives. Borgen et al. [BK85, BDMO86] ont proposé une formulation géométrique du cas de plusieurs sources et particulièrement l'étude des régions admissibles des paramètres de la matrice de transformation.

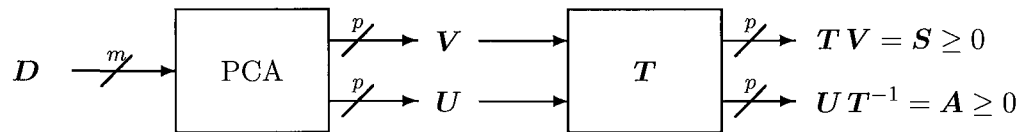


FIG. 1.6 – Approche de Lawton-Sylvestre pour la séparation de sources spectrales

**Remarque 1** Cette approche à deux étapes est similaire aux méthodes algébriques de séparation de sources. La différence réside dans le fait que la matrice de transformation n'est plus une matrice de rotation unitaire mais plutôt une transformation linéaire permettant de préserver la norme  $L_1$  ou  $L_2$  des sources, afin de lever l'indétermination d'échelle.

##### 1.3.1.1 Minimisation d'un critère de positivité

Afin de rechercher une matrice de transformation permettant d'assurer la positivité des sources et des coefficients de mélange, Sasaki et al. [SKM83] proposent une méthode fondée sur la minimisation du critère suivant

$$C_{pos}(\mathbf{T}) = \sum_{j=1}^p \sum_{k=1}^n f(s_{(j,k)}) + \sum_{i=1}^m \sum_{j=1}^p f(a_{i,j}) \quad (1.30)$$

sous la contrainte  $\sum_{k=1}^n s_{(j,k)} = 1$ , pour lever l'indétermination d'échelle. La fonction de coût  $f$  est définie de telle sorte à pénaliser les valeurs négatives

$$f(x) = \begin{cases} x^2 & \text{si } x < 0, \\ 0 & \text{sinon.} \end{cases} \quad (1.31)$$

Bien que le minimum de ce critère permet d'obtenir une solution non-négative ou la plus non-négative possible, sa minimisation n'est pas aisée et requiert l'utilisation d'une méthode d'optimisation non-linéaire. Dans le papier de Sasaki *et al.* [SKM83] un algorithme du simplexe est employé pour obtenir toutes les solutions qui minimisent ce critère. Ces solutions permettent de retrouver les résultats de Lawton et Sylvestre à savoir l'existence de plusieurs solutions admissibles et de rechercher les solutions admissibles d'une façon différente de celle de Borgen *et al.* [Oht73, BK85].

### 1.3.1.2 Minimisation d'un critère de positivité et de minimum d'entropie

Vu que la seule prise en compte de la non-négativité ne permet pas d'assurer l'unicité de la solution et afin de réduire l'espace des solutions admissibles, des critères de régularisation supplémentaires sont ajoutés. Une approche plus intéressante est celle de Sasaki *et al.* [SKM84, SKM89], qui consiste à minimiser un critère de la forme suivante

$$C_{tot}(\mathbf{T}) = C_{pos}(\mathbf{T}) + \gamma \sum_{j=1}^p \sum_{k=1}^n h(s_{(j,k)}), \quad (1.32)$$

où la fonction de coût  $h(s_{(j,k)})$  est définie par

$$h(s_{(j,k)}) = -\frac{|s_{(j,k)}^{(d)}|}{\sum_{k=1}^n |s_{(j,k)}^{(d)}|} \log \frac{|s_{(j,k)}^{(d)}|}{\sum_{k=1}^n |s_{(j,k)}^{(d)}|} \quad (1.33)$$

où l'exposant  $d$  signifie l'ordre de la différentiation. Le coefficient  $\gamma$  permet d'ajuster le compromis entre les deux critères. Le choix du critère entropique signifie que l'on cherche à obtenir une solution dans laquelle les signaux sources sont les plus simples possible et la contrainte de douceur est introduite par le calcul du critère entropique sur la dérivée du signal source. La minimisation de ce critère par rapport à la matrice de transformation optimale est réalisée en utilisant un algorithme du simplexe sous la contrainte  $\sum_{k=1}^n s_{(j,k)}^2 = 1$ . Récemment cette approche a été réexaminée par [NSK89, SBG99] et dans [YG98, WG03] un algorithme de recuit simulé est employé pour l'optimisation. Néanmoins, l'étude de convergence n'est pas complètement établie car aucune discussion sur la convexité du critère n'est présentée.

### 1.3.2 Séparation par moindres carrés sous contrainte de non-négativité

Cette approche se fonde sur une estimation de moindres carrés sous la contrainte de non-négativité. Il s'agit donc de la résolution du problème

$$(\hat{\mathbf{S}}, \hat{\mathbf{A}}) = \arg \min_{\mathbf{S}, \mathbf{A}} \mathcal{Q}(\mathbf{A}, \mathbf{S}), \quad \text{sous les contraintes : } \mathbf{S} \geq 0, \mathbf{A} \geq 0, \quad (1.34)$$

où

$$\mathcal{Q}(\mathbf{A}, \mathbf{S}) = \sum_{i=1}^m \sum_{k=1}^n \left( x_{(i,k)} - \sum_{j=1}^p a_{ij} s_{(j,k)} \right)^2, \quad (1.35)$$

est le critère des moindres carrés. En fonction de la manière avec laquelle cette contrainte de non-négativité est introduite durant l'optimisation on distingue deux algorithmes.

### 1.3.2.1 Méthode ALS

Cet algorithme (dit ALS pour *Alternating Least Squares*), présenté par [TKF93, TIRC93], réalise la décomposition en alternant la minimisation par rapport à la matrice de mélange et par rapport aux signaux sources. A chaque nouvelle itération, le critère des moindres carrés est minimisé sous la contrainte de non-négativité [LH74, BD97]. L'initialisation de ALS peut être réalisée en utilisant une méthode de décomposition sans contrainte de non-négativité (PCA, ICA). Une initialisation possible consiste à réaliser la décomposition en utilisant une méthode d'analyse en composantes indépendantes et ensuite d'utiliser l'ALS pour assurer la positivité des estimations. Cette approche a été utilisée par Nuzillard [NN99]. Dans le cadre du traitement des signaux spectroscopiques, plusieurs techniques chimiométriques sont employées. Nous pouvons citer la méthode dite SIMPLISMA [WG91] (SIMPLe to use Interactive Self-Modeling Mixture Analysis) qui se base sur la notion de variable pure, qui est définie comme étant une fréquence pour laquelle une seule source est observée. Ainsi pour  $p$  sources,  $p$  variables pures sont nécessaires pour réaliser la séparation. Une autre approche dite OPA [STvdBM96] (pour *Orthogonal Projection Approach*) suppose que les observations les plus décorrélatées correspondent aux sources. Cette hypothèse suppose donc que pour chaque signal source, il existe au moins une observation dans laquelle uniquement cette source est présente. Nous signalons qu'il existe plusieurs variantes de méthodes issues des approches d'analyse factorielle. Pour une présentation plus détaillée nous pouvons consulter [dYTM97, Mal02, JLO04].

### 1.3.2.2 Méthode NMF

Cet algorithme (dit NMF pour *Non-negative Matrix Factorization*) [LS99] minimise le critère des moindres carrés en utilisant un algorithme de descente du gradient alterné par rapport aux sources et par rapport aux coefficients de mélange. Un choix particulier du pas de mise à jour transforme cet algorithme du gradient en algorithme du point fixe tout en préservant la non-négativité des estimations. Cet algorithme est repris dans le cas d'observations bruitées par Sajda et al. [SDB<sup>+</sup>03, SSB<sup>+</sup>04]. La méthode NMF présente l'avantage de simplicité d'implantation et la rapidité de convergence, mais souffre du problème de non-unicité de la solution car le critère de moindres carrés sous contrainte de non-négativité admet plusieurs solutions possibles.

## 1.3.3 Séparation par moindres carrés pénalisés

Cette approche consiste à estimer directement les signaux sources et la matrice de mélange en minimisant un critère composite dans lequel un premier terme mesure l'adéquation aux données et un

deuxième terme permet de prendre en compte l'information *a priori* sur les sources et sur les coefficients de mélange. En d'autres termes, le critère minimisé est de la forme

$$\mathcal{C}(\mathbf{A}, \mathbf{S}) = \mathcal{Q}(\mathbf{A}, \mathbf{S}) + \mathcal{R}(\mathbf{A}) + \mathcal{R}(\mathbf{S}). \quad (1.36)$$

où le critère d'adéquation aux données est exprimé par l'équation (1.35).

**Remarque 2** Les méthodes fondées sur une estimation par moindres carrés pénalisés et la prise en compte explicite de la non-négativité peuvent être formulées à partir d'une approche bayésienne pour la séparation de sources.

Les différentes méthodes de cette famille diffèrent par les choix des critères de régularisation ainsi que par l'algorithme choisi pour l'optimisation du critère.

### 1.3.3.1 Méthode PMF

Dans cette approche (dite PMF pour *Positive Matrix factorization*), est présentée par Paatero [PT94, Paa97], minimise un critère composite dans lequel la partie régularisation se présente sous la forme

$$\mathcal{R}_{pmf}(\mathbf{A}) = \alpha_1 \sum_{j=1}^p \sum_{i=1}^m \log a_{ij} + \alpha_2 \sum_{j=1}^p \sum_{i=1}^m a_{ij}^2 \quad (1.37)$$

pour les coefficients de mélange et

$$\mathcal{R}_{pmf}(\mathbf{S}) = \alpha_3 \sum_{j=1}^p \sum_{k=1}^n \log s_{(j,k)} + \alpha_4 \sum_{j=1}^p \sum_{k=1}^n s_{(j,k)}^2, \quad (1.38)$$

pour les signaux sources. Dans la version originale de cette méthode une méthode de Gauss-Newton est utilisée pour l'optimisation. Dans le chapitre 4, nous discutons plus en détail cette approche et nous présentons une formulation bayésienne de ce critère.

### 1.3.3.2 Méthode NNSC

Dans cette approche (dite NNSC pour *Non-Negative Sparse Coding*) [Hoy02], on cherche à minimiser un critère composite dans lequel seulement les sources sont pénalisées. Le critère de régularisation favorise les solutions parcimonieuses et s'écrit sous la forme

$$\mathcal{R}_{nns}(\mathbf{S}) = \lambda \sum_{j=1}^p \sum_{k=1}^n s_{(j,k)}. \quad (1.39)$$

Le choix d'un tel critère de régularisation correspond à supposer une densité exponentielle à support positif, comme loi de distribution des signaux sources. Nous revenons également à ce point plus en détail dans le chapitre 4.

## 1.4 Conclusions

Dans ce chapitre nous avons présenté succinctement quelques méthodes de séparation de sources. L'objectif étant d'étudier la possibilité d'appliquer ces méthodes à la séparation de sources spectroscopiques. Deux points cruciaux ont été relevés durant cette analyse : compte tenu de la similarité des structures moléculaires de certaines substances, les sources spectroscopiques peuvent présenter des corrélations mutuelles importantes, ce qui met en échec les méthodes d'analyse en composantes indépendantes. De plus, le problème de séparation sous la seule contrainte de non-négativité n'admet pas une solution unique. Afin de réaliser une analyse plus précise du problème, nous étudions dans le chapitre 2 les conditions sous lesquelles la séparation sous contrainte de non-négativité admet une solution unique. Nous avons également souligné que l'approche du maximum de vraisemblance permet de relâcher la contrainte d'orthogonalité des sources. Rappelons que cette contrainte est le point faible des méthodes d'analyse en composantes indépendantes. Nous présentons alors dans le chapitre 3 une extension de cette approche à la séparation de sources spectroscopiques. Le résultat intéressant est de montrer que cette approche peut être appliquée avec succès, sous la condition d'inclure dans l'étape de sélection de la distribution des sources, la contrainte de non-négativité via le choix d'une densité à support non-négatif. Bien que cette approche du maximum de vraisemblance soit applicable dans un contexte non bruité, elle présente la limitation de ne pas imposer la non-négativité des coefficients de mélange. Dans le chapitre 4, nous développons une méthode fondée sur l'approche estimation bayésienne qui permet de prendre en compte conjointement la non-négativité des signaux sources et des coefficients de mélange.



## Chapitre 2

# Solutions admissibles et conditions d'unicité de la solution

Ce chapitre présente une étude des conditions d'unicité de la solution du problème de séparation de sources sous la seule contrainte de non-négativité. Dans un premier temps, nous analysons le cas de deux sources pour lequel nous déterminons les solutions admissibles et nous formulons des conditions nécessaires et suffisantes pour l'unicité de la solution. Ensuite, nous considérons le cas de plus de deux sources et nous présentons des conditions nécessaires et des conditions suffisantes pour l'unicité de la solution. Les solutions admissibles sont également déterminées dans le cas où la solution n'est pas unique. Finalement, nous discutons l'adéquation des méthodes de séparation de sources en fonction de la satisfaction ou non des conditions d'unicité et des hypothèses d'orthogonalité.

### 2.1 Introduction

Le problème de séparation de sources suppose que  $m$  signaux mesurés  $\{x_{(i,k)}, k = 1, \dots, n\}_{i=1}^m$ , sur  $n$  échantillons, sont une combinaison linéaire de  $p$  composantes non-observables  $\{s_{(j,k)}, k = 1, \dots, n\}_{j=1}^p$ . Le modèle de mélange est exprimé par

$$\mathbf{X} = \mathbf{A} \mathbf{S}, \quad (2.1)$$

où  $\mathbf{X}$  est la matrice ( $m \times n$ ) des données,  $\mathbf{A}$  est la matrice de mélange, de dimension ( $m \times p$ ),  $\mathbf{S}$  est matrice ( $p \times n$ ) des signaux sources. Dans ce chapitre, nous considérons que le niveau de bruit sur les observations est négligeable et les contraintes principales sont

$$\forall i, j, a_{ij} \geq 0 \text{ et } \forall j, k, s_{(j,k)} \geq 0, \quad (2.2)$$

qui seront notées par

$$\mathbf{A} \geq 0 \text{ et } \mathbf{S} \geq 0. \quad (2.3)$$

En considérant ces contraintes, ce problème inverse peut être formulé comme suit : à partir des données  $\mathbf{X}$ , estimer conjointement les matrices  $\mathbf{A}$  et  $\mathbf{S}$  qui suivent le modèle (2.1) et qui respectent les contraintes (2.3). Dans un contexte où l'on cherche à reproduire exactement la matrice des données selon le modèle (2.1), cette formulation rejoint la définition du problème de factorisation d'une matrice

non-négative en un produit de matrices non-négatives. Les premières publications sur ce sujet sont apparues essentiellement dans le domaine de l'algèbre linéaire [Mar72, Tho74, CP81, Che84, CR93]. Plus récemment, ce problème est repris avec une formulation géométrique dans [VV99] où la terminologie de factorisation en matrices positives (*Positive Matrix Factorization*) est utilisée, même si un peu auparavant Paatero a employé la même terminologie (PMF) [PT94, Paa97] pour une méthode destinée à la factorisation approchée d'une matrice de données spectrales selon le modèle (2.1). Dans un papier récent, Lee et Seung [LS99] ont également présenté deux algorithmes de factorisation en matrices non-négatives.

Avant d'appliquer ou de développer un algorithme pour une telle factorisation, il est important de répondre aux questions liées aux indéterminations du modèle et à l'unicité de la solution sous la seule contrainte de non-négativité. De plus, lorsque la solution n'est pas unique il est utile d'exprimer toutes les solutions admissibles. Une condition nécessaire et suffisante pour que la factorisation d'une matrice non-négative en matrices non-négatives soit unique est formulée dans [Che84], mais celle-ci est inexploitable pour vérifier *a priori* si une matrice non-négative admet une factorisation non-négative unique. Park et al. [PSH02] et Smilde et al [SHK<sup>+</sup>01] ont exprimé des conditions suffisantes pour l'unicité de la solution, mais ces conditions sont très restrictives et ne seront satisfaites que dans des contextes très particuliers. Le résultat le plus complet est celui de Donoho et Stodden [DS03] qui ont étudié les situations dans lesquelles la factorisation en matrices non-négatives n'admet pas une solution unique et ont formulé une condition suffisante plus flexible.

L'objectif de ce chapitre est d'analyser de plus près les méthodes de séparation fondées sur la prise en compte de la non-négativité pour essayer de spécifier les situations dans lesquelles ces méthodes peuvent être appliquées avec succès et aussi analyser les cas où leur application serait infructueuse. Pour atteindre ces objectifs, nous étudions dans un premier temps les solutions admissibles sous la seule contrainte de non-négativité et nous discutons, ensuite, les conditions (nécessaires et/ou suffisantes) pour l'unicité de la solution.

## 2.2 Formulation du problème

Considérons le modèle de mélange et introduisons une matrice non-singulière, notée  $\mathbf{T}$  et de dimension  $(p \times p)$ . Pour chaque couple  $(\mathbf{A}, \mathbf{S})$ , un nouveau couple  $(\tilde{\mathbf{A}}, \tilde{\mathbf{S}})$  peut être défini par

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}^{-1}, \quad (2.4)$$

$$\tilde{\mathbf{S}} = \mathbf{T}\mathbf{S}, \quad (2.5)$$

de sorte à ne pas modifier la matrice de données, c'est-à-dire  $\mathbf{X} = \tilde{\mathbf{A}}\tilde{\mathbf{S}}$ . Dans le cas où aucune contrainte n'est imposée, cette transformation indique l'existence d'une infinité de factorisations possibles de la matrice de données. Mais, dans le cadre de la séparation de sources non-négatives une transformation n'est possible que si les matrices transformées  $\tilde{\mathbf{A}}$  et  $\tilde{\mathbf{S}}$  obéissent les deux conditions

$$\tilde{\mathbf{A}} \geq 0 \text{ et } \tilde{\mathbf{S}} \geq 0. \quad (2.6)$$

Ainsi, les questions qui se posent sont les suivantes : (i) Quelles sont les conditions sur les sources et sur les coefficients de mélange pour que la décomposition (2.1) qui respecte les contraintes de non-négativité (2.3) soit unique? (ii) Si la factorisation n'est pas unique quelles sont les solutions admissibles? (iii) Peut-t-on définir parmi ces solutions admissibles la solution la plus plausible?

Dans ce qui suit nous allons essayer d'apporter des éléments de réponses aux deux premières questions. Le chapitre 4 est une tentative de répondre à la troisième question. Mais avant d'aller plus loin, nous formulons les indéterminations d'ordre et d'échelle dans le cas de la séparation de sources non-négatives.

### 2.2.1 Indétermination d'échelle

Soit la transformation définie par  $\mathbf{T} = \text{Diag}(t_1, \dots, t_p)$ , avec  $t_i > 0$ , et  $\mathbf{T}^{-1} = \text{Diag}(1/t_1, \dots, 1/t_p)$ . L'application des matrices  $\mathbf{T}$  et  $\mathbf{T}^{-1}$  à  $\mathbf{S}$  et  $\mathbf{A}$ , respectivement, selon (2.5) et (2.4) correspond juste à une dilatation d'échelle. Il s'en suit que  $\tilde{\mathbf{A}} \geq 0$ ,  $\tilde{\mathbf{S}} \geq 0$  si  $\mathbf{A} \geq 0$ ,  $\mathbf{S} \geq 0$ . Généralement, cette indétermination est levée en supposant des sources à variance unité. Dans ce qui suit, et sans perdre en généralité, l'indétermination d'échelle est levée en supposant que toutes les sources possèdent la même norme  $L_1$ .

### 2.2.2 Indétermination d'ordre

Soit  $\mathbf{T}$  la transformation qui correspond à la permutation de la  $i$ ème ligne avec la  $j$ ème ligne de la matrice  $\mathbf{S}$ . La matrice  $\mathbf{T}^{-1}$  correspond à la permutation de la  $i$ ème colonne avec la  $j$ ème colonne de la matrice de mélange. Ces matrices sont composées de 1 et 0, alors  $\mathbf{A} \geq 0$ ,  $\mathbf{S} \geq 0$  impliquent  $\tilde{\mathbf{A}} \geq 0$ ,  $\tilde{\mathbf{S}} \geq 0$ . L'indétermination d'ordre signifie simplement qu'on ne connaît pas *a priori* l'ordre des signaux sources.

### 2.2.3 Définitions

**Définition 1** Une paire  $(\mathbf{A}, \mathbf{S})$  est dite solution admissible si conjointement  $\mathbf{A}$  et  $\mathbf{S}$  satisfont (2.3) et donnent une factorisation exacte de  $\mathbf{X}$  selon le modèle (2.1).

**Définition 2** Pour une matrice de donnée  $\mathbf{X}$ , la solution de (2.1) sous les contraintes (2.3) est dite unique si et seulement si les seules sources d'ambiguïtés sont les indéterminations d'ordre et d'échelle. On peut également parler de solution "quasi-unique".

## 2.3 Solutions admissibles et conditions d'unicité dans le cas de deux sources

Dans le cas de deux sources, les matrices  $\mathbf{A}$  et  $\mathbf{S}$  peuvent être représentées par

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2] \quad \text{et} \quad \mathbf{S} = [\mathbf{s}_1 \ \mathbf{s}_2]^T, \quad (2.7)$$

où  $\{\mathbf{a}_1, \mathbf{a}_2\}$  sont des vecteurs de dimension  $(m \times 1)$  et  $\{\mathbf{s}_1, \mathbf{s}_2\}$  sont des vecteurs de dimension  $(n \times 1)$ .

### 2.3.1 Solutions admissibles

Afin de s'affranchir de l'indétermination d'échelle, les sources sont supposées de même norme  $L_1$ , fixée à l'unité. Ce choix permet de simplifier les calculs dans ce chapitre, alors que la convention de sources à variance unité est souvent utilisée en séparation de sources. La convention qu'avant et après transformation tous les signaux sources ont la une norme  $L_1$ , fixée à l'unité,

$$\sum_{k=1}^n s_{(i,k)} = 1, \quad \forall i \quad (2.8)$$

permet d'écrire

$$\sum_{k=1}^n \tilde{s}_{(j,k)} = \sum_{k=1}^n \sum_{i=1}^p t_{ji} s_{(i,k)} = \sum_{i=1}^p \left( t_{ji} \sum_{k=1}^n s_{(i,k)} \right), \quad (2.9)$$

où  $t_{ji}$  sont les éléments de la matrice  $\mathbf{T}$ . Par conséquent, nous déduisons la condition suivante sur la matrice de transformation

$$\sum_{i=1}^p t_{ji} = 1, \quad \forall j = 1, \dots, p. \quad (2.10)$$

Afin d'analyser les indéterminations du modèle de mélange dans le cas de deux sources, nous introduisons la matrice suivante

$$\mathbf{T}(\alpha, \beta) = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}, \quad (2.11)$$

dont l'inverse est

$$\mathbf{T}^{-1}(\alpha, \beta) = \frac{1}{1 - \alpha - \beta} \begin{bmatrix} 1 - \beta & -\alpha \\ -\beta & 1 - \alpha \end{bmatrix}. \quad (2.12)$$

La matrice  $\mathbf{T}(\alpha, \beta)$  transforme les sources  $\mathbf{s}_1, \mathbf{s}_2$  en  $\tilde{\mathbf{s}}_1, \tilde{\mathbf{s}}_2$  selon

$$\tilde{\mathbf{s}}_1 = (1 - \alpha)\mathbf{s}_1 + \alpha\mathbf{s}_2 \quad \text{et} \quad \tilde{\mathbf{s}}_2 = \beta\mathbf{s}_1 + (1 - \beta)\mathbf{s}_2.$$

Afin d'éviter l'indetermination d'ordre, nous considérons que la première source transformée est déduite du rajout d'une partie de  $\mathbf{s}_2$  à  $\mathbf{s}_1$  et de façon similaire, la deuxième source transformée résulte de l'ajout d'une partie de  $\mathbf{s}_1$  à  $\mathbf{s}_2$ . Cette convention revient à supposer pour la première source que  $(1 - \alpha) > \beta$  et pour la seconde source que  $\alpha < (1 - \beta)$ , ce qui mène donc à la contrainte  $(\alpha + \beta) < 1$ . De plus, une telle contrainte assure que  $\mathbf{T}$  soit inversible.

En utilisant la transformation définie par (2.11), la contrainte de non-négativité des deux sources transformées permet d'écrire

$$\forall k = 1, \dots, n, \quad \tilde{s}_{(1,k)} = (1 - \alpha)s_{(1,k)} + \alpha s_{(2,k)} \geq 0, \quad (2.13)$$

$$\forall k = 1, \dots, n, \quad \tilde{s}_{(2,k)} = (1 - \beta)s_{(2,k)} + \beta s_{(1,k)} \geq 0. \quad (2.14)$$

La non-négativité des coefficients de mélange transformés permet d'écrire aussi

$$\forall \ell = 1, \dots, m, \quad \tilde{a}_{\ell 1} = \frac{1}{1 - \alpha - \beta} [(1 - \beta)a_{\ell 1} - \beta a_{\ell 2}] \geq 0, \quad (2.15)$$

$$\forall \ell = 1, \dots, m, \quad \tilde{a}_{\ell 2} = \frac{1}{1 - \alpha - \beta} [(1 - \alpha)a_{\ell 2} - \alpha a_{\ell 1}] \geq 0. \quad (2.16)$$

La résolution de ces inégalités sous la contrainte  $(\alpha + \beta) < 1$  permet d'obtenir les bornes inférieures et supérieures des valeurs admissibles des deux paramètres  $\alpha$  et  $\beta$ .

Nous introduisons d'abord les ensembles  $\mathbb{K}_1 = \{k; s_{(2,k)} > s_{(1,k)}\}$  et  $\mathbb{K}_2 = \{k; s_{(1,k)} > s_{(2,k)}\}$ . Les inégalités (2.13) et (2.14) définissent les bornes inférieures sur les valeurs de  $\alpha$  et  $\beta$ ,

$$\alpha \geq \alpha_{\min} = -\min_{k \in \mathbb{K}_1} \left\{ \frac{s_{(1,k)}}{s_{(2,k)} - s_{(1,k)}} \right\}, \quad (2.17)$$

et

$$\beta \geq \beta_{\min} = -\min_{k \in \mathbb{K}_2} \left\{ \frac{s_{(2,k)}}{s_{(1,k)} - s_{(2,k)}} \right\}. \quad (2.18)$$

Les inégalités (2.13) et (2.14) définissent également,

$$\alpha \leq \min_{k \in \mathbb{K}_2} \left\{ \frac{s_{(1,k)}}{s_{(1,k)} - s_{(2,k)}} \right\} \Rightarrow \alpha \leq 1 - \min_{k \in \mathbb{K}_2} \left\{ \frac{s_{(2,k)}}{s_{(1,k)} - s_{(2,k)}} \right\} \Rightarrow \alpha \leq 1 - \beta_{\min}, \quad (2.19)$$

et

$$\beta \leq \min_{k \in \mathbb{K}_1} \left\{ \frac{s_{(2,k)}}{s_{(2,k)} - s_{(1,k)}} \right\} \Rightarrow \beta \leq 1 - \min_{k \in \mathbb{K}_1} \left\{ \frac{s_{(1,k)}}{s_{(2,k)} - s_{(1,k)}} \right\} \Rightarrow \beta \leq 1 - \alpha_{\min} \quad (2.20)$$

Compte tenu de la contrainte  $(\alpha + \beta) < 1$ , les relations (2.19) et (2.20) sont toujours satisfaites et donc ne sont pas utilisées pour définir les bornes supérieures sur les paramètres. Par ailleurs, les inégalités (2.15) et (2.16) définissent

$$\beta \leq \beta_{\max} = \min_{\ell} \left\{ \frac{a_{\ell 1}}{a_{\ell 1} + a_{\ell 2}} \right\}, \quad (2.21)$$

et

$$\alpha \leq \alpha_{\max} = \min_{\ell} \left\{ \frac{a_{\ell 2}}{a_{\ell 1} + a_{\ell 2}} \right\}. \quad (2.22)$$

Finalement, l'ensemble des solutions admissibles correspond à

$$\tilde{\mathbf{S}} = \mathbf{T}(\alpha, \beta) \mathbf{S}, \quad \text{et} \quad \tilde{\mathbf{A}} = \mathbf{A} \mathbf{T}^{-1}(\alpha, \beta), \quad (2.23)$$

pour  $\alpha \in [\alpha_{\min}, \alpha_{\max}]$  et  $\beta \in [\beta_{\min}, \beta_{\max}]$ .

### 2.3.2 Conditions d'unicité de la solution

**Proposition 1** *La factorisation de  $\mathbf{X}$  sous la forme*

$$\mathbf{X} = \mathbf{A} \mathbf{S}, \quad \text{avec} \quad \mathbf{A}, \mathbf{S} \geq 0, \quad (2.24)$$

*est unique si et seulement si  $\exists (k_1, k_2, \ell_1, \ell_2)$ , avec  $k_1 \neq k_2$  et  $\ell_1 \neq \ell_2$ , tel que*

$$\begin{cases} s_{(1,k_1)} = 0, \text{ et } s_{(2,k_1)} \neq 0, \\ s_{(2,k_2)} = 0, \text{ et } s_{(1,k_2)} \neq 0, \\ a_{(\ell_1,1)} = 0, \text{ et } a_{(\ell_1,2)} \neq 0, \\ a_{(\ell_2,2)} = 0, \text{ et } a_{(\ell_2,1)} \neq 0. \end{cases} \quad (2.25)$$

**Démonstration 1** La démonstration de cette proposition découle directement de la section 2.3.1. Selon les bornes définie par (2.17)–(2.22), l'ensemble des valeurs possibles de  $\alpha$  et  $\beta$  est réduit à  $\alpha = 0$  et  $\beta = 0$  si et seulement si les conditions (2.25) sont satisfaites.

**Remarque 3** Le résultat de l'unicité de la proposition 1 correspond à

**R1** il existe une sous-matrice de  $\mathbf{S}$  qui est diagonale ou antidiagonale et de dimension  $(2 \times 2)$ ,

**R2** il existe une sous-matrice de  $\mathbf{A}$  qui est diagonale ou antidiagonale et de dimension  $(2 \times 2)$ .

### 2.3.3 Illustration des solutions admissibles

Les signaux sources sont simulés pour ( $p = 2$ ,  $n = 500$ ) et les coefficients de mélange pour ( $m = 10$ ) avec des formes similaires aux données spectroscopiques. La figure 2.1 montre les signaux sources et les coefficients de mélange utilisés.

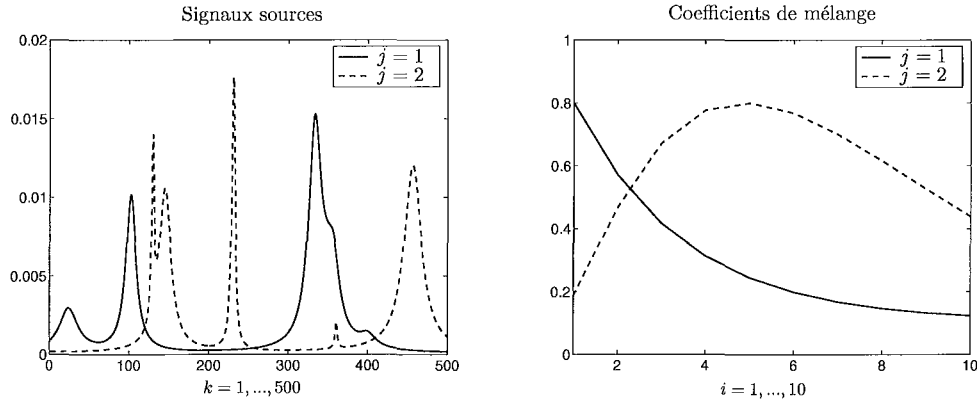


FIG. 2.1 – Signaux sources et profils d'évolution des coefficients de mélange

En utilisant les résultats précédents, on note que les conditions d'unicité ne sont pas satisfaites et les intervalles des valeurs admissibles des paramètres sont

$$-0.0163 \leq \alpha \leq 0.1898, \quad (2.26)$$

$$-0.0198 \leq \beta \leq 0.1924. \quad (2.27)$$

Les solutions admissibles en termes de signaux sources et de coefficients de mélange sont montrées dans la figure 2.2.

## 2.4 Solutions admissibles et conditions d'unicité dans la cas de plus de deux sources

Dans cette section, nous considérons le cas de  $p$  signaux sources. Notons

$$\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_p] \quad \text{et} \quad \mathbf{S} = [\mathbf{s}_1 \cdots \mathbf{s}_p]^T, \quad (2.28)$$

où  $\{\mathbf{a}_i\}_{i=1}^p$  sont des vecteurs de dimension  $(m \times 1)$  et  $\{\mathbf{s}_j\}_{j=1}^p$  sont des vecteurs de dimension  $(n \times 1)$ . Nous utilisons les notations suivantes :  $\mathbf{a}_i(k)$  est le  $k$ ème élément du vecteur  $\mathbf{a}_i$  et de façon similaire,  $\mathbf{s}_j(k)$  est le  $k$ ème élément du vecteur  $\mathbf{s}_j$ .

### 2.4.1 Conditions nécessaires pour l'unicité

Dans ce cas, nous pouvons uniquement donner une condition nécessaire pour l'unicité de la solution.

**Proposition 2** *Si la décomposition de  $\mathbf{X}$  en  $\mathbf{A}$  et  $\mathbf{S}$  selon*

$$\mathbf{X} = \mathbf{A} \mathbf{S} \quad \text{avec} \quad \mathbf{A} \geq 0, \mathbf{S} \geq 0, \quad (2.29)$$

*est unique, alors les conditions suivantes sont satisfaites :*

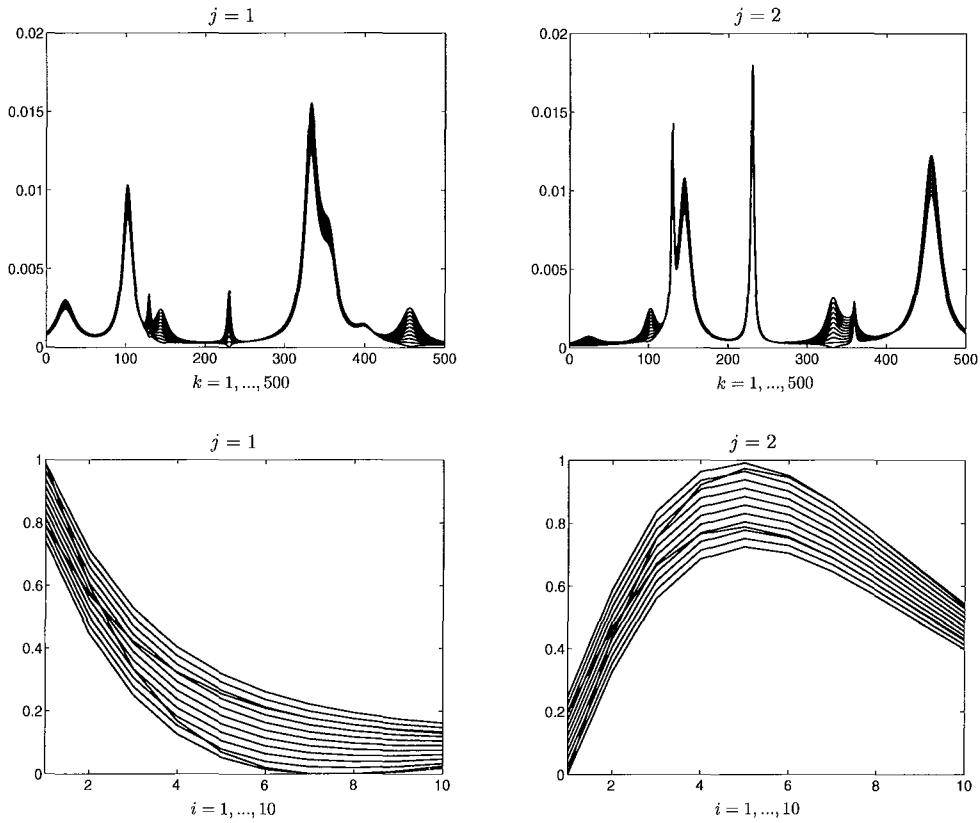


FIG. 2.2 – Quelques solutions admissibles des sources et des coefficients de mélange

(A1)  $\exists k_1, \dots, k_p$  tel que :

$$\forall i \neq j, k_i \neq k_j, \mathbf{s}_i(k_i) = 0, \text{ et } \mathbf{s}_j(k_i) \neq 0. \quad (2.30)$$

(A2)  $\exists \ell_1, \dots, \ell_p$  tel que :

$$\forall i \neq j, \ell_i \neq \ell_j, \mathbf{a}_i(\ell_i) = 0, \text{ et } \mathbf{a}_j(\ell_i) \neq 0. \quad (2.31)$$

**Démonstration 2** La démonstration de cette proposition est réalisée par l'absurde : supposons que les conditions (A1)-(A2) ne sont pas satisfaites et que la décomposition  $\mathbf{X} = \mathbf{A} \mathbf{S}$  avec  $\mathbf{A}, \mathbf{S} \geq 0$ , est unique.

Supposons que la condition (A1) n'est pas satisfaite. Soit la matrice de transformation élémentaire  $\mathbf{T}_{ij}(\lambda)$  de dimension  $(p \times p)$  et définie par

$$\forall k = 1, \dots, p, \forall \ell = 1, \dots, p, \begin{cases} t_{kk} = 1, \\ t_{k\ell} = \lambda \text{ si } (k, \ell) = (i, j); \\ t_{k\ell} = 0 \text{ si } (k, \ell) \neq (i, j) \text{ et } (k, \ell) \neq (k, k). \end{cases} \quad (2.32)$$

On note que  $\mathbf{T}^{-1}(\lambda) = \mathbf{T}(-\lambda)$ . Définissons

$$\mathbb{K} = \{k, \mathbf{s}_i(k) \neq 0, \mathbf{s}_j(k) \neq 0\}.$$

Les échantillons où les sources  $\mathbf{s}_i$  et  $\mathbf{s}_j$  sont simultanément nulles puisqu'ils ne sont pas affectés par la transformation  $\mathbf{T}_{ij}(\lambda)$ . La décomposition suivante

$$\mathbf{X} = \mathbf{A} \mathbf{T}_{ij}(\lambda) \mathbf{T}_{ij}(-\lambda) \mathbf{S} \quad (2.33)$$

avec  $\lambda > 0$  assure la non-négativité de  $\tilde{\mathbf{A}} = \mathbf{A} \mathbf{T}_{ij}(\lambda)$  et la matrice  $\mathbf{T}_{ij}(-\lambda)$  conduit à une transformation du  $i$ ème signal source selon

$$\tilde{\mathbf{s}}_i = \mathbf{s}_i - \lambda \mathbf{s}_j, \quad (2.34)$$

alors que les autres sources restent inchangées. En définissant  $(\underline{s}_i, \underline{s}_j)$  par  $\underline{s}_i = \min_{k \in \mathbb{K}} s_i(k)$  et  $\bar{s}_j = \max_{k \in \mathbb{K}} s_j(k)$ . Il existe alors  $0 < \lambda < \underline{s}_i / \bar{s}_j$  tel que

$$\forall k, s_i(k) - \lambda s_j(k) \geq \underline{s}_i - \lambda \bar{s}_j \geq 0 \implies \mathbf{T}_{ij}(\lambda) \mathbf{S} \geq 0.$$

Donc, la décomposition n'est pas unique, ce qui est en contradiction avec les hypothèses.

Concernant la condition (A2) sur la matrice de mélange, le même raisonnement est employé avec  $\lambda < 0$ .

**Remarque 4** Les conditions (A1) et (A2) montrent que, dans la plupart des cas, la séparation en utilisant uniquement la contrainte de non-négativité ne peut fournir la solution correcte.

**Remarque 5** Les conditions (A1) et (A2) ne sont pas suffisantes pour l'unicité de la solution. Considérons l'exemple suivant :

$$\mathbf{S} = \begin{bmatrix} 0.53 & 0 & 0.24 & 0.23 \\ 0.13 & 0.52 & 0 & 0.35 \\ 0 & 0.32 & 0.34 & 0.34 \end{bmatrix},$$

et soit  $\mathbf{T}^{-1}$  la matrice définie par

$$\mathbf{T}^{-1} = \begin{bmatrix} 0.66 & 0 & 0.34 \\ 0.16 & 0.84 & 0 \\ 0 & 0.52 & 0.49 \end{bmatrix}.$$

Nous avons :

$$\mathbf{T} = \begin{bmatrix} 1.37 & 0.60 & -0.97 \\ -0.26 & 1.08 & 0.18 \\ 0.28 & -1.15 & 1.87 \end{bmatrix},$$

d'où

$$\left\{ \begin{array}{l} \tilde{\mathbf{A}} = \mathbf{A} \mathbf{T}^{-1} \geq 0, \forall \mathbf{A} \geq 0; \\ \tilde{\mathbf{S}} = \mathbf{T} \mathbf{S} = \begin{bmatrix} 0.80 & 0 & 0 & 0.20 \\ 0 & 0.62 & 0 & 0.38 \\ 0 & 0 & 0.70 & 0.30 \end{bmatrix} \geq 0. \end{array} \right.$$



### 2.4.2 Conditions suffisantes pour l'unicité

Dans cette section, nous présentons des conditions suffisantes pour l'unicité de la solution, mais avant de présenter les résultats, certaines définitions et propriétés sont rappelées [PVJV95, VV99].

**Définition 3** Une matrice positive  $\mathbf{T}$  de dimension  $(p, p)$  est dite monomiale si chaque ligne et chaque colonne de cette matrice contient un seul élément non-nul; i.e :

$$\forall i = 1, \dots, p, \forall j = 1, \dots, p, i \neq j, \exists k \text{ tel que } \begin{cases} t_{ik} > 0, \\ t_{jk} = 0. \end{cases}$$

**Propriété 1** Soit  $\mathbf{T}$  une matrice positive de dimension  $(p, p)$ . La matrice  $\mathbf{T}$  possède une inverse positive si et seulement si  $\mathbf{T}$  est une matrice monomiale. Alors  $\mathbf{T}^{-1}$  est aussi monomiale.

**Propriété 2** Chaque matrice monomiale  $\mathbf{T}$  peut être décomposée sous la forme  $\mathbf{T} = \mathbf{D}\mathbf{P}$  ou  $\mathbf{D}$  est une matrice diagonale positive et  $\mathbf{P}$  est une matrice de permutation, qui est une matrice monomiale dont les éléments non-nuls valent 1. En séparation de sources une telle transformation lorsqu'elle est appliquée à  $\mathbf{S}$  correspond aux indéterminations d'ordre et d'échelle.

**Proposition 3** La décomposition de  $\mathbf{X}$  en  $\mathbf{A}$  et  $\mathbf{S}$  suivant

$$\mathbf{X} = \mathbf{A}\mathbf{S} \text{ avec } \mathbf{A}, \mathbf{S} \geq 0, \quad (2.35)$$

est unique si les conditions suivantes sont satisfaites :

(B1) Il existe une sous matrice de  $\mathbf{S}$  monomiale et de dimension  $(p, p)$ .

(B2) Il existe une sous matrice de  $\mathbf{A}$  monomiale et de dimension  $(p, p)$ .

**Démonstration 3** Supposons que les conditions (B1) et (B2) sont satisfaites. Après permutation de ses colonnes, la matrice  $\mathbf{S}$  peut être écrite selon :

$$\mathbf{S} = \begin{bmatrix} s_{(1,1)} & & 0 & s_{(1,p+1)} & s_{(1,p+2)} & \cdots & s_{(1,n)} \\ & \ddots & & & & & \vdots \\ 0 & & s_{(p,p)} & s_{(p,p+1)} & s_{(p,p+2)} & \cdots & s_{(p,n)} \end{bmatrix}.$$

De façon similaire, après permutation de ses lignes, la matrice  $\mathbf{A}$  peut être réécrite selon :

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & & & & 0 \\ & \ddots & & & \\ & & 0 & & a_{p,p} \\ a_{p+1,1} & & & & a_{p+1,p} \\ a_{p+2,1} & & & & a_{p+2,p} \\ \vdots & & & & \vdots \\ a_{m,1} & \cdots & & & a_{m,p} \end{bmatrix}.$$

Considérons alors une matrice régulière  $\mathbf{T} = [t_{ij}]$  dont l'inverse est notée  $\mathbf{T}^{-1} = [t_{i,j}^\#]$ . Alors :

$$\mathbf{TS} = \begin{bmatrix} t_{11}s_{(1,1)} & \cdots & t_{1p}s_{(p,p)} & \cdots \\ \vdots & & \vdots & \\ t_{p1}s_{(1,1)} & \cdots & t_{pp}s_{(p,p)} & \cdots \end{bmatrix} \quad \text{et} \quad \mathbf{AT}^{-1} = \begin{bmatrix} a_{11}t_{11}^\# & \cdots & a_{11}t_{1p}^\# \\ \vdots & & \vdots \\ a_{pp}t_{1p}^\# & \cdots & a_{pp}t_{pp}^\# \\ \vdots & & \vdots \end{bmatrix}$$

d'où on déduit que  $\mathbf{TS} \geq 0$  et  $\mathbf{AT}^{-1} \geq 0$  si et seulement si  $\mathbf{T} > 0$  et  $\mathbf{T}^{-1} > 0$ . A partir de la propriété 1, ceci est équivalent à dire que  $\mathbf{T}$  est une matrice monomiale. A partir de la propriété 2, on conclut que la solution est unique car la seule transformation possible correspond soit à une permutation des sources ou à un changement d'échelle.

A partir de ce résultat, on déduit immédiatement le corollaire suivant qui donne une condition suffisante sur une matrice  $\mathbf{X}$  pour l'unicité de sa factorisation en matrices non-négatives :

**Corollaire 1** *La décomposition de  $\mathbf{X}$  en  $\mathbf{A}$  et  $\mathbf{S}$  selon*

$$\mathbf{X} = \mathbf{AS} \text{ avec } \mathbf{A}, \mathbf{S} \geq 0, \quad (2.36)$$

*est unique si la condition suivante est satisfaite :*

(C1) *Il existe une sous matrice monomiale et de dimension  $(p \times p)$ .*

### 2.4.3 Formulation géométrique

Dans ce paragraphe, nous utilisons la notion de cône simplicial pour donner une interprétation géométrique des résultats de l'unicité de la section 2.4.2. Pour cela, quelques définitions sont nécessaires [VV99, DS03].

**Définition 4 (Cône convexe)** *Un sous-ensemble  $\mathcal{C} \subseteq \mathbb{R}^n$  est un cône convexe si :*

- $v_1, v_2 \in \mathcal{C} \Rightarrow v_1 + v_2 \in \mathcal{C}$
- $v \in \mathcal{C}, \lambda \in \mathbb{R}_+ \Rightarrow \lambda v \in \mathcal{C}$

**Définition 5 (Cône simplicial)** *Le cône simplicial généré par un nombre fini de vecteurs indépendants  $\{v_1, \dots, v_p\} \in \mathbb{R}_+^n$  est l'ensemble :*

$$\mathcal{C} = \left\{ \mathbf{x} \mid \mathbf{x} = \sum_{j=1}^p a_j \mathbf{v}_j, a_j > 0 \right\}$$

$\{v_1, \dots, v_p\}$  sont les arêtes du cône simplicial et  $p$  représente la dimension du cône :  $p = \dim \mathcal{C}$ . En rassemblant les vecteurs dans la matrice  $\mathbf{V} = [v_1 \cdots v_p]^T$ , on peut écrire  $\mathcal{C}_{\mathbf{V}} = \text{cone}(\mathbf{V})$ .

A partir de ces deux définitions, il apparaît que la décomposition de  $\mathbf{X} = \mathbf{AS}$ , avec  $\mathbf{A} \geq 0, \mathbf{S} \geq 0$ , signifie que chaque ligne de  $\mathbf{X}$  satisfait  $\mathbf{x}_j \in \mathcal{C}_{\mathbf{S}} = \text{cone}(\mathbf{S})$ .

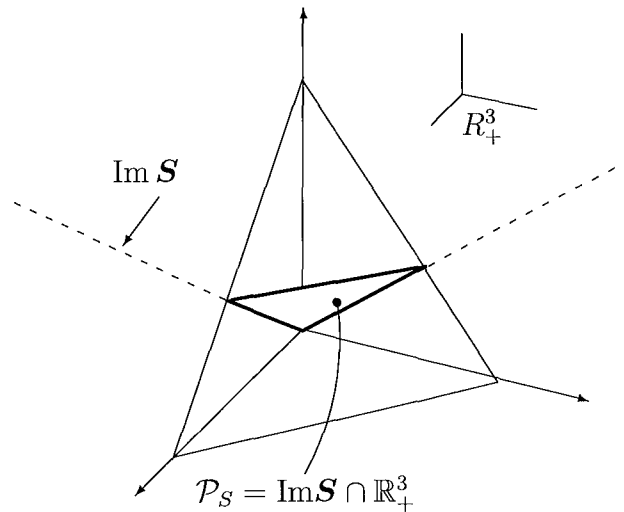


FIG. 2.3 – Intersection de  $\text{Im } \mathbf{S}$  avec  $\mathbb{R}_+^n$ , dans le cas  $p = 2$  et  $n = 3$ .

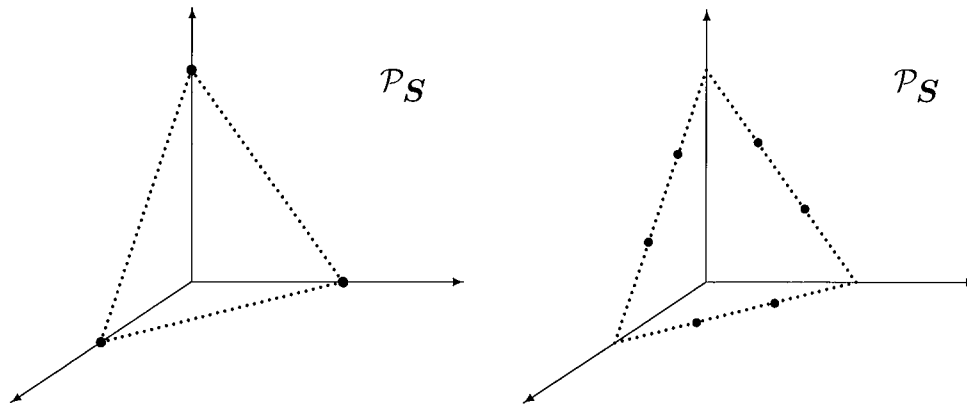
**Condition de Chen :** En notant  $\mathcal{K}$  l'enveloppe convexe de la matrice  $\mathbf{X}$ , les conditions d'unicité de la factorisation en matrices non-négatives sont données par le théorème suivant [Che84]

**Theorem 1** *La décomposition de  $\mathbf{X}$  en  $\mathbf{A}$  et  $\mathbf{S}$  est unique si et seulement si le cône simplicial  $\mathcal{C}$  tel que  $\mathcal{K} \subset \mathcal{C}$  est unique.*

Le point critique est que ce théorème ne donne aucun moyen numérique pour vérifier à partir de la matrice de données si sa décomposition en matrices non-négatives est unique ou pas. Par contre, les conditions de la proposition 3 et celles de Donoho-Stodden [DS03] permettent de vérifier numériquement l'unicité de la solution et admettent une formulation géométrique en accord avec le théorème de Chen.

**Conditions proposées :** On définit le cône  $\mathcal{P}_{\mathbf{S}} = \text{Im } \mathbf{S} \cap \mathbb{R}_+^n$ , qui résulte de l'intersection de l'image de  $\mathbf{S}$  avec  $\mathbb{R}_+^n$  (voir figure 2.3). La condition (B1) de la proposition 3 correspond à  $\mathcal{C}_{\mathbf{S}} = \mathcal{P}_{\mathbf{S}}$  alors que la condition (B2) impose qu'au plus  $p$  vecteurs ligne de  $\mathbf{X}$  sont colinéaires avec les  $p$  arêtes du cône  $\mathcal{P}_{\mathbf{S}}$ .

**Conditions de Donoho-Stodden :** La condition suffisante fournie par Donoho-Stodden est formulée sur une classe particulière d'images et se fonde sur la notion de famille factorielle d'articulations séparables. La première condition de ce théorème est une condition générique et suppose l'existence de la factorisation. La condition de séparabilité est équivalente à la condition (B1) de la proposition 3, c'est à dire  $\mathcal{C}_{\mathbf{S}} = \mathcal{P}_{\mathbf{S}}$ . La dernière condition (échantillonnage factoriel complet) impose qu'il existe au moins  $(p-1)$  vecteurs lignes différents de  $\mathbf{X}$  sur chacune des  $p$  facettes (de dimension  $(p-1)$ ) du cône  $\mathcal{P}_{\mathbf{S}}$ . Dans ce cas, il est naturel que le seul cône de dimension  $p$  qui inclut tous les vecteurs de  $\mathbf{X}$  est le cône  $\mathcal{P}_{\mathbf{S}}$  lui même.



Cas 1 : Conditions de la proposition 3    Cas 2 : Conditions de Donoho-Stodden

FIG. 2.4 – Différents cas d'unicité de la factorisation non-négative. Les cercles représentent les vecteurs de  $\mathbf{X}$  nécessaire pour assurer l'unicité de la factorisation.

**Discussion :** La condition suffisante de la proposition 3 peut être retrouvée en se déplaçant sur les  $p$  vecteurs générateurs du cône. Chaque facette du cône est de dimension  $(p - 1)$  et chaque vecteur générateur du cône est inclus dans  $(p - 1)$  facettes distinctes. Ainsi, nous avons toujours les conditions que  $(p - 1)$  vecteurs différents de  $\mathbf{X}$  appartiennent aux  $(p - 1)$  facettes du cône. Mais, à ce moment nous n'avons besoin que de  $p$  vecteurs différents au lieu de  $p(p - 1)$  de la condition de Donoho-Stodden. A partir de cette formulation géométrique, nous pouvons déduire les différentes situations dans lesquelles l'unicité de la factorisation en matrices non-négatives est unique. Ces situations sont représentées dans la figure 2.4.

**Remarque 6** Nous avons étudié dans cette section des conditions nécessaires et des conditions suffisantes pour l'unicité de la solution. Mais, nous n'avons pas pu déterminer des conditions suffisantes moins restrictives pour garantir l'unicité de la solution. Ce problème reste donc ouvert.

#### 2.4.4 Détermination des solutions admissibles

Une autre question aussi importante est celle liée à la détermination des solutions admissibles lorsque la factorisation n'est pas unique. La recherche analytique des solutions admissibles est assez délicate pour un nombre de sources supérieur à deux et nous n'avons pas pu trouver une solution explicite. Nous utilisons alors une méthode fondée sur la minimisation d'un critère de positivité. Pour cela, nous reprenons l'algorithme de Sasaki et al. [SKM83]. Les solutions admissibles sont déduites par des transformations linéaires des vraies sources. On recherche toutes les matrices  $\mathbf{T}$  qui, appliquées aux sources et aux coefficients de mélange, permettent de préserver la non-négativité. Une approche possible, est de rechercher toutes les matrices qui minimisent un critère de positivité  $\mathcal{C}_{pos}(\tilde{\mathbf{S}}, \tilde{\mathbf{A}})$ ,

$$\hat{\mathbf{T}} = \arg \min_{\mathbf{T}} \mathcal{C}_{pos}(\tilde{\mathbf{S}}, \tilde{\mathbf{A}}) \quad (2.37)$$

où  $\tilde{\mathbf{S}} = \mathbf{T}\mathbf{S}$  et  $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}^{-1}$ . L'affectation d'un coût quadratique pour les valeurs négatives

$$f(x) = \begin{cases} x^2 & \text{si } x < 0, \\ 0 & \text{sinon,} \end{cases} \quad (2.38)$$

permet de déduire le critère suivant :

$$C_{pos}(\tilde{\mathbf{S}}, \tilde{\mathbf{A}}) = \sum_{j=1}^p \sum_{k=1}^n f(\tilde{s}_{(j,k)}) + \sum_{j=1}^p \sum_{i=1}^m f(\tilde{a}_{ij}). \quad (2.39)$$

L'optimisation de ce critère est réalisée par un algorithme du simplexe avec une initialisation aléatoire de la matrice de transformation. Seules les solutions avec une valeur nulle du critère sont retenues. En effectuant une simulation de Monte Carlo, nous retrouvons les régions des valeurs admissibles des paramètres de la matrice de transformation. Afin d'illustrer cet algorithme de recherche des solutions admissibles, considérons le cas de trois sources ayant une norme  $L_1$  qui vaut l'unité. Les sources utilisées sont illustrées dans la figure 2.5.

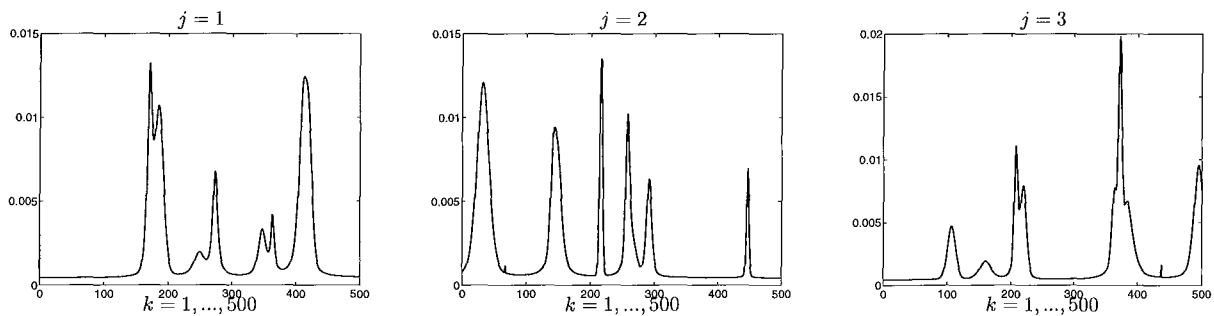


FIG. 2.5 – Signaux sources et coefficients de mélange simulés.

Afin de conserver cette norme  $L_1$ , la matrice de transformation est écrite sous la forme

$$\mathbf{T}(t_1, \dots, t_6) = \begin{bmatrix} (1 - t_1 - t_2) & t_1 & t_2 \\ t_3 & (1 - t_3 - t_4) & t_4 \\ t_5 & t_6 & (1 - t_5 - t_6) \end{bmatrix}. \quad (2.40)$$

Afin de minimiser le critère (2.39) par rapport à cette transformation, l'algorithme de simplexe est appliqué avec plusieurs initialisations différentes. Les valeurs initiales des paramètres de la matrice  $\mathbf{T}$  sont simulées à partir d'une loi normale de moyenne nulle et de variance unité. En ne retenant que les solutions pour lesquelles le critère est nul, et après remise dans le bon ordre des sources, nous obtenons les valeurs possibles des paramètres de la matrice de transformation et les solutions admissibles associées à ces paramètres. La figure 2.6 montre quelques valeurs admissibles des couples  $(t_1, t_2)$ ,  $(t_3, t_4)$  et  $(t_5, t_6)$ . La figure 2.7 montre quelques solutions admissibles en termes de signaux sources et coefficients de mélange.

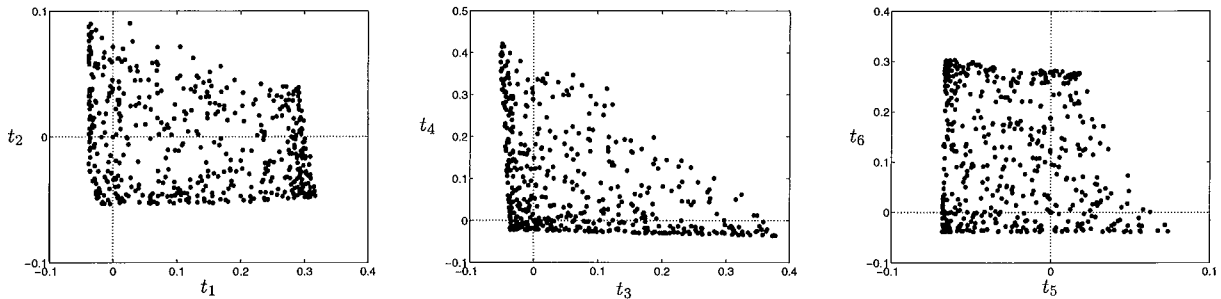


FIG. 2.6 – Valeurs des paramètres des matrices de transformation permettant d'obtenir des solutions admissibles.

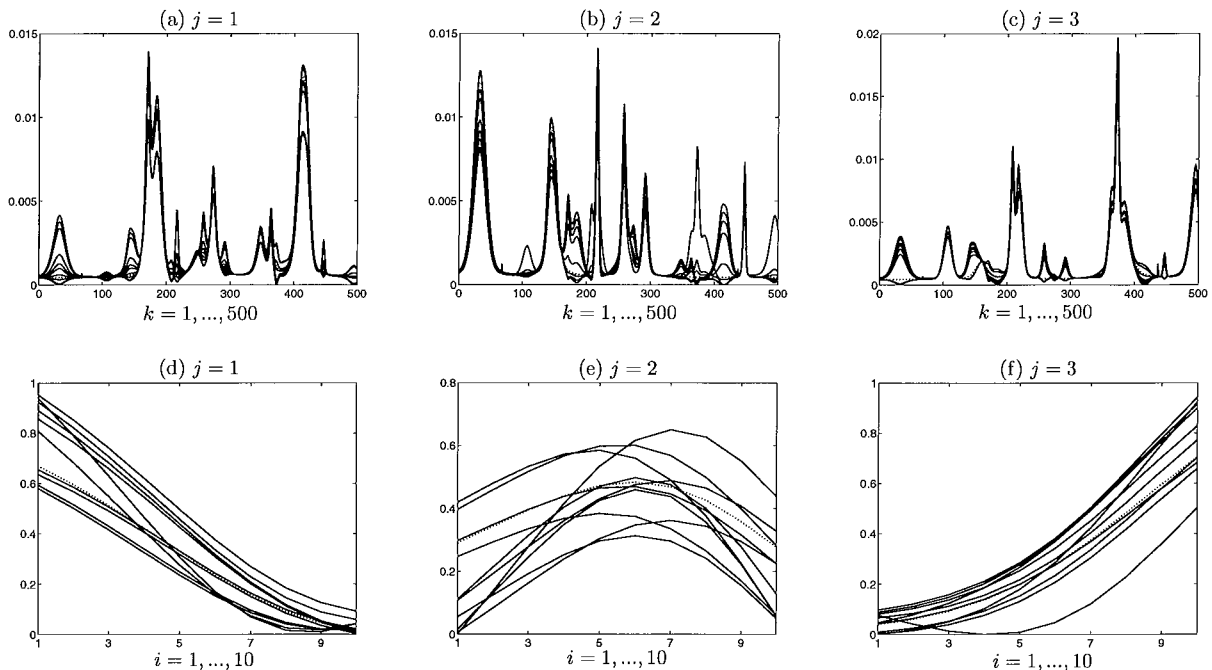


FIG. 2.7 – Illustration de quelques solutions admissibles en termes de signaux sources (a–c) et coefficients de mélange (d–f). La solution correcte est tracée en pointillés.

## 2.5 Séparation de sources non-négatives : quel algorithme ?

Afin d'illustrer la difficulté de séparer des sources non-négatives nous utilisons un exemple simplifié avec deux sources de 1000 échantillons et à distribution uniforme ( $p = 2$ ,  $n = 1000$ ). Les sources sont simulées à partir d'une loi uniforme sur l'intervalle  $[0, 1[$  puis normalisés à une variance unité. Une matrice de dimension ( $m = 10$ ,  $p = 2$ ) et de distribution uniforme sur l'intervalle  $[0, 1[$  est appliquée pour la synthèse des observations. Les méthodes NMF, NNICA et JADE sont appliquées pour la séparation. Les résultats sont affichés dans les plans  $(s_1, s_2)$  et  $(a_1, a_2)$  et pour une meilleure lisibilité, seulement les 100 premiers échantillons des sources sont tracés. On distingue principalement quatre cas différents dont les résultats sont analysés séparément et montrés dans les figures (2.8–2.11).

### 2.5.1 Coefficients non-négatifs et sources non-négatives et indépendantes

Dans ce cas toutes les méthodes fournissent des résultats corrects, car les conditions requises par les différentes méthodes sont satisfaites.

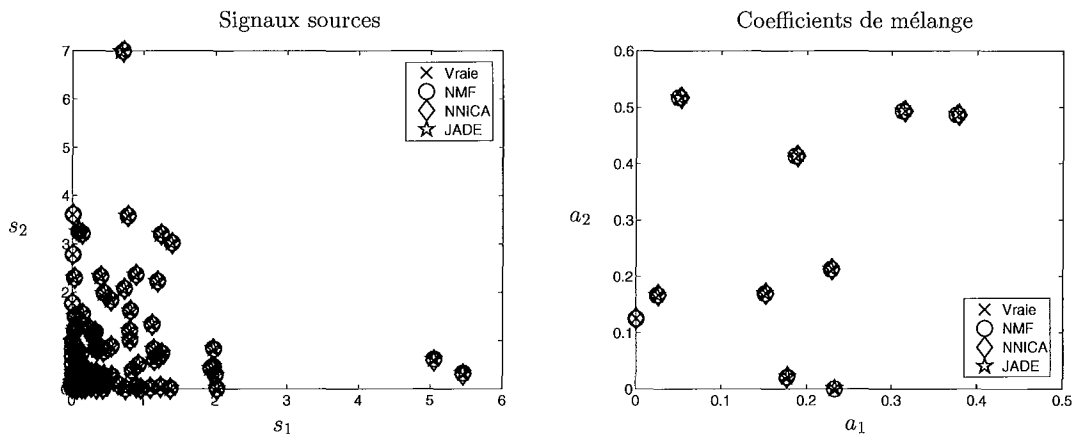


FIG. 2.8 – Cas 1 : Coefficients de mélange non-négatifs et sources mutuellement indépendantes et non-négatives

### 2.5.2 Coefficients non-négatifs et sources indépendantes et strictement positives

Dans ce cas, un offset est rajouté aux sources, et par conséquent la condition nécessaire pour l'unicité de la solution par NMF n'est plus satisfaite, de plus la condition de sources *well grounded* requise par NNICA n'est pas satisfaite aussi, ce qui fait que seul JADE fournit un résultat correct dans ce cas.

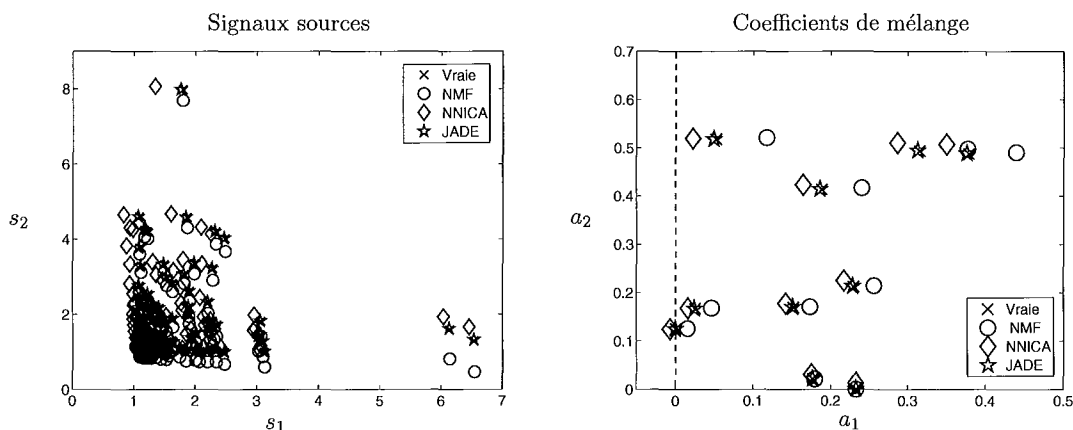


FIG. 2.9 – Cas 2 : Coefficients de mélange non-négatifs et sources strictement positives et mutuellement indépendantes

### 2.5.3 Coefficients non-négatifs et sources non-négatives et corrélées

Les sources dépendantes sont obtenues en appliquant d'abord une transformation linéaire aux deux sources indépendantes et en introduisant un décalage de telle sorte que la condition d'unicité soit satisfaite. Etant donné que l'hypothèse d'indépendance n'est plus respectée, les méthodes NNICA et JADE échouent à donner un résultat correct alors que la méthode NMF fournit une solution correcte, puisque les conditions d'unicité sont vérifiées.

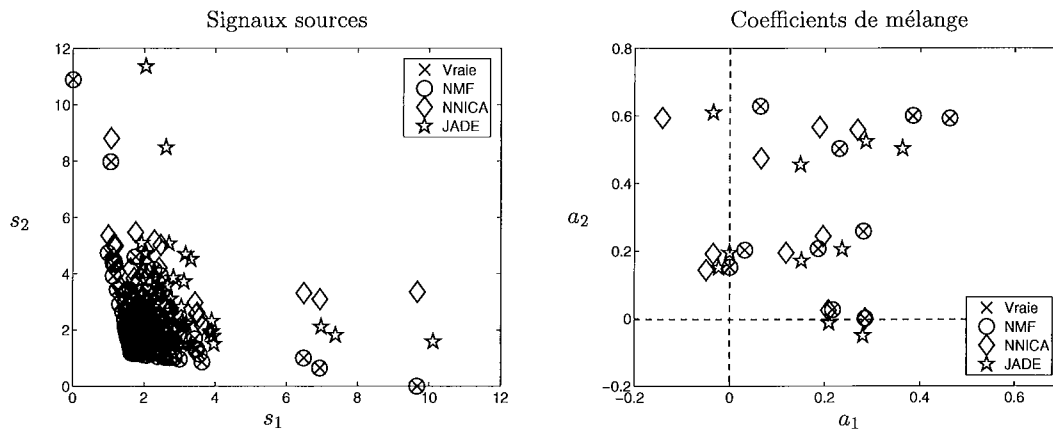


FIG. 2.10 – Cas 3 : Coefficients non-négatifs et sources non-négatives mais mutuellement dépendantes

### 2.5.4 Coefficients strictement positifs et sources strictement positives et corrélées

Ce cas est plus délicat car ni l'hypothèse d'indépendance et ni les conditions d'unicité ne sont satisfaites. Dans ce cas aucune des méthodes considérées ne permettent de faire la séparation avec succès. Cet exemple illustre la nécessité de développer des algorithmes capables de réaliser la séparation en incluant des hypothèses supplémentaires sur les sources et sur les coefficients de mélange.

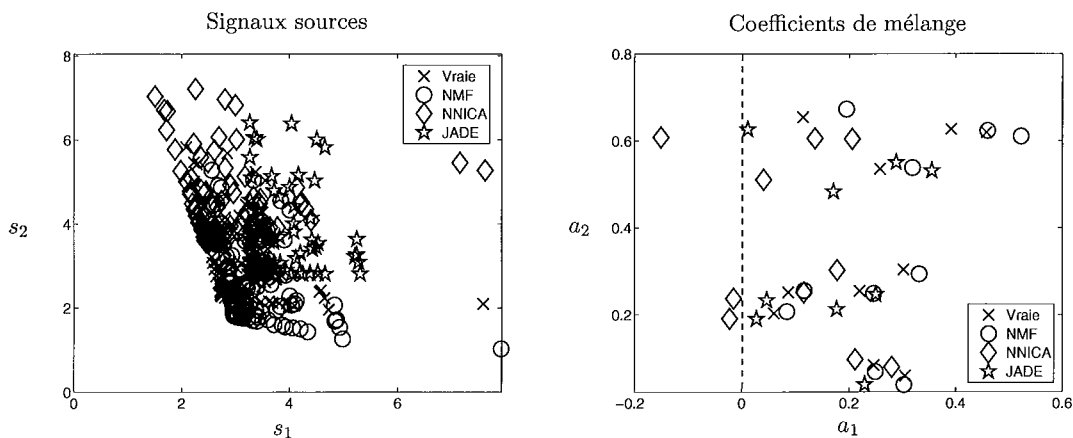


FIG. 2.11 – Cas 4 : coefficients de mélange strictement positifs et sources corrélées et strictement positives



## 2.6 Conclusion

Dans ce chapitre, nous avons montré que sous certaines conditions, le problème de la séparation de sources non-négatives admet une solution unique. Ces conditions sont plus flexibles que celles présentées dans [SHK<sup>+</sup>01, PSH02]. Un lien avec les travaux de Donoho-Stodden [DS03] a été établi pour la formulation géométrique des conditions suffisantes pour l'unicité de la factorisation. Les performances d'un algorithme de séparation de sources ou de factorisation en matrices non-négatives dépendent de la satisfaction des conditions d'unicité ou de la validité de l'hypothèse d'indépendance. Un résultat intéressant est que si les conditions d'unicité sont satisfaites, l'hypothèse d'indépendance n'est plus nécessaire pour réaliser la séparation. Par contre, les méthodes NMF échouent dans le cas où ces conditions ne sont pas satisfaites. De façon similaire appliquer les méthodes ICA dans le cas où les sources ne sont pas indépendantes mène à une solution incorrecte. Finalement, le cas le plus délicat est celui où les conditions d'unicité et l'hypothèse d'indépendance ne sont pas satisfaites. C'est principalement le cas avec les signaux de spectroscopie.



## Deuxième partie

# Développement d'algorithmes pour la séparation de sources spectrales



## Chapitre 3

# Séparation de sources spectrales par maximum de vraisemblance

Ce chapitre est consacré à la séparation de sources non-négatives en utilisant l'approche du maximum de vraisemblance. Après une brève présentation des travaux antérieurs concernant l'application de cette approche à la séparation de sources. Afin d'adapter cette approche au cas de sources spectrales, nous représentons la densité de probabilité des signaux sources par une distribution à support non-négatif afin de prendre en compte explicitement la non-négativité. Outre leurs non-négativité, les sources spectrales sont parcimonieuses. Nous utilisons alors un modèle de distribution gamma qui permet de prendre en compte conjointement ces deux informations. Nous discutons deux techniques d'optimisation possibles et nous illustrons à l'aide de signaux synthétiques les algorithmes qui en résultent.

### 3.1 Introduction

Considérons le modèle de mélange linéaire instantané exprimé par

$$\mathbf{x}_k = \mathbf{A} \mathbf{s}_k + \mathbf{e}_k, \quad k = 1, \dots, n \quad (3.1)$$

où les  $m$  composantes du vecteur  $\mathbf{x}_k$  représentent les signaux observés, les  $p$  signaux sources sont représentés par le vecteur  $\mathbf{s}_k$ , alors que les coefficients de mélange  $\{a_{ij}, i = 1, \dots, m\}_{j=1}^p$  forment une matrice notée  $\mathbf{A}$ , de dimension  $(m \times p)$ , appelée matrice de mélange. Chaque colonne de cette matrice représente le profil d'évolution de la contribution de la source correspondante dans les mélanges observés. On note  $\mathbf{e}_k$  le vecteur dont les composantes correspondent à un bruit additif d'autocovariance  $\mathbf{R}_e$  et indépendant des signaux sources. En utilisant une notation matricielle, le modèle de mélange s'écrit

$$\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{E}, \quad (3.2)$$

où  $\mathbf{X} \in \mathbb{R}^{(m \times n)}$  est la matrice des observations,  $\mathbf{A} \in \mathbb{R}^{(m \times p)}$  la matrice de mélange,  $\mathbf{S} \in \mathbb{R}^{(p \times n)}$  est la matrice des signaux sources et  $\mathbf{E} \in \mathbb{R}^{(m \times n)}$  est la matrice du bruit additif.

Le problème de séparation consiste à estimer les signaux sources et les coefficients de mélange à partir de la seule mesure des observations sous la contrainte de non-négativité des signaux sources et des coefficients de mélange. Dans ce chapitre, nous abordons ce problème en utilisant l'approche du maximum de vraisemblance pour la séparation. Il est à noter, qu'en adoptant une telle approche pour la séparation, aucune contrainte de positivité n'est imposée sur la matrice de mélange. Cette contrainte ne pourra donc qu'être vérifiée *a posteriori*. Après une présentation succincte de l'approche du maximum de vraisemblance en séparation de sources, nous discutons l'impact du choix de la distribution des sources sur le résultat de la séparation. Partant de ce constat, nous présentons une fonction non-linéaire dérivée à partir de la modélisation de la distribution des sources spectrales par des densités gamma afin de prendre en compte leur non-négativité et de représenter au mieux leur parcimonie. Pour résoudre le problème de maximisation de la vraisemblance, nous utilisons l'algorithme du gradient dans le cas d'un niveau de bruit négligeable et les algorithmes EM pour le cas d'observations contaminées par un bruit non-négligeable.

## 3.2 Séparation de sources par maximum de vraisemblance

L'application de l'approche du maximum de vraisemblance à la séparation de sources a été considérée par plusieurs auteurs, nous citons à titre d'exemples [GL90b, PGJ92, BC95, Mac96, Car97, PG97, DZ04, HJP03]. Néanmoins, à notre connaissance, son application au cas de sources positives n'a pas été envisagée. Avant de proposer un nouvel algorithme de séparation dans le cas de sources spectrales, quelques rappels sur les principes fondamentaux de cette approche sont nécessaires.

### 3.2.1 Formulation de la vraisemblance

#### 3.2.1.1 Cas sans bruit

Considérons dans un premier temps le cas "carré" ( $m = p$ ) et sans bruit. Pour  $m > p$ , une réduction de dimension de la matrice des données est nécessaire, soit par simple transformation linéaire aléatoire de dimension ( $p \times m$ ) ou par analyse en composantes principales [Bin03]. Sous l'hypothèse d'indépendance mutuelle des signaux sources et compte tenu du modèle de mélange  $\mathbf{x}_k = \mathbf{A} \mathbf{s}_k$ , dans le cas carré et en l'absence de bruit, la vraisemblance s'exprime par :

$$p(\mathbf{x}_k | \mathbf{A}) = \frac{1}{\det(\mathbf{A})} \prod_{j=1}^p p_j \left( [\mathbf{A}^{-1} \mathbf{x}_k]_j \right), \quad (3.3)$$

où  $p_j(\cdot)$  est la densité de probabilité de la  $j$ ème source. En introduisant la matrice séparante,  $\mathbf{B} = \mathbf{A}^{-1}$ , la vraisemblance devient [GL90b, PGJ92]

$$p(\mathbf{x}_k | \mathbf{B}) = |\det(\mathbf{B})| \prod_{j=1}^p p_j \left( [\mathbf{B} \mathbf{x}_k]_j \right). \quad (3.4)$$

Sous l'hypothèse de signaux sources i.i.d, l'estimation de la matrice séparante par maximum de vraisemblance est obtenue par

$$\hat{\mathbf{B}} = \arg \max_{\mathbf{B}} \left\{ |\det \mathbf{B}|^n \prod_{k=1}^n \prod_{j=1}^p p_j([\mathbf{B} \mathbf{x}_k]_j) \right\}, \quad (3.5)$$

Ce problème de maximisation est équivalent à celui de la minimisation du critère défini par

$$\mathcal{L}_n(\mathbf{X}|\mathbf{B}) = -\log |\det(\mathbf{B})| - \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^p \log p_j([\mathbf{B} \mathbf{x}_k]_j). \quad (3.6)$$

### 3.2.1.2 Cas avec bruit

En supposant un bruit i.i.d gaussien de moyenne nulle et de matrice de covariance  $\mathbf{R}_e$ , la vraisemblance s'exprime par

$$p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \mathbf{R}_e) = \prod_{k=1}^n \mathcal{N}(\mathbf{A} \mathbf{s}_k, \mathbf{R}_e). \quad (3.7)$$

Sous l'hypothèse de sources i.i.d et mutuellement indépendantes

$$p(\mathbf{S}) = \prod_{k=1}^n \prod_{j=1}^p p_j(s_{(j,k)}), \quad (3.8)$$

et en considérant le problème d'identification de la matrice de mélange à partir des sources et des observations, la vraisemblance sera alors exprimée par

$$p(\mathbf{X}, \mathbf{S}|\mathbf{A}, \mathbf{R}_e) = \prod_{k=1}^n \mathcal{N}(\mathbf{A} \mathbf{s}_k, \mathbf{R}_e) \times p(\mathbf{s}_k). \quad (3.9)$$

Le problème se ramène donc à celui de l'estimation conjointe de la matrice du mélange et de la covariance du bruit par maximisation de la log-vraisemblance complète

$$(\hat{\mathbf{A}}, \hat{\mathbf{R}}_e) = \arg \max_{\mathbf{A}, \mathbf{R}_e} \log p(\mathbf{X}, \mathbf{S}|\mathbf{A}, \mathbf{R}_e) \quad (3.10)$$

### 3.2.1.3 Discussion

La séparation effective par maximum de vraisemblance est réalisée en utilisant une méthode de maximisation directe de la vraisemblance ou par minimisation de l'anti log-vraisemblance. On distingue principalement deux solutions. La première est l'algorithme du gradient relatif [CL96] appelé aussi gradient naturel [ACY96]. La seconde consiste à utiliser les algorithmes *Expectation-Maximization* (EM) [DLR77, MK97]. En général, la méthode du gradient est utilisée dans le cas d'un mélange inversible et d'un bruit d'observation négligeable, alors que les algorithmes EM sont privilégiés dans le cas de mélanges bruités ou dans le cas d'une matrice de mélange rectangulaire ( $m > p$ ).

### 3.2.2 Optimisation en utilisant l'algorithme du gradient

En l'absence de bruit sur les observations, le critère à minimiser est défini par (3.6). En calculant le gradient de ce critère par rapport à  $\mathbf{B}$ , il en résulte

$$\frac{\partial}{\partial \mathbf{B}} \mathcal{L}_n(\mathbf{X}|\mathbf{B}) = -[\mathbf{B}^T]^{-1} + \frac{1}{n} \sum_{k=1}^n \varphi(\mathbf{B}\mathbf{x}_k) \mathbf{x}_k^T, \quad (3.11)$$

$$= \left[ \frac{1}{n} \sum_{k=1}^n \varphi(\mathbf{B}\mathbf{x}_k) (\mathbf{B}\mathbf{x}_k)^T - \mathbf{I} \right] [\mathbf{B}^T]^{-1}, \quad (3.12)$$

où  $\varphi(\cdot)$  est une fonction vectorielle, définie de  $\mathbb{R}^p \mapsto \mathbb{R}^p$  tel que

$$\forall z \in \mathbb{R}^p; \quad \varphi_j(z_j) = -\frac{\partial}{\partial z_j} \log p_j(z_j), \quad \text{pour } j = 1, \dots, p \quad (3.13)$$

En post-multipliant le gradient par  $\mathbf{B}^T \mathbf{B}$  et en notant  $\mathbf{y}_k = \mathbf{B}\mathbf{x}_k$ , on obtient le gradient relatif/naturel, exprimé par

$$\nabla(\mathbf{B}) = \left[ \frac{1}{n} \sum_{k=1}^n \varphi(\mathbf{y}_k) (\mathbf{y}_k)^T - \mathbf{I} \right] \mathbf{B}, \quad (3.14)$$

et pour un pas d'adaptation  $\mu_b$ , la mise à jour à chaque itération devient

$$\hat{\mathbf{B}}^{(r+1)} = \left[ (1 + \mu_b) \mathbf{I} - \mu_b \frac{1}{n} \sum_{k=1}^n \varphi(\hat{\mathbf{y}}_k^{(r)}) (\hat{\mathbf{y}}_k^{(r)})^T \right] \hat{\mathbf{B}}^{(r)}, \quad (3.15)$$

**Récapitulatif :** La séparation de sources par l'approche du maximum de vraisemblance, en utilisant le gradient relatif/naturel, se décompose en six étapes :

1. Suppression de la moyenne empirique des signaux de mélange

$$\overline{\mathbf{X}} = \mathbf{X} - (\boldsymbol{\mu}_x \otimes \mathbf{1}_{1 \times n}),$$

où  $\boldsymbol{\mu}_x = \frac{1}{n} \sum_{k=1}^n \mathbf{x}(k)$  et  $\overline{\mathbf{X}}$  est la matrice de données centrées,  $\otimes$  représente le produit de kronecker et  $\mathbf{1}_{m \times n}$  une matrice de dimension  $(m \times n)$  dont tous les éléments valent 1.

2. Réduction de dimension si  $m > p$ ,

$$\widetilde{\mathbf{X}} = \mathbf{W} \overline{\mathbf{X}},$$

où  $\widetilde{\mathbf{X}}$  est de dimension  $(p \times n)$  et  $\mathbf{W}$  est de dimension  $(p \times m)$ . Cette matrice peut être choisie de façon à effectuer un blanchiment spatial des sources ou par transformation aléatoire positive.

3. Initialisations :  $\mathbf{B}^{(0)}, \boldsymbol{\theta}^{(0)}$ ,

4. Mise à jour de la matrice de séparation  $\mathbf{B}$  et des signaux sources  $\mathbf{Y}$ ,

$$\begin{cases} \hat{\mathbf{B}}^{(r+1)} = \hat{\mathbf{B}}^{(r)} - \mu_b^{(r)} \nabla(\mathbf{B}|\hat{\boldsymbol{\theta}}^{(r)}), \\ \hat{\mathbf{Y}}^{(r+1)} = \hat{\mathbf{B}}^{(r+1)} \widetilde{\mathbf{X}}. \end{cases}$$



5. Mise à jour des paramètres des densités de probabilité des sources,

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} - \mu_{\boldsymbol{\theta}}^{(r)} \nabla \left( \boldsymbol{\theta} | \mathbf{Y}^{(r+1)} \right).$$

Tant que l'algorithme n'a pas convergé vers un point stationnaire, les deux dernières étapes (4) et (5) sont répétées.

6. Estimation finale des sources est obtenue en rajoutant la moyenne estimée

$$\hat{\mathbf{S}} = \hat{\mathbf{Y}} + \hat{\mathbf{B}} \mathbf{W} (\boldsymbol{\mu}_x \otimes \mathbf{1}_{1 \times n}),$$

**Remarque :** cette méthode d'estimation n'impose pas de façon stricte l'orthogonalité des sources et par conséquent permet de séparer des signaux dont la matrice d'autocovariance empirique est différente d'une matrice diagonale. Signalons aussi que l'estimateur du maximum de vraisemblance peut être appliqué également sous la contrainte d'orthogonalité des sources. Nous pouvons consulter à ce propos [CL96].

### 3.2.3 Optimisation en utilisant un algorithme EM

En présence de bruit dans les observations, la maximisation de la vraisemblance est réalisée en utilisant un algorithme EM.

#### 3.2.3.1 Algorithme EM et ses variantes stochastiques

L'algorithme EM (*Expectation Maximization*) [DLR77] permet la recherche de la solution du maximum de vraisemblance. Considérons les données  $\mathbf{x}$  et les variables dites *cachées*,  $\mathbf{s}$ , qui dans un contexte de séparation de sources correspondent, respectivement, aux observations et aux sources. Les données complètes sont notées  $\mathbf{z} = (\mathbf{x}, \mathbf{s})$  et notons  $\boldsymbol{\theta}$  le vecteur de paramètres à estimer par la maximisation de  $p(\mathbf{x}|\boldsymbol{\theta})$ .

L'algorithme EM à chaque itération ( $r$ ), est constitué de deux étapes :

1. Étape *Expectation*, notée E, dans laquelle la fonctionnelle EM est calculée selon

$$Q \left( \boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(r)} \right) \triangleq \mathbb{E} \left[ \log p(\mathbf{s}, \mathbf{x} | \boldsymbol{\theta}) | \mathbf{x}, \hat{\boldsymbol{\theta}}^{(r)} \right], \quad (3.16)$$

où  $\boldsymbol{\theta}^{(r)}$  est la valeur du paramètre obtenue à l'itération précédente.

2. Étape *Maximization*, notée M, dans laquelle la valeur du paramètre d'intérêt est remise à jour selon

$$\hat{\boldsymbol{\theta}}^{(r+1)} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(r)}) \quad (3.17)$$

Ces deux étapes sont alternées jusqu'à convergence de l'algorithme vers un point stationnaire de la vraisemblance. Lorsque l'une de ces deux étapes (E ou M) est difficile à mettre en œuvre, des variantes de cet algorithme sont alors adoptées. Dans le cas où la maximisation de la fonctionnelle admet une solution explicite, alors que l'étape d'évaluation de la fonctionnelle  $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(r)})$  nécessite des calculs analytiques complexes, une intégration de Monte Carlo sur des réalisations des données manquantes

obtenues par simulation de la densité conditionnelle  $p(\mathbf{s}|\mathbf{x}, \hat{\boldsymbol{\theta}})$  permettra de surmonter cette difficulté. C'est le principe de l'algorithme MCEM [WT90] pour *Monte Carlo EM*. A chaque itération, l'étape E est divisée en deux étapes

1. Simulation de  $M(k)$  réalisations des données manquantes

$$\mathbf{s}^{(\ell)} \sim p\left(\mathbf{s}|\mathbf{x}, \hat{\boldsymbol{\theta}}^{(r)}\right), \quad \text{pour } \ell = 1, \dots, M(r), \quad (3.18)$$

2. Calcul de la fonctionnelle EM via une intégration de Monte Carlo

$$Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}\right) \approx \frac{1}{M(r)} \sum_{\ell=1}^{M(r)} \log p\left(\mathbf{x}, \mathbf{s}^{(\ell)}|\hat{\boldsymbol{\theta}}^{(r)}\right) \quad (3.19)$$

Le point crucial avec cet algorithme réside dans le choix du nombre de réalisations  $M(r)$  qui permet d'obtenir un meilleur compromis entre coût de calcul, bonne approximation de la fonctionnelle EM et temps de convergence de l'algorithme. Un grand nombre de simulations réduirait les erreurs d'approximation de la fonctionnelle EM mais augmenterait sensiblement le coût de calcul. Aucune étude théorique n'est disponible sur le choix de ce paramètre. Néanmoins, le nombre de réalisations peut être fixé à une grande valeur au début de l'algorithme et réduit au cours des itérations [DLM99]. Cet algorithme peut être vu comme une généralisation de l'algorithme SEM (*Stochastic EM*) [CD85] qui a été présenté auparavant avec un nombre de réalisations  $M(r)$  fixé à 1. Afin d'éviter des simulations trop nombreuses tout en conservant de bonnes propriétés de convergence, une approximation stochastique de la fonctionnelle EM a été suggérée dans [DLM99]. La différence de cet algorithme par rapport aux algorithmes MCEM et SEM réside dans le fait que le calcul de la fonctionnelle EM prend en compte aussi sa valeur calculée durant les itérations précédentes. L'étape d'approximation stochastique introduit une suite décroissante de coefficients positifs  $\{\gamma_r\}$  et réalise l'approximation de la fonctionnelle EM selon

$$\hat{Q}^{(r+1)}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}\right) = \hat{Q}^{(r)}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}\right) + \gamma_r \left[ Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}\right) - \hat{Q}^{(r)}\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}\right) \right]. \quad (3.20)$$

où la fonction  $Q\left(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(r)}\right)$  est calculée selon la relation (3.19). Une étude de convergence de cet algorithme est présentée dans [DLM99].

### 3.2.3.2 Application des algorithmes EM en séparation de sources

Plusieurs auteurs ont utilisé ces algorithmes dans le cadre de la séparation de sources par maximum de vraisemblance [Bel95, MCG97, Att99, Ber00]. Les paramètres d'intérêt dans ce cas sont les coefficients de mélange et la matrice de covariance du bruit  $\mathbf{R}_e$ ; autrement dit :  $\boldsymbol{\theta} = (\mathbf{A}, \mathbf{R}_e)$ . Les signaux sources sont formulés comme étant des données manquantes. Comme la maximisation de la vraisemblance est équivalente à la maximisation de la fonctionnelle EM définie par

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = \int \log p(\mathbf{x}, \mathbf{s}|\boldsymbol{\theta}) \cdot p(\mathbf{s}|\mathbf{x}, \hat{\boldsymbol{\theta}}) d\mathbf{s}, \quad (3.21)$$

on exprime donc la log-vraisemblance complète par

$$\log p(\mathbf{x}, \mathbf{s}|\boldsymbol{\theta}) = \log p(\mathbf{x}|\boldsymbol{\theta}) + \log p(\mathbf{s}|\mathbf{x}, \hat{\boldsymbol{\theta}}), \quad (3.22)$$

où la densité de probabilité conditionnelle des signaux sources est donnée par  $p(\mathbf{s}|\mathbf{x}, \hat{\boldsymbol{\theta}})$ .

Dans le cadre de la séparation de sources, à chaque itération ( $r$ ), le calcul de la fonctionnelle EM donne [Bel95, Ber00] :

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(r)}) = -\frac{1}{2} \sum_{k=1}^n \mathbb{E} \left[ (\mathbf{x}_k - \mathbf{A}\mathbf{s}_k)^T \left[ \hat{\mathbf{R}}_e^{(r)} \right]^{-1} (\mathbf{x}_k - \mathbf{A}\mathbf{s}_k) | \mathbf{x}_k, \hat{\boldsymbol{\theta}}^{(r)} \right], \quad (3.23)$$

qui, en introduisant les matrices de corrélation empirique *a posteriori* des sources conditionnellement à la valeur courante du paramètre  $\boldsymbol{\theta}$ ,

$$\hat{\mathbf{R}}_{xx} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \mathbf{x}_k^T, \quad (3.24)$$

$$\hat{\mathbf{R}}_{xs}^{(r)} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k \mathbb{E} \left[ \mathbf{s}_k^T | \mathbf{x}_k, \hat{\boldsymbol{\theta}}^{(r)} \right], \quad (3.25)$$

$$\hat{\mathbf{R}}_{ss}^{(r)} = \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[ \mathbf{s}_k \mathbf{s}_k^T | \mathbf{x}_k, \hat{\boldsymbol{\theta}}^{(r)} \right], \quad (3.26)$$

prend alors la forme

$$Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(r)}) = -\frac{n}{2} \text{Trace} \left\{ \left[ \hat{\mathbf{R}}_e^{(r)} \right]^{-1} \left( \hat{\mathbf{R}}_{xx} - \mathbf{A} \left[ \hat{\mathbf{R}}_{xs}^{(r)} \right]^T - \hat{\mathbf{R}}_{xs}^{(r)} \mathbf{A}^T - \mathbf{A} \hat{\mathbf{R}}_{ss}^{(r)} \mathbf{A}^T \right) \right\} - \frac{n}{2} \log |\det \hat{\mathbf{R}}_e^{(r)}|. \quad (3.27)$$

En dérivant par rapport à  $\mathbf{A}$  et  $\mathbf{R}_e$  respectivement, on déduit la solution

$$\hat{\mathbf{A}}^{(r+1)} = \hat{\mathbf{R}}_{xs}^{(r)} \left[ \hat{\mathbf{R}}_{ss}^{(r)} \right]^{-1}, \quad (3.28)$$

$$\hat{\mathbf{R}}_e^{(r+1)} = \hat{\mathbf{R}}_{xx} - \hat{\mathbf{A}}^{(r+1)} \left[ \hat{\mathbf{R}}_{xs}^{(r)} \right]^T. \quad (3.29)$$

**Récapitulatif :** les principales étapes de séparation de sources par l'approche du maximum de vraisemblance en utilisant l'algorithme EM peuvent être résumées en cinq étapes :

1. Calcul de  $\hat{\mathbf{R}}_x$  et initialisations de  $\hat{\mathbf{A}}^{(0)}$ ,  $\hat{\mathbf{R}}_e^{(0)}$  ;
2. Calcul des statistiques,  $\hat{\mathbf{R}}_{xs}^{(r)}$  et  $\hat{\mathbf{R}}_{ss}^{(r)}$  ;
3. Maximisation de la fonctionnelle EM,
4. Tant que l'algorithme n'a pas convergé vers un point stationnaire, les deux dernières étapes (2) et (3) sont répétées.
5. Après convergence, réaliser l'estimation des signaux sources à partir de leurs densités *a posteriori*  $p(\mathbf{S}|\mathbf{X}, \hat{\mathbf{A}}, \hat{\mathbf{R}}_e, \hat{\boldsymbol{\theta}}_s)$ .

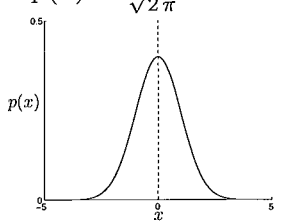
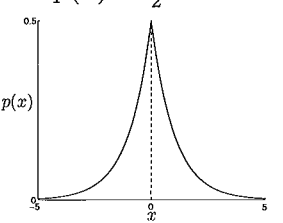
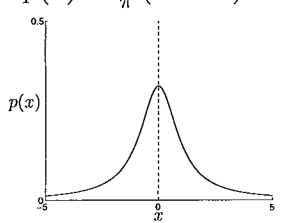
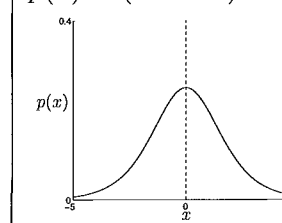
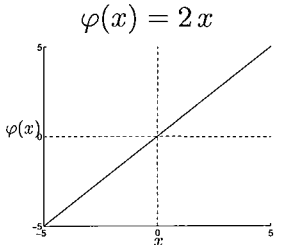
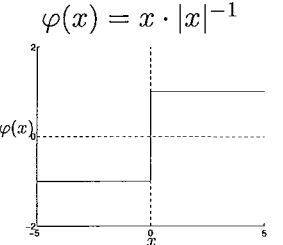
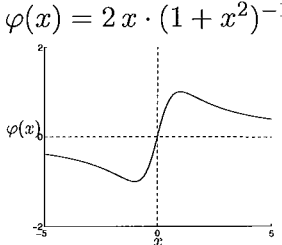
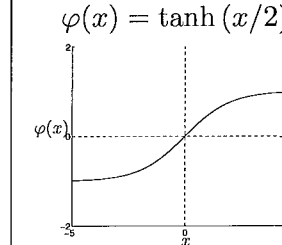
**Discussion :** l'estimation de la matrice de mélange et de la matrice de covariance du bruit via l'algorithme EM nécessite la connaissance des statistiques (3.25) et (3.26) qui requièrent l'estimation des moments  $\mathbb{E}[\mathbf{s}^T | \mathbf{x}, \hat{\boldsymbol{\theta}}^{(r)}]$  et  $\mathbb{E}[\mathbf{s}\mathbf{s}^T | \mathbf{x}, \hat{\boldsymbol{\theta}}^{(r)}]$ . Hormis le cas d'un modèle de mélange de gaussiennes, le calcul de ces statistiques n'est pas aisé. Une possibilité pour pallier cette difficulté est de simuler

les signaux sources à partir de leurs lois conditionnelles *a posteriori* en utilisant les techniques de Monte Carlo, ce qui revient à utiliser les algorithmes SEM, MCEM ou SAEM. Néanmoins, en plus de ce problème de calcul, se pose la question, plus cruciale à notre sens, du choix de la structure de la densité de probabilité des signaux sources.

### 3.2.4 Choix de la densité de probabilité des sources

Les solutions obtenues lors de la maximisation de la vraisemblance sont liées au choix des densités de probabilité des signaux sources. Dans le cas de l'algorithme du gradient, la fonction de séparation doit, idéalement, correspondre exactement à la densité de probabilité des sources [PGJ92, PG97, Car98]. Le tableau 3.1 illustre quelques densités et les fonctions non-linéaires correspondantes. Cependant, à l'exception de certaines applications comme les communications numériques [BC94, BC95], ces densités de probabilité ne sont pas connues *a priori*. Il faudra alors les estimer en même temps que la matrice de séparation. Le problème de séparation se ramène donc à une estimation conjointe de la matrice de séparation et des distributions des signaux sources. La distribution des sources est approchée soit en utilisant les statistiques d'ordre supérieur [GL90b, Com94] ou par un modèle paramétrique [PP96, MCG97].

TAB. 3.1 – Quelques densités de probabilité et fonctions non-linéaires correspondantes.

Gaussienne	Laplacienne	Cauchy	Logistic
$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ 	$p(x) = \frac{1}{2} e^{- x }$ 	$p(x) = \frac{1}{\pi} (1 + x^2)^{-1}$ 	$p(x) = (1 + e^{-x})^{-2} e^{-x}$ 
$\varphi(x) = 2x$ 	$\varphi(x) = x \cdot  x ^{-1}$ 	$\varphi(x) = 2x \cdot (1 + x^2)^{-1}$ 	$\varphi(x) = \tanh(x/2)$ 

La première solution consiste à utiliser les statistiques d'ordre supérieur à deux pour l'approximation des densités de probabilité des sources. Gaeta et Lacoume [GL90b, GL90a] ont obtenu une approximation de la log-vraisemblance en utilisant un développement de Gram-Charlier et de façon similaire, Comon [Com94] a proposé une approximation par un développement d'Edgeword. Une alternative à cette approche est d'utiliser un modèle paramétrique de la distribution des sources [CM99]. Le choix d'un modèle de distributions est effectué de sorte à prendre en compte certaines informations préalables sur la distribution des sources recherchées. Le tableau 3.2 présente quelques exemples de distributions.

TAB. 3.2 – Quelques modèles de densité de probabilité et leurs fonctions de séparation associées

distribution	paramètres $\theta$	$p(x; \theta)$	$\varphi(x; \theta)$
Gaussienne généralisée	$\alpha > 0, \beta > 0$	$\frac{\alpha}{2} \cdot \frac{\beta^{1/\alpha}}{\Gamma(1/\alpha)} e^{-\beta x ^\alpha}$	$\alpha\beta x \cdot  x ^{\alpha-2}$
Student	$\alpha > 0, \beta > 0$	$\frac{\beta^2/\sqrt{\alpha}}{\mathbf{B}(\alpha/2, 1/2)} \left[1 + \frac{\beta}{\alpha}x^2\right]^{-(\alpha+1)/2}$	$\frac{(1+\alpha)\beta}{\alpha + \beta x^2}$
Mélange de gaussiennes	$\alpha > 0, \beta > 0$	$\sqrt{\frac{\beta}{2\pi}} \left(e^{-\beta(x+\alpha)^2} + e^{-\beta(x-\alpha)^2}\right)$	$\beta x - \alpha \tanh(\beta x)$

### 3.2.4.1 Distribution gaussienne généralisée

Cette famille a été utilisée dans [CCA00, CM99]. Le paramètre  $\alpha$  permet d'ajuster la forme de la distribution et  $\beta$  est un paramètre de dispersion. Pour  $\alpha = 1$ , la distribution correspond à une loi laplacienne, pour  $\alpha = 2$  la distribution correspond à une loi normale et lorsque  $\alpha$  tend vers l'infini la distribution tend vers une loi uniforme.

### 3.2.4.2 Distribution de Student

Cette famille a été utilisée dans [CM99, Mac96] et la distribution de Cauchy en est un cas particulier pour  $\alpha = 1$ .

### 3.2.4.3 Mélange de distributions

Une premier mélange est celui d'une distribution sous-gaussienne avec une distribution sur-gaussienne [LGS99]. La fonction non-linéaires sont définie par

$$\varphi(x) = \begin{cases} \varphi_+(x) = x - \tanh(x) & \text{pour une source sur-gaussienne,} \\ \varphi_-(x) = x + \tanh(x) & \text{pour une source sous-gaussienne.} \end{cases} \quad (3.30)$$

Le choix entre l'une des deux fonction non-linéaire est réalisée soit selon le signe du kurtosis de chaque source estimée ou encore de façon à satisfaire la condition suffisante de stabilité de l'algorithme du gradient [Car00]

$$\mathbb{E}[\varphi'(x)x^2] - \mathbb{E}[\varphi(x)x] > 0. \quad (3.31)$$

Un modèle plus général consiste à approcher le densité de probabilité des sources à l'aide d'un mélange de gaussiennes [MCG97, Att99]. Ce modèle est très avantageux, notamment dans le cas de sources qui possèdent des distributions multimodales [Ber00] ou encore dans le cas de sources non-stationnaires [Sno03].

### 3.2.4.4 Système de Pearson

Ce système représente une famille de distributions dont la densité  $p(x)$  est solution de l'équation différentielle

$$\frac{dp(x)}{dx} = \frac{(x-a)}{b_0 + b_1x + b_2x^2} p(x). \quad (3.32)$$

Les paramètres de cette densité sont estimés en utilisant la méthode des moments. La fonction de séparation résultante est exprimée par

$$\varphi(x) = \frac{a - x}{b_0 + b_1x + b_2x^2}. \quad (3.33)$$

Ce modèle a été utilisé pour la séparation de sources par [KK02].

### 3.2.4.5 Discussion

En plus de l'aspect simplification des calculs lors de l'estimation, le choix d'un modèle de densité de probabilité des signaux sources permet une formulation explicite d'informations préalables sur les signaux sources. On peut donc dire que la séparation par maximum de vraisemblance consiste à trouver les composantes indépendantes dans la classe de distributions choisie. Par exemple, les modèles de gaussienne généralisée et de Student supposent implicitement que les sources sont à distributions symétriques et unimodales. D'autre part, le système de Pearson et le modèle de mélange de gaussiennes permettent de considérer des densités asymétriques ou multi-modales. Dans le cadre de notre application, tous ces modèles ne prennent pas en compte explicitement la non-négativité des signaux sources. Dans le paragraphe suivant, nous proposons une densité de probabilité utilisable dans le cas de sources spectrales et qui permet de prendre en compte explicitement la non-négativité des sources et l'asymétrie de leur distributions.

## 3.3 Séparation de sources spectrales par l'approche du maximum de vraisemblance

### 3.3.1 Modélisation de la distribution des sources

Afin de séparer des sources non-négatives, il est nécessaire de représenter la distribution des signaux sources par des densités de probabilité à support non-négatif. Plusieurs modèles statistiques peuvent être utilisés, néanmoins en prenant compte d'autres informations supplémentaires sur les signaux sources, certaines densités seront privilégiées. Dans le cadre de l'analyse des signaux spectroscopiques, les données présentent une parcimonie qui se traduit par une concentration de la densité de probabilité des sources proche du zéro. Une bonne loi candidate pour prendre compte ces information est la loi gamma à deux paramètres.

La densité Gamma à deux paramètres est exprimée par

$$s \sim \mathcal{G}(\alpha, \beta) \Leftrightarrow p(s) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} \exp[-\beta s] \times \mathbb{I}_{[0,+\infty[}(s), \quad (3.34)$$

où  $\Gamma(a)$  est la fonction gamma et  $\mathbb{I}_{[0,+\infty[}(s)$  est la fonction indicatrice, définie par

$$\mathbb{I}_{[0,+\infty[}(s) = \begin{cases} 1 & \text{si } s \in [0, +\infty[, \\ 0 & \text{sinon.} \end{cases} \quad (3.35)$$

La densité gamma appartient à la famille exponentielle et englobe la densité exponentielle, la densité de la loi du khi-2 et la densité d'Erlang comme cas particuliers [SO94]. Cette distribution présente

deux caractéristiques qui justifient son utilisation : (i) celle-ci prend en compte explicitement la non-négativité ; (ii) ses paramètres  $\alpha$  et  $\beta$ , permettent d'adapter sa forme à plusieurs situations. La figure 3.1 illustre les trois formes typiques de cette distribution. Le premier cas est bien adapté pour coder la parcimonie vu la concentration de sa distribution au voisinage de zéro (pour  $\alpha < 1$ ), alors que le deuxième cas permet de prendre en compte le cas de sources parcimonieuses superposées à un arrière-plan. Le troisième cas montre que la densité Gamma peut tendre vers une distribution uniforme à support positif. Bien qu'aucun choix n'est totalement objectif, nous pensons que la densité gamma est la distribution la plus flexible en terme d'adaptativité aux différentes situations.

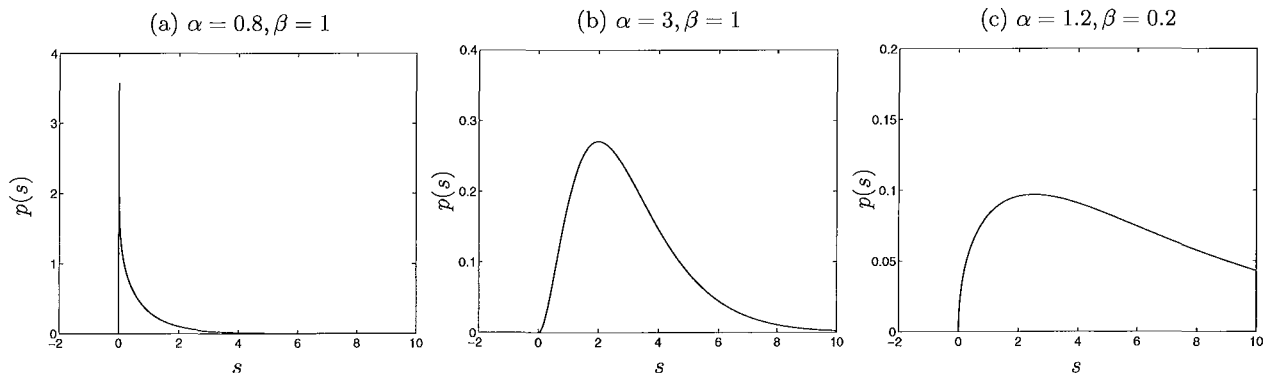


FIG. 3.1 – Illustration de la forme de la densité de probabilité Gamma.

### 3.3.2 Séparation en utilisant l'algorithme du gradient

#### 3.3.2.1 Synthèse d'une fonction non-linéaire

En choisissant une densité gamma comme modèle de distribution des sources, la fonction non-linéaire correspondante est exprimée par

$$\varphi_{gam}(s; \alpha, \beta) = -\frac{\partial}{\partial s} \log p(s; \alpha, \beta) \quad (3.36)$$

$$= \beta - (\alpha - 1)/s \quad \text{pour } s > 0. \quad (3.37)$$

Cette fonction non-linéaire est définie pour des sources positives alors que durant l'optimisation les estimations intermédiaires des sources peuvent présenter des valeurs négatives. Il faut alors redéfinir cette fonction non-linéaire pour que celle-ci soit également définie pour  $s \leq 0$ . Une possibilité consiste à introduire une rectification selon

$$\varphi_{rect}(s; \alpha, \beta) = \beta - (\alpha - 1)/\max(s, \epsilon), \quad (3.38)$$

où le seuil  $\epsilon$  est fixé à une faible valeur. L'introduction d'une telle rectification peut être formulée également à partir d'une modification de la loi *a priori* attribuée aux signaux sources de telle sorte à prendre en compte la possibilité d'obtenir des valeurs négatives mais avec une très faible probabilité. En effet, si l'on associe aux signaux sources la loi suivante

$$p(s) \propto \begin{cases} s^{\alpha-1} \exp\{-\beta s\} & \text{pour } s \geq \epsilon \\ \exp\{\lambda(s - \epsilon)\} & \text{pour } s < \epsilon, \end{cases} \quad (3.39)$$

où  $\lambda = (\alpha - 1)/\epsilon - \beta$ . Le calcul de la fonction non-linéaire associée à cette distribution permet d'obtenir la fonction définie par l'équation (3.38). Cette loi, que nous appellerons dans ce qui suit *densité gamma modifiée*, peut être vue comme un mélange de distributions

$$p(s) = \rho p_+(s) + (1 - \rho) p_-(s), \quad (3.40)$$

où les deux densités  $p_+(s)$  et  $p_-(s)$  sont définies par

$$\begin{cases} p_+(s) = \frac{\beta^\alpha}{\Gamma(\alpha) - \gamma(\beta\epsilon, \alpha)} s^{\alpha-1} \exp\{-\beta s\} \times \mathbb{I}_{[\epsilon, +\infty[}(s), \\ p_-(s) = \lambda \exp\{\lambda(s - \epsilon)\} \times \mathbb{I}_{]-\infty, \epsilon]}(s), \end{cases} \quad (3.41)$$

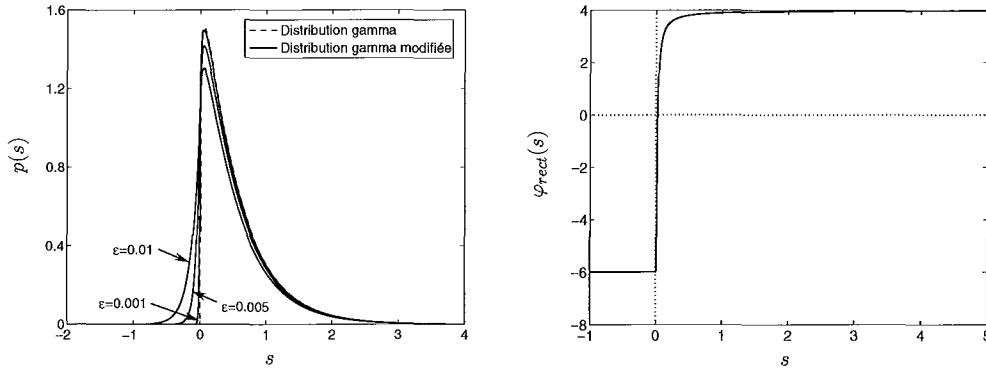
où  $\gamma(x, a)$  est la fonction gamma incomplète

$$\gamma(x, a) = \int_0^x t^{(a-1)} \exp\{-t\} dt.$$

Le coefficient  $\rho$  est choisi de telle sorte à assurer la continuité de la fonction  $p(s)$  en  $s = \epsilon$ . Ce qui donne

$$\rho p_+(\epsilon) = (1 - \rho) p_-(\epsilon) \Rightarrow \rho = \frac{1}{1 + p_+(\epsilon)/p_-(\epsilon)}. \quad (3.42)$$

La figure 3.2(a) illustre l'allure de la distribution gamma modifiée pour  $(\alpha = 1.1, \beta = 2)$  et trois valeurs différentes de  $\epsilon$ . Nous pouvons noter que lorsque ce seuil tend vers zero, la distribution gamma modifiée se rapproche de plus en plus de la distribution gamma. La figure 3.2(b) illustre la forme de la fonction non-linéaire résultante pour  $\epsilon = 0.01$ .



(a) distribution gamma modifiée pour  $\alpha = 1.1$  et  $\beta = 2$

(b) fonction non-linéaire associée pour  $\epsilon = 0.01$ ,  $\alpha = 1.1$  et  $\beta = 2$ .

FIG. 3.2 – Illustration de distribution gamma modifiée et de la fonction non-linéaire résultante.

### 3.3.2.2 Stabilité de l'algorithme

Des conditions suffisantes pour la stabilité de l'algorithme du gradient relatif/naturel sont exprimées par [ACC97, Car00]

$$\mathbb{E}[\varphi'(\tilde{s}_j) \tilde{s}_j^2] - \mathbb{E}[\varphi(\tilde{s}_j) \tilde{s}_j] > 0, \quad \forall j = 1, \dots, p \quad (3.43)$$



où  $E[f(\tilde{s}_j)]$  sont les moments non-linéaires des sources centrées

$$\tilde{s}_j = s_j - E[s_j] = s_j - \alpha_j/\beta_j. \quad (3.44)$$

Le calcul de ces moments donne

$$E[\varphi'(\tilde{s}_j)\tilde{s}_j^2] = E\left[\frac{(\alpha_j - 1)}{\tilde{s}_j^2} \cdot \tilde{s}_j^2\right] = (\alpha_j - 1), \quad (3.45)$$

et

$$E[\varphi(\tilde{s}_j)\tilde{s}_j] = E[\beta_j\tilde{s}_j - (\alpha_j - 1)] = (1 - \alpha_j). \quad (3.46)$$

Par conséquent, une condition suffisante pour la stabilité de l'algorithme est

$$\alpha_j > 1, \quad j = 1, \dots, p. \quad (3.47)$$

A partir de cette condition de stabilité, il apparaît que seules des densités gamma ayant un paramètre de forme  $\alpha > 1$  peuvent être utilisées.

### 3.3.2.3 Estimation des paramètres des sources

Les paramètres des densités gamma  $\boldsymbol{\theta} = [\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p]^T$  sont estimés conjointement avec la matrice de séparation par la maximisation de la fonction de vraisemblance conjointe

$$(\hat{\mathbf{B}}, \hat{\boldsymbol{\theta}}) = \arg \max_{\mathbf{B}, \boldsymbol{\theta}} p(\mathbf{X}|\mathbf{B}, \boldsymbol{\theta}), \quad (3.48)$$

La minimisation du critère qui en résulte est réalisée par un algorithme de gradient alterné mettant à jour, à chaque itération, la matrice de séparation en utilisant la dernière estimation du vecteur de paramètres et en utilisant la nouvelle valeur de la matrice de séparation pour mettre à jour les paramètres. La mise à jour de la matrice de séparation étant réalisée selon l'équation (3.15) alors que les paramètres des densités gamma sont mis à jour à partir de la dernière estimation des sources, en utilisant un algorithme du gradient. Soit  $s_{(j,1:n)}$  l'estimation la plus récente de la  $j$ ème source qui est supposée distribuée selon une loi gamma dont on veut estimer les paramètres  $(\alpha_j, \beta_j)$ . La vraisemblance s'exprime selon

$$p(s_{(j,1:n)}|\alpha_j, \beta_j) = \left(\frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\right)^n \prod_{k=1}^n \left(s_{(j,k)}^{\alpha_j-1}\right) \exp\left\{-\sum_{k=1}^n \beta_j s_{(j,k)}\right\}, \quad (3.49)$$

et le critère associé est défini par

$$\mathcal{J}(s_{(j,1:n)}|\alpha_j, \beta_j) = -\frac{1}{n} \log p(s_{(j,1:n)}|\alpha_j, \beta_j), \quad (3.50)$$

$$= -\alpha_j \log \beta_j + \log \Gamma(\alpha_j) - (\alpha_j - 1) \frac{1}{n} \sum_{k=1}^n \log s_{(j,k)} + \frac{1}{n} \sum_{k=1}^n \beta_j s_{(j,k)}. \quad (3.51)$$

En utilisant l'algorithme du gradient, la mise à jour des paramètres à chaque itération  $r$  est réalisée par

$$\begin{cases} \alpha_j^{(r+1)} = \alpha_j^{(r)} - \mu_\alpha \nabla_{\alpha_j} \mathcal{J}(s_{(j,1:n)}|\alpha_j^{(r)}, \beta_j^{(r)}), \\ \beta_j^{(r+1)} = \beta_j^{(r)} - \mu_\beta \nabla_{\beta_j} \mathcal{J}(s_{(j,1:n)}|\alpha_j^{(r+1)}, \beta_j^{(r)}), \end{cases} \quad (3.52)$$

où  $(\mu_\alpha, \mu_\beta)$  sont des pas d'adaptation et les gradients du critère par rapport aux deux paramètres sont donnés par

$$\begin{cases} \nabla_{\alpha} \mathcal{J}(s_{(j,1:n)} | \alpha_j^{(r)}, \beta_j^{(r)}) = -\log \beta_j^{(r)} + \psi(\alpha_j^{(r)}) - \frac{1}{n} \sum_{k=1}^n \log s_{(j,k)} \\ \nabla_{\beta} \mathcal{J}(s_{(j,1:n)} | \alpha_j^{(r+1)}, \beta_j^{(r)}) = -\alpha_j^{(r+1)} / \beta_j^{(r)} + \frac{1}{n} \sum_{k=1}^n s_{(j,k)}, \end{cases}$$

où  $\psi(\alpha) = \frac{d}{d\alpha} \Gamma(\alpha)$  est la fonction digamma [AS72]. En tenant compte des conditions de stabilité (3.47) et de la définition de la densité gamma, les valeurs de  $\alpha_j$  et  $\beta_j$  sont contraintes à  $\alpha_j > 1$  et  $\beta_j > 0$ .

### 3.3.3 Séparation en utilisant l'algorithme MCEM

#### 3.3.3.1 Formulation de la densité conditionnelle *a posteriori* des données manquantes

On donne l'écriture des densités conditionnelles *a posteriori* des données manquantes. La vraisemblance est écrite selon

$$p(\mathbf{X} | \mathbf{A}, \mathbf{S}, \mathbf{R}_e) = \prod_{k=1}^n \mathcal{N}(\mathbf{A} \mathbf{s}_k, \mathbf{R}_e); \quad (3.53)$$

La densité conditionnelle *a posteriori* des signaux sources est donnée par

$$p(\mathbf{S} | \mathbf{X}, \mathbf{A}, \mathbf{R}_e) \propto p(\mathbf{X} | \mathbf{A}, \mathbf{S}, \mathbf{R}_e) \times p(\mathbf{S} | \boldsymbol{\theta}_s); \quad (3.54)$$

où la densité *a priori* est donnée par

$$p(\mathbf{S} | \boldsymbol{\theta}_s) = \prod_{k=1}^n \prod_{j=1}^p \mathcal{G}(\alpha_j, \beta_j); \quad (3.55)$$

avec  $\boldsymbol{\theta}_s = [\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p]^T$ . La densité conditionnelle *a posteriori* des données manquantes est exprimée par

$$p(\mathbf{S} | \mathbf{X}, \mathbf{A}, \mathbf{R}_e, \boldsymbol{\theta}_s) = \prod_{k=1}^n \mathcal{N}(\boldsymbol{\mu}_{s_k}, \mathbf{R}_s) \times \prod_{k=1}^n \prod_{j=1}^p \mathcal{G}(\alpha_j, \beta_j). \quad (3.56)$$

où  $\mathbf{R}_s = (\mathbf{A}^T \mathbf{R}_e^{-1} \mathbf{A})^{-1}$  et  $\boldsymbol{\mu}_{s_k} = \mathbf{R}_s \mathbf{A}^T \mathbf{R}_e^{-1} \mathbf{x}_k$ . Cette densité conditionnelle *a posteriori* dépend aussi de la valeur des paramètres de la densité *a priori* des signaux sources. Pour une estimation non-supervisée, ces hyperparamètres doivent aussi être inférés. Ainsi, la densité conditionnelle *a posteriori* des données manquantes est réécrite selon

$$p(\mathbf{S}, \boldsymbol{\theta}_s | \mathbf{X}, \mathbf{A}, \mathbf{R}_e) \propto p(\mathbf{S} | \mathbf{X}, \mathbf{A}, \mathbf{R}_e, \boldsymbol{\theta}_s) \times p(\boldsymbol{\theta}_s), \quad (3.57)$$

où  $p(\boldsymbol{\theta}_s)$  est la densité *a priori* des hyperparamètres  $\boldsymbol{\theta}_s$ .

#### 3.3.3.2 Simulation de la densité conditionnelle *a posteriori* des données manquantes

Afin de simuler la densité (3.57), une version scalaire de l'algorithme de Gibbs est utilisée. A chaque itération ( $\ell = 1, \dots, M(r)$ ), l'algorithme de Gibbs consiste à

1. simuler, pour  $k = 1, \dots, n$ , les variables

$$s_{(1,k)}^{(\ell+1)} \sim p \left( s_{(1,k)} | s_{(2:p,k)}^{(\ell)}, \theta_{s_{(1,k)}}^{likel}, \theta_{s_1}^{prior} \right), \quad (3.58)$$

⋮

$$s_{(j,k)}^{(\ell+1)} \sim p \left( s_{(1,k)} | s_{(1:j-1,k)}^{(\ell+1)}, s_{(j+1:p,k)}^{(\ell)}, \theta_{s_{(j,k)}}^{likel}, \theta_{s_j}^{prior} \right), \quad (3.59)$$

⋮

$$s_{(p,k)}^{(\ell+1)} \sim p \left( s_{(1,k)} | s_{(1:p-1,k)}^{(\ell+1)}, \theta_{s_{(1,1:n)}}^{likel}, \theta_{s_p}^{prior} \right), \quad (3.60)$$

où  $\theta_{s_{(j,1:n)}}^{likel} = \{x_{(1:m,1:n)}, s_{(1:j-1,1:n)}^{(r+1)}, s_{(j+1:p,1:n)}^{(r)}, \mathbf{A}, \mathbf{R}_e\}$  et  $\theta_{s_j}^{prior} = \{\alpha_j^{(r)}, \beta_j^{(r)}\}$ .

2. simuler, pour  $j = 1, \dots, p$ , les paramètres

$$\alpha_j^{(r+1)} \sim p \left( \alpha_j | s_{(j,1:n)}^{(\ell+1)}, \beta_j^{(\ell)} \right), \quad (3.61)$$

$$\beta_j^{(r+1)} \sim p \left( \beta_j | s_{(j,1:n)}^{(\ell+1)}, \alpha_j^{(\ell+1)} \right). \quad (3.62)$$

L'annexe A présente les détails du calcul des densités conditionnelles *a posteriori* des sources et les techniques proposées pour la simulation de ces densités sont présentées dans l'annexe B. L'annexe C décrit la méthode proposée pour la simulation des paramètres des densités gamma.

**Illustration :** Cet algorithme est appliqué pour simuler la densité (3.56) avec des paramètres :

$$\boldsymbol{\mu}_s = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix}, \mathbf{R}_s = \begin{bmatrix} 0.1 & \rho_{12} \\ \rho_{12} & 0.1 \end{bmatrix}, \boldsymbol{\alpha}_s = \begin{bmatrix} 0.5 \\ 2 \end{bmatrix} \text{ et } \boldsymbol{\beta}_s = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad (3.63)$$

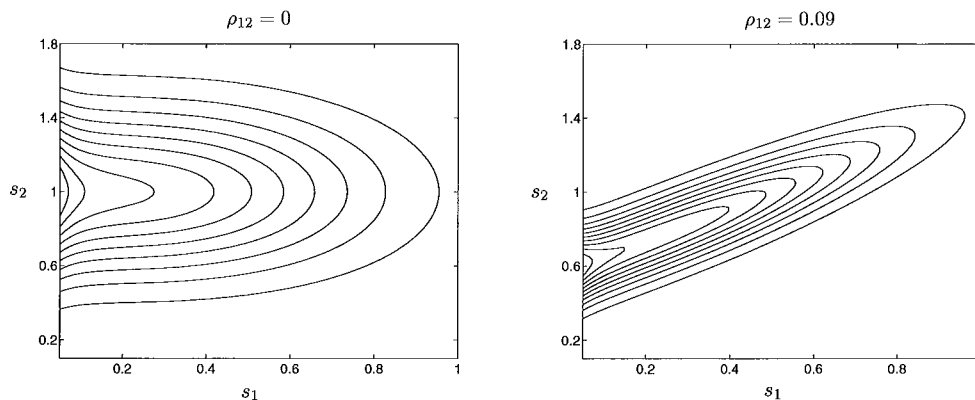
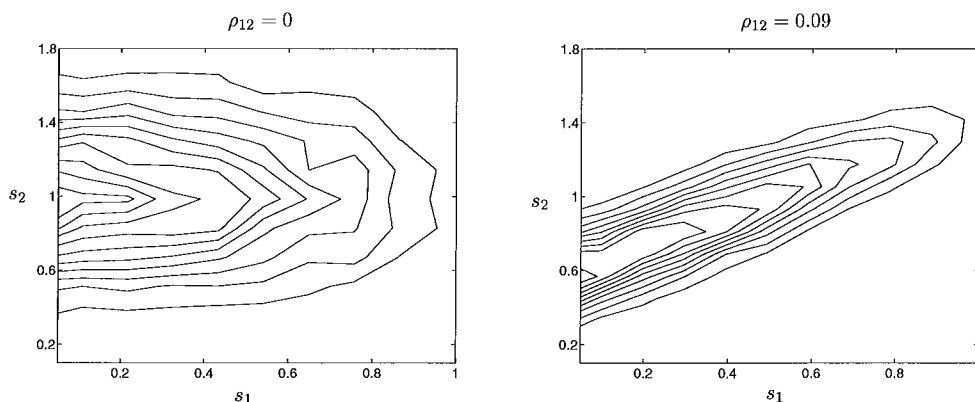
où le coefficient  $\rho_{12}$  permet d'envisager le cas de sources indépendantes et le cas de sources corrélées. Le nombre de réalisations simulées est  $M = 5000$  et le nombre d'itérations de chauffage est fixé à 10 itérations. Nous pouvons constater sur la figure 3.3 que les échantillons possèdent des distributions similaires à celles des distributions cibles. Cet exemple illustre que l'algorithme proposé permet de réaliser la simulation de la densité conditionnelle *a posteriori* des sources.

### 3.3.3.3 Estimation des statistiques requises pour la maximisation

A partir de réalisation simulées à partir de la densité conditionnelle des sources, les statistiques nécessaires pour l'étape de maximisation dans l'algorithme EM sont approchées par

$$\mathbb{E} \left[ \mathbf{s}_k^T | \mathbf{x}_k, \hat{\boldsymbol{\theta}}^{(r)} \right] \approx \frac{1}{M^{(r)}} \sum_{\ell=1}^{M^{(r)}} \left[ \mathbf{s}_k^{(\ell)} \right]^T, \quad (3.64)$$

$$\mathbb{E} \left[ \mathbf{s}_k \mathbf{s}_k^T | \mathbf{x}_k, \hat{\boldsymbol{\theta}}^{(r)} \right] \approx \frac{1}{M^{(r)}} \sum_{\ell=1}^{M^{(r)}} \mathbf{s}_k^{(\ell)} \left[ \mathbf{s}_k^{(\ell)} \right]^T, \quad (3.65)$$

(a) Densité conditionnelle *a posteriori* des sources pour  $\rho_{12} = 0$  et pour  $\rho_{12} = 0.9$ (b) Distribution des échantillons simulés à partir des densités conditionnelles *a posteriori* des sources pour  $\rho_{12} = 0$  et pour  $\rho_{12} = 0.9$ FIG. 3.3 – Illustration de l'algorithme de simulation de la densité conditionnelle *a posteriori* des sources.

et les matrices sont mises à jour par

$$\hat{\mathbf{R}}_{xs}^{(r+1)} = (1 - \gamma_r) \hat{\mathbf{R}}_{xs}^{(r)} + \gamma_r \left[ \frac{1}{nM(r)} \sum_{k=1}^n \sum_{\ell=1}^{M(r)} \mathbf{x}_k \left[ \mathbf{s}_k^{(\ell)} \right]^T \right], \quad (3.66)$$

$$\hat{\mathbf{R}}_{ss}^{(r+1)} = (1 - \gamma_r) \hat{\mathbf{R}}_{ss}^{(r)} + \gamma_r \left[ \frac{1}{nM(r)} \sum_{k=1}^n \sum_{\ell=1}^{M(r)} \mathbf{s}_k^{(\ell)} \left[ \mathbf{s}_k^{(\ell)} \right]^T \right]. \quad (3.67)$$

Les quantités  $M(r)$  et  $\gamma_r$  permettent de choisir le type d'algorithme (SEM, MCEM ou SAEM). L'organigramme de la méthode qui en résulte est illustré par la figure 3.4.

### 3.4 Illustration des algorithmes sur des données synthétiques

Cette section a pour but d'illustrer l'approche de séparation de sources non-négatives par maximum de vraisemblance. Les signaux sources et les coefficients de mélange sont simulés avec des formes similaires aux données spectroscopiques. La figure 3.5 montre un exemple de signaux sources qui sont normalisés à une variance unité afin de lever l'indétermination d'échelle. Ces sources possèdent des

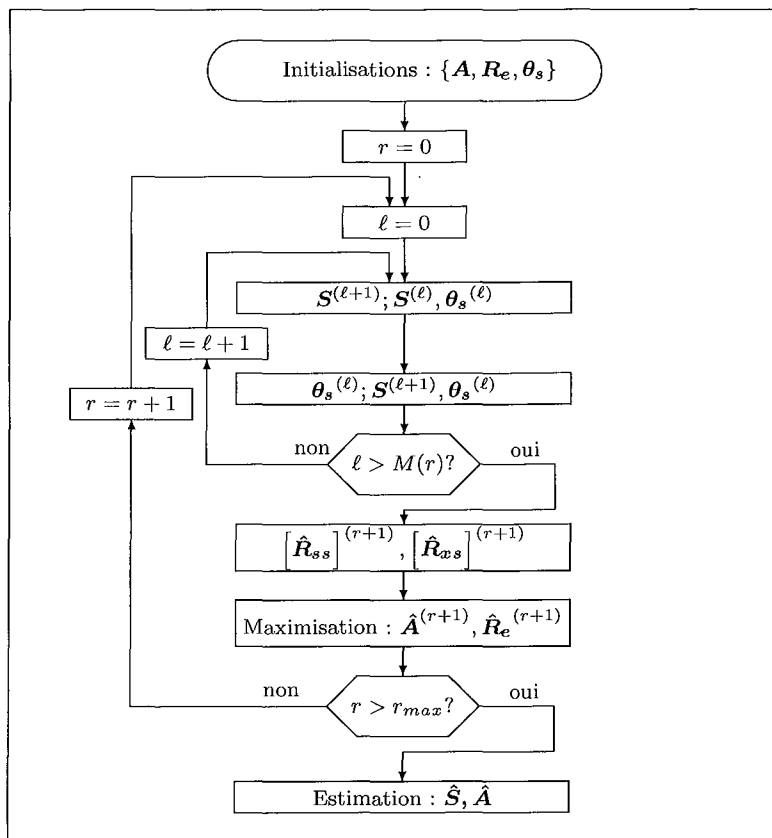


FIG. 3.4 – Organigramme simplifié de la méthode de séparation par MCEM

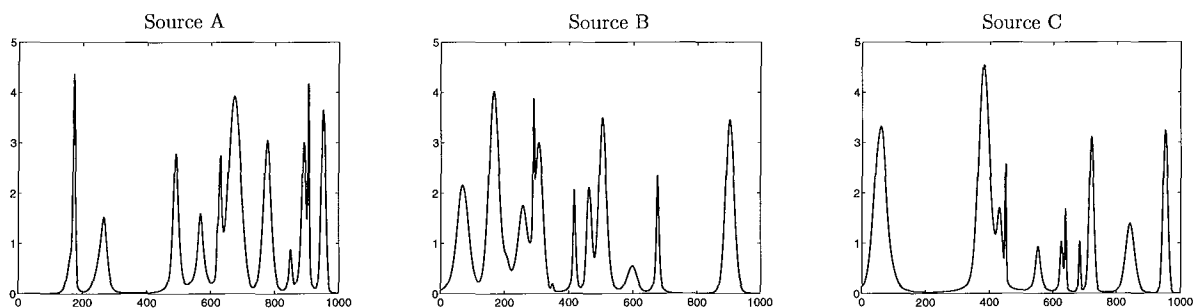


FIG. 3.5 – Signaux sources simulés.

coefficients de corrélation mutuelle  $\{\rho_{12} = 0.13, \rho_{13} = -0.18, \rho_{23} = -0.22\}$  et la matrice de mélange est arbitrairement choisie de la forme

$$\mathbf{A} = \begin{bmatrix} 0.150 & 0.250 & 0.600 \\ 0.350 & 0.450 & 0.200 \\ 0.600 & 0.300 & 0.100 \end{bmatrix}. \quad (3.68)$$

### 3.4.1 Séparation en utilisant l'algorithme du gradient relatif

Des observations avec un faible niveau de bruit sont considérées. Un bruit gaussien i.i.d est rajouté aux observations de telle sorte à avoir un RSB de 40 dB sur chaque observation. L'approche de séparation par maximum de vraisemblance et gradient relatif a été appliquée en utilisant la fonction non-linéaire déduite à partir de la distribution gamma modifiée. Une analyse de Monte Carlo est réalisée avec 100 simulations dans lesquelles les valeurs initiales de la matrice de séparation sont choisis à partir d'une loi normale de moyenne nulle et variance unité. La figure 3.6 résume le résultat d'une analyse de Monte Carlo avec 100 simulations l'évolution de l'anti log-vraisemblance au cours des itérations ainsi que la courbe de convergence des coefficients de la matrice du système globale  $\hat{\mathbf{B}}\mathbf{A}$  durant l'optimisation. On note, d'une part, une stabilisation de la vraisemblance après 5000 itérations et d'autre part, la convergence de la matrice séparante vers l'inverse de la matrice de mélange. Après 10000 itérations et après correction des indéterminations d'ordre et d'échelle, la valeur de cette matrice est

$$\hat{\mathbf{B}}\mathbf{A} = \begin{bmatrix} 1.0006 & 0.0023 & 0.0036 \\ 0.0038 & 1.0003 & 0.0053 \\ 0.0053 & 0.0011 & 1.0012 \end{bmatrix}. \quad (3.69)$$

Les termes diagonaux de cette matrice sont très petits, ce qui signifie une bonne estimation de la matrice de mélange et par conséquent une séparation des composantes non-négatives avec succès, même si ces sources présentent des corrélations mutuelles significatives.

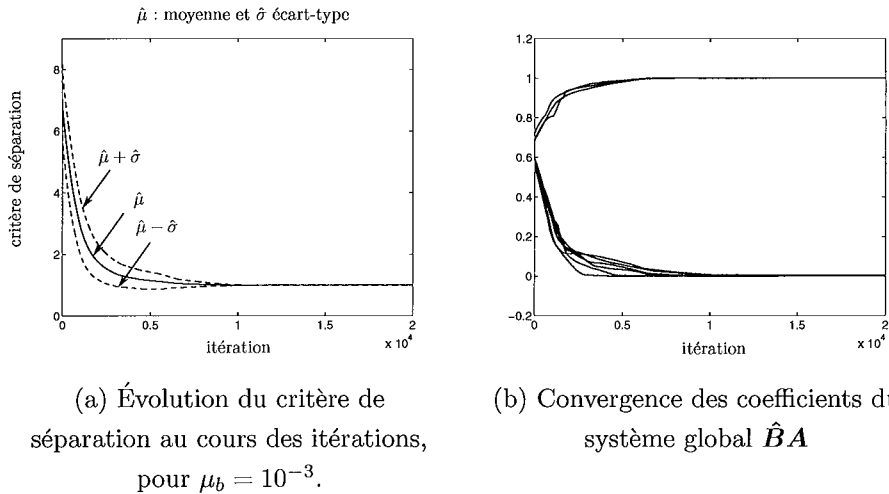


FIG. 3.6 – Séparation par maximum de vraisemblance et gradient relatif.

Afin d'illustrer la pertinence de ce résultat, nous avons appliqué également la méthode du maximum de vraisemblance avec une fonction non-linéaire tangente hyperbolique et sans blanchiment des données. Le système global moyen résultant d'une analyse de Monte Carlo avec 100 simulation,

$$\hat{\mathbf{B}}\mathbf{A} = \begin{bmatrix} 1.0184 & -0.1611 & 0.1626 \\ 0.1079 & 0.9916 & 0.0558 \\ 0.0070 & 0.1698 & 1.0238 \end{bmatrix}, \quad (3.70)$$

montre que, comparée à la méthode fondée sur le modèle de distribution gamma modifiée, la qualité de séparation dans ce cas est médiocre.

### 3.4.2 Séparation en utilisant l'algorithme MCEM

Un bruit additif gaussien i.i.d est rajouté de telle sorte à obtenir un RSB de 20 dB sur chaque observation. La séparation dans ce cas est réalisée en utilisant l'algorithme SAEM avec  $M = 10$  et  $\gamma = 0.95$ . Une simulation de Monte Carlo avec 100 initialisations différentes est réalisée. La valeur initiale de la matrice de mélange est obtenue arbitrairement à partir d'une loi uniforme  $\mathcal{U}(0, 1)$ . Les résultats de cette analyse sont montrés dans la figure 3.7. La figure 3.7(a) montre l'évolution moyenne de la log-vraisemblance durant l'optimisation par l'algorithme SAEM et la figure 3.7(b) montre l'erreur de reconstruction des sources en fonction du nombre d'itérations. Nous pouvons noter la convergence de l'algorithme après une quarantaine d'itérations vers une solution où les sources sont reconstruites avec une faible erreur quadratique moyenne. D'autre part, la visualisation des coefficients de mélange estimés pour une initialisation particulière montre la présence d'un faible écart par rapport au valeurs vraies. Cet écart s'explique, d'une part, par la présence du bruit et l'erreur d'approximation des statistiques des données manquantes et, d'autre part, par le fait que l'estimateur du maximum de vraisemblance perd ses propriétés statistiques pour un nombre fini d'échantillons.

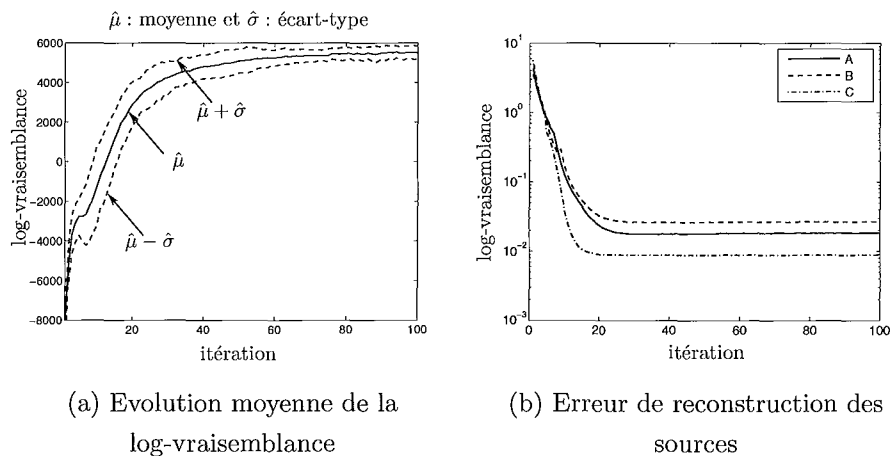


FIG. 3.7 – Séparation par maximum de vraisemblance et algorithme SAEM.

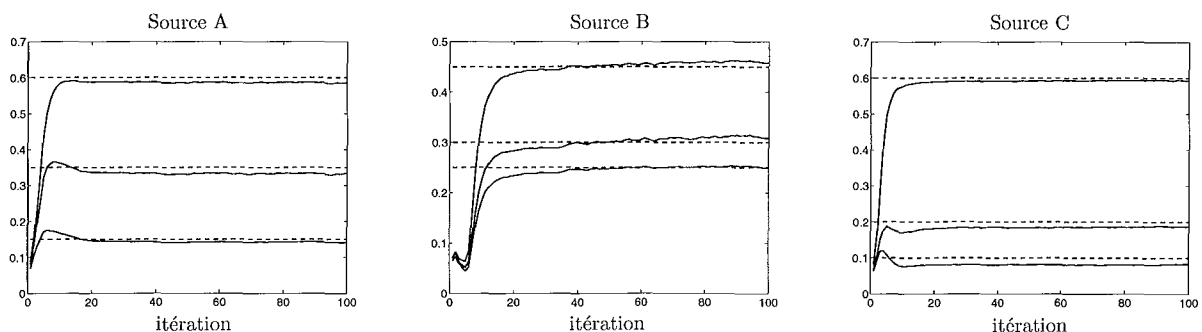


FIG. 3.8 – Évolution des estimations des coefficients de mélange au cours des itérations de l'algorithme MCEM.

### 3.5 Conclusion

Dans ce chapitre nous avons proposé deux algorithmes de séparation de sources non-négatives en utilisant l'approche du maximum de vraisemblance. Cette approche peut être appliquée à la séparation, à condition d'utiliser un modèle de distribution des sources qui permet de prendre en compte explicitement la non-négativité. La première contribution de ce chapitre est la synthèse d'une fonction non-linéaire qui permet d'appliquer les techniques d'analyse en composantes indépendantes à la séparation de sources non-négatives et non-orthogonales. A un niveau de bruit non-négligeable, l'utilisation de cette première méthode fondée sur l'optimisation par gradient relatif n'est plus adéquate. Une deuxième contribution consiste en l'utilisation d'un algorithme MCEM pour la maximisation de la vraisemblance en présence de bruit. Néanmoins, l'approche du maximum de vraisemblance ne garantit pas la positivité des coefficients de mélange et dès que la distribution des vraies sources s'éloigne d'une densité gamma ou encore dans le cas d'un niveau de bruit important des coefficients de mélange à valeurs négatives peuvent être estimés. Afin de résoudre le problème de séparation sous les contraintes de non-négativité des signaux sources et des coefficients de mélange, nous proposons dans le chapitre 4 une méthode fondée sur l'approche bayésienne.



## Chapitre 4

# Séparation de sources spectrales par approche bayésienne

Ce chapitre traite le problème de séparation de sources spectrales en utilisant une approche bayésienne. La méthodologie de séparation de sources par inférence bayésienne est rappelée succinctement dans un premier temps, et par la suite une inférence fondée sur la modélisation de la distribution des signaux sources et des coefficients de mélange par des densités gamma est présentée. La méthode proposée est illustrée à l'aide de signaux synthétiques.

### 4.1 Introduction

Le modèle de mélange linéaire instantané est exprimé par

$$\mathbf{x}_k = \mathbf{A} \mathbf{s}_k + \mathbf{e}_k, \text{ pour } k = 1, \dots, n, \quad (4.1)$$

où  $\mathbf{s}_k$  représente le vecteur  $(n \times 1)$  des signaux sources  $\{s_{(j,k)}\}_{j=1}^p$ ,  $\mathbf{x}_k$  le vecteur  $(m \times 1)$  des signaux mesurés  $\{x_{(i,k)}\}_{i=1}^m$ ,  $\mathbf{e}_k$  le vecteur  $(m \times 1)$  représentant un bruit additif et  $\mathbf{A}$  est la matrice de mélange de dimension  $(m \times p)$  et contenant les coefficients de mélange  $\{a_{ij}; i = 1, \dots, m\}_{j=1}^p$ , et  $k$  est un indice de la variable d'observation (temps, fréquence, nombre d'onde, etc.). Ayant toutes les observations et en utilisant une notation matricielle, le modèle de mélange s'exprime par

$$\mathbf{X} = \mathbf{A} \mathbf{S} + \mathbf{E}, \quad (4.2)$$

où les matrices  $\mathbf{X} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{S} \in \mathbb{R}^{p \times n}$  et  $\mathbf{E} \in \mathbb{R}^{m \times n}$  contiennent, respectivement, les observations, les signaux sources et le bruit. En supposant que le nombre de sources connu, le problème de séparation peut être formulé comme suit : à partir des observations, estimer conjointement les signaux sources et les coefficients de mélange.

Dans le chapitre 3, deux méthodes d'analyses issues de l'approche de séparation par maximum de vraisemblance. La particularité de cette approche est de pouvoir choisir la densité de probabilité des sources. Ainsi le choix d'une distribution à support positif permet de prendre en compte explicitement la non-négativité des signaux sources. Cependant, la non-négativité des coefficients de mélange n'est

pas assurée et n'est vérifiée qu'*a posteriori*. Afin de prendre en compte cette information conjointement sur les signaux sources et sur les coefficients de mélange le recours à une approche bayésienne pour la séparation est plus adéquat.

## 4.2 Approche bayésienne pour la séparation de sources

La formulation du problème de séparation de sources dans un cadre d'estimation bayésienne est assez récent. Les premiers travaux remontent à Rowe [Row98], Roberts [Rob98] et Knuth [Knu98, Knu99] qui ont présenté des algorithmes dans lesquels, en plus d'un modèle *a priori* sur les signaux sources, des informations sur les coefficients de mélange sont introduites. L'article de Mohammad-Djafari [MD99] présente les liens avec les méthodes d'analyse en composantes indépendantes et le cadre général de la méthodologie de séparation de sources par approche bayésienne. En fonction des informations considérées et des modèles probabilistes utilisés, plusieurs orientations ont été suivies. Par exemple, le cas de communications numériques où les signaux sources possèdent des distributions discrètes et connues *a priori* a été abordé dans [SA00]. Dans le cas de sources à distribution continue, les contributions se focalisent sur la modélisation de la distribution des sources par des mélanges de gaussiennes [SMD01b, SD02] de façon à considérer la multi-modalité et la non-stationnarité des signaux sources. Plus récemment, les travaux de [FGW04, SI05] montrent que la parcimonie est une information supplémentaire qui peut être introduite pour la séparation, même dans le cas sous-déterminé. Néanmoins, seulement quelques auteurs se sont focalisés sur le cas particulier de la séparation de sources non-négatives mélangées avec des coefficients non-négatifs [OSAMB99, MM01, RC03, MMDBC04, MBCMD04, MBMDC05, MCBMD05].

L'avantage principal de l'approche estimation bayésienne est qu'elle offre un cadre théorique favorable pour traiter un tel problème car la non-négativité des coefficients de mélange peut être prise en compte explicitement d'une part, et d'autre part, la formulation probabiliste du problème permet d'utiliser des méthodes Monte Carlo par Chaînes de Markov [GRS99, Rob99, CSI00] qui offrent une alternative à la résolution analytique des problèmes d'optimisation, notamment dans le cas de densités de probabilité *a posteriori* complexes ou multidimensionnelles. Des rapports très complets sur l'approche bayésienne pour la séparation de sources sont fournis dans les rapports de thèses de Sénécal [Sén02, part. 2] et Snoussi [Sno03, chap. 2].

### 4.2.1 Principe de la séparation bayésienne

L'idée principale de l'approche bayésienne en séparation de sources est de prendre en compte toute information préalablement disponible sur les signaux sources et les coefficients de mélange. Le problème est formulé dans un cadre probabiliste en affectant des densités de probabilité aux différentes variables. Le modèle direct est décrit par la vraisemblance qui résulte de la modélisation du processus d'observation et le modèle inverse est caractérisé par la traduction des informations disponibles sur les signaux sources et les coefficients de mélange par des densités de probabilité *a priori*. Le point fondamental dans l'estimation bayésienne est le théorème de Bayes qui permet de fusionner toutes ces informations afin de déduire une description probabiliste *a posteriori* des variables recherchées. En séparation de

sources, le théorème de Bayes permet d'écrire

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}) = p(\mathbf{X}|\mathbf{S}, \mathbf{A}) \times p(\mathbf{S}, \mathbf{A}) \div p(\mathbf{X}), \quad (4.3)$$

où  $p(\mathbf{X}|\mathbf{S}, \mathbf{A})$  est la vraisemblance,  $p(\mathbf{S}, \mathbf{A})$  est la densité de probabilité conjointe *a priori* de  $\mathbf{S}$  et  $\mathbf{A}$  et  $p(\mathbf{X})$  est la "densité de probabilité" des observations, qui peut être vue comme un terme de normalisation. Étant donné que les variables recherchées sont  $\mathbf{A}$  et  $\mathbf{S}$ , et afin de simplifier l'écriture en ne gardant que les termes dans lesquels apparaissent les variables d'intérêt, l'équation (4.3) peut être exprimée par une relation de proportionnalité

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}) \times p(\mathbf{S}, \mathbf{A}). \quad (4.4)$$

D'une façon générale, il est totalement justifié de supposer que la matrice de mélange est indépendante des signaux sources et par conséquent on peut écrire

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}) \times p(\mathbf{S}) \times p(\mathbf{A}). \quad (4.5)$$

En notant  $\boldsymbol{\theta}$  l'ensemble des hyperparamètres du problème,  $\boldsymbol{\theta} = \{\boldsymbol{\theta}_e, \boldsymbol{\theta}_s, \boldsymbol{\theta}_a\}$ , qui sont les paramètres inconnus des densités de probabilité affectées aux différentes variables du modèle. Ainsi, la relation (4.5) devient

$$p(\mathbf{S}, \mathbf{A}|\mathbf{X}, \boldsymbol{\theta}) \propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \boldsymbol{\theta}_e) \times p(\mathbf{S}|\boldsymbol{\theta}_s) \times p(\mathbf{A}|\boldsymbol{\theta}_a). \quad (4.6)$$

Si l'on se place dans la situation où l'on ne connaît pas les valeurs de ces hyperparamètres, il est nécessaire de les estimer également. Il faut alors les prendre en compte dans (4.6) en écrivant

$$p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{X}) \propto p(\mathbf{X}|\mathbf{S}, \mathbf{A}, \boldsymbol{\theta}_e) \times p(\mathbf{S}|\boldsymbol{\theta}_s) \times p(\mathbf{A}|\boldsymbol{\theta}_a) \times p(\boldsymbol{\theta}) \quad (4.7)$$

C'est principalement à partir de cette densité de probabilité *a posteriori* que les différents estimateurs sont construits. Les estimateurs qui peuvent être utilisés sont le maximum conjoint *a posteriori* (JMAP pour *joint maximum a posteriori*), le maximum marginal *a posteriori* (MMAP pour *marginal maximum a posteriori*) et la moyenne marginale *a posteriori* (MPM pour *marginal posterior mean*). En pratique, le choix d'un estimateur particulier est très lié à la technique de calcul utilisée pour la résolution du problème d'optimisation.

#### 4.2.2 Techniques d'estimation bayésienne pour la séparation de sources

La résolution du problème de séparation de sources par une approche bayésienne nécessite deux étapes importantes dans lesquelles certains choix doivent être faits. La première concerne la description probabiliste des différentes variables. En général les erreurs d'observation sont décrites par un modèle gaussien, bien que dans d'autres cas des modèles poissonniens seront plus adéquats. Par ailleurs, le point crucial est celui du choix du modèle *a priori* sur la distribution des sources et les coefficients de mélange. Pour une information disponible, plusieurs descriptions probabilistes peuvent être employées par différents statisticiens, d'où le caractère subjectif de cette étape [Rob01]. Néanmoins, la complexité des calculs bayésiens et les prédispositions du statisticien sont des facteurs importants qui favorisent certains modèles par rapport à d'autres. Le deuxième point concerne la

définition du risque bayésien dont résulte le critère à optimiser pour l'estimation. En général, cette étape de définition de l'estimateur est très conditionnée par la complexité des calculs nécessaires pour l'optimisation numérique. On en distingue les algorithmes EM et les méthodes MCMC. Les algorithmes EM sont généralement utilisés pour la recherche du maximum *a posteriori* alors que les algorithmes MCMC permettent de simuler des chaînes de Markov ayant comme distribution stationnaire la densité conjointe *a posteriori*. A partir de ces échantillons, des approximations de ces statistiques peuvent être calculées en utilisant la méthode de Monte Carlo. En général l'estimateur de la moyenne marginale *a posteriori* est choisi pour cause de simplicité de calcul. Un couplage entre les algorithmes MCMC et les techniques du recuit simulé fournit une méthode d'obtention d'une estimation du maximum marginal *a posteriori* (MMAP) [DGR02].

Plus concrètement, dans le cadre de séparation de sources, l'étape cruciale est celle de la recherche de l'information préalable sur les signaux sources et les coefficients de mélange susceptible d'assurer une bonne séparation. Idéalement, cette information doit non seulement assurer l'unicité de la solution mais aussi doit permettre d'obtenir une solution correcte. Suit alors une étape de codage de cette information via un choix de la densité de probabilité *a priori* la plus adéquate. Dans le chapitre précédent, nous avons discuté les différents modèles qui ont été employés en littérature pour représenter les signaux sources afin de prendre en compte des informations telles que la multi-modalité, la non-stationnarité et la parcimonie. Nous avons montré que pour séparer des sources non-négatives il faut utiliser des modèles *a priori* qui consistent en des distributions à support non-négatif. Cette orientation a été également retenue pour la séparation par une approche bayésienne.

### 4.3 Séparation de sources non-négatives par approche bayésienne

En adoptant une approche statistique pour la séparation, et afin de prendre en compte explicitement cette information de non-négativité, il est nécessaire de représenter la distribution des signaux sources et des coefficients de mélange par des densités à support non-négatif. Cependant, il existe une multitude de densités à support non-négatif et le choix d'une distribution particulière est sujet à discussion. Considérer un modèle en adéquation avec d'autres informations ou hypothèses supplémentaires rendrait ce choix mieux justifié. Nous présentons le modèle probabiliste pour la séparation de sources non-négatives et les calculs bayésiens nécessaires pour la résolution du problème d'estimation via une simulation Monte Carlo par chaînes de Markov.

#### 4.3.1 Modélisation probabiliste

##### 4.3.1.1 Observations

Le bruit additif est supposé i.i.d, gaussien, de moyenne nulle et de matrice de covariance  $\mathbf{R}_e = \text{diag}\{\sigma_1^2, \dots, \sigma_m^2\}$ . En utilisant le modèle de mélange (4.1), la vraisemblance sera exprimée par

$$p(\mathbf{X}|\mathbf{A}, \mathbf{S}, \boldsymbol{\theta}_e) = \prod_{i=1}^m \prod_{k=1}^n \mathcal{N}\left(\sum_{j=1}^p a_{ij} s_{(j,k)}, \sigma_i^2\right). \quad (4.8)$$

où  $\boldsymbol{\theta}_e = [\sigma_1^2, \dots, \sigma_m^2]^T$ .

#### 4.3.1.2 Signaux sources

Les signaux sources sont supposés mutuellement indépendants et les échantillons de chaque signal sont supposés indépendants et identiquement distribués. En d'autres termes,

$$p(\mathbf{S}) = \prod_{j=1}^p \prod_{k=1}^n p(s_{(j,k)}). \quad (4.9)$$

Cette hypothèse d'indépendance temporelle des signaux permet de spécifier simplement qu'on ne tient pas compte d'une éventuelle structure "temporelle" des signaux sources. Afin de coder la non-négativité, chaque échantillon d'un signal source  $\{s_{(j,k)}, k = 1, \dots, n\}$  doit être représenté par une densité à support non-négatif. Nous utilisons une densité gamma de paramètres  $\alpha_j, \beta_j$  comme loi *a priori* pour chaque source, mais ces paramètres peuvent être différents d'une source à une autre. La densité de probabilité *a priori* des signaux sources est exprimée par

$$p(\mathbf{S}|\boldsymbol{\theta}_s) = \prod_{j=1}^p \prod_{k=1}^n \mathcal{G}(\alpha_j, \beta_j), \quad (4.10)$$

où  $\mathcal{G}(\alpha, \beta)$  représente la densité Gamma de paramètres  $\alpha$  et  $\beta$ . Le vecteur  $\boldsymbol{\theta}_s = [\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_p]^T$  contient les paramètres de toutes les distributions *a priori*.

#### 4.3.1.3 Coefficients de mélange

Afin de prendre en compte la non-négativité des coefficients de mélange, chaque  $j$ -ème colonne de la matrice de mélange est considérée distribuée selon une densité gamma de paramètres  $(\gamma_j, \delta_j)$ . Ces paramètres sont supposés constants pour chaque profil d'évolution mais peuvent être différents pour des sources distinctes. Sous ces hypothèses, la densité de probabilité *a priori* des coefficients de mélange est exprimée par

$$p(\mathbf{A}|\boldsymbol{\theta}_a) = \prod_{i=1}^m \prod_{j=1}^p \mathcal{G}(\gamma_j, \delta_j), \quad (4.11)$$

où  $\boldsymbol{\theta}_a = [\gamma_1, \dots, \gamma_p, \delta_1, \dots, \delta_p]^T$ .

**Discussion :** ce modèle *a priori* sur les sources peut être simplifié en considérant le cas particulier d'une densité exponentielle en fixant la valeur de  $\alpha$  à l'unité. Une deuxième possibilité dans le cas de réactions chimiques évolutives comme les cinétiques chimiques une contrainte de douceur peut être rajoutée sur les profils d'évolutions des coefficients de mélange.

#### 4.3.1.4 Densité conjointe *a posteriori*

En utilisant le théorème de Bayes et notant  $\boldsymbol{\theta}$  le vecteur qui contient les variances du bruit et les paramètres des densités *a priori* des signaux sources et des coefficients de mélange  $\boldsymbol{\theta} = [\boldsymbol{\theta}_e; \boldsymbol{\theta}_s; \boldsymbol{\theta}_a]$ , la

densité de probabilité conjointe *a posteriori* des sources et des coefficients de mélange s'exprime par

$$p(\mathbf{S}, \mathbf{A} | \mathbf{X}, \boldsymbol{\theta}) \propto \prod_{i=1}^m \prod_{k=1}^n \mathcal{N} \left( \sum_{j=1}^p a_{ij} s_{(j,k)}, \sigma_i^2 \right) \times \prod_{k=1}^n \prod_{j=1}^p \mathcal{G}(\alpha_j, \beta_j) \times \prod_{i=1}^m \prod_{j=1}^p \mathcal{G}(\gamma_j, \delta_j). \quad (4.12)$$

Puisque les paramètres des densités *a priori* des sources et les variances des séquences de bruit ne sont pas connues au préalable, il faut alors les inférer en même temps que les signaux sources et les coefficients de mélange. Pour cela, la densité conjointe *a posteriori* est ré-écrite comme suit

$$p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta} | \mathbf{X}) \propto p(\mathbf{S}, \mathbf{A} | \mathbf{X}, \boldsymbol{\theta}) \times p(\boldsymbol{\theta}), \quad (4.13)$$

où des densités *a priori* sont spécifiées pour tous les hyperparamètres  $\boldsymbol{\theta} = [\boldsymbol{\theta}_e; \boldsymbol{\theta}_s; \boldsymbol{\theta}_a]$ .

### 4.3.2 Interprétation du critère et liens avec les méthodes existantes

#### 4.3.2.1 Estimation par JMAP

A partir de la densité *a posteriori* (4.13), plusieurs estimateurs bayésiens peuvent être utilisés pour l'estimation des signaux sources, coefficients de mélange et les hyperparamètres. La maximisation conjointe de cette densité conduit à l'estimateur du maximum conjoint *a posteriori* JMAP. Cet estimateur correspond également à la minimisation conjointe par rapport  $\mathbf{S}$  et  $\mathbf{A}$  du critère défini par

$$\Phi(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta} | \mathbf{X}) = -\log p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta} | \mathbf{X}). \quad (4.14)$$

Ce critère est composé de trois parties,

$$\Phi(\mathbf{S}, \mathbf{A} | \mathbf{X}, \boldsymbol{\theta}) = \Phi_L(\mathbf{S}, \mathbf{A} | \mathbf{X}, \boldsymbol{\theta}_e) + \Phi_{P_1}(\mathbf{S} | \boldsymbol{\theta}_s) + \Phi_{P_2}(\mathbf{A} | \boldsymbol{\theta}_a) \quad (4.15)$$

où  $\Phi_L$ ,  $\Phi_{P_1}$  et  $\Phi_{P_2}$  sont donnés par

$$\Phi_L = \sum_{i=1}^m \sum_{k=1}^n \frac{1}{2\sigma_i^2} \left( x_{(i,k)} - \sum_{j=1}^p a_{ij} s_{(j,k)} \right)^2, \quad (4.16)$$

$$\Phi_{P_1} = \sum_{j=1}^p (1 - \alpha_j) \sum_{k=1}^n \log s_{(j,k)} + \sum_{j=1}^p \sum_{k=1}^n \beta_j s_{(j,k)}, \quad (4.17)$$

$$\Phi_{P_2} = \sum_{j=1}^p (1 - \gamma_j) \sum_{i=1}^m \log a_{ij} + \sum_{j=1}^p \sum_{i=1}^m \delta_j a_{ij}. \quad (4.18)$$

Le premier terme  $\Phi_L$  de ce critère est l'erreur quadratique moyenne alors que les deux derniers termes sont des critères de régularisation qui permettent d'éviter des valeurs négatives de  $\mathbf{S}$  et  $\mathbf{A}$  respectivement.

Concernant la résolution du problème d'optimisation, nous avons préalablement présenté une approche d'estimation JMAP [MMDBC04], où l'estimation est réalisée par une procédure de gradient alternée. A chaque itération, les signaux sources sont mis à jour en utilisant la dernière estimation des

signaux sources, ensuite ces nouvelles estimations sont utilisées pour mettre à jour les coefficients de mélange

$$\begin{cases} \hat{\mathbf{S}}^{(r+1)} = \hat{\mathbf{S}}^{(r)} - \mu_s^{(r+1)} \nabla_{\mathbf{S}} \Phi \left( \mathbf{S}^{(r)}, \hat{\mathbf{A}}^{(r)} \right) \odot \hat{\mathbf{S}}^{(r)}, \\ \hat{\mathbf{A}}^{(r+1)} = \hat{\mathbf{A}}^{(r)} - \mu_a^{(r+1)} \nabla_{\mathbf{A}} \Phi \left( \hat{\mathbf{S}}^{(r+1)}, \mathbf{A}^{(r)} \right) \odot \hat{\mathbf{A}}^{(r)}, \end{cases}$$

où  $\odot$  représente la multiplication terme à terme,  $\mu_s^{(r+1)}$  et  $\mu_a^{(r+1)}$  sont les paramètres d'adaptation qui contrôlent la convergence de l'algorithme.  $\nabla_{\mathbf{S}} \Phi$  et  $\nabla_{\mathbf{A}} \Phi$  sont les gradients du critère par rapport à  $\mathbf{S}$  et  $\mathbf{A}$ , respectivement. Notons que la multiplication du gradient par  $\hat{\mathbf{S}}^{(r)}$  et  $\hat{\mathbf{A}}^{(r)}$  permet de satisfaire les conditions de Kuhn-Tucker pour la minimisation d'une fonction sous la contrainte de non-négativité [Kau93, BTT91]. Les hyperparamètres sont estimés par maximisation de leurs densité conditionnelle *a posteriori* à chaque itération, quant au pas d'adaptation une procédure de recherche par la méthode du nombre d'or [PTVF92] permet de fixer le meilleur pas d'adaptation. Le point critique avec cette procédure d'optimisation est assez classique car il est largement fréquent de constater la convergence de l'algorithme du gradient vers un minimum local et par conséquent une bonne initialisation est nécessaire pour obtenir des solutions adéquates. Un choix possible est de considérer les mesures les plus décorréelées comme initialisation pour les signaux sources. Afin de réduire cette dépendance vis-à-vis de l'initialisation, qui constitue la principale limitation de cette approche, nous pensons que les algorithmes stochastiques sont bien adaptés.

#### 4.3.2.2 Liens avec les méthodes existantes

Cette approche peut être liée avec quelques méthodes existantes : (i) si l'on considère un *a priori* uniforme positif et l'estimateur du JMAP, on obtient le critère des moindres carrés contraints minimisé par les approches NMF et ALS ; (ii) le cas  $\{\alpha_j = 1\}_{j=1}^p, \{\beta_j = \beta\}_{j=1}^p, \{\gamma_j = 1\}_{j=1}^p, \{\delta_j = 0\}_{j=1}^p$  correspond à considérer un modèle exponentiel sur la distribution des sources et mène à un critère de régularisation similaire à celui minimisé par la méthode NNSC ; (iii) dans le cas de la méthode PMF [PT94, Paa97], le critère minimisé est donné par

$$J_2(\mathbf{S}, \mathbf{A} | \mathbf{X}, \boldsymbol{\theta}) = \sum_{i=1}^m \sum_{k=1}^n \frac{1}{2\sigma_{(i,k)}^2} \left( x_{(i,k)} - \sum_{j=1}^p a_{ij} s_{(j,k)} \right)^2 - \alpha \sum_{j=1}^p \sum_{k=1}^n \log s_{(j,k)} + \beta \sum_{j=1}^p \sum_{k=1}^n s_{(j,k)}^2 - \gamma \sum_{j=1}^p \sum_{i=1}^m \log a_{ij} + \delta \sum_{j=1}^p \sum_{i=1}^m a_{ij}^2. \quad (4.19)$$

où  $(\alpha, \beta, \gamma, \delta)$  sont des paramètres de régularisation. Cette méthode de moindres carrés pénalisés utilise une fonction de pénalité de la forme

$$f(z; \alpha, \beta) = -\alpha \log z + \beta z^2, \text{ with } z \geq 0, \quad (4.20)$$

qui correspond à utiliser une densité

$$p(z; \alpha, \beta) = K(\alpha, \beta) \exp[-f(z; \alpha, \beta)] = K(\alpha, \beta) z^\alpha \exp[-\beta z^2] \quad \text{avec } z \geq 0, \quad (4.21)$$

où  $K(\alpha, \beta) = 2\beta^{(\alpha+1)/2} / \Gamma((\alpha+1)/2)$ . Par transformation de variable aléatoire, on peut montrer que cette loi correspond à supposer une distribution gamma sur  $z^2$  avec des paramètres  $([\alpha+1]/2, \beta)$ .

La figure 4.1 illustre trois formes typiques de cette densité et montre sa ressemblance avec la densité Gamma montrée dans la figure 3.1. D'autre part dans les méthodes NNSC et PMF, les valeurs des hyperparamètres doivent être fixées *a priori* et sont supposés similaires pour tous les signaux sources et pour tous les coefficients de mélange, ce qui correspond à utiliser la même densité *a priori* sur toutes les sources et sur tous les profils de coefficients de mélange. Par contre, l'approche proposée suppose que ces paramètres peuvent être différents et sont estimés conjointement avec les sources et les coefficients de mélange, ce qui rend le modèle *a priori* plus flexible. Un autre point est celui de la non convexité des deux critères  $J_1$  et  $J_2$ , ce qui rend l'utilisation des méthodes d'optimisation déterministes plus délicat. Afin d'éviter ces difficultés, nous utilisons par la suite une méthode d'optimisation stochastique fondée sur les méthodes MCMC.

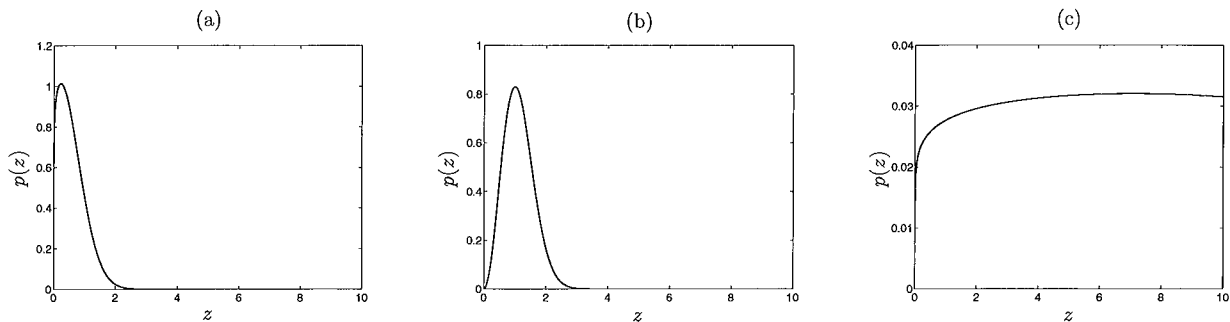


FIG. 4.1 – Illustration de la densité *a priori* associée à la méthode PMF pour (a)  $\alpha = 0, \beta = 1$ , (b)  $\alpha = 0.1, \beta = 1$  et (c)  $\alpha = 2, \beta = 1$

### 4.3.3 Estimation par simulation MCMC

L'objectif des méthodes de Monte Carlo par chaînes de Markov est de simuler des données distribuées selon la densité conjointe *a posteriori*, et ensuite de construire des estimateurs à partir de ces données. Pour ce qui est de l'estimation, nous utilisons l'estimateur de la moyenne marginale *a posteriori* (MPM),

$$\hat{\mathbf{S}} = \mathbb{E}[\mathbf{S}|\mathbf{X}], \quad \hat{\mathbf{A}} = \mathbb{E}[\mathbf{A}|\mathbf{X}] \text{ et } \hat{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta}|\mathbf{X}]. \quad (4.22)$$

L'objectif de l'algorithme de Gibbs est de simuler une chaîne de Markov stationnaire ergodique dont les échantillons suivent asymptotiquement la densité de probabilité *a posteriori*  $p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{X})$ . Les estimations des signaux sources et coefficients de mélange sont alors calculées à partir des échantillons simulés. La procédure de simulation MCMC dans le cas général est rappelée dans un premier temps, et par la suite nous ne détaillerons pas les étapes de simulation pour le cas de la séparation de sources non-négatives avec des *a priori* gamma sur les signaux sources et coefficients de mélange.

#### 4.3.3.1 Méthodologie

La figure 4.2 donne l'organigramme de la méthodologie de séparation de sources par approche bayésienne et simulation MCMC : afin de simuler  $p(\mathbf{S}, \mathbf{A}, \boldsymbol{\theta}|\mathbf{X})$ , à chaque itération  $r$ , les étapes principales de l'algorithme de Gibbs consistent à



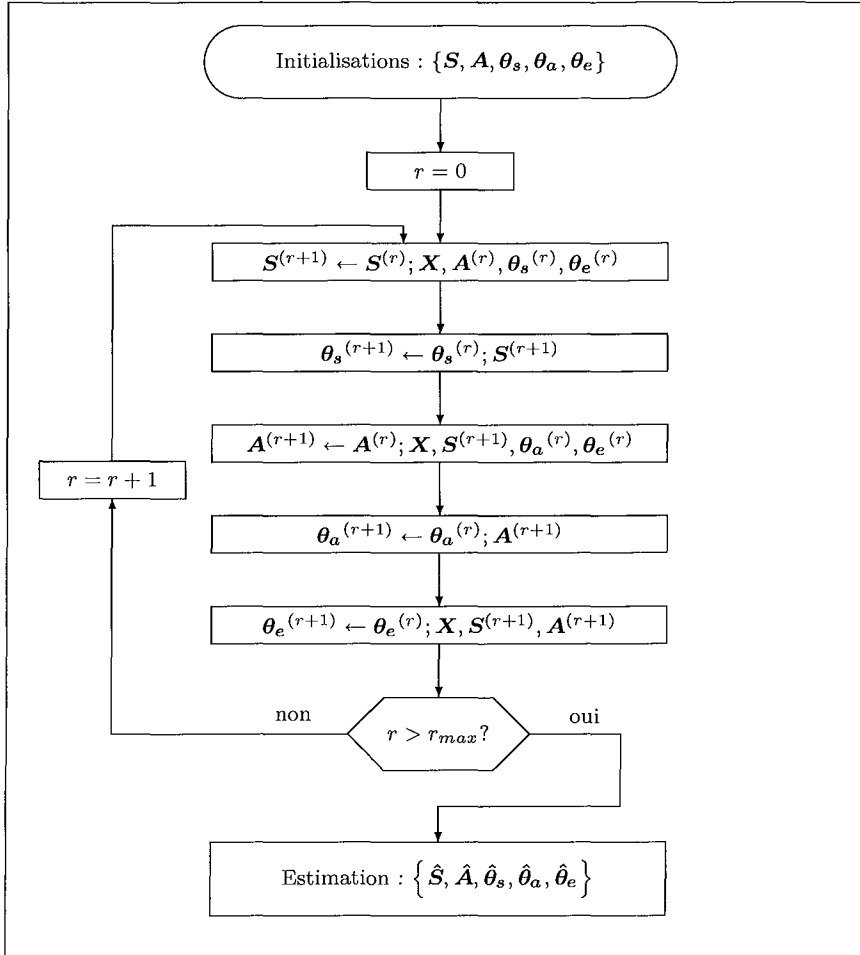


FIG. 4.2 – Organigramme simplifié de la méthode de séparation par simulation MCMC

1. simuler les signaux sources  $\mathbf{S}^{(r+1)}$  à partir de

$$p\left(\mathbf{S}|\mathbf{X}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \propto p\left(\mathbf{X}|\mathbf{S}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \times p\left(\mathbf{S}|\boldsymbol{\theta}^{(r)}\right); \quad (4.23)$$

2. simuler les coefficients de mélange  $\mathbf{A}^{(r+1)}$  à partir de

$$p\left(\mathbf{A}|\mathbf{X}, \mathbf{S}^{(r+1)}, \boldsymbol{\theta}^{(r)}\right) \propto p\left(\mathbf{X}|\mathbf{S}^{(r+1)}, \mathbf{A}, \boldsymbol{\theta}^{(r)}\right) \times p\left(\mathbf{A}|\boldsymbol{\theta}^{(r)}\right); \quad (4.24)$$

3. simuler les hyperparamètres  $\boldsymbol{\theta}^{(r+1)}$  à partir de

$$p\left(\boldsymbol{\theta}|\mathbf{X}, \mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}\right) \propto p\left(\mathbf{X}|\mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}, \boldsymbol{\theta}\right) \times p\left(\mathbf{S}^{(r+1)}|\boldsymbol{\theta}\right) \times p\left(\mathbf{A}^{(r+1)}|\boldsymbol{\theta}\right) \times p(\boldsymbol{\theta}). \quad (4.25)$$

Il existe principalement trois types d'hyperparamètres,  $\boldsymbol{\theta}_e, \boldsymbol{\theta}_s$  et  $\boldsymbol{\theta}_a$ , qui sont supposés indépendants *a priori*. La troisième étape de l'algorithme peut donc être divisée en trois parties :

- 3.1 simuler les hyperparamètres du bruit  $\boldsymbol{\theta}_e^{(r+1)}$  à partir de

$$p\left(\boldsymbol{\theta}_e|\mathbf{X}, \mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}\right) \propto p\left(\mathbf{X}|\mathbf{S}^{(r+1)}, \mathbf{A}^{(r+1)}, \boldsymbol{\theta}_e\right) \times p(\boldsymbol{\theta}_e); \quad (4.26)$$

- 3.2 simuler les hyperparamètres des signaux sources  $\boldsymbol{\theta}_s^{(r+1)}$  à partir de

$$p\left(\boldsymbol{\theta}_s|\mathbf{S}^{(r+1)}\right) \propto p\left(\mathbf{S}^{(r+1)}|\boldsymbol{\theta}_s\right) \times p(\boldsymbol{\theta}_s); \quad (4.27)$$

3.3 simuler les hyperparamètres des coefficients de mélange  $\theta_a^{(r+1)}$  à partir de

$$p\left(\theta_a | \mathbf{A}^{(r+1)}\right) \propto p\left(\mathbf{A}^{(r+1)} | \theta_a\right) \times p\left(\theta_a\right). \quad (4.28)$$

#### 4.3.3.2 Schéma détaillé de la simulation

Nous décrivons maintenant d'une façon détaillée l'implantation de l'algorithme de Gibbs pour la séparation de sources non-négatives. Nous utilisons la version scalaire de l'algorithme et afin de simplifier les équations, nous introduisons les notations suivantes  $y_{(1:n)}, z_{(1:m,1:p)}$  qui représentent respectivement  $\{y_i\}_{i=1}^n$  et  $\{z_{(i,j)}\}_{(i=1,j=1)}^{(m,p)}$ . Nous introduisons également, les notations  $\theta_z^{likel}$  et  $\theta_z^{prior}$  qui représentent les paramètres des lois  $p(x|z, \theta_z^{likel})$  et  $p(z|\theta_z^{prior})$ .

Après une initialisation aléatoire de toutes les variables, chaque itération  $r$  de l'algorithme consiste à

1. pour  $j = 1, \dots, p$  et  $t = 1, \dots, n$ , simuler  $s_{(j,k)}^{(r+1)}$  à partir de la densité *a posteriori* conditionnelle

$$p\left(s_{(j,k)} | x_{(1:m,k)}, \theta_{s_{(j,k)}}^{likel}, \theta_{s_{(j,k)}}^{prior}\right), \quad (4.29)$$

avec

$$\begin{cases} \theta_{s_{(j,t)}}^{likel} = \left\{ s_{(1:j-1,k)}^{(r+1)}, s_{(j+1:p,t)}^{(r)}, a_{(1:m,1:p)}^{(r)}, \sigma_{(1:m)}^{(r)} \right\}, \\ \theta_{s_{(j,k)}}^{prior} = \left\{ \alpha_j^{(r)}, \beta_j^{(r)} \right\}. \end{cases} \quad (4.30)$$

2. pour  $i = 1, \dots, m$  et  $j = 1, \dots, p$ , simuler  $a_{ij}^{(r+1)}$  à partir de la densité *a posteriori* conditionnelle

$$p\left(a_{ij} | x_{(1:m,k)}, \theta_{a_{ij}}^{likel}, \theta_{a_{ij}}^{prior}\right), \quad (4.31)$$

avec

$$\begin{cases} \theta_{a_{ij}}^{likel} = \left\{ a_{(i,1:j-1)}^{(r+1)}, a_{(i,j+1:p)}^{(r)}, s_{(1:p,1:n)}^{(r+1)}, \sigma_i^{(r)} \right\}, \\ \theta_{a_{ij}}^{prior} = \left\{ \gamma_j^{(r)}, \delta_j^{(r)} \right\}. \end{cases} \quad (4.32)$$

3. pour  $i = 1, \dots, m$ , simuler  $(\sigma_i^2)^{(r+1)}$  à partir de la densité *a posteriori* conditionnelle

$$p\left(1/\sigma_i^2 | x_{(i,1:n)}, a_{(i,1:p)}^{(r+1)}, s_{(1:p,1:n)}^{(r+1)}\right); \quad (4.33)$$

4. pour  $j = 1, \dots, p$ , simuler  $\alpha_j^{(r+1)}$  à partir de la densité *a posteriori* conditionnelle

$$p\left(\alpha_j | s_{(j,1:n)}^{(r+1)}, \beta_j^{(r)}\right); \quad (4.34)$$

5. pour  $j = 1, \dots, p$ , simuler  $\beta_j^{(r+1)}$  à partir de la densité *a posteriori* conditionnelle

$$p\left(\beta_j | s_{(j,1:n)}^{(r+1)}, \alpha_j^{(r+1)}\right); \quad (4.35)$$

6. pour  $j = 1, \dots, p$ , simuler  $\gamma_j^{(r+1)}$  à partir de la densité *a posteriori* conditionnelle

$$p\left(\gamma_j | a_{(1:m,j)}^{(r+1)}, \delta_j^{(r)}\right); \quad (4.36)$$

7. pour  $j = 1, \dots, p$ , simuler  $\delta_j^{(r+1)}$  à partir de la densité *a posteriori* conditionnelle

$$p\left(\delta_j | a_{(1:m,j)}^{(r+1)}, \gamma_j^{(r+1)}\right). \quad (4.37)$$

Après  $r_{max}$  itérations de l'algorithme, les estimations des signaux sources et des coefficients de mélange sont obtenues par

$$\begin{cases} \hat{s}_{(j,k)} = \frac{1}{r_{max} - r_{min}} \sum_{r=r_{min}+1}^{r_{max}} s_{(j,k)}^{(r)}; \\ \hat{a}_{ij} = \frac{1}{r_{max} - r_{min}} \sum_{r=r_{min}+1}^{r_{max}} a_{ij}^{(r)}; \end{cases} \quad (4.38)$$

pour  $i = 1, \dots, m, j = 1, \dots, p$  et  $k = 1, \dots, n$ .

La valeur  $r_{min}$  est le nombre d'itérations qui correspondent au temps de chauffe de l'échantillonneur. Les échantillons simulés pendant ces itérations ne sont pas pris en compte pour l'estimation. A partir de ces échantillons d'autres statistiques *a posteriori* peuvent être calculées et les histogrammes peuvent être présentés.

### 4.3.4 Densités conditionnelles *a posteriori*

Toutes les densités conditionnelles requises par l'algorithme de Gibbs sont détaillées dans ce paragraphe.

#### 4.3.4.1 Signaux sources

La densité conditionnelle *a posteriori* de chaque échantillon des signaux sources  $s_{(j,k)}$  est obtenue selon

$$p\left(s_{(j,k)} \mid x_{(1:m,t)}, \theta_{s_{(j,k)}}^{likel}, \theta_{s_{(j,k)}}^{prior}\right) \propto p\left(x_{(1:m,k)} \mid s_{(j,k)}, \theta_{s_{(j,k)}}^{likel}\right) \times p\left(s_{(j,k)} \mid \theta_{s_{(j,k)}}^{prior}\right), \quad (4.39)$$

où  $\theta_{s_{(j,k)}}^{likel}$  et  $\theta_{s_{(j,k)}}^{prior}$  sont définis dans (4.30). La vraisemblance est déduite à partir de la vraisemblance jointe définie dans l'équation (4.8).

$$p\left(x_{(1:m,k)} \mid s_{(j,k)}, \theta_{s_{(j,k)}}^{likel}\right) \propto \exp\left\{-\frac{\left(s_{(j,k)} - \mu_{s_{(j,k)}}^{likel}\right)^2}{2 \left[\sigma_{s_j}^{likel}\right]^2}\right\}, \quad (4.40)$$

où

$$\begin{cases} \left[\sigma_{s_j}^{likel}\right]^2 = \left[\sum_{i=1}^m \frac{[a_{ij}^{(r)}]^2}{[\sigma_i^{(r)}]^2}\right]^{-1}, \\ \mu_{s_{(j,k)}}^{likel} = \frac{1}{\left[\sigma_{s_j}^{likel}\right]^2} \sum_{i=1}^m \frac{a_{ij}^{(r)} \varepsilon_{(i,k)}^{(-j)}}{[\sigma_i^{(r)}]^2}, \end{cases}$$

et  $\varepsilon_{(i,k)}^{(-j)} = x_{(i,k)} - \sum_{\ell=1}^{j-1} a_{i\ell}^{(r)} s_{(\ell,k)}^{(r+1)} - \sum_{\ell=j+1}^n a_{i\ell}^{(r)} s_{(\ell,t)}^{(r)}$ . En prenant en compte de la loi *a priori*

$$p\left(s_{(j,k)} \mid \theta_{s_{(j,k)}}^{prior}\right) \propto s_{(j,k)}^{\alpha_j^{(r)} - 1} \exp\left\{-\beta_j^{(r)} s_{(j,k)}\right\} \times \mathbb{I}_{[0, +\infty[}\left(s_{(j,k)}\right), \quad (4.41)$$

la densité *a posteriori* conditionnelle est donnée par

$$p\left(s_{(j,k)} \mid x_{(1:m,k)}, \theta_{s_{(j,k)}}^{likel}, \theta_{s_{(j,k)}}^{prior}\right) \propto s_{(j,k)}^{\alpha_j^{(r)} - 1} \exp\left\{-\frac{\left(s_{(j,k)} - \mu_{s_{(j,k)}}^{likel}\right)^2}{2 \left[\sigma_{s_j}^{likel}\right]^2} - \beta_j^{(r)} s_{(j,k)}\right\} \times \mathbb{I}_{[0, +\infty[}\left(s_{(j,k)}\right). \quad (4.42)$$

Cette densité n'est pas une loi usuelle, sa simulation nécessite donc l'utilisation d'une méthode de simulation par rejection [Dev86], tel que l'algorithme de Metropolis-Hastings. Nous décrivons dans l'annexe B l'algorithme proposé pour la simulation de cette densité.

#### 4.3.4.2 Coefficients de mélange

La densité *a posteriori* conditionnelle de chaque coefficient de mélange  $a_{ij}$  est exprimée par

$$p\left(a_{ij} | x_{(i,1:n)}, \theta_{a_{ij}}^{likel}, \theta_{a_{ij}}^{prior}\right) \propto p\left(x_{(i,1:n)} | a_{ij}, \theta_{a_{ij}}^{likel}\right) \times p\left(a_{ij} | \theta_{a_{ij}}^{prior}\right), \quad (4.43)$$

où  $\theta_{a_{ij}}^{likel}$  et  $\theta_{a_{ij}}^{prior}$  sont définies par (4.32).

La vraisemblance est déduite à partir de (4.8) et exprimée par

$$p\left(x_{(i,1:n)} | a_{ij}, \theta_{a_{ij}}^{likel}\right) \propto \exp\left\{-\frac{1}{2(\sigma_{a_{ij}}^{likel})^2} \left(a_{ij} - \mu_{a_{ij}}^{likel}\right)^2\right\}, \quad (4.44)$$

avec

$$\begin{cases} [\sigma_{a_{ij}}^{likel}]^2 = \frac{[\sigma_i^{(r)}]^2}{\sum_{t=1}^n s_{(j,t)}^{(r+1)}}; \\ \mu_{a_{ij}}^{likel} = \frac{1}{[\sigma_{a_{ij}}^{likel}]^2} \sum_{t=1}^n s_{(j,t)}^{(r+1)} \varepsilon_{(i,k)}^{(-j)}, \end{cases}$$

et  $\varepsilon_{(i,k)}^{(-j)} = x_{(i,k)} - \sum_{\ell=1}^{j-1} a_{i\ell}^{(r+1)} s_{(\ell,k)}^{(r+1)} - \sum_{\ell=j+1}^p a_{i\ell}^{(r)} s_{(\ell,k)}^{(r+1)}$ . En prenant compte la densité *a priori* du coefficient de mélange

$$p\left(a_{ij} | \theta_{a_{ij}}^{prior}\right) \propto a_{ij}^{\gamma_j^{(r)} - 1} \exp\left\{-\beta_j^{(r)} a_{ij}\right\} \times \mathbb{I}_{[0,+\infty[}(a_{ij}), \quad (4.45)$$

cette densité conditionnelle *a posteriori* sera alors donnée par

$$p\left(a_{ij} | x_{(1:m,t)}, \theta_{a_{ij}}^{likel}, \theta_{a_{ij}}^{prior}\right) \propto a_{ij}^{\gamma_j^{(r)} - 1} \exp\left\{-\frac{\left(a_{ij} - \mu_{a_{ij}}^{likel}\right)^2}{2(\sigma_{a_{ij}}^{likel})^2} - \delta_j^{(r)} a_{ij}\right\} \times \mathbb{I}_{[0,+\infty[}(a_{ij}). \quad (4.46)$$

Cette densité conditionnelle n'étant pas usuelle, sa simulation est réalisée en utilisant les mêmes développements que pour les signaux sources.

#### 4.3.4.3 Variances du bruit

La densité *a posteriori* conditionnelle de chaque variance des séquences de bruit est obtenue par

$$\begin{aligned} p\left(1/\sigma_i^2 | x_{(i,1:n)}, a_{(i,1:m)}^{(r+1)}, s_{(1:p,1:n)}^{(r+1)}\right) \\ \propto (1/\sigma_i^2)^{\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_i^2} \sum_{k=1}^n \left(x_{(i,k)} - \sum_{j=1}^p a_{ij}^{(r+1)} s_{(j,k)}^{(r+1)}\right)^2\right\} \times p(1/\sigma_i^2). \end{aligned} \quad (4.47)$$

L'estimation des variances en utilisant une loi *a priori* uniforme ou une loi de Jeffreys sur  $(\sigma_i^2)$ , donne une solution dégénérée [SMD01a]. Ce problème de dégénérescence est éliminé en prenant une loi inverse

gamma comme loi *a priori* sur ces variances, ce qui est aussi une loi conjuguée [Rob01]. A partir de l'expression de la vraisemblance, nous pouvons constater que cette loi conjuguée est une distribution gamma pour  $(1/\sigma_i^2)$

$$p(1/\sigma_i^2) = \mathcal{G}\left(\alpha_{\sigma_i^2}^{prior}, \beta_{\sigma_i^2}^{prior}\right), \quad (4.48)$$

et mène à une loi *a posteriori* conditionnelle donnée par

$$p\left(1/\sigma_i^2 | x_{(i,1:n)}, a_{(i,1:n)}^{(r+1)}, s_{(i,1:n)}^{(r+1)}\right) = \mathcal{G}\left(1/\sigma_i^2; \alpha_{\sigma_i^2}^{post}, \beta_{\sigma_i^2}^{post}\right), \quad (4.49)$$

avec

$$\begin{cases} \alpha_{\sigma_i^2}^{post} = \frac{n}{2} + \alpha_{\sigma_i^2}^{prior}, \\ \beta_{\sigma_i^2}^{post} = \frac{1}{2} \sum_{k=1}^n \left( x_{(i,k)} - \sum_{j=1}^p a_{ij}^{(r+1)} s_{(j,k)}^{(r+1)} \right)^2 + \beta_{\sigma_i^2}^{prior}. \end{cases} \quad (4.50)$$

Dans le cas de l'estimation de la variance, les paramètres  $\alpha_{\sigma_i^2}^{prior}$  et  $\beta_{\sigma_i^2}^{prior}$  sont choisis de sorte à ce que cet *a priori* couvre l'intervalle des valeurs possibles de cette variance.

#### 4.3.4.4 Hyperparamètres des signaux sources

La densité conditionnelle *a posteriori* de chaque hyperparamètre  $\alpha_j$  est donnée par

$$p\left(\alpha_j | s_{(j,1:n)}^{(r+1)}, \beta_j^{(r)}\right) \propto \prod_{k=1}^n \frac{\beta_j^{\alpha_j}}{\Gamma(\alpha_j)} s_{(j,k)}^{\alpha_j-1} \times p(\alpha_j), \quad (4.51)$$

$$\propto \frac{1}{\Gamma(\alpha_j)^n} \exp\left\{\left(n \log \beta_j^{(r)} + \sum_{k=1}^n \log s_{(j,k)}^{(r+1)}\right) \alpha_j\right\} \times p(\alpha_j). \quad (4.52)$$

Afin de prendre en compte la non-négativité du paramètre  $\alpha_j$ , une loi *a priori* exponentielle de paramètre  $\delta_{\alpha_j}^{prior}$  est utilisée comme loi *a priori*, ce qui conduit à une densité conditionnelle *a posteriori* donnée par

$$p(\alpha_j | s_j, \mu_j) \propto \left(\frac{1}{\Gamma(\alpha_j)} \exp\left\{\delta_{\alpha_j}^{post} \alpha_j\right\}\right)^n, \quad (4.53)$$

où

$$\delta_{\alpha_j}^{post} = \log \beta_j^{(r)} + \frac{1}{n} \sum_{k=1}^n \log s_{(j,k)}^{(r+1)} - \frac{1}{n} \delta_{\alpha_j}^{prior}.$$

Cette densité n'est pas usuelle, donc sa simulation nécessite l'utilisation d'un algorithme Metropolis-Hastings. La technique utilisée pour simuler cette densité est décrite en annexe.

Concernant les paramètres  $\beta_{(1:p)}$ , la densité conditionnelle *a posteriori* de chaque paramètre est donnée par

$$p\left(\beta_j | s_{(j,1:n)}^{(r+1)}, \alpha_j^{(r+1)}\right) \propto \beta_j^{n\alpha_j^{(r+1)}} \exp\left\{-\beta_j \sum_{k=1}^n s_{(j,k)}^{(r+1)}\right\} \times p(\beta_j). \quad (4.54)$$

On peut noter rapidement qu'une loi conjuguée pour le paramètre  $\beta_j$  est une densité gamma,

$$\beta_j \sim \mathcal{G}\left(\alpha_{\beta_j}^{prior}, \beta_{\beta_j}^{prior}\right), \quad (4.55)$$

ce qui mène à une densité conditionnelle *a posteriori* qui est aussi une loi gamma

$$\left(\beta_j^{(r+1)} \mid s_{(j,1:n)}^{(r+1)}, \alpha_j^{(r+1)}\right) \sim \mathcal{G}\left(\alpha_{\beta_j}^{post}, \beta_{\beta_j}^{post}\right), \quad (4.56)$$

avec des paramètres

$$\begin{cases} \alpha_{\beta_j}^{post} &= 1 + n\alpha_j^{(r+1)} + \alpha_{\beta_j}^{prior}, \\ \beta_{\beta_j}^{post} &= \sum_{k=1}^n s_{(j,k)} + \beta_{\beta_j}^{prior}. \end{cases} \quad (4.57)$$

#### 4.3.4.5 Hyperparamètres des coefficients de mélange

Les coefficients de mélange sont aussi représentés par des densités gamma comme loi *a priori*, alors les paramètres  $\gamma_{(1:p)}$  et  $\delta_{(1:p)}$  sont simulés de la même façon que les hyperparamètres des signaux sources.

### 4.4 Illustration de l'algorithme

Le fonctionnement de la méthode proposée est illustré sur un exemple synthétique permettant de discuter toutes les étapes de l'estimation par MCMC. Dans ce qui suit, nous utiliserons le sigle BPSS (pour *Bayesian Positive Source Separation*) afin de désigner le méthode de séparation proposée avec un *a priori* gamma sur les sources et sur les coefficients de mélange. Les signaux sources sont simulés de telle sorte à obtenir des mélanges de formes similaires aux signaux de spectroscopie. La figure 4.3 montre les signaux sources et les coefficients de mélange simulés ( $m = 3$ ,  $p = 3$  et  $n = 1000$ ). Un bruit i.i.d gaussien de moyenne nulle et de variance  $\sigma^2 = 0.01$  est rajouté à chaque observation.

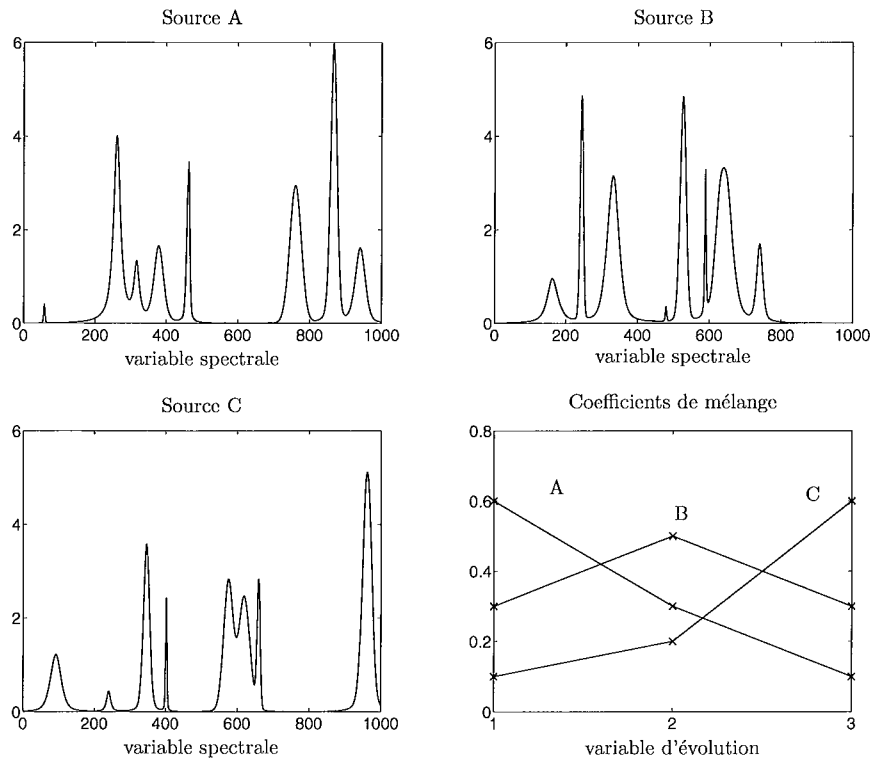


FIG. 4.3 – Signaux sources et coefficients de mélange simulés.

L'algorithme de séparation par approche bayésienne est appliqué en utilisant les valeurs de paramètres données dans le tableau 4.1. Une particularité de ce choix des paramètres des densités *a priori* des hyperparamètres est que les densités résultants tendent vers des loi quasi-uniformes sur l'intervalle  $]0, +\infty[$ . La seule information qui est introduite par ce choix est la positivité des paramètres  $\{\sigma_i^2, \alpha_j, \beta_j, \gamma_j, \delta_j\}$ . Par ailleurs, après supervision des chaînes de Markov simulées le nombre d'itérations de chauffe et le nombre maximal d'itérations sont fixés respectivement à 500 et 10000.

Paramètre	$\alpha_{\sigma_i^2}^{prior}$	$\beta_{\sigma_i^2}^{prior}$	$\lambda_{\alpha_j}^{prior}$	$\alpha_{\beta_j}^{prior}$	$\beta_{\beta_j}^{prior}$	$\lambda_{\gamma_j}^{prior}$	$\alpha_{\delta_j}^{prior}$	$\beta_{\delta_j}^{prior}$
Valeur	2	0.001	0.001	2	0.001	0.001	2	0.001

TAB. 4.1 – Valeurs des paramètres des lois *a priori* associées aux variances du bruit, aux signaux sources et aux coefficients de mélange.

#### 4.4.1 Analyse des chaînes de Markov

La figure 4.4 montre les 1000 premiers échantillons des chaînes de Markov simulées en utilisant l'algorithme proposé. L'analyse des chaînes simulées permet d'approcher les distributions marginales *a posteriori* des différentes variables.

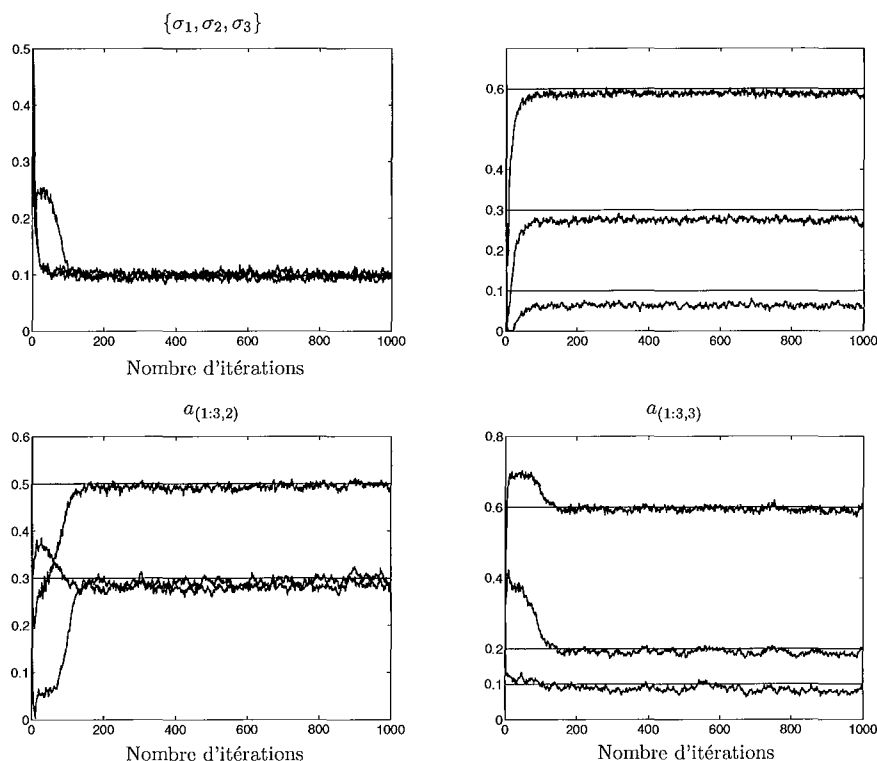


FIG. 4.4 – Chaînes de Markov des variances du bruit et des coefficients de mélange

L'un des apports de cette méthode par rapport aux algorithmes NNSC et PMF réside dans la possibilité d'estimer les hyperparamètres durant la simulation MCMC. La Figure 4.5 montre les histogrammes des hyperparamètres. Il est à noter que les valeurs de ces paramètres sont différentes et

que les paramètres  $\alpha_1, \alpha_2$  et  $\alpha_3$  sont inférieures à l'unité, ce qui correspond à utiliser des lois *a priori* gamma dont la forme est illustrée par le premier cas de la figure 3.1. La figure 4.6 résume les histogrammes des coefficients de mélange et la figure 4.7 montre que les sources estimées sont très similaires avec les vraies sources.

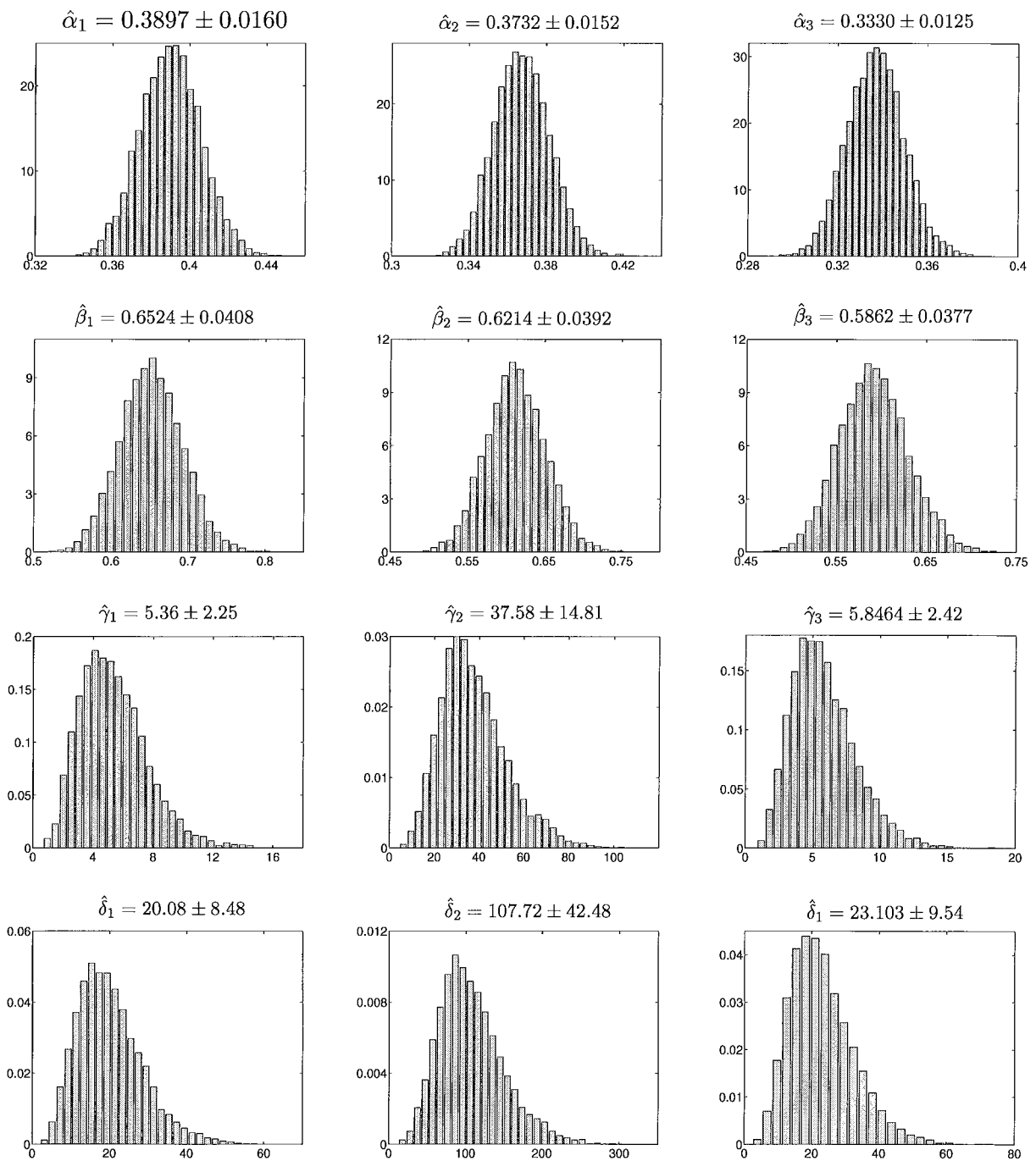


FIG. 4.5 – Histogrammes des paramètres des lois *a priori* des signaux sources et des coefficients de mélange.



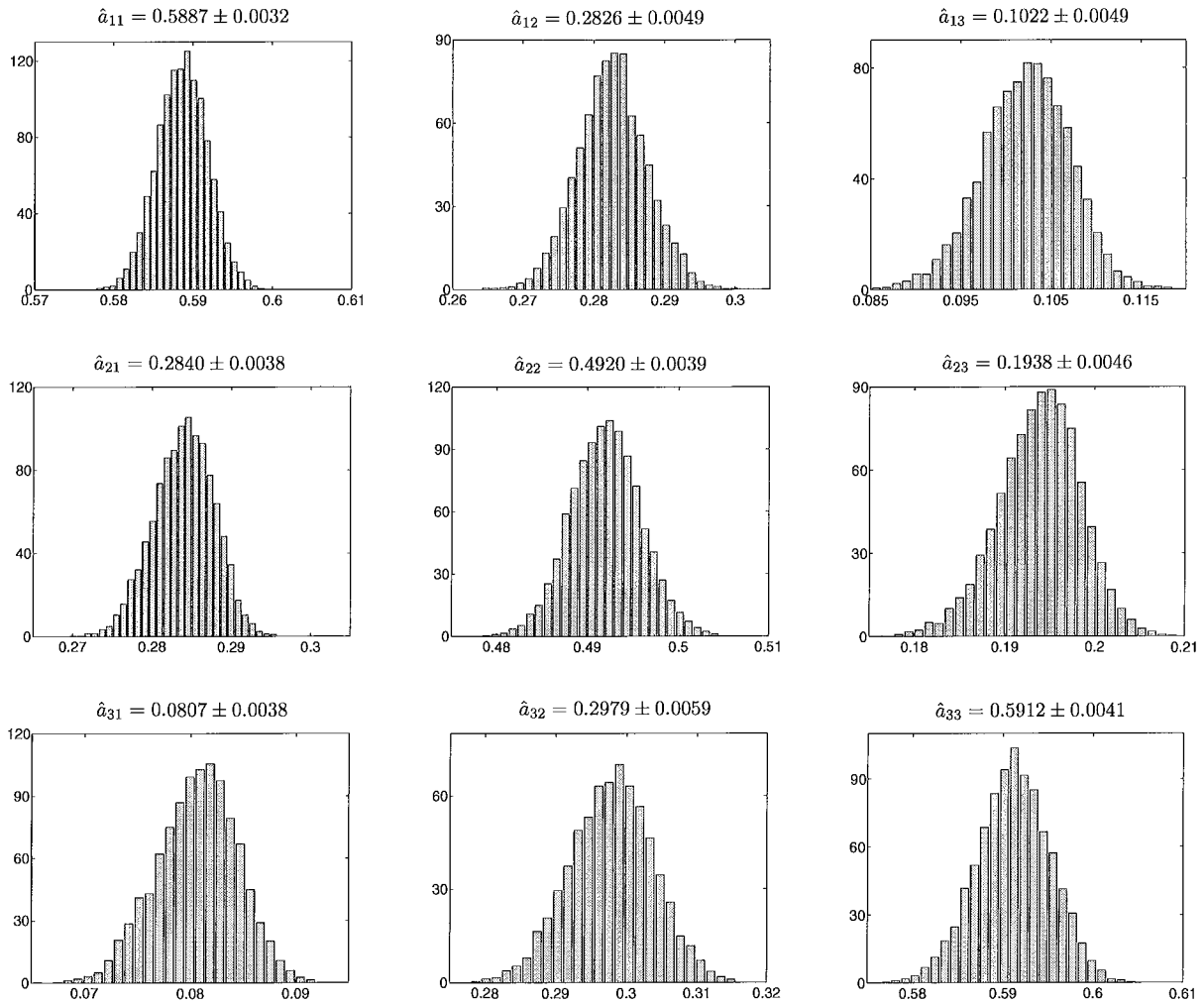


FIG. 4.6 – Histogrammes des coefficients de mélange simulés.

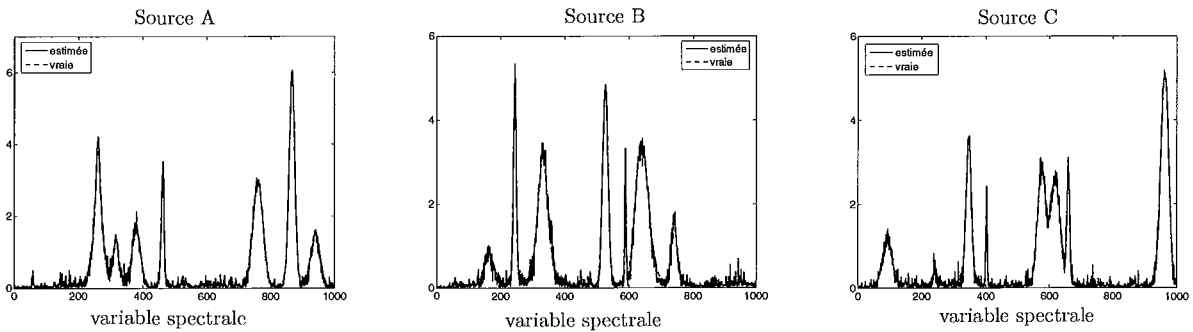


FIG. 4.7 – Signaux sources estimés par approche bayésienne.

### 4.4.2 Temps de calcul

Le temps de calcul de l'algorithme dépend du nombre d'échantillons, du nombre d'observations et du nombre de sources. Celui-ci est aussi fonction du nombre total d'itérations de la simulation de Monte Carlo. Afin d'analyser l'influence de ces facteurs, une étude du temps de calcul est réalisée en faisant varier les ordres  $(m, n, p)$ . La figure 4.8 montre que la durée de traitement augmente de façon

linéaire avec ces trois paramètres. Pour des dimensions de taille moyenne, le calcul peut être réalisé en une durée raisonnable. Par exemple, pour  $(m = 10, n = 1000, p = 4)$  le temps nécessaire pour 5000 itérations est de 8 minutes. En revanche, pour des dimensions plus larges ou pour un nombre d'itérations plus élevé, une durée plus longue sera nécessaire pour le traitement.

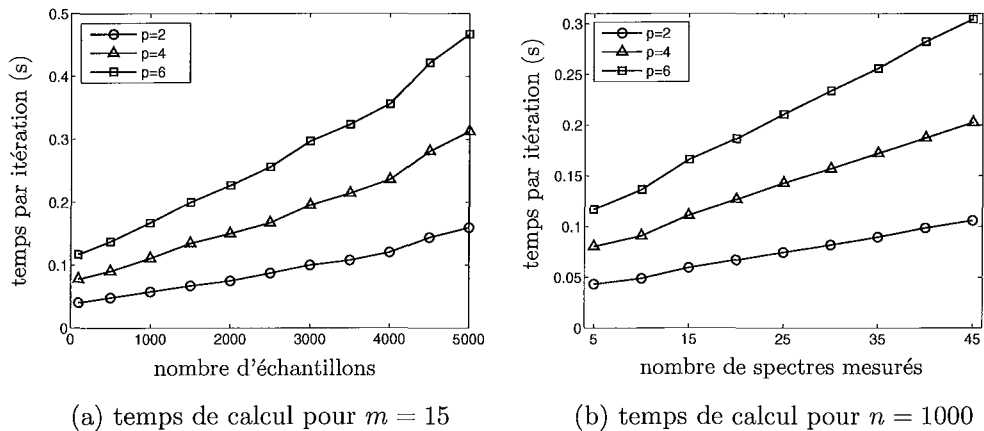


FIG. 4.8 – Temps de calcul par itération.

## 4.5 Conclusions

Dans ce chapitre, nous avons abordé le problème de séparation de sources non-négatives en utilisant une approche bayésienne. La méthode proposée consiste à utiliser une densité gamma comme modèle *a priori* sur la distribution des sources et des coefficients de mélange. L'estimation conjointe des signaux sources, des coefficients de mélange et des hyperparamètres est réalisée en utilisant les méthodes MCMC. Par contre, une question fondamentale est celle de l'adéquation du modèle de distribution gamma pour le cas de sources spectrales. Néanmoins, nous pouvons dire que le choix d'une telle distribution permet d'obtenir une solution particulière parmi les solutions admissibles qui soient le plus indépendantes possibles et dont la distribution est la plus proche d'une densité gamma. Ainsi les performances de la méthode BPSS vont dépendre de la capacité de la densité gamma à représenter la distribution des sources recherchées. Dans le cas où les signaux d'intérêt possèdent des densités très éloignées d'une densité gamma, il faut alors chercher des hypothèses ou contraintes supplémentaires en plus de la non-négativité.

## Chapitre 5

# Analyse des performances et comparaisons

Ce chapitre a pour objectifs l'analyse des performances des algorithmes développés aux chapitres 3 et 4 et l'étude comparative avec quelques méthodes usuelles de séparation de sources et de factorisation en matrices non-négatives. L'analyse est réalisée avec des signaux synthétiques de formes similaires aux données spectroscopiques. Nous discutons l'influence de trois facteurs sur les performances des algorithmes : (i) la corrélation mutuelle des sources, (ii) le niveau de bruit dans les observations et (iii) la présence d'un arrière-plan sur les sources.

### 5.1 Mise au point des simulations

#### 5.1.1 Simulation des spectres purs

Afin d'obtenir des signaux sources de formes similaires aux spectres de composantes chimiques, nous utilisons un modèle simplifié dans lequel chaque source est représentée par la superposition de plusieurs motifs élémentaires de formes lorentzienne et gaussienne, dont les amplitudes et les paramètres de position et de dispersion sont choisis de façon pseudo-aléatoire. Ce modèle est décrit par

$$s_j(k) = \sum_{\ell=1}^L \alpha_{j\ell} f_{j\ell}(k; \nu_{j\ell}, \gamma_{j\ell}, \rho_{j\ell}), \quad (5.1)$$

où  $\alpha_{j\ell} \geq 0; \forall \ell$  et les  $L$  motifs élémentaires  $f_{j\ell}(k; \nu_{j\ell}, \gamma_{j\ell}, \rho_{j\ell})$  sont définis par

$$f_{j\ell}(k; \nu_{j\ell}, \gamma_{j\ell}, \rho_{j\ell}) = \rho_{j\ell} \frac{1}{\gamma_{j\ell}} \exp \left[ -\frac{(k - \nu_{j\ell})^2}{2\gamma_{j\ell}^2} \right] + (1 - \rho_{j\ell}) \frac{1}{\sqrt{\pi}\gamma_{j\ell}} \frac{\gamma_{j\ell}^2}{\gamma_{j\ell}^2 + (k - \nu_{j\ell})^2}. \quad (5.2)$$

Noter que ce modèle est utilisé en spectroscopie pour la modélisation des spectres [GM82, IAT00]. Pour simuler un spectre de  $n$  points, à l'aide de ce modèle, les paramètres  $\{\alpha_{j\ell}, \nu_{j\ell}, \gamma_{j\ell}, \rho_{j\ell}\}$  sont choisis arbitrairement selon

$$\alpha_{j\ell} \sim \mathcal{N}_+(0, 1), \quad \nu_{j\ell} \sim \mathcal{U}(0, n-1), \quad \rho_{j\ell} \sim \mathcal{U}(0, 1), \quad \gamma_{j\ell}^2 \sim \mathcal{U}(\gamma_{min}^2, \gamma_{max}^2). \quad (5.3)$$

où  $\gamma_{min}$  et  $\gamma_{max}$  contrôlent la largeur minimale et la largeur maximale des motifs élémentaires. La figure 5.1 montre un exemple de sources normalisées (variance unité) obtenues à l'aide de ce modèle.

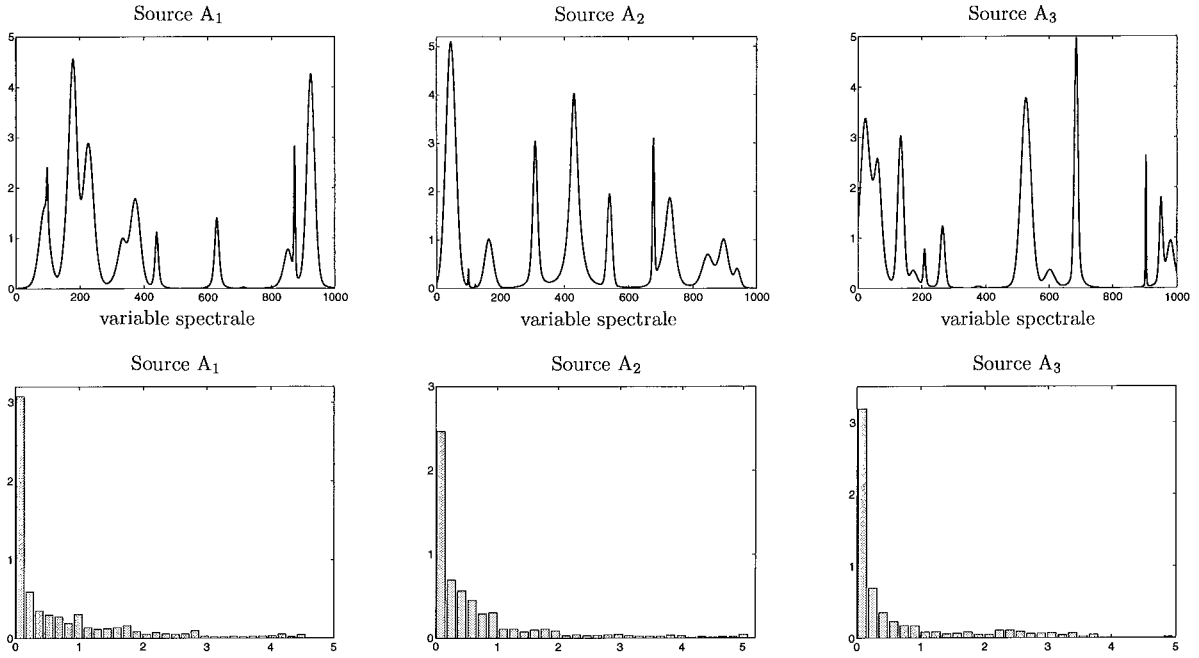


FIG. 5.1 – Allure des signaux sources simulés et leurs histogrammes. Les paramètres de synthèse sont  $K = 15$ ,  $\gamma_{max} = 10$ ,  $\gamma_{min} = 2n$ . Ces sources présentent les coefficients de corrélation mutuelle suivants  $\text{corr}(s_1, s_2) = -0.16$ ,  $\text{corr}(s_1, s_3) = -0.24$ , et  $\text{corr}(s_2, s_3) = +0.27$ .

### 5.1.2 Simulation des coefficients de mélange

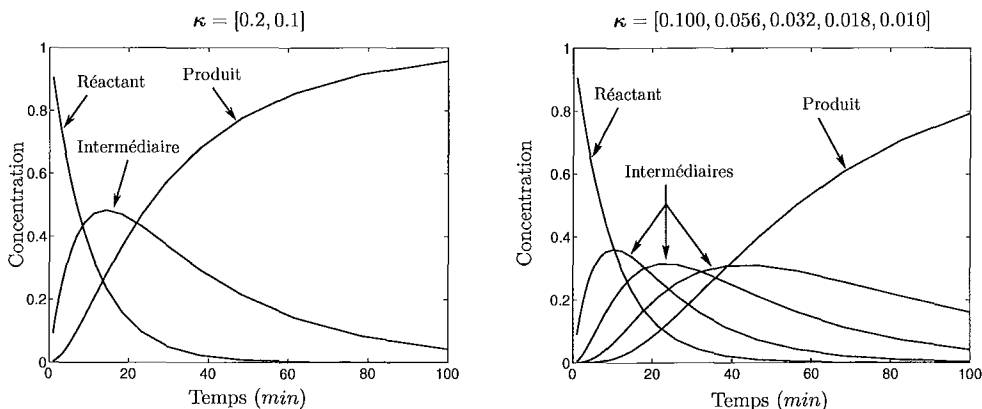
Les coefficients de mélange peuvent être simulés arbitrairement à partir d'une loi uniforme à support positif. Cependant, afin de se rapprocher des signaux expérimentaux, nous avons choisi de les simuler à partir d'un modèle d'évolution des concentrations des espèces lors d'une réaction chimique. Pour cela, considérons une cinétique dans le cas d'un processus de réactions successives de premier ordre,  $A_1 \xrightarrow{\kappa_1} A_2 \xrightarrow{\kappa_2} A_3$ , où  $\kappa_1, \kappa_2$  sont les constantes de vitesse des réactions. Dans ce cas, la réaction possède un seul intermédiaire réactionnel avant d'atteindre un état quasi-stationnaire où seulement la troisième espèce sera présente. Les concentrations des différentes espèces sont données par les relations suivantes [Nic05]

$$\begin{cases} C_{(i,1)} &= e^{-\kappa_1 t_i}; \\ C_{(i,2)} &= \frac{\kappa_1}{\kappa_2 - \kappa_1} (e^{-\kappa_1 t_i} - e^{-\kappa_2 t_i}); \\ C_{(i,3)} &= 1 - C_{(i,1)} - C_{(i,2)}, \end{cases} \quad (5.4)$$

où  $C_{(i,j)}$  est la concentration de l'espèce  $A_j$  à l'instant  $t_i$ . Dans le cas de plusieurs sources  $p > 3$ , le modèle général est un peu plus complexe. Un modèle simplifié est défini par les équations suivantes

$$\begin{cases} C_{(i,1)} &= e^{-\kappa_1 t_i}, \\ C_{(i,j)} &= \prod_{\ell=1}^{j-1} (1 - e^{-\kappa_\ell t_i}) e^{-\kappa_j t_i}, \text{ pour } j = 2, \dots, p-1, \\ C_{(i,p)} &= 1 - \sum_{j=1}^{p-1} C_{(i,j)}, \end{cases} \quad (5.5)$$

dans lesquelles nous supposons que  $\kappa_1 > \kappa_2 > \dots > \kappa_p$ .



(a) Cinétique avec un seul intermédiaire réactionnel

(b) Cinétique avec trois intermédiaires réactionnels

FIG. 5.2 – Coefficients de mélange simulés à partir d’une cinétique chimique.

### 5.1.3 Synthèse des observations

A partir des coefficients de mélange et des signaux sources simulés, nous construisons les observations selon le modèle de mélange linéaire avec un bruit additif, blanc gaussien i.i.d et spatialement décorrélé. Un exemple de mélanges de  $p = 3$  sources avec  $m = 10$  observations, dans le cas sans bruit, est illustré par la figure 5.3.

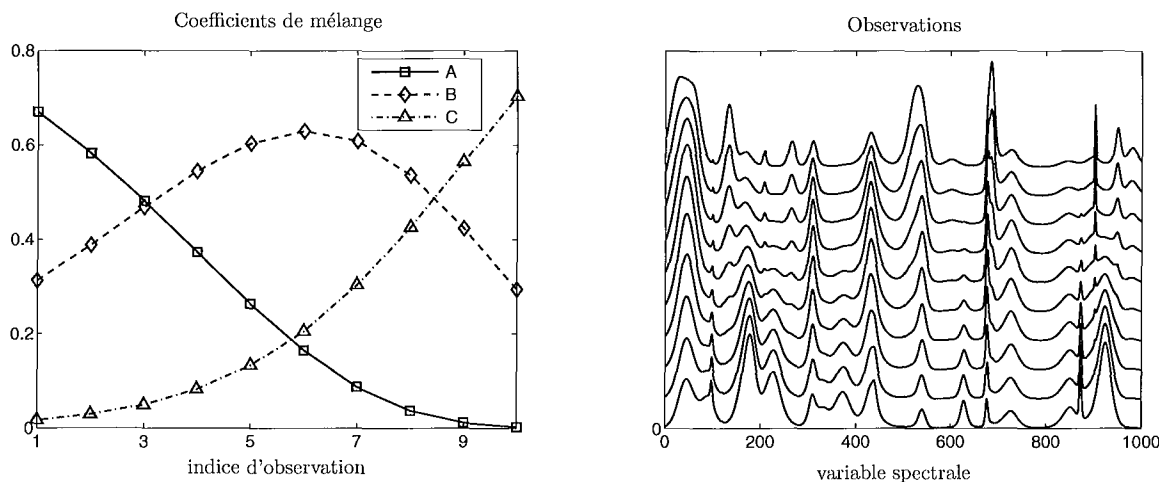


FIG. 5.3 – Coefficients de mélange et observations simulées pour  $m = 10, p = 3$  et  $n = 1000$ .

### 5.1.4 Indices de performances

Le niveau de bruit dans chaque observation, mesuré en terme de rapport signal à bruit (RSB), est exprimé par

$$RSB_i(\text{dB}) = 10 \log_{10} \left( \frac{\sum_{k=1}^n x_{(i,k)}^2}{\sum_{k=1}^n e_{(i,k)}^2} \right), \tag{5.6}$$

où  $x_{(i,k)}$  correspond à l'observation non-bruitée.

Afin de comparer les résultats, nous utilisons l'indice de performance, noté  $\mathcal{IP}$ , défini par [MM94, CA02]

$$\mathcal{IP} = \frac{1}{2p(p-1)} \sum_{i=1}^p \left\{ \left( \sum_{k=1}^p \frac{|g_{ik}|^2}{\max_{\ell} |g_{i\ell}|^2} - 1 \right) + \left( \sum_{k=1}^p \frac{|g_{ki}|^2}{\max_{\ell} |g_{\ell i}|^2} - 1 \right) \right\}, \quad (5.7)$$

où  $g_{ij}$  est  $(i, j)$ -ème élément de la matrice d'indétermination  $\mathbf{G} = \hat{\mathbf{B}}\mathbf{A}$ ,  $\max_{\ell} g_{i\ell}$  représente la valeur maximale sur les éléments de la  $i$ -ème ligne de  $\mathbf{G}$  et  $\max_{\ell} g_{\ell i}$  représente la valeur maximale sur la  $i$ -ème colonne de  $\mathbf{G}$ . Cet indice mesure les performances globales de la séparation et vaut zéro pour une parfaite estimation de la matrice de mélange. En pratique, des petites valeurs, inférieures à 0.01, de cet indice indiquent une bonne séparation. D'autre part, il est également important de mesurer la qualité de reconstruction des sources. Pour cela, nous utilisons le Rapport Signal à Distorsion ( $\mathcal{RSD}$ ) sur chaque signal source, défini par

$$\mathcal{RSD}_j(\text{dB}) = 10 \log_{10} \left( \frac{\sum_{k=1}^n s_{(j,k)}^2}{\sum_{k=1}^n (s_{(j,k)} - \hat{s}_{(j,k)})^2} \right), \quad (5.8)$$

où  $\hat{s}_j$  est l'estimation de la  $j$ -ème source  $s_j$ , les deux signaux étant à variance unité. Tout au long de ce document, tous ces critères ( $\mathcal{RSB}$ ,  $\mathcal{IP}$  et  $\mathcal{RSD}$ ) sont exprimés en décibels.

## 5.2 Comparaisons avec des méthodes usuelles

A travers cette section nous illustrons les avantages des méthodes proposées pour l'analyse de signaux décrits dans le paragraphe 5.1. Nous distinguons deux situations :

- le cas d'un faible niveau de bruit où l'algorithme de séparation par maximum de vraisemblance avec estimation par gradient naturel est employé ;
- le cas d'un niveau de bruit plus important où la séparation est effectuée par l'approche bayésienne et méthodes MCMC.

Les résultats de ces deux méthodes sont comparés à celles présentées dans le chapitre 1. L'ensemble de ces algorithmes est implanté sous Matlab 7.

### 5.2.1 Cas d'un faible niveau de bruit

Un bruit additif est choisi de telle sorte à avoir une RSB de 50 dB sur chaque observation. Il s'agit donc d'un cas où le niveau de bruit est très faible.

#### 5.2.1.1 Séparation par analyse en composantes indépendantes

Une première approche est de séparer les composantes en utilisant un algorithme de séparation par analyse en composantes indépendantes. Le calcul de la matrice de covariance empirique des sources donne

$$\hat{\mathbf{R}}_s = \frac{1}{n} \sum_{k=1}^n [\mathbf{s}_k - \boldsymbol{\mu}_s][\mathbf{s}_k - \boldsymbol{\mu}_s]^T = \begin{bmatrix} 1.00 & -0.16 & -0.23 \\ -0.16 & 1.00 & 0.27 \\ -0.23 & 0.27 & 1.00 \end{bmatrix}, \quad (5.9)$$

Au vu des coefficients de corrélations mutuelles des sources, l'utilisation des méthodes de séparation qui forcent l'orthogonalité des sources n'est pas adéquat. La figure 5.4 montre le résultat de la séparation par JADE, où nous pouvons noter l'apparition de pics inexistantes dans les vraies sources A et B et des pics négatifs au niveau de la source C. D'autre part, des coefficients de mélange négatifs ont été également estimés.

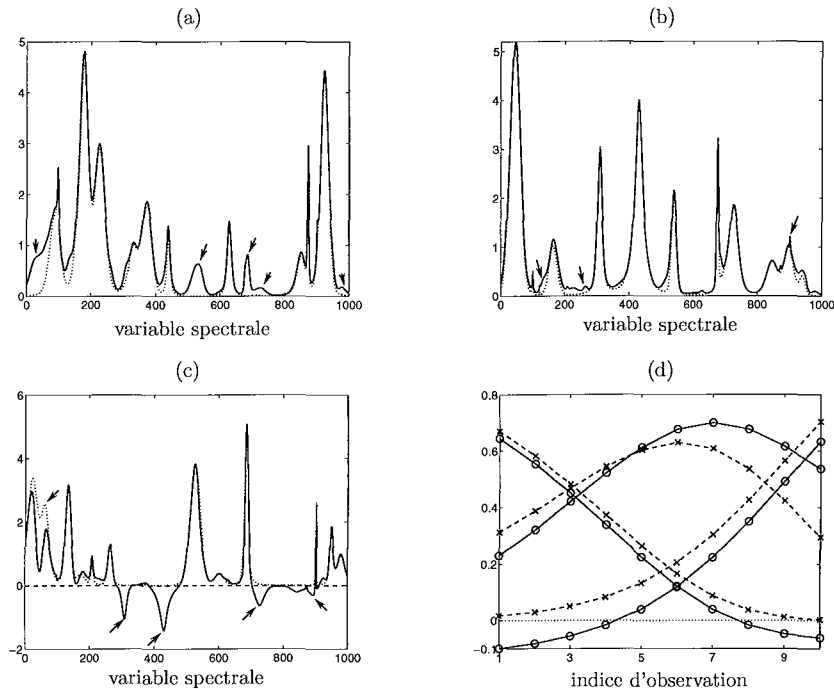


FIG. 5.4 – Résultat de la séparation en utilisant JADE : (a,b,c) sources originales (discontinues) et sources estimées (continues). Les flèches indiquent les endroits où apparaissent des composantes inexistantes ou aberrantes. (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles).

### 5.2.1.2 Séparation par analyse en composantes indépendantes non-négatives

La méthode NNICA présente l'avantage de chercher des composantes indépendantes en tenant compte de l'information de non-négativité. Mais, la réalisation d'une étape de blanchiment des observations est une limitation de cette méthode, car comme illustré par l'exemple considéré, les sources peuvent ne pas être orthogonales. La figure 5.5 montre le résultat de la séparation en utilisant cette méthode, et nous pouvons noter l'apparition de composantes négatives que ce soit sur les signaux sources ou sur les coefficients de mélange. Néanmoins, l'amplitude des valeurs négatives apparaissant sur les sources est plus faible que dans le cas de l'algorithme JADE.

### 5.2.1.3 Séparation par factorisation en matrices non-négatives

Une autre approche consiste à utiliser des méthodes de type NMF sur la matrice de données. Néanmoins, comme nous pouvons le remarquer sur l'allure des coefficients de mélange, la condition nécessaire A2 du théorème 1 du chapitre 2 n'est pas satisfaite et par conséquent la séparation en utilisant seulement la contrainte de non-négativité n'admet pas une solution unique. La figure 5.6 montre le

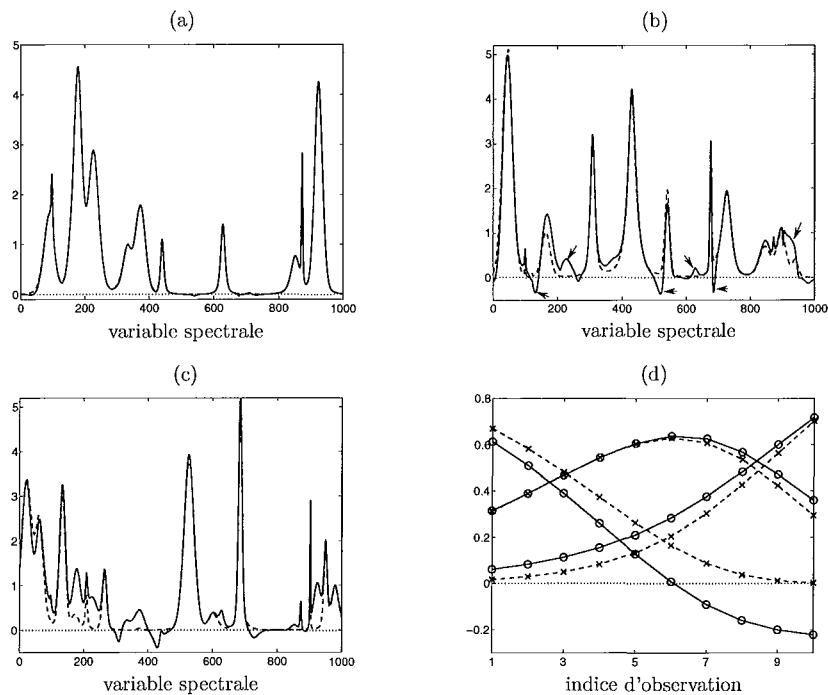


FIG. 5.5 – Résultat de la séparation en utilisant NNICA : (a,b,c) sources originales (discontinu) et sources estimées (continu). (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles).

résultat de séparation par NMF, où nous pouvons noter que pour plusieurs initialisations différentes de l'algorithme, nous obtenons autant de solutions différentes. Par ailleurs, le même constat reste valable pour la méthode ALS qui appliquée à l'analyse avec comme valeurs initiales les résultat de l'estimation par JADE, OPA ou SIMPLISMA fournit des résultats différents. Rappelons que les méthodes OPA et SIMPLISMA sont utilisées en chimométrie.

#### 5.2.1.4 Séparation par analyse en composantes parcimonieuses non-négatives

Cette approche consiste en la méthode NNSC qui est une méthode d'estimation par minimisation d'un critère pénalisé. Le critère de régularisation favorise une solution particulière parmi toutes les solutions admissibles. Cette solution particulière dépend fortement du choix du coefficient de régularisation qui est un paramètre de synthèse de la méthode. En ajustant ce paramètre de façon à obtenir la meilleure reconstruction des sources, nous obtenons le résultat illustré par la figure 5.7. Nous pouvons constater que la spécification de l'hypothèse de parcimonie des sources permet d'améliorer le résultat de la séparation, notamment par rapport à la méthode NMF. Néanmoins, il reste un point crucial qui est celui du choix du coefficient de régularisation. En outre, comme mentionné dans le chapitre 4, la méthode NNSC suppose que chacune des sources est distribuée selon une loi exponentielle de même paramètre.



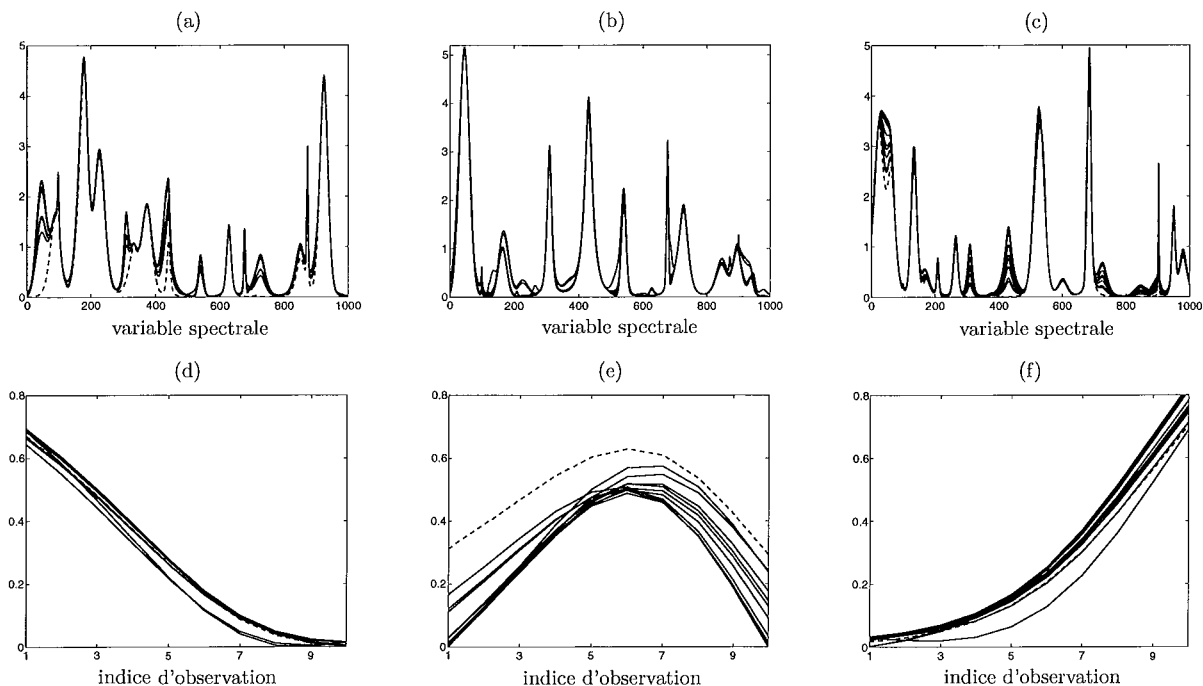


FIG. 5.6 – Résultat de la séparation en utilisant NMF : (a,b,c) sources originales (discontinu) et sources estimées (continu). (d) coefficients de mélange originaux (discontinu) et coefficients identifiés (continu).

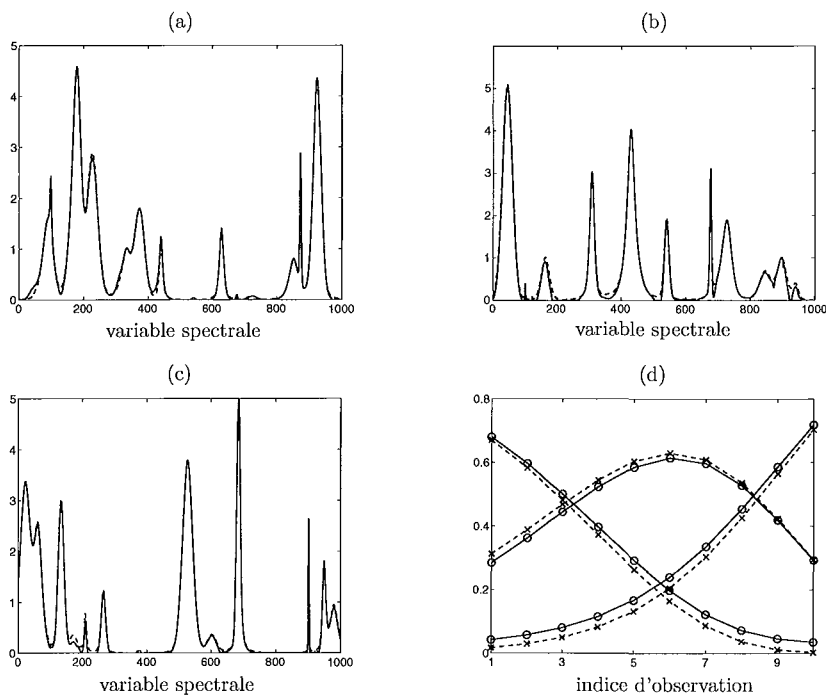


FIG. 5.7 – Résultat de la séparation en utilisant NNSC : (a,b,c) sources originales (discontinu) et sources estimées (continu). (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles).

### 5.2.1.5 Séparation par maximum de vraisemblance et gradient relatif

L'approche de séparation par maximum de vraisemblance offre l'avantage de ne pas imposer de façon stricte l'orthogonalité des sources et permet de spécifier *a priori* les formes de leurs distributions. Nous pouvons constater sur la figure 5.8 la quasi-parfaite reconstruction des sources par cette approche. On peut attribuer cela à la possibilité d'avoir une loi différente pour chacune des sources. En outre, cette approche permet l'estimation des hyperparamètres ; cependant il convient de souligner que l'estimation conjointe de la matrice de séparation et des hyperparamètres peut se heurter à un problème de convergence vers un minimum local. Cette remarque est également valable pour l'algorithme MCEM. De ce point de vue, l'approche par méthodes MCMC est plus adéquate mais peut, dans le cas d'un faible niveau de bruit, converger très lentement.

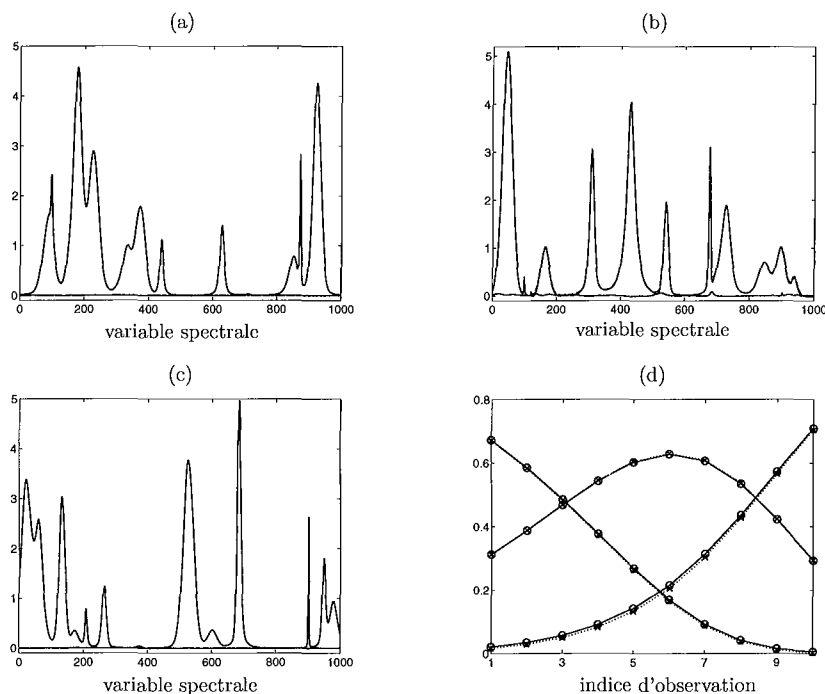


FIG. 5.8 – Résultat de la séparation en utilisant l'approche MLPS-gamma : (a,b,c) signaux sources originales (discontinu) et sources estimées (continu). Le résidu de l'estimation est aussi tracé pour évaluer l'erreur de reconstruction. (d) coefficients de mélange originaux (croix) et coefficients identifiés (cercles).

### 5.2.1.6 Récapitulatif des résultats

Le tableau 5.1 donne une évaluation numérique des performances des différents algorithmes. Nous pouvons noter la supériorité de l'approche du maximum de vraisemblance par rapport à toutes les méthodes testées, que ce soit en terme de qualité de séparation globale ou encore en terme de qualité de reconstruction des sources.

	JADE	JADE-ALS	OPA-ALS	SIMPLISMA-ALS	NMF	NNSC	NNICA	MLPSS
$\mathcal{RSD}_A$	13.50	11.25	7.59	26.69	18.31	28.12	37.69	53.78
$\mathcal{RSD}_B$	20.28	15.74	14.96	34.63	17.84	23.32	14.79	34.54
$\mathcal{RSD}_C$	9.40	35.51	10.29	26.02	10.16	28.29	11.97	49.45
$\mathcal{IP}$	-12.74	-13.97	-8.91	-28.91	-11.70	-26.44	-14.66	-38.81

TAB. 5.1 – Comparaison des performances de la séparation dans le cas d’une un RSB de 50 dB.

### 5.2.2 Cas d’un niveau de bruit non-négligeable

Un bruit gaussien i.i.d et spatialement indépendant est ajouté aux observations de telle sorte à obtenir un RSB de 10 dB sur chaque observation. Le tableau 5.2 donne un récapitulatif des performances de quelques méthodes. En présence d’un bruit non-négligeable sur les observations, l’approche bayésienne est la plus adéquate pour l’analyse. D’autre part, l’algorithme NMF de Lee et Seung est remplacé par cNMF [SDB<sup>+</sup>03, SSB<sup>+</sup>04] dans lequel un seuillage des estimations des sources et des coefficients de mélange est introduit afin d’éviter l’estimation de valeurs négatives. Nous pouvons constater que l’approche BPSS donne des meilleures performances, d’autant plus que la contrainte de non-négativité est explicitement prise en compte lors de la simulation MCMC.

	JADE	ALS	cNMF	NNSC	NNICA	BPSS
$\mathcal{RSD}_A$	9.05	6.44	6.72	14.25	8.12	14.54
$\mathcal{RSD}_B$	11.94	11.94	11.55	13.62	9.24	15.74
$\mathcal{RSD}_C$	5.22	8.87	8.85	13.22	11.52	13.46
$\mathcal{IP}$	-12.70	-8.40	-8.56	-19.94	-15.87	-23.90

TAB. 5.2 – Comparaison des performances de la séparation dans le cas d’une un RSB de 10 dB.

## 5.3 Analyse des performances

### 5.3.1 Influence de la corrélation mutuelle des sources

Afin d’analyser l’effet de la corrélation mutuelle des sources sur les performances de la séparation, nous utilisons un mélange de deux sources et nous étudions la variation des performances de la séparation en fonction de leur coefficient de corrélation. Des coefficients de mélange similaires sont retenus durant cette simulation. Afin de s’affranchir de l’influence de la distribution des sources, la deuxième source est une réplique décalée de la première et le coefficient d’inter-corrélation entre les sources correspond au  $\ell$ -ème coefficient d’autocovariance de la première source,  $\ell$ -ème étant le nombre d’échantillons de décalage. Durant cette analyse le pas d’adaptation de l’algorithme du gradient est fixé à 0.001 et le nombre d’itérations maximal est fixé à 10000. Nous pouvons noter sur les résultats illustrés par la figure 5.9 que l’augmentation du coefficient de corrélation des sources engendre une diminution des performances des deux algorithmes (NNICA et MLPSS). Néanmoins, ces performances restent acceptables même pour un coefficient de corrélation de 0.6.

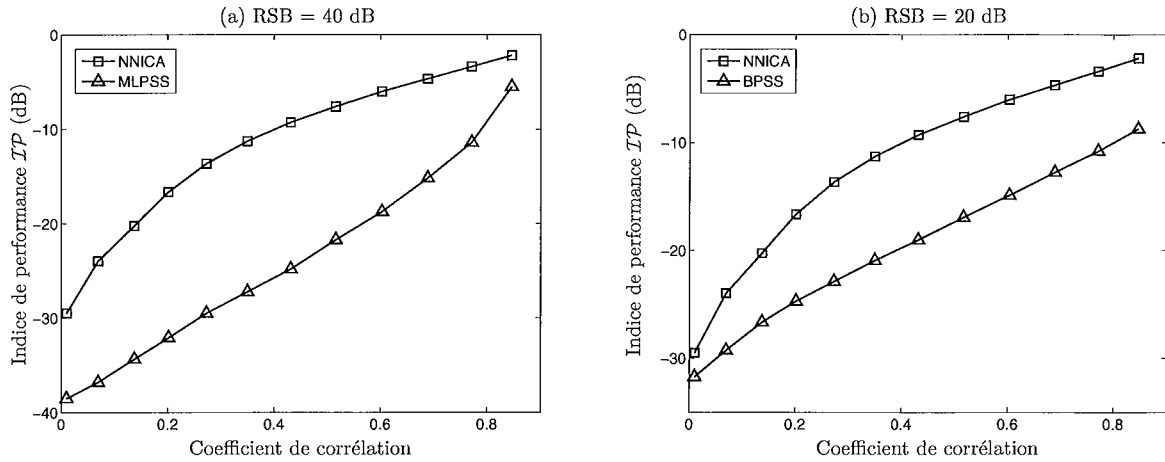


FIG. 5.9 – Influence de la corrélation mutuelle des sources sur les performances de la séparation.

La même expérience est réalisée pour un RSB de 20 dB et dans ce cas nous avons appliqué l'approche BPSS pour la séparation. L'algorithme MCMC est appliqué pour 10000 itérations et les 1000 dernières réalisations sont retenues pour l'estimation. Les résultats de cette expérience sont résumés dans la figure 5.9 qui exprime la variation des indices de performances des méthodes BPSS et NNICA en fonction du coefficient de corrélation des sources. On note que les deux méthodes fournissent de meilleures performances pour des sources faiblement corrélées mutuellement. Par contre, on peut noter que BPSS présente de meilleures performances que NNICA et que l'approche proposée est moins sensible à la corrélation des sources que NNICA. Par exemple, pour un coefficient de corrélation de 0.6 l'indice de performance de la méthode est de l'ordre de -20 dB. Ceci s'explique par le fait que contrairement aux algorithmes ICA, la méthode BPSS n'optimise pas une mesure explicite de l'indépendance mutuelle entre les signaux sources et n'impose pas la contrainte d'orthogonalité. Par ailleurs, nous avons remarqué que l'algorithme MCMC converge plus lentement et nécessite donc de plus en plus d'itérations lorsque la corrélation mutuelle des sources augmente.

### 5.3.2 Influence du niveau de bruit

Afin d'étudier l'influence du niveau de bruit sur les performances de l'approche bayésienne, nous utilisons le mélange synthétique précédent et nous faisons varier progressivement le niveau de bruit sur chaque signal d'observation. Les résultats de cette analyse sont rapportés dans la figure 5.10. Pour un RSB très fort, l'utilisation de la distribution gamma et l'estimation soit par maximum de vraisemblance ou par approche bayésienne donne de meilleures performances que NNICA. D'autre part, en dépit du fait que MLPSS et BPSS utilisent le même modèle *a priori* sur la distribution des sources, les performances de la séparation par approche bayésienne sont meilleures, ce qui s'explique par la restriction du paramètre de forme de la densité gamma à une valeur supérieure à 1 dans le cas de l'algorithme du gradient afin d'assurer sa stabilité et aussi la prise en compte de la non-négativité des coefficients de mélange.



En fait, cette simulation permet de mettre en évidence les limites de la distribution gamma pour représenter les sources spectrales. En effet, dans cette situation, se pose clairement la question de l'adéquation du modèle et ceci de façon d'autant plus cruciale que l'arrière-plan présentera un offset (décalage par rapport à la valeur nulle) important. Dans ce cadre, une généralisation des méthodes proposées serait de considérer des distributions gamma décalées ou généralisées. Une approche alternative serait d'inclure une étape préalable de suppression de l'arrière-plan ou encore de considérer un modèle *a priori* dans lequel chaque source serait la superposition d'un arrière-plan et d'un signal parcimonieux.

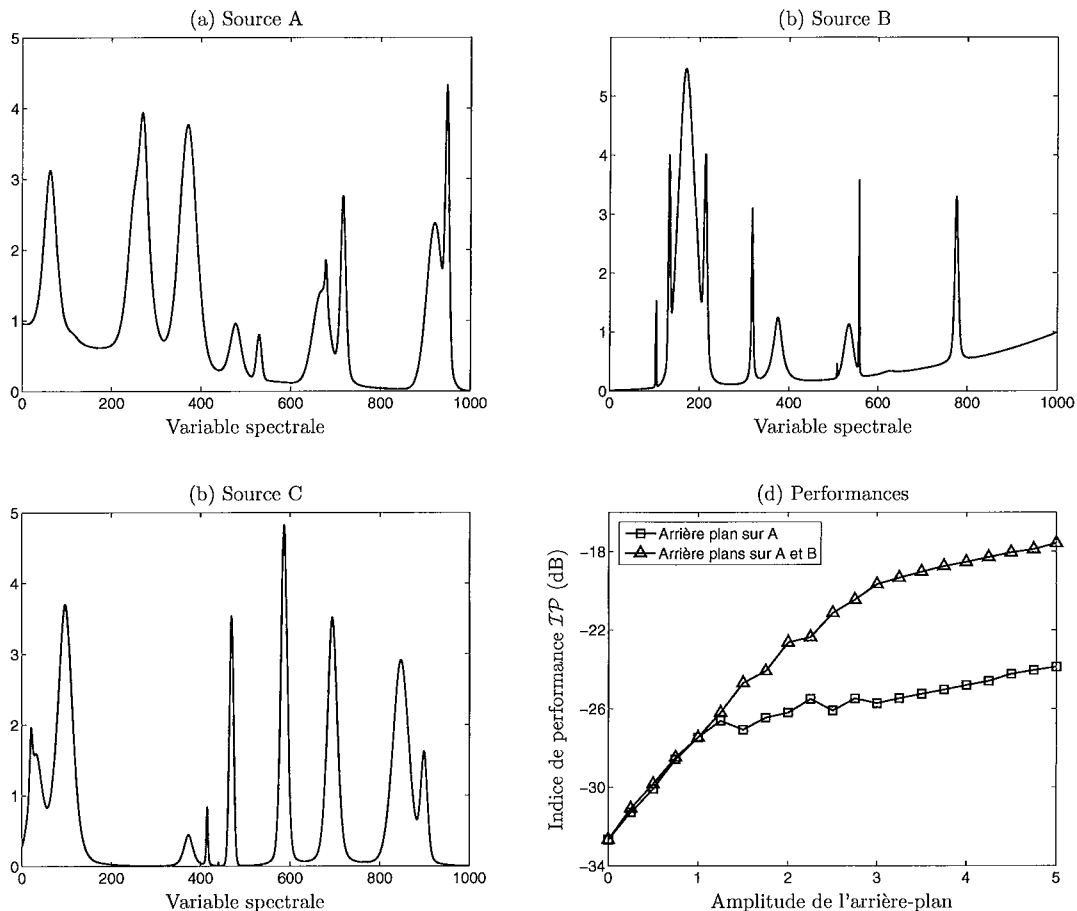


FIG. 5.11 – (a,b,c) Allure des signaux sources pour un arrière-plan d'amplitude égal à l'unité et (d) évolution des performances en fonction de l'amplitude de l'arrière-plan pour un RSB de 20 dB.

## 5.4 Conclusions

Dans ce chapitre, nous avons présenté une analyse des performances des algorithmes proposés à l'aide de quelques simulations numériques qui sont mises au point de telle sorte à obtenir des signaux similaires aux données spectroscopiques. Les résultats de ces simulations ont permis de montrer l'avantage des méthodes proposées (MLPSS et BPSS) par rapport aux approches classiques de séparation de sources non-négatives. L'approche du maximum de vraisemblance est bien adaptée pour le cas de

mélanges où les observations sont faiblement bruitées, car elle permet d'atteindre des niveaux de performances très satisfaisants. Le cas avec un niveau de bruit important est traité en utilisant l'approche bayésienne qui permet de prendre en compte la positivité des coefficients de mélange. Il faut signaler que dans le cas d'un faible niveau de bruit, la méthode de simulation MCMC converge lentement, ce qui implique un grand nombre d'itérations pour l'estimation. C'est pour cette raison que pour un faible niveau de bruit et en l'absence d'erreur de modèle, l'approche du maximum de vraisemblance est plus adéquate car elle permet d'atteindre des performances quasi-similaires à celles de l'approche bayésienne, avec une complexité arithmétique plus faible. L'analyse des performances des algorithmes en fonction de la corrélation mutuelle des sources a mis en évidence la dégradation de la qualité de séparation lorsque les sources deviennent fortement corrélées. Néanmoins, pour des coefficients de corrélations de l'ordre de 0.6, les performances des deux méthodes restent nettement meilleures que les approches classiques. Nous avons également constaté que la présence d'un arrière-plan sur plusieurs sources engendre une dégradation sensible des performances. La prise en compte de ce type de situation nécessite des développements supplémentaires.





Troisième partie

Application à des données  
expérimentales



## Chapitre 6

# Application au traitement de signaux spectroscopiques

Les méthodes spectroscopiques permettent de suivre les cinétiques de réactions, de déterminer les mécanismes réactionnels sur des surfaces, de comprendre l'interaction entre molécules ou encore faciliter la synthèse d'une nouvelle molécule. Ce chapitre présente une introduction à la spectroscopie et illustre l'apport des techniques de séparation de sources pour l'interprétation qualitative et quantitative de données spectroscopiques. Après une étude des performances sur des données expérimentales, nous décrivons des applications à trois problèmes d'analyse physico-chimique : (i) le suivi de l'hydratation d'une silice par spectroscopie infrarouge, (ii) l'étude des polymorphes du carbonate de calcium par spectroscopie Raman et (iii) l'analyse des complexes cations métalliques - biomolécules par spectroscopie UV-Visible.

### 6.1 Introduction à la spectroscopie

La spectroscopie permet l'observation, la compréhension et l'interprétation de phénomènes physico-chimiques. Elle fournit une caractérisation qualitative et quantitative de la matière à partir de l'étude de son interaction avec une onde électromagnétique. Cette interaction se traduit par un phénomène d'absorption, d'émission ou de diffusion de photons, et induit des perturbations moléculaires au niveau de la matière. L'absorption consiste en un transfert d'énergie du rayonnement électromagnétique incident vers la matière absorbante, l'émission est la génération d'une onde électromagnétique après excitation par un faisceau lumineux (fluorescence) ou par un apport d'énergie en température (émission atomique), alors que la diffusion est un phénomène de redirection d'une partie de l'onde émise soit à la même fréquence que l'onde incidente (diffusion élastique Rayleigh), soit à une fréquence différente (diffusion inélastique Raman).

#### 6.1.1 Rayonnement électromagnétique et types de spectroscopie

Les méthodes spectroscopiques peuvent être classées en fonction des caractéristiques des ondes électromagnétiques utilisées pour l'excitation de la matière et de la bande spectrale sur laquelle le

spectre est mesuré. Ces caractéristiques sont exprimées en termes de fréquence  $\nu$ , de longueur d'onde  $\lambda = c_0/\nu$  ( $c_0$  étant la célérité de la lumière) ou du nombre d'onde ( $\bar{\nu} = 1/\lambda$ ). La nature des mouvements moléculaires, électroniques ou atomiques provoqués par l'interaction de la matière avec l'onde incidente dépend fortement de l'énergie de cette onde et donc de sa fréquence. En fonction de la bande spectrale d'analyse, on peut distinguer différentes spectroscopies. La figure 6.1 donne un récapitulatif de quelques régions du spectre électromagnétique qui présentent un intérêt pour l'analyse spectroscopique ainsi que le type de mouvements associés à chaque région. L'absorption de la radiation électromagnétique par une molécule induit un changement d'état du spin nucléaire, s'il s'agit d'une onde radio, ce qui est la base de la spectroscopie par résonance magnétique nucléaire (RMN), un mouvement vibrationnel, s'il s'agit d'une radiation infrarouge, ce qui est la base de la spectroscopie infrarouge (IR), ou encore une transition électronique, s'il s'agit d'une radiation visible ou ultraviolet, ce qui est la base de la spectroscopie d'absorption ultraviolet et visible (UV-Vis). Noter que l'excitation par un photon visible peut s'accompagner d'un phénomène de diffusion inélastique (spectroscopie Raman) caractérisant des mouvements vibrationnels.

Outre, l'aspect lié à la préparation des échantillons et à l'instrumentation en vue de l'acquisition des spectres, l'étape finale de l'analyse spectroscopique est l'interprétation qualitative et quantitative des spectres.

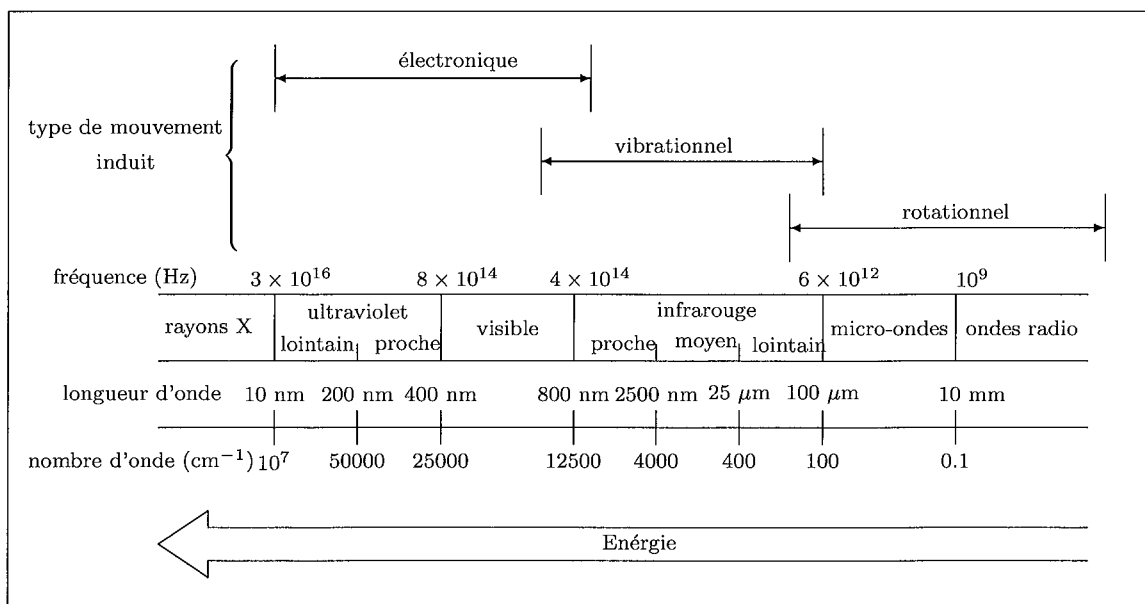


FIG. 6.1 – Régions du spectre électromagnétique présentant un intérêt pour l'analyse spectroscopique et type de mouvements associés.

### 6.1.2 Interprétation qualitative des spectres

L'analyse qualitative consiste à déterminer à partir d'un spectre mesuré les liaisons chimiques et groupements fonctionnels qui constituent la matière ou la molécule analysée. Le spectre d'une molécule

est une superposition des bandes caractéristiques de toutes les liaisons où chaque bande particulière est caractérisée par son intensité, sa fréquence centrale, sa forme et sa largeur. Partant de ce modèle d'observation et d'un point de vue physico-chimique, le problème inverse d'interprétation qualitative des spectres est un problème d'identification de la structure moléculaire de la substance analysée. Pour résoudre ce problème, une ou plusieurs techniques spectroscopiques peuvent être employées. L'analyse du spectre mesuré, consiste à inférer, à partir de la mesure des bandes spectrales importantes et des connaissances théoriques sur les propriétés d'absorption de la plupart des liaisons et groupements fonctionnels, la structure moléculaire la plus plausible. En utilisant cette approche, le spectroscopiste est généralement capable de déduire des informations précises sur la structure des molécules analysées. En fonction de la technique spectroscopique, un protocole d'analyse des raies les plus significatives permettra d'atteindre cet objectif [BD00, Hol98]. Afin de simplifier l'interprétation, l'analyse peut être combinée avec une méthode de traitement des signaux pour la détection et la caractérisation des raies spectrales. Nous pouvons consulter à ce propos [Dje03] pour le cas de signaux RMN et [Maz05] pour le cas de spectres Raman et Infrarouge.

### 6.1.3 Interprétation quantitative des spectres

Lorsque des radiations lumineuses entrent en contact avec la matière, elles peuvent aussi bien être absorbées que transmises ou réfléchies. Selon le principe de conservation de l'énergie, l'intensité des radiations incidentes  $I_0$  doit être égale à la somme des intensités des radiations absorbées  $I_A$ , transmises  $I_T$  et réfléchies  $I_R$ . L'application de ce principe en spectroscopie permet de calculer les proportions d'énergie lumineuse absorbée par les molécules à partir des résultats de mesure de l'intensité transmise  $I_T$ , ou de l'intensité réfléchie  $I_R$ . Considérons les conditions selon lesquelles un faisceau de lumière monochromatique de longueur d'onde donnée traverse un échantillon d'épaisseur  $d$  et de concentration  $C$  en substance absorbant spécifiquement la lumière du faisceau. Du fait de l'absorption de la lumière par l'échantillon, l'intensité du faisceau est atténuée depuis la valeur  $I_0$  à l'entrée de l'échantillon jusqu'à la valeur  $I_T$  à la sortie.

**Définition 6** *La transmittance est définie comme étant la fraction d'énergie lumineuse traversant l'échantillon de part en part. Elle est égale au rapport d'intensité  $I_T/I_0$ .*

**Définition 7** *L'absorbance est définie comme étant le logarithme décimal de l'inverse de la transmittance  $T$  :  $A = \log I_0/I_T$ .*

Si on suppose que l'intensité du faisceau lumineux diminue d'autant plus que la concentration de la substance absorbante est élevée et que l'échantillon est homogène, on peut écrire

$$A = \log(1/T) = \varepsilon \cdot d \cdot c, \quad (6.1)$$

où  $A$  est l'absorbance et  $\varepsilon$  est l'absorptivité de l'échantillon [BD00]. Cette équation  $A = \varepsilon \cdot d \cdot c$  est connue sous le nom de la loi de Beer-Lambert. On peut distinguer la loi de Beer de la loi de Lambert en considérant séparément l'effet de la concentration  $c$  de celui de l'épaisseur  $d$  sur l'absorbance  $A$ . La loi de Beer concerne uniquement l'effet de la concentration et s'exprime par la formule  $A = c \cdot s$

dans laquelle  $s$  est le produit de  $d$  et  $\varepsilon$ . De son côté, la loi de Lambert, concerne uniquement l'effet de l'épaisseur  $d$  sur l'absorbance lorsque  $c$  ne varie pas. Elle s'exprime par la relation  $A = k' \cdot d$ , dans laquelle  $k' = \varepsilon \cdot c$ . La loi de Beer qui énonce une relation linéaire simple entre l'absorbance d'un échantillon et la concentration d'analyte selon l'équation  $A = k \cdot c$  constitue la base physique et mathématique de l'analyse chimique quantitative par spectrophotométrie d'absorption. Lorsque les échantillons analysés sont de simples mélanges de deux constituants dont, l'un absorbe la lumière du faisceau et l'autre n'absorbe pas la lumière dans la gamme de longueur d'onde de ce faisceau, le calibrage du spectrophotomètre peut être réalisé grâce à la loi de Beer. La loi de Beer est donc l'expression la plus simple du principe de proportionnalité entre l'absorbance et la concentration de molécules absorbant la lumière. Cette loi est applicable dans toute la gamme spectrale d'analyse et donc pour toutes les spectroscopies d'absorption (UV-Vis, MIR, PIR). En spectroscopie de diffusion Raman, un modèle linéaire similaire à la loi de Beer peut être utilisé. L'intensité diffusée est linéairement proportionnelle à la concentration de la substance diffusante.

Cependant, la majorité des substances chimiques sont des liquides ou des solides complexes composés de nombreuses molécules de structures distinctes qui absorbent la lumière simultanément avec des absorptivités différentes à plusieurs longueurs d'ondes du spectre. Du fait de ce recouvrement, l'exploitation quantitative des spectres devient beaucoup plus compliquée dans le cas de substances où plusieurs composantes absorbent simultanément la lumière incidente. On s'aperçoit donc que, pour l'analyse de ce type de substances, il faut recourir au principe d'additivité des absorbances, en plus de la loi de proportionnalité de Beer. Pour analyser des mélanges de plusieurs constituants, il faut utiliser des données à plusieurs longueurs d'ondes de façon à pouvoir corriger l'effet d'interférence spectrale entre les différents constituants. Ainsi l'absorbance d'une substance de  $p$  constituants absorbants à une longueur d'onde  $\lambda$  s'écrit sous la forme

$$A(\lambda) = \sum_{j=1}^p c_j s_j(\lambda) \quad (6.2)$$

où  $C_j$  est la concentration du  $j$ ème constituant et  $s_j$  est égal au produit du coefficient d'absorptivité et de l'épaisseur de l'échantillon. Cette formule définit l'équation fondamentale pour l'interprétation des spectres issus de l'analyse de substances à plusieurs constituants.

La linéarité exprimée par la loi de Beer est limitée pour des raisons instrumentales ou chimiques. Pour des fortes concentrations, la distance entre molécules absorbantes diminue et des interactions électrostatiques entre molécules apparaissent. Cette interaction moléculaire peut modifier l'absorptivité. D'autre part, des déviations par rapport à la loi de Beer peuvent être observées dans le cas d'une diffraction ou diffusion du faisceau lumineux à l'intérieur de la substance. Toutes ces imperfections du modèle peuvent être prises en compte par un modèle plus général

$$A(\lambda) = \sum_{j=1}^p c_j s_j(\lambda) + E(\lambda) \quad (6.3)$$

où  $E(\lambda)$  représente la déviation de l'absorbance par rapport à la loi de Beer, à la longueur d'onde  $\lambda$ .

### 6.1.4 Application de la séparation de sources à l'analyse de mélanges

L'équation 6.2 établit la relation directe entre l'absorbance d'une substance à plusieurs composantes et les absorbances individuelles de ces composantes. Dans le cas où ces composantes et leurs concentrations ne sont pas connues, l'analyse de cette substance à partir d'un seul spectre est quasiment impossible, sauf si les spectres des différentes espèces ne se recouvrent pas. La procédure usuelle pour l'analyse est l'acquisition de plusieurs spectres dans des conditions physico-chimiques variables, de telle sorte à faire évoluer différemment les concentrations de toutes les composantes. La caractérisation des composantes de la substance et la détermination de leurs proportions à partir de tous les spectres mesurés est un problème de séparation de sources qui, en chimométrie, est connu sous le nom de "résolution de courbes" (*curve resolution*) [LS71], "analyse de mélanges" (*mixture analysis*) [WG91] ou encore "analyse factorielle" (*factor analysis*) [Mal02]. Par ailleurs, compte tenu des contraintes de non-négativité des sources et des coefficients de mélange, l'interprétation qualitative et quantitative d'une collection de spectres est un problème de séparation de sources non-négatives.

## 6.2 Application à des données expérimentales

Cette section présente quelques exemples d'application des méthodes de séparation de sources non-négatives à l'analyse de quelques mélanges synthétisés à partir de spectres réels. Ces simulations permettent de tester les méthodes en essayant de résoudre un problème d'analyse de mélanges dans lesquels les spectres purs et leurs proportions dans les observations sont connus au préalable.

### 6.2.1 Signaux de spectroscopie Raman

#### 6.2.1.1 Synthèse des mélanges

Dans cette expérience, nous avons considéré quatre sources spectrales qui correspondent aux spectres de diffusion Raman de quatre composés organiques qui possèdent des structures moléculaires très proches (amaltose, fructose, lactose et ribose). Les spectres Raman de ces molécules sont montrés sur la figure 6.2 et nous pouvons constater que ces spectres possèdent des formes très similaires.

A la lecture de la matrice de corrélation empirique de ces sources,

$$\hat{\mathbf{R}}_s = \begin{bmatrix} 1.00 & 0.35 & 0.55 & 0.59 \\ 0.35 & 1.00 & 0.38 & 0.50 \\ 0.55 & 0.38 & 1.00 & 0.45 \\ 0.59 & 0.50 & 0.45 & 1.00 \end{bmatrix}, \quad (6.4)$$

nous pouvons remarquer une corrélation mutuelle très importante, ce qui rend la séparation de ce type de signaux très difficile. D'autre part, les profils d'évolution des coefficients de mélange, montrés dans la figure 6.3, sont choisis de sorte que les conditions nécessaires pour l'unicité de la solution ne soient pas satisfaites. Les observations sont obtenues en considérant un modèle de mélange linéaire dans lequel un bruit gaussien i.i.d et de moyenne nulle est ajouté à chaque observation pour un RSB de 30 dB.

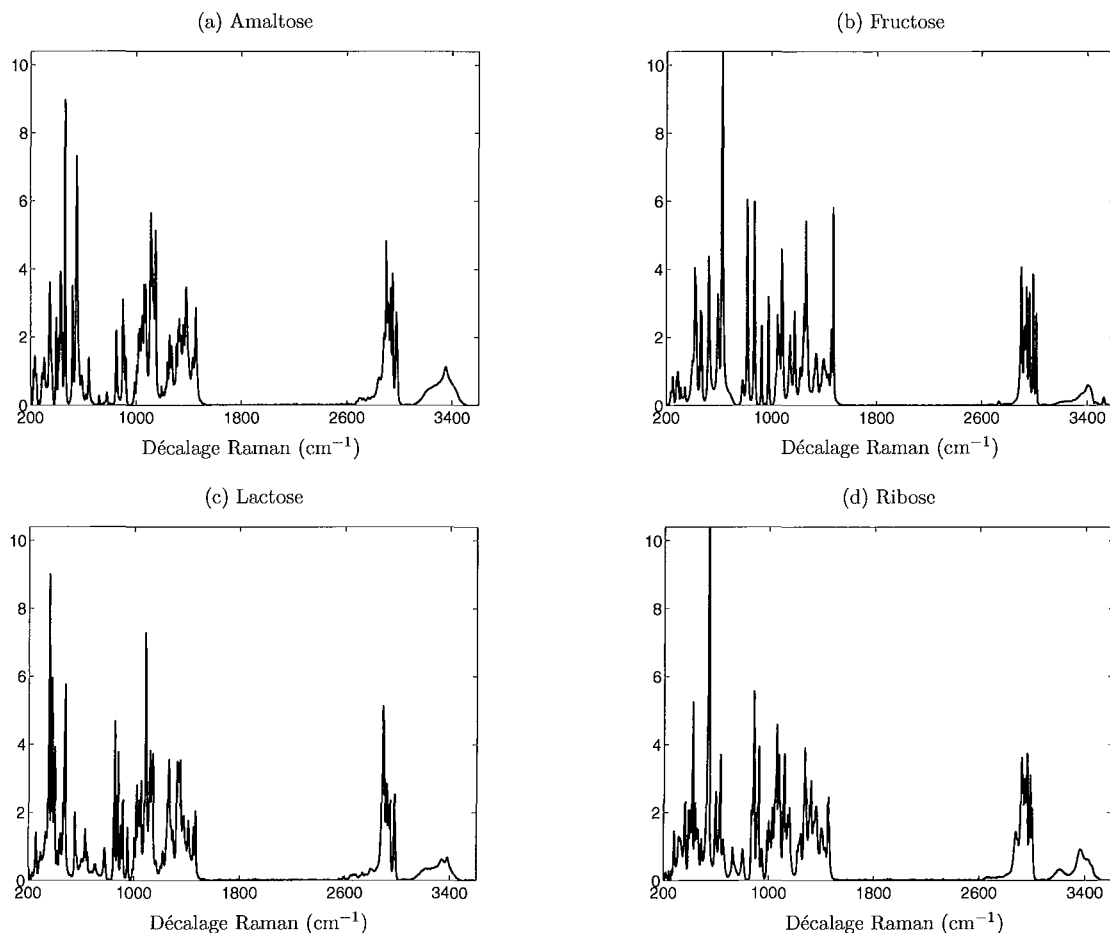


FIG. 6.2 – Spectres Raman des quatre composés organiques ( $p=4$  et  $n = 3400$ ).

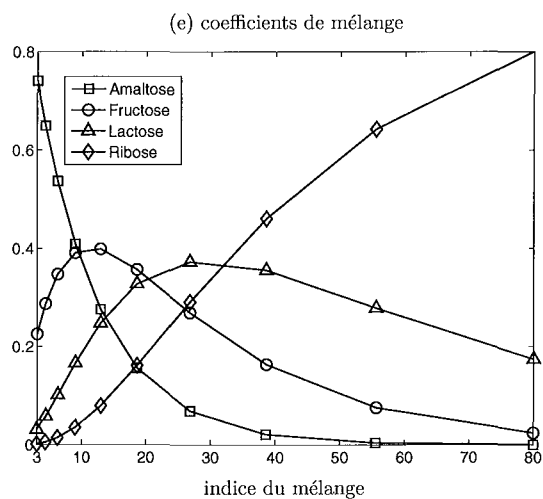


FIG. 6.3 – Profils d'évolution des coefficients de mélange ( $m=10$ ).

### 6.2.1.2 Analyse des résultats de la séparation

Le tableau 6.1 résume les performances des différentes approches appliquées à l'analyse de ces signaux. Compte tenu de la forte corrélation mutuelle des sources, les méthodes FastICA et NNICA fournissent



des composantes indépendantes qui sont différentes des sources recherchées. Une amélioration des performances est observée lorsque le résultat de la séparation par la méthode FastICA est utilisé pour initialiser la méthode ALS. Néanmoins, nous pouvons constater que la méthode BPSS fournit les meilleurs résultats que ce soit en termes de reconstruction des spectres purs ou d'estimation des coefficients de mélange.

	FastICA	FastICA-ALS	NMF	NNICA	NNSC	BPSS
$RSD_{\text{amaltose}}$	8.81	23.71	34.63	8.95	20.55	30.72
$RSD_{\text{fructose}}$	8.43	19.88	9.50	9.8	17.72	24.16
$RSD_{\text{lactose}}$	9.96	16.49	11.42	8.52	14.28	21.13
$RSD_{\text{ribose}}$	5.91	15.90	27.88	6.95	13.4	28.01
$IP$	-11.11	-18.83	-11.12	-11.45	-16.35	-25.74

TAB. 6.1 – Comparaison des performances de quelques méthodes de séparation pour un RSB de 30 dB.

## 6.2.2 Signaux de spectroscopie Infrarouge

### 6.2.2.1 Synthèse des mélanges

Afin de réaliser une première expérience dans laquelle les spectres purs et les coefficients de mélange sont connus mais les observations sont mesurées à l'aide d'un spectromètre, nous avons opté pour un mélange d'alcane. Avant qu'on ne les appelle les alcanes, cette famille de produits organiques était appelée paraffine (*para* : sans, *affinis* : affinité) : les molécules sans affinité, qui ne réagissent pas. On les destinait donc à l'utilisation combustible. D'un point de vue structure atomique, les alcanes sont la famille des molécules composées uniquement de carbone et d'hydrogène. On distingue généralement les molécules linaires et ramifiées (alcane) des molécules cycliques (cycloalcanes). Cette expérience est réalisée comme suit : trois alcanes (cyclopentane, cyclohexane et n-pentane) sont mélangés dans des proportions différentes. Ensuite, les spectres d'absorption des différents mélanges sont enregistrés avec un spectromètre infrarouge. Ces espèces sont des solvants inertes, ce qui fait qu'aucune réaction n'a lieu dans les mélanges. Par conséquent, les spectres purs sont mesurés individuellement et, de plus, les proportions de chaque espèce seront connues avec exactitude.

La figure 6.4 montre les spectres purs normalisés (variance unité) des trois espèces et les coefficients de mélange utilisés. Les spectres infrarouge de ces trois alcanes sont constitués de bandes spectrales très fines et en fort recouvrement car toutes ces molécules contiennent des liaisons de type C-H et C-C, dont on observe les harmoniques de vibration et bandes de combinaison (couplage entre les modes de vibration) dans le proche infrarouge. De plus, un léger arrière-plan peut être observé sur chacune des sources. Toutes ces remarques expliquent l'origine de la forte corrélation entre ces spectres purs,

$$\mathbf{R}_s = \begin{bmatrix} 1.00 & 0.77 & 0.84 \\ 0.77 & 1.00 & 0.86 \\ 0.84 & 0.86 & 1.00 \end{bmatrix}. \quad (6.5)$$

La figure 6.5 montre les observations dans la gamme spectrale (3000-6000  $\text{cm}^{-1}$ ).

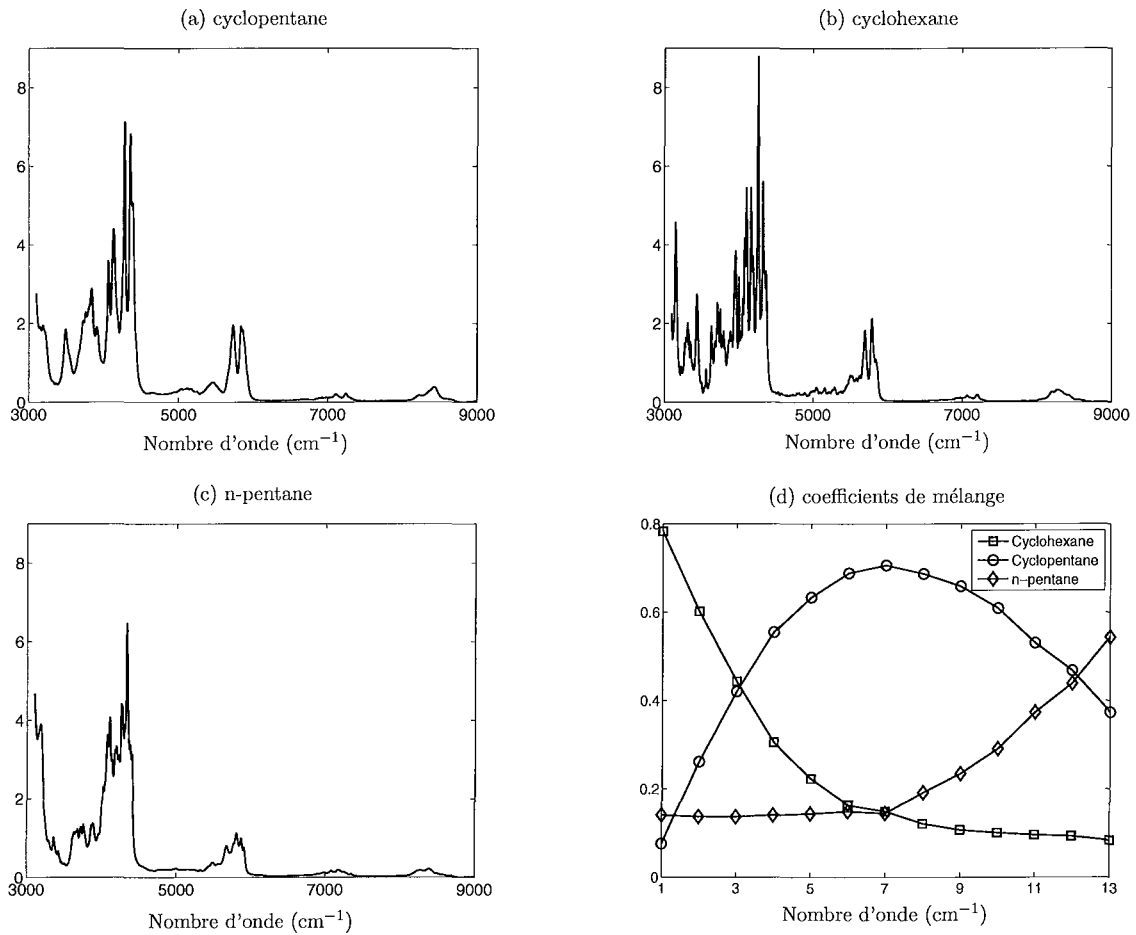


FIG. 6.4 – Spectres infrarouge des trois alcanes ( $p = 3$  et  $n = 5200$ ).

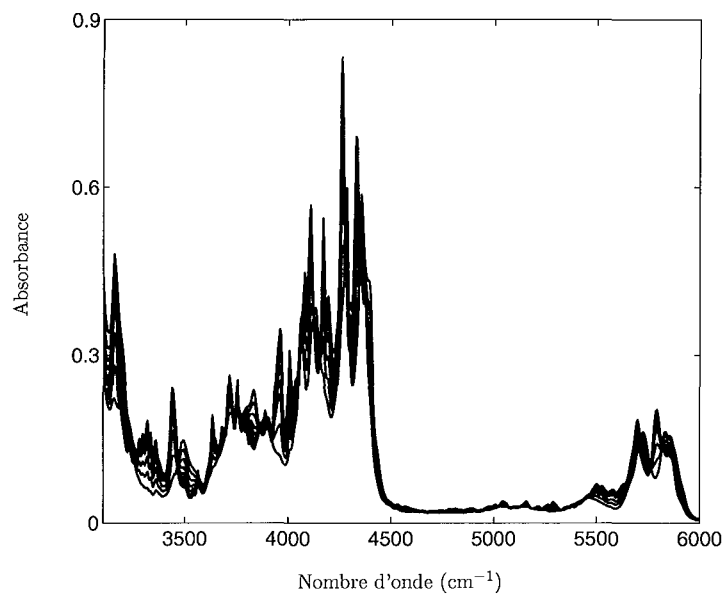


FIG. 6.5 – Spectres infrarouges mesurés dans la gamme spectrale (3000-6000 cm<sup>-1</sup>).

### 6.2.2.2 Analyse et interprétation des résultats de la séparation

Le tableau 6.2 donne un récapitulatif des performances des approches utilisées, en termes d'indice de performance global et d'erreur de reconstruction des spectres purs. De façon similaire au cas précédent, nous pouvons constater que les méthodes qui imposent l'orthogonalité produisent des composantes indépendantes différentes des composantes recherchées, alors que l'approche bayésienne produit un résultat nettement meilleur. Signalons, également que l'application de l'approche bayésienne nécessite un nombre d'itérations très important (de l'ordre de  $10^6$ ) dans ce cas.

	FastICA	FastICA-ALS	NMF	NNICA	NNSC	BPSS
$\mathcal{RSD}_{\text{cyclopentane}}$	6.90	14.20	15.18	4.82	10.69	33.23
$\mathcal{RSD}_{\text{cyclohexane}}$	5.06	17.50	23.43	5.65	15.35	24.98
$\mathcal{RSD}_{\text{n-pentane}}$	3.52	17.88	14.01	4.77	13.98	26.05
$IP$	-6.17	-11.60	-8.10	-5.77	-10.64	-19.22

TAB. 6.2 – Comparaison des performances de quelques algorithmes de séparation.

Si on se place dans la situation où la composition chimique des mélanges synthétisés dans cette expérience est inconnue, il faut alors une étape d'interprétation des résultats en vue de l'identification des constituants et la détermination de leurs proportions dans les mélanges. L'objectif de notre interprétation de ces résultats n'est pas de décrire la procédure à suivre pour l'identification mais plutôt l'étude de la similarité des sources reconstruites avec les spectres des trois alcanes. La même étude est réalisée sur les coefficients de mélange.

La figure 6.6 donne une comparaison visuelle entre les sources reconstruites et les spectres des trois alcanes sur deux sous bandes où les trois alcanes possèdent de fortes absorbances. Nous pouvons constater la forte similarité entre les spectres des alcanes et les sources estimées qui permet de réaliser l'identification aisée des espèces. Une comparaison entre les coefficients de mélange estimés et les proportions des trois alcanes dans les mélanges est montrée sur la figure 6.7 où nous pouvons constater un faible écart d'estimation. Nous pensons que cet écart est dû à la présence de l'arrière-plan sur les sources et à l'erreur de modélisation. Néanmoins, les allures des profils d'évolution des concentrations estimées sont correctes, ce qui garantit une interprétation qualitative cohérente mais une caractérisation quantitative légèrement biaisée.

## 6.3 Application à des problèmes d'analyse physico-chimique

Dans cette section nous présentons deux applications de la séparation de sources non-négatives au traitement de problèmes d'analyse physico-chimique à l'aide des techniques spectroscopiques. Avant d'appliquer une méthode de séparation de sources, une étape de pré-traitement des données est nécessaire.

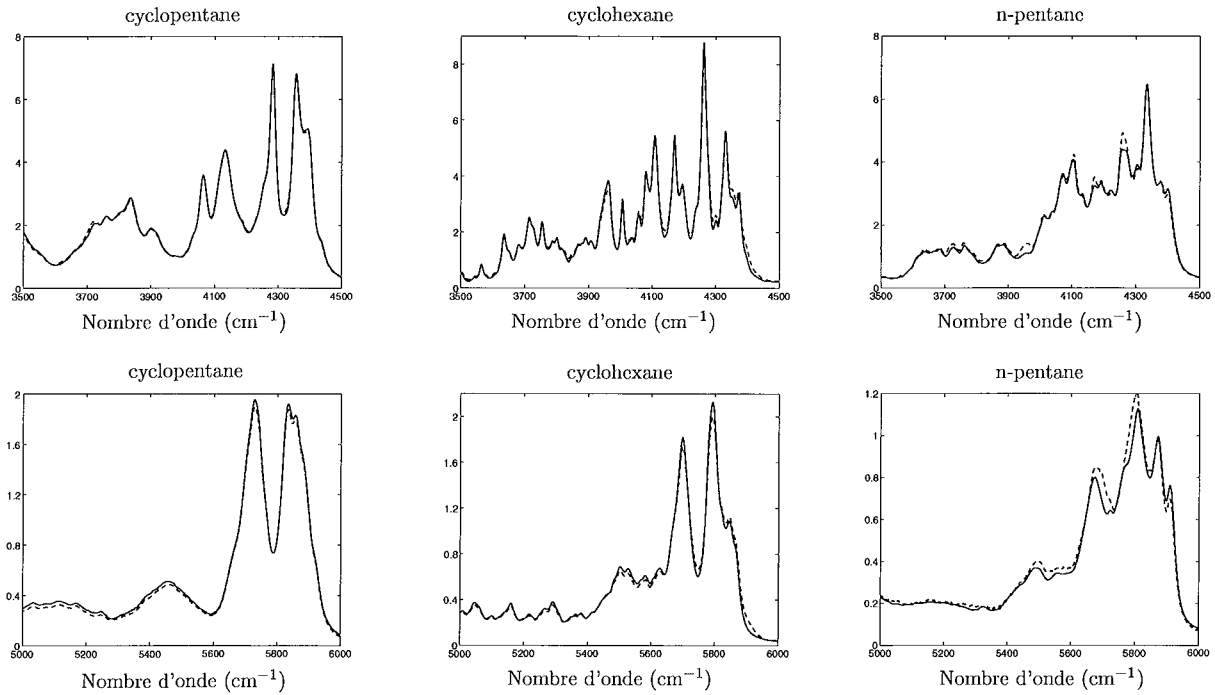


FIG. 6.6 – Comparaison visuelle entre les sources reconstruites (continu) et les spectres purs (discontinu) sur deux régions spectrales).

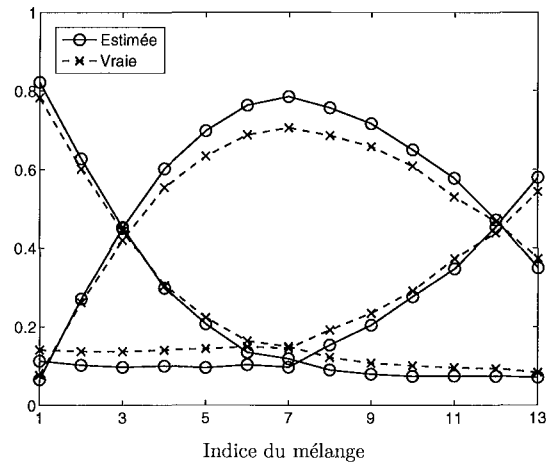


FIG. 6.7 – Comparaisons entre les coefficients de mélange estimés (discontinu) et les proportions des trois alcanes dans les mélanges.

### 6.3.1 Méthodologie de traitement

#### 6.3.1.1 Pré-traitements des données

Les pré-traitements concernent essentiellement la suppression d'un éventuel arrière-plan qui se rajoute aux observations. Cette étape fait l'hypothèse que le spectre observé est la somme d'un signal utile contenant l'information spectrale et d'un signal de variation incontrôlée (fond, arrière-plan). La suppression consiste donc à modéliser cet arrière-plan et à le soustraire du signal observé. En général,

l'arrière plan est représenté par un polynôme, dont l'estimation se fait à partir de quelques points du spectre mesuré pour lesquels on n'observe pas de bande d'absorption. Ces points sont ensuite utilisés pour estimer les coefficients du polynôme. Nous pouvons consulter à ce propos [MCB<sup>+</sup>05], où des méthodes de sélection des points et d'estimation des coefficients du polynôme sont présentées. Un deuxième pré-traitement qui est aussi important est celui de la mise à l'échelle des données [BS03]. La mise à l'échelle est réalisée par normalisation de chaque spectre mesuré à une variance unité. Autrement dit les données brutes sont transformées selon

$$\tilde{\mathbf{X}} = \mathbf{K}^{-1} \mathbf{X}, \quad (6.6)$$

où  $\mathbf{K} = \text{diag}([\sigma_1, \dots, \sigma_m])$  et  $\{\sigma_i\}_{i=1}^m$  sont les variances des différentes observations. Nous pouvons signaler que cette opération n'affecte ni la linéarité du modèle, ni la forme des sources à reconstruire. L'algorithme de séparation est alors appliqué aux données  $\tilde{\mathbf{X}}$ .

### 6.3.1.2 Traitement des résultats

Cette étape de post-traitement concerne la prise en compte de la normalisation des observations. Afin de décrire ces opérations, rappelons que le résultat de la séparation à partir des données transformées permet d'écrire

$$\tilde{\mathbf{X}} \approx \tilde{\mathbf{A}} \mathbf{S} \quad (6.7)$$

où les sources contenues dans la matrice  $\mathbf{S}$  sont supposées de variance unité afin d'éliminer l'indétermination d'échelle. Ainsi les données brutes seront exprimées selon

$$\mathbf{X} \approx \mathbf{A} \mathbf{S} \quad (6.8)$$

où  $\mathbf{A} = \mathbf{K} \tilde{\mathbf{A}}$ .

Par ailleurs, afin de déterminer la proportion de chaque composante  $j$  dans les mélanges  $i$  est obtenue par

$$C_{ij} = \frac{a_{ij}}{\sum_{\ell=1}^p a_{i\ell}}. \quad (6.9)$$

Finalement, après séparation et post-traitements, les observations sont exprimées selon le modèle

$$\mathbf{X} \approx \mathbf{N} \mathbf{C} \mathbf{S} \quad (6.10)$$

où  $\mathbf{N}$  est une matrice diagonale dont les éléments sont donnés par  $\left\{ N_{ii} = 1 / \sum_{j=1}^p a_{ij}; i = 1, \dots, m \right\}$  et  $\mathbf{C}$  contient les proportions de chaque source estimée dans les observations.

## 6.3.2 Analyse de l'hydratation de matériaux poreux et/ou lamellaires

### 6.3.2.1 Problématique

Certains solides présentent des espaces dont au moins une dimension est de l'ordre de quelques nanomètres. Dans cette définition d'espace entrent des solides non déformables poreux (silices) aussi

bien que des systèmes gonflants (argiles). L'eau confinée dans une telle matière solide présente des propriétés différentes de celles du liquide. Ceci a notamment pour conséquence de perturber les propriétés macroscopiques classiques d'acido - basicité, d'échange d'ions, de solvatation d'espèces chargées ou non, d'oxydoréduction et d'adsorption - désorption aux surfaces. De plus, le statut de l'eau dans un milieu poreux et/ou lamellaire conditionne l'interprétation d'expériences d'altération géochimique, de rétention et de thermo - hydro - mécanique (THM). De même, les modélisations géochimiques nécessitent la connaissance du comportement physico-chimique de l'eau dans les ouvrages en argiles de forte densité : la notion "d'eau liée" (molécules d'eau en interaction avec le solide) et "d'eau libre" (molécules d'eau autoassociées). Enfin, pour répondre aux préoccupations liées à la diffusion des polluants à travers des milieux poreux, il est fondamental de décrire correctement les états de l'eau, vecteur des transports. L'adsorption contrôlée d'eau permet de caractériser la morphologie du matériau, l'hétérogénéité de sa surface, ses propriétés hydrophiles/hydrophobes. La mesure des isothermes d'adsorption - quantité d'eau adsorbée par le matériau en fonction de l'humidité relative - constitue une bonne méthode macroscopique pour suivre l'hydratation. Une meilleure compréhension de ces équilibres dans de tels systèmes nécessite des approches d'analyse à l'échelle moléculaire. Susceptible d'interagir avec des sites spécifiques, l'eau se révèle une excellente sonde moléculaire pour révéler des hétérogénéités à l'échelle nanométrique d'une surface d'un solide. Au cours de trois des dernières thèses de l'équipe de Chimie et Spectrochimie des Interfaces du Laboratoire de Chimie Physique et Microbiologie pour l'Environnement (LCPME) ont été développés des systèmes expérimentaux couplant spectrométries vibrationnelles infrarouge et/ou Raman à des mesures d'isothermes d'adsorption d'eau ou d'humidité relative [Car99b, Rin04].

### 6.3.2.2 Acquisition des données

Nous reprendrons dans ce rapport de thèse des données acquises dans le domaine spectral du proche infrarouge par Carteret [Car99b] et Rinnert [Rin04] sur des silices poreuses. Les séries de spectres dans le proche infrarouge sont obtenues en faisant varier, pour un même échantillon de silice, l'humidité relative ou taux d'hydratation (variable  $P/P_0$ ) dans la cellule d'analyse. La silice passe progressivement d'un état sec ( $P/P_0 = 0$ ) à un état saturé en eau ( $P/P_0 = 1$ ). Les travaux antérieurs ont permis de mettre en évidence différents types d'eau lors du processus d'hydratation. Toutefois, le spectre de l'eau étant rendu complexe par la grande variété de sites il convient d'utiliser les méthodes de séparation de sources pour extraire à la fois les spectres caractéristiques et les profils de concentration des différents types d'eau. Les signaux obtenus pour différentes valeurs de  $P/P_0$  sont tracés dans la figure 6.8.

### 6.3.2.3 Traitement des données et interprétation des résultats

Les spectres des différents signaux sont analysés en utilisant la méthode BPSS. Les signaux sources et les coefficients de mélange estimés sont représentés sur la figure 6.9. Les coefficients de mélange sont exprimés en termes de proportions de chaque source dans les mélanges observés. Une première analyse qualitative des profils d'évolution des trois sources dans les mélanges permet de déduire que la première source estimée correspondrait à l'eau liée car c'est la première espèce qui apparaît après

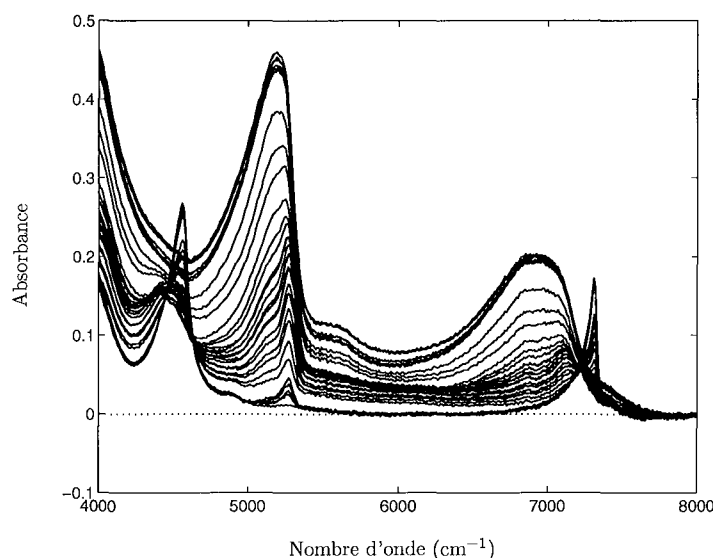


FIG. 6.8 – Observations enregistrées pour plusieurs valeurs du  $P/P_0$ . ( $m = 33$  et  $n = 4000$ ).

rajout de l'eau. La deuxième source estimée correspondrait au support, c'est à dire la silice pure, car sa proportion dans les mélanges diminue au fur et à mesure que l'eau est rajoutée. La troisième source estimée correspondrait à l'eau libre car celle-ci est majoritaire en fin d'hydratation (saturation en eau). Il est à noter que le décrochement des proportions de l'eau liée à partir de  $P/P_0 = 0.6$  s'explique par la proportion d'eau libre qui devient très importante et l'eau à la surface de la silice devient donc en interaction avec le support mais également avec les couches d'eau supérieures. Cette pression relative est exploitée lors de l'analyse du phénomène d'hydratation. Par ailleurs, l'interprétation des sources estimées se focalise sur les différentes bandes spectrales. La deuxième source fait apparaître essentiellement deux pics de forte intensité aux nombre d'ondes  $4560, 7310 \text{ cm}^{-1}$  et un pic de faible intensité à  $5270 \text{ cm}^{-1}$ . Les deux premiers pics sont caractéristiques des groupements de surface de la silice (groupements silanol, SiOH), alors que le troisième correspond à une erreur d'estimation. La troisième source estimée possède essentiellement deux motifs spectraux très larges et de forte intensité localisés autour des nombres d'onde  $5170$  et  $6900 \text{ cm}^{-1}$  et une bande moins intense à  $5600 \text{ cm}^{-1}$ . Ces bandes spectrales sont caractéristiques de l'eau libre (autoassociation forte des molécules d'eau). La première source estimée présente plusieurs pics aux nombre d'onde  $4410, 5150, 6250, 5300, 6840, 6920$  et  $7100 \text{ cm}^{-1}$ . Toutes ces bandes de vibrations sont les bandes de combinaison (couplage) des vibrations dues à l'interaction entre les groupements silanol (SiOH) et l'eau (OH). C'est donc la source caractéristique de l'interaction de l'eau avec la silice. La bande spectrale se trouvant à  $4410 \text{ cm}^{-1}$  caractérise les silanol perturbés par l'eau alors que les nombres d'onde  $5160, 5250, 5300, 6880$  et  $7100 \text{ cm}^{-1}$  sont l'empreinte spectrale des molécules d'eau en interaction avec des groupements silanol. Nous signalons que les résultats de cette analyse et notamment les valeurs des pics caractéristiques sont en accord avec les analyses présentées par Carteret [Car99b]. Cependant, la contribution la plus significative de cette analyse est l'obtention des spectres complets des systèmes "eau libre" et "eau liée" ainsi que la description des profils de concentrations en fonction du taux d'hydratation de la

silice, ce qui n'avait jamais été fait jusqu'à présent. Ce résultat est très important car il permet de distinguer le spectre de l'eau liée de celui de l'eau libre, malgré le fort recouvrement entre les bandes spectrales caractéristiques de ces deux espèces.

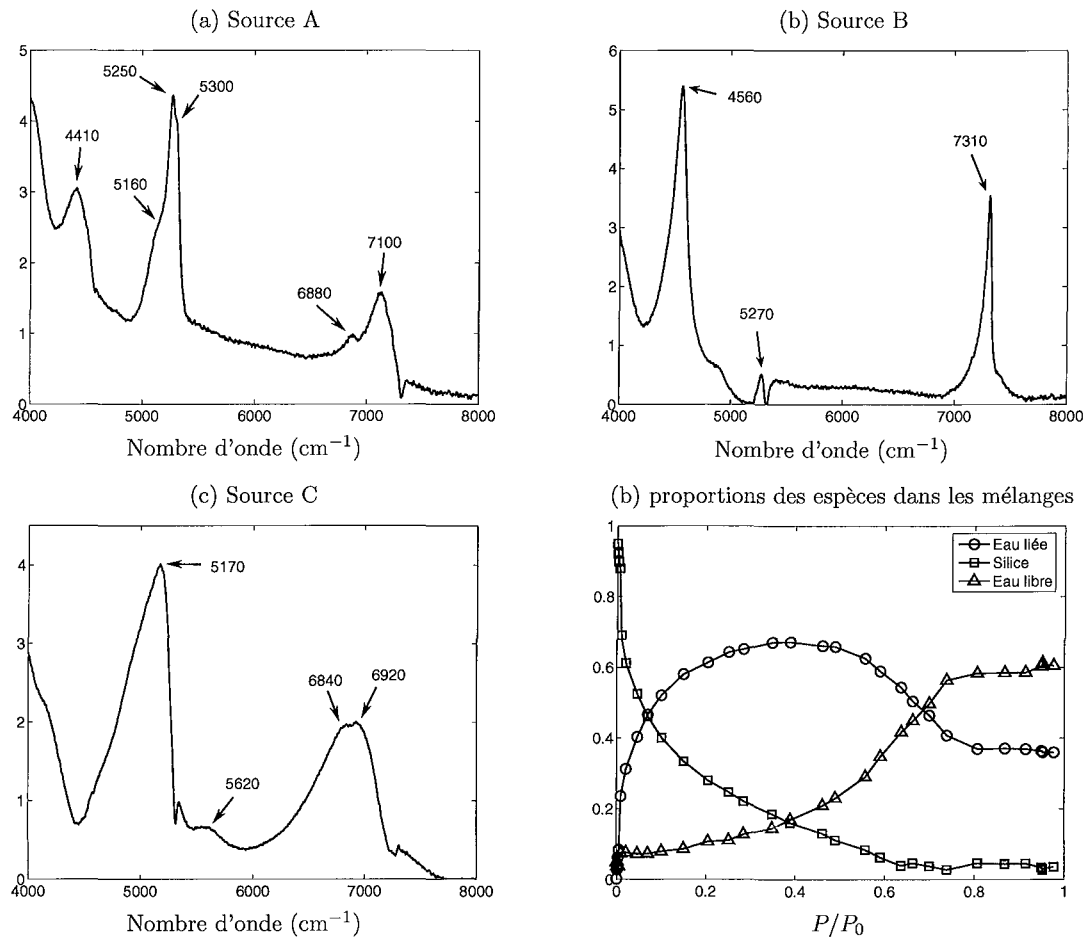


FIG. 6.9 – Résultats de la séparation en termes de spectres purs et proportions des différentes sources dans les mélanges.

### 6.3.3 Analyse des polymorphes du carbonate de calcium

#### 6.3.3.1 Problématique

Le carbonate de calcium ( $\text{CaCO}_3$ ) a suscité beaucoup d'attention dans la littérature en raison de ses nombreuses applications dans des domaines aussi variés que l'industrie du papier, l'industrie pharmaceutique et alimentaire. Il peut se présenter sous trois formes polymorphiques, qui sont dans l'ordre de stabilité thermodynamique croissante : vatérite, aragonite, calcite. Les propriétés physiques du carbonate de calcium, utilisé comme additif, sont fortement liées à la quantité de chaque polymorphe présent dans le mélange. C'est pourquoi une méthodologie de détermination quantitative des formes polymorphiques est essentielle. Par ailleurs, le suivi quantitatif des transformations de phase est une information importante permettant d'élucider les mécanismes de formation et de dissolution des polymorphes du carbonate de calcium. Trois techniques d'analyse permettent de caractériser les différents



polymorphes. En spectroscopie Infra Rouge (IR), il est possible de différencier les trois polymorphes par leur mode de vibration situé à bas nombre d'onde ( $< 400 \text{ cm}^{-1}$ ), qui ne sont pas faciles d'accès en IR classique. Par conséquent, la spectroscopie Raman apparaît comme l'outil le plus adapté pour l'analyse de la composition d'un mélange de polymorphes sur un procédé. Cependant, la quantification précise reste encore délicate selon les systèmes polymorphiques considérés. De façon quantitative, pour  $\text{CaCO}_3$ , la spectroscopie Raman a été jusqu'ici principalement réservée aux mélanges de deux polymorphes. Lorsque les trois polymorphes sont présents, la superposition des bandes caractéristiques complique le traitement des spectres. Il est par conséquent difficile de trouver une méthode d'analyse qualitative et quantitative précise de mélanges de trois polymorphes en raison de la difficulté à différencier les différentes phases. L'objectif de ce travail est de développer une méthodologie d'exploitation des spectres afin de faire de la spectroscopie Raman la technique d'analyse quantitative la plus précise, fiable et rapide de la composition d'un mélange des trois polymorphes de carbonate de calcium. C'est donc associée aux méthodes de séparation de sources que la spectroscopie Raman devient une méthode analytique quantitative précise de détermination de la composition d'un mélange polymorphique. Les propriétés physiques du produit cristallisé dépendent en grande partie du pourcentage de chaque présent de polymorphe. Par conséquent, il est nécessaire de mesurer les concentrations relatives de ces polymorphes dans les mélanges. La formation des polymorphes de carbonate de calcium par le mélange de deux solutions contenant respectivement les ions calcium (solution  $\text{CaCl}_2$ ) et carbonate (solution  $\text{Na}_2\text{CO}_3$ ) se déroule en deux périodes distinctes. La première période est la précipitation. Cette période est très rapide (quelques minutes) et elle fournit un mélange des polymorphes de carbonate de calcium. La deuxième période (un processus lent de quelques heures à plusieurs jours) représente la transformation de phase des polymorphes instables vers le polymorphe le plus stable (calcite). Comprendre le mécanisme de leur transformation est d'une importance primordiale pour élucider des mécanismes de formation et de dissolution de carbonate de calcium. De plus, dans les conditions industrielles, la calcite est le produit désiré, ainsi un objectif du travail est d'étudier les mécanismes de la transformation de phase conduisant à la calcite.

### 6.3.3.2 Acquisition des données

Le spectre Raman est caractéristique de l'échantillon étudié. C'est pourquoi il est possible de distinguer les polymorphes entre eux. Le spectromètre Raman utilisé dans ce travail est un dispositif Jobin-Yvon T64000 couplé à une caméra CDD (Coupled Device Detector) refroidi à l'azote liquide et à un microscope confocal (objectif X50). L'échantillon est positionné sous le microscope qui focalise le rayon laser (550 nm) et qui collecte la lumière rétrodiffusée. Dans les conditions de travail choisies, le temps d'analyse est très court puisque l'acquisition d'un spectre se fait en moins de deux minutes. Afin de vérifier l'homogénéité des mélanges étalons, l'acquisition est réalisée sur au moins cinq régions différentes du même échantillon. Cependant, pour des mélanges obtenus au cours d'une synthèse par précipitation simultanée des trois polymorphes, il est inutile de faire l'acquisition sur autant de points parce que les mélanges sont relativement homogènes. Les expériences commencent par l'introduction de 100mL d'une solution contenant  $\text{Na}_2\text{CO}_3$  0.625 mol/L + NaCl 5 mol/L dans 2.5L d'une solution contenant  $\text{CaCl}_2$  0.025 mol/L + NaCl 5 mol/L introduite précédemment dans un

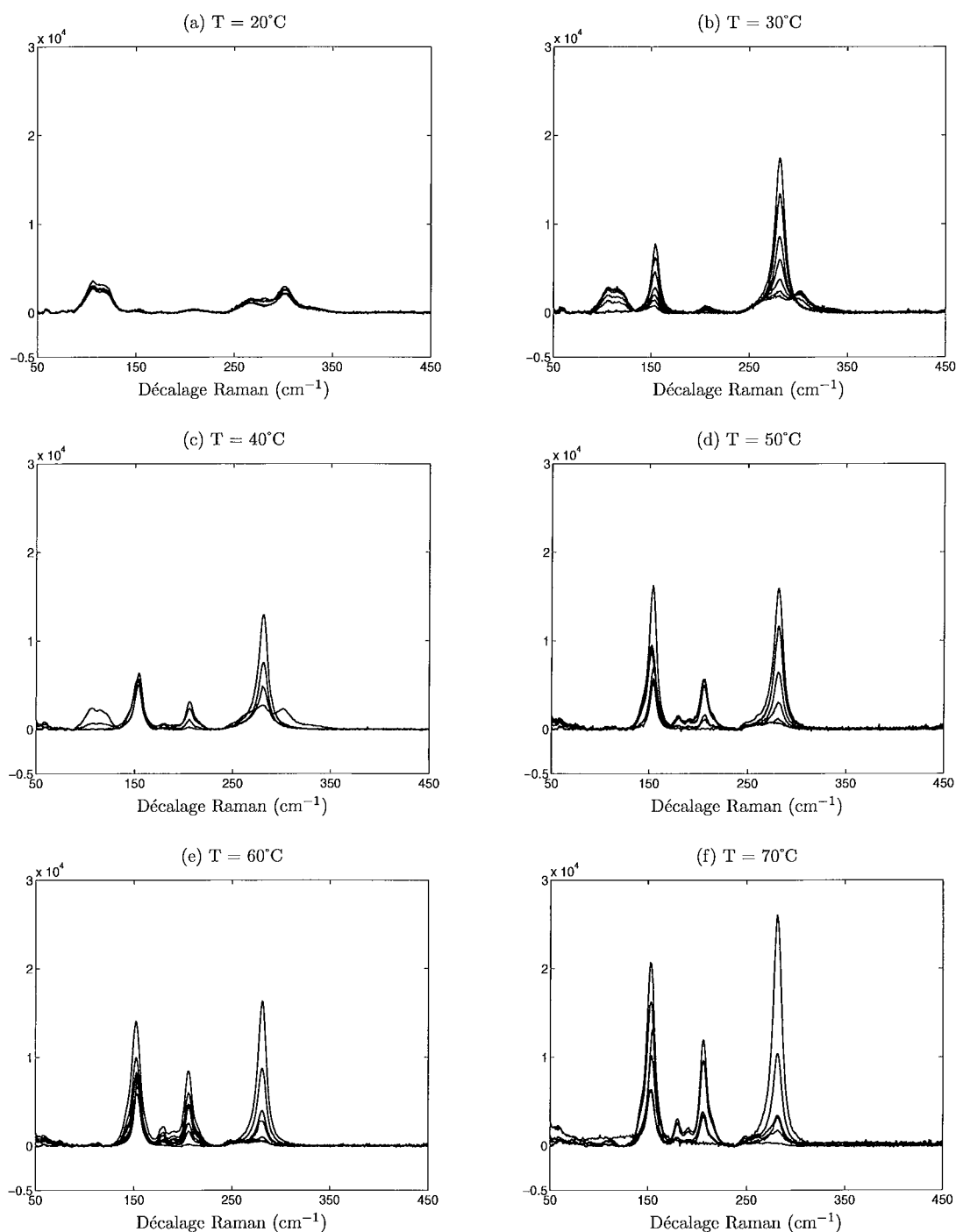


FIG. 6.10 – Spectres mesurés à plusieurs températures.

réacteur thermostaté. Après l'étape initiale de précipitation (correspondant à l'échantillon prélevé à 2 minutes), la transformation de phase est suivie par des prélèvements réguliers dans le temps. Chaque prélèvement est ensuite analysé par spectrométrie Raman. Afin de fournir une étude complète de ces polymorphes, nous avons également tenu compte d'un autre paramètre pouvant influençant les mécanismes : la température. Nous avons ainsi mené des expériences à 20, 30, 40, 50, 60 et 70°C et les spectres enregistrés sont montrés dans la figure 6.10.

### 6.3.3.3 Traitement des données et interprétation des résultats

Les spectres des différents mélanges sont analysés en utilisant la méthode BPSS. Les signaux sources et les coefficients de mélange estimés sont représentés dans la figure 6.11. Nous pouvons constater que le résultat obtenu fournit une séparation assez nette des spectres purs des différents polymorphes. Afin d'étudier la validité des résultats fournis par la méthode d'analyse, des mesures ont été effectuées de telle sorte à mesurer indépendamment les spectres purs des trois amorphes. La figure 6.12 montre une analyse de corrélation entre les sources reconstruites et les spectres purs. Nous pouvons constater la très forte corrélation entre les sources estimées et les spectres purs des trois amorphes.

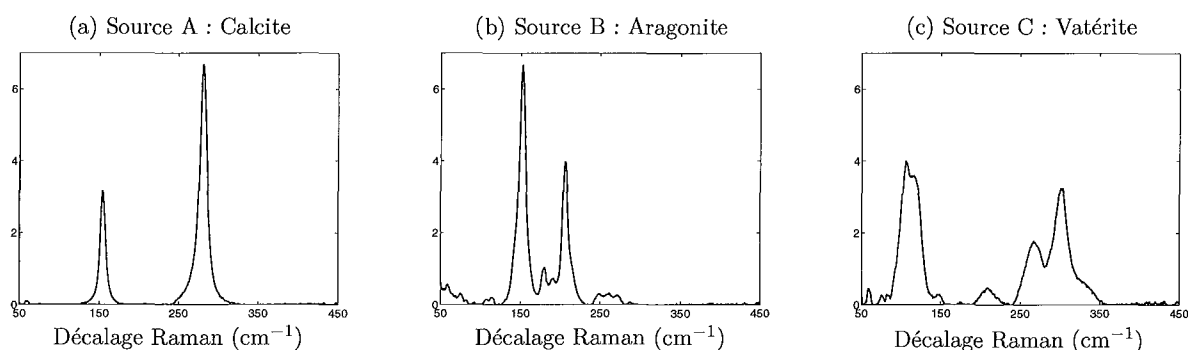


FIG. 6.11 – Spectres purs estimés par BPSS.

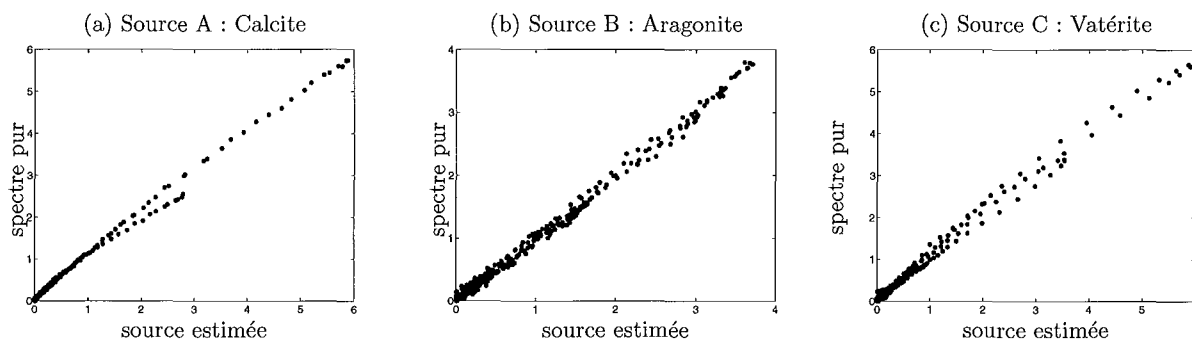


FIG. 6.12 – Analyse de corrélation entre sources estimées et spectres purs.

De plus, les profils des coefficients de mélange (figure 6.13) permettent de déduire les proportions de chaque espèce pour une température et un instant donnés. Ainsi, nous pouvons constater que la température joue un rôle très important sur le polymorphes formé après précipitation (2 min) : à 20 °C le  $\text{CaCO}_3$  précipite sous la forme de vatérite pure alors qu'à 70 °C l'aragonite pure est formée. Par ailleurs, le suivi au cours du temps, nous informe sur le processus de transformation de phase : le système évolue vers la phase cristalline la plus stable, la calcite. Par exemple à 40 °C, il faut une durée de 7 heures pour que le système évolue vers la calcite pure.

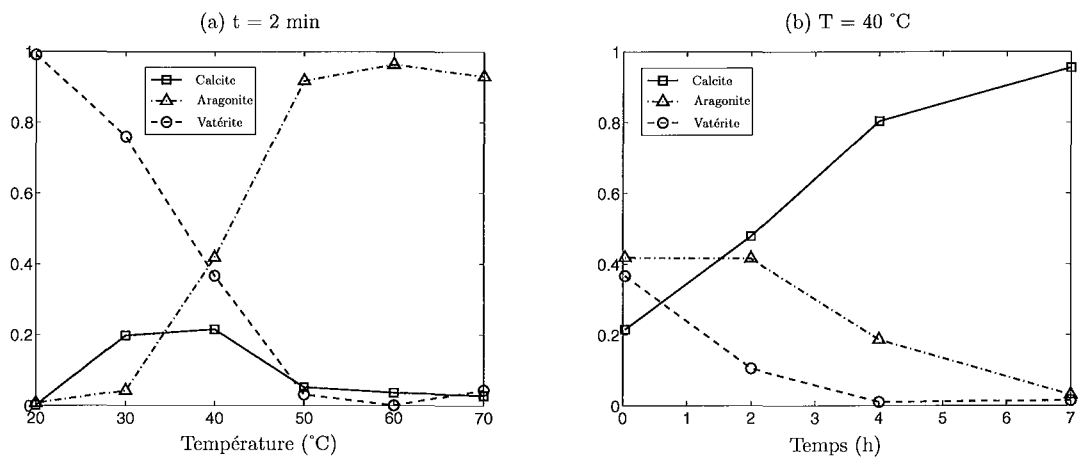


FIG. 6.13 – Proportions des trois espèces dans les mélanges : (a) suivi la transformation de phase à 40°C et (b) polymorphes après précipitation.

### 6.3.4 Analyse des complexes cations métalliques - biomolécules

#### 6.3.4.1 Problématique

L'objectif essentiel de cette analyse est l'étude physico-chimique du comportement des métaux, tels que le fer, le cuivre, le mercure, le nickel, en solution et aux interfaces, en vue de la compréhension de leur impact dans l'environnement. Cet objectif peut être atteint par la description qualitative et quantitative des relations structure-propriétés-réactivité des édifices moléculaires, libres en solution ou présents aux interfaces solides-liquides à l'aide des spectroscopies UV-Vis. Notre application portera sur le comportement du cuivre en solution. Le cation  $\text{Cu}^{2+}$  est connu pour sa capacité complexante avec des biomolécules. Ces biomolécules servent ainsi de vecteurs dans le transport de ce cation. La structure des complexes formés dans l'eau est déterminée par potentiométrie et spectroscopies en fonction du pH. Une technique spectroscopique systématiquement employée est la spectroscopie d'absorption UV-Visible. Les données acquises s'avèrent souvent difficiles à interpréter car les spectres des différents complexes se recouvrent très fortement à cause de la largeur des bandes d'absorption électroniques. Leur analyse quantitative nécessite donc le recours aux méthodes de séparation de sources non-négatives.

#### 6.3.4.2 Acquisition des données

La complexation du ligand alanylglycylhistamine (Agha) avec le cation  $\text{Cu}^{2+}$  a été suivie par spectroscopie UV-Visible en fonction du pH. Les données ont été fournies par S. Giroux<sup>1</sup> et présentées dans [GHR<sup>+</sup>05]. Les spectres ont été enregistrés pour un mélange dont le rapport cation sur ligand est de 1/5, avec une concentration de cuivre constante de  $1.07 \cdot 10^{-2} \text{ mol/L}^{-1}$ . Le pH varie au cours des mesures de 2 à 8. La figure 6.14 montre les spectres observés dans la bande spectrale du visible.

<sup>1</sup>Université Henri Poincaré, Nancy 1, Laboratoire SRSMC, UMR CNRS 7565.

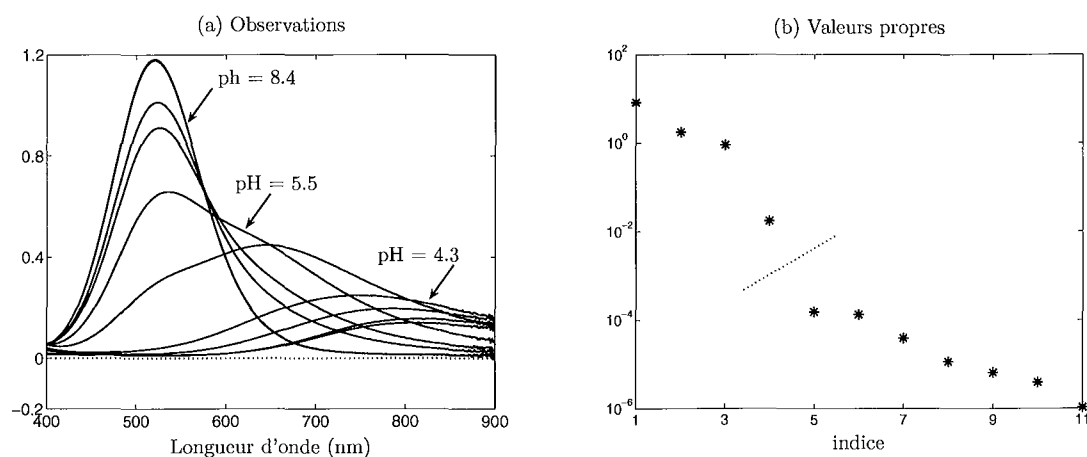
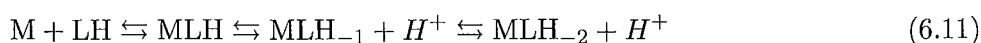


FIG. 6.14 – (a) Observations enregistrées pour plusieurs valeurs du pH et (b) valeurs propres de la matrice de covariance empirique des observations.

### 6.3.4.3 Traitement des données et interprétation des résultats

Les spectres enregistrés pour différentes valeurs du pH sont analysés en utilisant la méthode BPSS. Afin de choisir le nombre de sources à estimer, le profil de valeurs propres permet de suspecter la présence de quatre espèces. La figure 6.15 montre les résultats de la séparation en utilisant l'approche BPSS. La source estimée d'indice 2 représente le cation libre  $M$  (source 2) et les sources numéro 1, 3 et 4 caractérisent les complexes  $MLH$ ,  $MLH_{-1}$  et  $MLH_{-2}$ , respectivement. Il s'agit des trois équilibres suivants



Le résultat de l'analyse par BPSS fournit également le diagramme de distribution des espèces en fonction du pH. Ce dernier est en accord avec celui obtenu par Gizzi et al. [GHR<sup>+</sup>05] à partir de données potentiométriques.

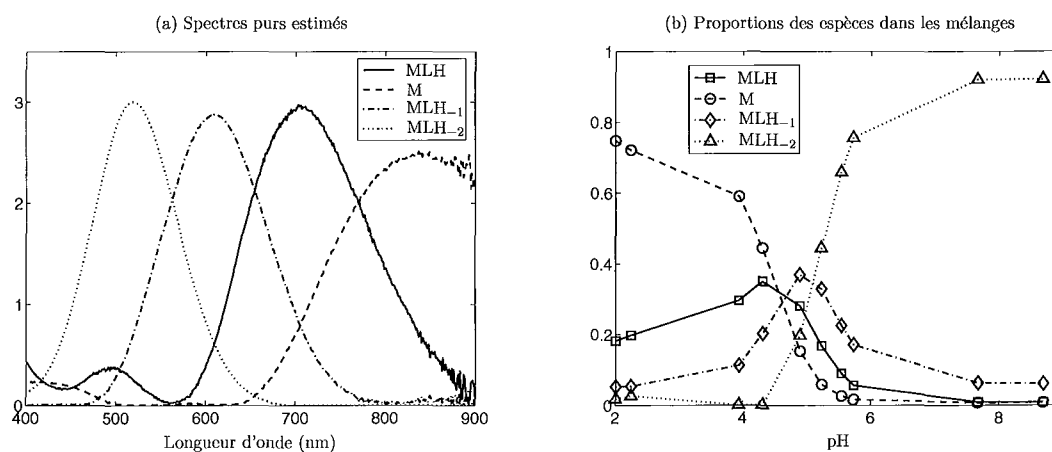


FIG. 6.15 – Résultats de la séparation en termes de spectres purs et proportions des différentes sources dans les mélanges.

## 6.4 Conclusion

Dans ce chapitre nous avons présenté une introduction à la spectroscopie et l'application des méthodes développées à des signaux expérimentaux. Les deux objectifs sont la validation des méthodes sur des données réelles et leur test sur des problématiques d'analyse physico-chimiques. Sur la totalité des expériences réalisées, l'approche BPSS a fourni les meilleurs résultats, ce qui signifie que le modèle de distribution gamma est assez réaliste, sauf dans le cas de sources avec un arrière-plan très grand. Par ailleurs, à partir des analyses présentées ainsi que le traitement d'autres signaux, nous avons établi plusieurs constatations qui révèlent les avantages et limitations des approches proposées et également la difficulté du traitement de données spectroscopiques. Un premier point est celui de l'estimation du nombre de composantes du mélange. En l'absence d'informations théoriques sur le nombre de sources, l'estimation de ce nombre est assez difficile sur des données réelles. Cette difficulté s'explique par le fait que les sources peuvent être de très faible proportion ou encore certains spectres purs sont très fortement corrélés, ce qui engendre la non détection de la composante dans le premier cas ou l'assimilation des deux spectres à une seule source dans le deuxième cas. Il est alors nécessaire de choisir un modèle *a priori* plus adapté ou de travailler sur les dérivées des spectres afin d'augmenter le contraste entre les sources recherchées. Dans certaines applications réelles, les profils des coefficients de mélange présentent des allures prévisibles. Outre la douceur d'évolution, ces profils peuvent présenter un seul mode pour chaque source (on parle alors d'unimodalité). Parfois, les espèces présentes au début et fin de réaction sont connues et seulement les intermédiaires réactionnels ne sont pas accessibles à la mesure. Toutes ces informations peuvent être prises en compte lors de la séparation et notamment dans une approche bayésienne et peut conduire à une amélioration des performances des algorithmes et surtout assurer la cohérence de l'interprétation des résultats. Lors d'une étude physico-chimique, les expériences sont réalisées en faisant varier plusieurs paramètres simultanément. Si on suppose que les signatures spectrales des différentes espèces restent inchangées pour toutes les valeurs des paramètres, uniquement leurs proportions changent, le problème peut être ramené au cas bilinéaire, séparation de sources avec mélange linéaire invariant. Par contre, si les signatures spectrales changent pour certains paramètres d'évolution, le problème devient multi-linéaire et dans ce cas les modèles PARAFAC (pour *Parallel factor analysis*) sont plus adéquats.

# Conclusion générale

Dans ce document nous avons étudié un problème de séparation de sources non-négatives dans un but d'application au traitement de signaux spectroscopiques.

## Bilan

La première étape de ce travail a été focalisée sur l'étude de la séparation de sources spectroscopiques en utilisant les méthodes d'analyse en composantes indépendantes. Nous avons réalisé une étude des situations dans lesquelles ces méthodes peuvent être appliquées avec succès et nous avons aussi formulé les raisons de l'échec de ces méthodes à séparer des sources spectroscopiques. Il s'avère que la difficulté principale provient de la non-orthogonalité des spectres de composantes pures et particulièrement lorsque celles-ci possèdent des structures moléculaires très proches. Compte tenu de la non-négativité des coefficients de mélange et des signaux sources, la séparation de sources spectroscopiques peut être vu comme un problème de factorisation en matrices non-négatives. Nous avons présenté une analyse des cas où les méthodes de factorisation ne donnent pas une solution correcte. La raison principale est due à la non-unicité des solutions admissibles sous la seule contrainte de non-négativité. Nous avons particulièrement analysé de plus près les conditions nécessaires et les conditions suffisantes pour que cette solution soit unique. Bien que les conditions suffisantes formulées soient restrictives, les conditions nécessaires permettent de vérifier *a priori* si les méthodes de factorisation en matrices non-négatives donnent une solution unique.

L'étude de l'approche du maximum de vraisemblance s'est révélée intéressante à plus d'un titre. D'une part, cette approche permet d'aborder le problème de séparation dans une optique de reconstruction et non de factorisation ou décomposition. D'autre part, la distribution des sources à reconstruire est prise en compte explicitement via un choix approprié de la fonction non-linéaire. Ainsi, nous avons montré que la modélisation de la distribution des sources par une densité gamma permet de synthétiser une fonction non-linéaire permettant d'aboutir à des sources non-négatives. Cependant, cette approche du maximum de vraisemblance n'impose aucune contrainte particulière sur les coefficients de mélange et donc, n'est pas totalement adaptée aux cas de signaux spectroscopiques. Par conséquent, le recours à une approche bayésienne pour la séparation de sources spectroscopiques s'avère comme étant la démarche la plus appropriée. Ainsi, nous avons présenté une méthode originale dans laquelle les sources et les coefficients de mélange sont représentés par des densités gamma et les calculs bayésiens sont effectués en utilisant les méthodes MCMC. L'analyse des performances des différentes approches dans

des conditions différentes et particulièrement l'étude de l'influence du niveau de bruit, de la corrélation des sources et la présence d'un arrière plan, montre que l'approche bayésienne reste supérieure aux autres approches, même si ses performances se dégradent plus au moins fortement dès que le modèle *a priori* s'éloigne des sources recherchées. L'analyse de données physico-chimiques par spectroscopie est une démarche très intéressante car elle permet d'accéder à des informations structurelle par une étude de l'interaction des molécules avec un faisceau électromagnétique. L'application de la méthode proposée au traitement de signaux issus de trois problèmes d'analyse physico-chimique a montré que les méthodes de séparation de sources non-négatives constituent un outil efficace pour l'aide à l'étude spectroscopique et parfois même indispensable pour l'analyse. Outre les aspects méthodologiques, nous avons réalisé durant ce travail une implantation d'une très large famille de méthodes de séparation de sources et ces programmes peuvent servir au développement d'une boîte à outils de séparation de sources non-négatives.

Par ailleurs, certains points nécessitent des analyses complémentaires. Nous avons décelé la problématique de l'estimation du nombre de sources sur des données réelles, car dans tous les algorithmes proposés le nombre de sources est supposé connu. D'autre part, l'analyse de convergence des chaînes de Markov est réalisé de façon interactive pour le choix du temps de chauffe et du nombre d'itérations nécessaires pour la convergence. Nous pouvons imaginer de spécifier des "indicateurs" de convergence ou encore définir des critères d'arrêt automatique.

## Perspectives

Le traitement des données expérimentales et les études menées durant ce travail ont permis un retour d'expérience très enrichissant. Partant de cette expérience, plusieurs perspectives à ce travail peuvent être envisagées. Une étape immédiate est l'application des méthodes proposées à l'analyse de données réelles et notamment leur mise à disposition des utilisateurs potentiels des techniques spectroscopiques. Sur le plan méthodologique, un problème ouvert est celui de l'estimation du nombre de sources. Habituellement, le nombre de sources en absence de bruit correspond au rang de la matrice de données. Cependant, dans le cas de la séparation de sources non-négatives, le nombre de sources correspond au rang positif, qui est défini comme étant le nombre minimum de sources non-négatives dont une combinaison linéaire non-négative reproduit les données. La question fondamentale est d'étudier si le nombre de sources correspond effectivement au rang non-négatif et de déterminer un algorithme permettant d'estimer ce rang. Ou encore, définir un modèle *a priori* adéquat pour appliquer les méthodes de Monte Carlo à sauts réversibles (RJMCMC) pour réaliser une estimation du nombre de sources en même temps que leur séparation. Un autre point concerne la recherche de conditions suffisantes pour l'unicité de la solution avec comme objectif, trouver des conditions flexibles et faciles à vérifier sur la matrice de données. Un tel résultat permettra de tester *a priori* le succès ou l'échec des méthodes de factorisation en matrices non-négatives. Dans le cas de l'échec de ces méthodes et compte tenu des résultats obtenus durant ce travail, l'approche bayésienne est très appropriée pour l'analyse. Cependant, afin de choisir un modèle *a priori* qui soit le plus proche possible des données, des



contraintes supplémentaires telles que la douceur des spectres, la continuité de l'évolution des profils de concentrations ou encore leur unimodalité méritent d'être prise en compte. Une perspective à plus long terme, concerne l'étude de l'identifiabilité du modèle si on utilise une représentation paramétrique de la forme des spectres purs qui consiste en une superposition de motifs élémentaires de type gaussien ou lorentzien auxquels se rajoute un arrière-plan. La même étude est à réaliser dans le cas où on suppose également un modèle paramétrique sur les profils d'évolution des concentrations.

Sur le plan de l'analyse de données réelles, une étude intéressante à mener concerne le traitement des signaux issues de l'analyse d'une substance multi-composantes par plusieurs types de spectromètres. L'objectif du traitement dans ce cas rejoint à la fois la problématique de séparation de sources ainsi que celle de la fusion de données. Une orientation importante concerne l'extension de tous ces développements au cas multilinéaire qui résulte du suivi d'un processus physico-chimiques où les spectres des composantes purs et leurs proportions dans les mélanges varient simultanément en fonction de plusieurs paramètres. Il s'agit des modèles Multiway et PARAFAC. Dans une telle situation, une question importante est celle de l'apport des méthodes d'estimation bayésienne pour la résolution d'un tel problème inverse.



Quatrième partie

**Annexes**



## Annexe A

# Calcul de la densité conditionnelle *a posteriori* des signaux sources

La densité conditionnelle *a posteriori* des signaux sources est exprimée sous la forme

$$p(\mathbf{s}|\boldsymbol{\mu}, \mathbf{R}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathcal{N}(\mathbf{s}; \boldsymbol{\mu}, \mathbf{R}) \times \prod_{j=1}^p \mathcal{G}(s_j; \alpha_j, \beta_j). \quad (\text{A.1})$$

Notons que l'indice temporel est volontairement omis pour simplifier l'écriture. Afin de simuler des réalisations à partir de cette densité multivariables, nous utilisons l'algorithme de Gibbs pour générer successivement les variables à partir des densités conditionnelles des composantes du vecteur  $\mathbf{s}$ . C'est-à-dire, à chaque itération  $\{\ell; \ell = 1, \dots, L\}$ , simuler

$$s_1^{(\ell+1)} \sim \mathcal{N}(s_1; \tilde{\mu}_1^{(\ell)}, \sigma_1^{(\ell)}) \times \mathcal{G}(s_1; \alpha_1, \beta_1), \quad (\text{A.2})$$

$$s_2^{(\ell+1)} \sim \mathcal{N}(s_2; \tilde{\mu}_2^{(\ell)}, \sigma_2^{(\ell)}) \times \mathcal{G}(s_2; \alpha_2, \beta_2), \quad (\text{A.3})$$

⋮

$$(\text{A.4})$$

$$s_p^{(\ell+1)} \sim \mathcal{N}(s_p; \tilde{\mu}_p^{(\ell)}, \sigma_p^{(\ell)}) \times \mathcal{G}(s_p; \alpha_p, \beta_p), \quad (\text{A.5})$$

avec

$$\tilde{\mu}_j^{(\ell)} = \mu_j + \mathbf{r}_{(j,-j)}^T [\mathbf{R}_{(-j,-j)}]^{-1} (\tilde{\mathbf{s}}_{(-j)} - \boldsymbol{\mu}_{(-j)}), \quad (\text{A.6})$$

$$\sigma_j^{(\ell)} = \sqrt{r_{(j,j)} - \mathbf{r}_{(j,-j)}^T [\mathbf{R}_{(-j,-j)}]^{-1} \mathbf{r}_{(j,-j)}}, \quad (\text{A.7})$$

où  $\mathbf{r}_{(j,-j)}$  est le vecteur obtenu à partir de la  $j$ -ème colonne de  $\mathbf{R}$  et dont le  $j$ -ème terme est supprimé.  $\mathbf{R}_{(-j,-j)}$  est la matrice obtenue à partir de  $\mathbf{R}$  en supprimant sa  $j$ -ème ligne et sa  $j$ -ème colonne.  $\boldsymbol{\mu}_{(-j)}$  est le vecteur des moyennes dont la  $j$ -ème composante est supprimée et  $\tilde{\mathbf{s}}_{(-j)} = [s_1^{(\ell+1)}, s_{(j-1)}^{(\ell+1)}, s_{(j+1)}^{(\ell)}, s_p^{(\ell)}]$ . Il apparaît ainsi que la densité conditionnelle *a posteriori* de chaque source est un produit d'une loi normale avec une loi gamma (loi normale-gamma), qui conduit à une densité de la forme

$$p(s_j|\alpha_j, \beta_j, \mu_j, \sigma_j) \propto s_j^{\alpha_j-1} \exp\left\{-\frac{1}{2\sigma_j^2}[s_j - \tilde{\mu}_j]^2 - \beta_j s_j\right\} \mathbb{I}_{[0,+\infty[}(s_j). \quad (\text{A.8})$$

La technique utilisée pour simuler cette distribution est décrite dans l'annexe B. Notons que cette densité a été également rencontrée dans [Gre90, Tsi00].

## Annexe B

# Simulation d'une densité normale-gamma

La densité conditionnelle *a posteriori* des sources et des coefficients de mélange est une densité normale-gamma qui se présente sous la forme

$$p(s|\mu, \sigma, \alpha, \beta) = K(\mu, \sigma, \alpha, \beta) s^{\alpha-1} \exp \left\{ -\frac{1}{2\sigma^2} [s - \mu]^2 - \beta s \right\} \mathbb{I}_{[0, +\infty[}(s), \quad (\text{B.1})$$

où  $K(\mu, \sigma, \alpha, \beta)$  est une constante de normalisation.

Étant donné que cette densité n'est pas usuelle, une méthode de simulation spécifique doit être appliquée. L'algorithme de Metropolis-Hastings (M-H) est utilisé. Afin d'optimiser le taux d'acceptation de l'algorithme M-H, la densité de proposition est calculée à partir du mode et du facteur de dispersion de cette densité.

La densité B.1 est réécrite sous la forme

$$p(s|\mu, \sigma, \alpha, \beta) \propto s^{\alpha-1} \exp \left\{ -\frac{(s - \tilde{\mu})^2}{2\tilde{\sigma}^2} \right\} \mathbb{I}_{[0, +\infty[}(s), \quad (\text{B.2})$$

où  $(\tilde{\mu} = \mu - \beta\sigma^2)$  et  $(\tilde{\sigma} = \sigma)$ . Le mode de cette densité est obtenu par résolution de l'équation caractéristique

$$s^2 - \mu_p s - \sigma_p^2(\alpha - 1) = 0, \quad (\text{B.3})$$

avec  $s > 0$ , ce qui donne

$$\nu_p^{max} = \begin{cases} 0 & \text{si } \Delta < 0; \\ \max \left\{ \frac{1}{2} (\mu_p + \sqrt{\Delta}), 0 \right\} & \text{sinon,} \end{cases} \quad (\text{B.4})$$

où  $\Delta = \mu_p^2 + 4\sigma_p^2(\alpha - 1)$ . Noter que la racine  $\left\{ \frac{1}{2} (\mu_p - \sqrt{\Delta}) \right\}$  ne correspond pas à un maximum de la densité cible. La figure B.2 illustre trois formes typiques de cette distribution

On peut considérer une loi normale à support positif et de paramètres  $(\mu_q = \nu_p^{max})$  et  $(\sigma_q = \sigma_p)$  comme densité instrumentale,

$$q(x) = \mathcal{N}_+(x; \mu_q, \sigma_q^2). \quad (\text{B.5})$$

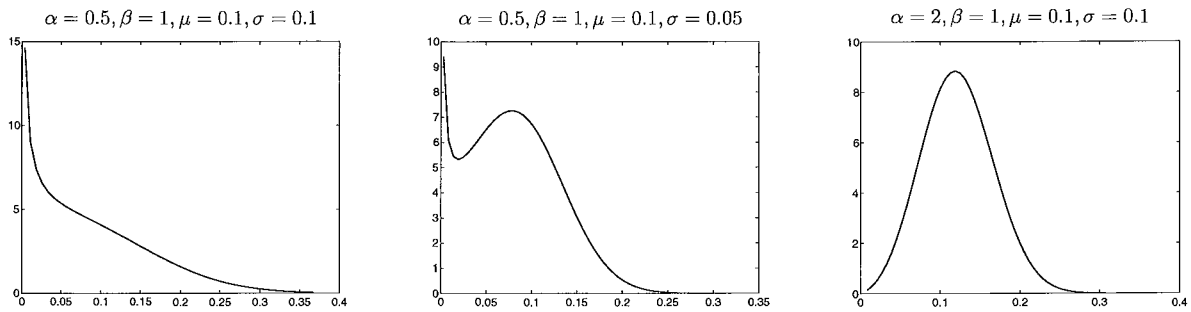


FIG. B.1 – Quelques formes typiques de la densité normale-gamma.

Cette loi offre l'avantage de proposer à chaque fois des échantillons qui peuvent être acceptés, car si l'on choisit une loi à support non-borné on sait que tout échantillon négatif sera rejeté avec une probabilité 1. La simulation de la loi normale tronquée est réalisée soit par inversion de la fonction de répartition [GSL92], soit par une méthode d'acceptation rejet [Rob95, MBI05a].

La figure B.2 illustre trois formes typiques de cette distribution et les histogrammes des variables simulées à partir de ces densités en utilisant l'algorithme proposé.

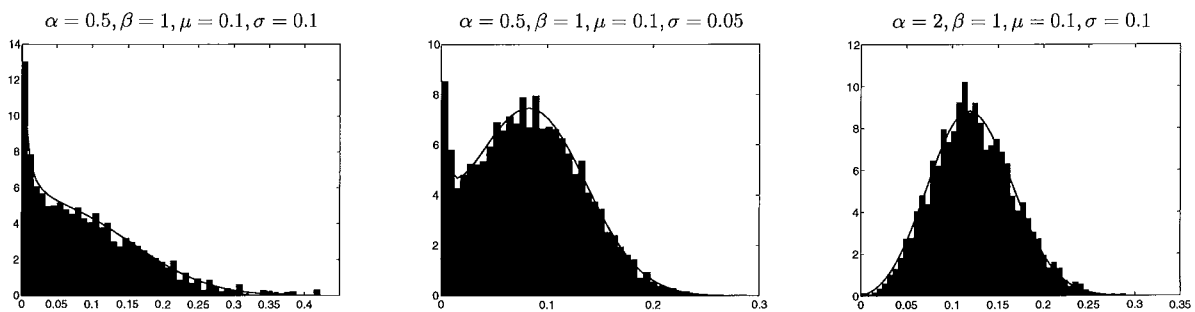


FIG. B.2 – Histogrammes de 5000 échantillons simulés en utilisant l'algorithme M-H indépendant.



## Annexe C

# Simulation des paramètres d'une densité gamma

Rappelons d'abord qu'une densité gamma s'exprime par

$$p(s|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} s^{\alpha-1} \exp\{-\beta s\}, \quad (\text{C.1})$$

où  $\alpha > 0$  est son paramètre de forme et  $\beta > 0$  est son paramètre d'échelle.

Le problème d'estimation des paramètres se pose comme suit : ayant  $n$  échantillons, d'un signal  $s_{(1:n)}$  supposé distribué selon une loi gamma, estimer les paramètres  $(\alpha, \beta)$  de cette loi.

La densité conjointe *a posteriori* de ces paramètres s'exprime par

$$p(\alpha, \beta | s_{(1:n)}, \beta) \propto \left( \frac{\beta^\alpha}{\Gamma(\alpha)} \right)^n \exp \left\{ \alpha \sum_{k=1}^n \log s_k - \beta \sum_{k=1}^n s_k \right\} \times p(\alpha, \beta). \quad (\text{C.2})$$

La simulation de cette densité en utilisant l'algorithme de Gibbs requiert les deux densités conditionnelles *a posteriori* de ces deux paramètres

$$p(\alpha | s_{(1:n)}, \beta) \propto \left( \frac{1}{\Gamma(\alpha)} \exp \left\{ \left[ \log \beta + \frac{1}{n} \sum_{k=1}^n \log s_k \right] \alpha \right\} \right)^n \times p(\alpha) \quad (\text{C.3})$$

$$p(\beta | s_{(1:n)}, \alpha) \propto \beta^{n\alpha} \exp \left\{ -\beta \sum_{k=1}^n s_k \right\} \times p(\beta). \quad (\text{C.4})$$

Premièrement, notons  $\lambda_\alpha = \log \beta + \frac{1}{n} \sum_{k=1}^n \log s_k$  et utilisons une densité Gamma

$$q(\alpha) \propto \alpha^{\alpha_q-1} \exp\{-\beta_q \alpha\} \mathbb{I}_{[0,+\infty[}(\alpha), \quad (\text{C.5})$$

pour approcher la fonction

$$g(\alpha) = \frac{1}{\Gamma(\alpha)} \exp\{\lambda_\alpha \alpha\}. \quad (\text{C.6})$$

Les paramètres  $(\alpha_q, \beta_q)$  de cette distribution sont déterminés de telle sorte à ce que les modes et les points d'inflexions des deux fonctions  $q(\alpha)$  et  $g(\alpha)$  coïncident. Les paramètres de la densité gamma

sont alors déduits par

$$\begin{cases} \alpha_q = 1 + \frac{\alpha_{mode}^2}{(\alpha_{mode} - \alpha_{infl})^2}, \\ \beta_q = \frac{\alpha_{mode}}{(\alpha_{mode} - \alpha_{infl})^2}, \end{cases} \quad (C.7)$$

où  $\alpha_{mode}$  et  $\alpha_{infl}$  sont respectivement le mode et le point d'inflexion supérieur de  $g(\alpha)$ . Le calcul des deux premières dérivées de  $g(z)$  donne les deux équations non-linéaires suivantes

$$\begin{cases} \psi(\alpha_{mode}) - \lambda_\alpha = 0, \\ \psi^{(1)}(\alpha_{infl}) - (\psi(\alpha_{infl}) - \lambda_\alpha)^2 = 0, \end{cases} \quad (C.8)$$

où  $\psi$  est la fonction psi, définie par  $\psi(z) = \frac{d}{dz} \log \Gamma(z)$ , connue aussi sous le nom de fonction digamma et  $\psi^{(1)}$  est sa première dérivée (fonction Trigamma) [AS72, p.253]. La résolution de ces deux équations est réalisée par une méthode de recherche des racines d'une fonction [PTVF92, ch.9] (Dichotomie, Newton-Raphson, Point fixe).

En définitive, la densité conditionnelle *a posteriori*  $p(\alpha|s_{(1:n)}, \beta) = g(\alpha)^n$  est simulée en utilisant un algorithme Metropolis-Hastings avec une loi gamma comme densité instrumentale avec des paramètres

$$\begin{cases} \alpha_\alpha = n(\alpha^q - 1) + 1, \\ \beta_\alpha = n\beta^q. \end{cases} \quad (C.9)$$

Pour ce qui est des paramètres  $\beta_{(1:p)}$ , la densité *a posteriori* conditionnelle de chaque paramètre est donnée par

$$p(\beta|s_{(1:n)}, \alpha) \propto \beta^{n\alpha} \exp\left\{-\beta \sum_{k=1}^n s_k\right\} \times p(\beta). \quad (C.10)$$

On peut noter rapidement qu'une loi conjuguée pour la paramètres  $\beta$  est une densité gamma,

$$\beta \sim \mathcal{G}(\alpha_\beta^{prior}, \beta_\beta^{prior}), \quad (C.11)$$

ce qui mène à une densité conditionnelle *a posteriori* qui est aussi une loi gamma

$$(\beta|s_{(1:n)}, \alpha) \sim \mathcal{G}(\alpha_\beta^{post}, \beta_\beta^{post}), \quad (C.12)$$

avec des paramètres

$$\begin{cases} \alpha_\beta^{post} = 1 + n\alpha + \alpha_\beta^{prior}, \\ \beta_\beta^{post} = \sum_{k=1}^n s_k + \beta_\beta^{prior}. \end{cases} \quad (C.13)$$

Afin d'illustrer la méthode proposée pour la simulation des paramètres d'une densité gamma, nous présentons un exemple avec une séquence de  $n = 200$  points et simulée à partir d'une densité gamma de paramètres  $\alpha_o = 3$  and  $\beta_o = 2$ . La figure C.1 montre une réalisation de la chaîne de Markov simulée à partir de la densité *a posteriori* du couple de paramètres  $(\alpha, \beta)$  et le taux d'acceptation de l'algorithme M-H. Le temps de chauffe étant fixé à 100 itérations et le résultat de l'estimation est exprimé en termes de moyennes et écart type calculés sur 10000 itérations.

Les résultats de l'estimation par la moyenne marginale *a posteriori* sont satisfaisants ( $\hat{\alpha} = 3.02 \pm 0.25$  et  $\hat{\beta} = 2.00 \pm 0.18$  et le taux d'acceptation de l'algorithme M-H est de l'ordre de 81 %).

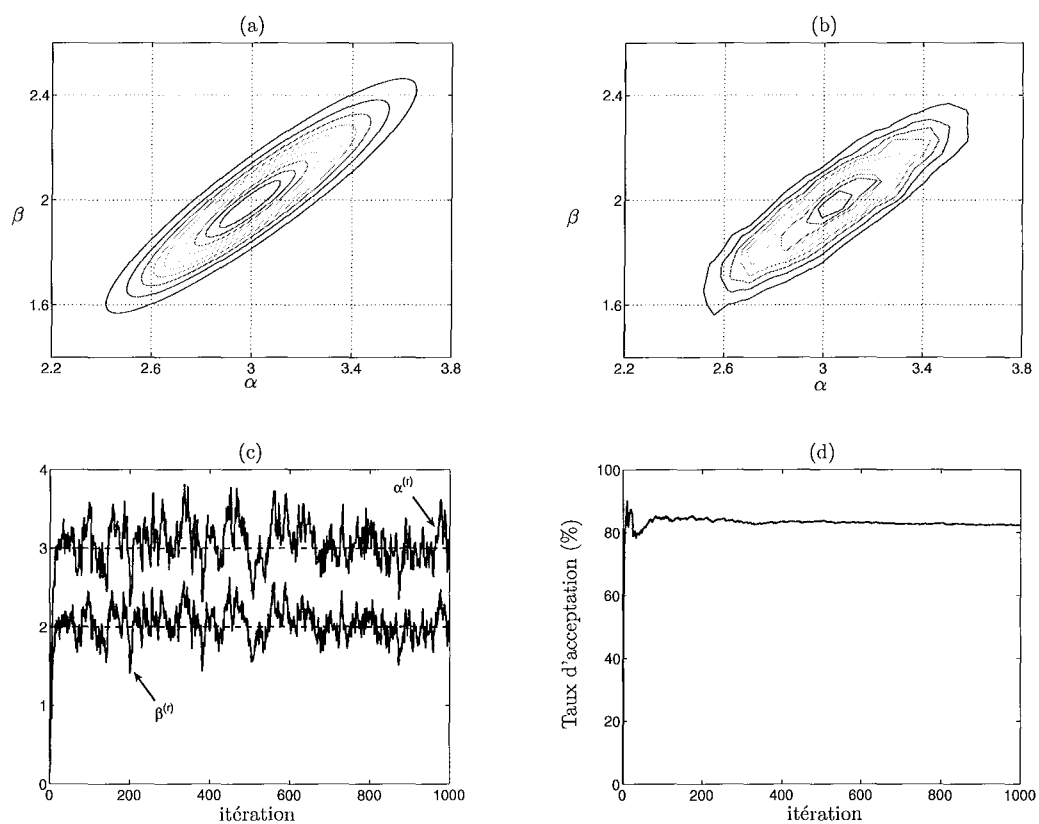


FIG. C.1 – (a) Distribution conjointe *a posteriori* des deux paramètres de la distribution gamma, (b) distribution des paramètres simulés par Gibbs, (c) chaînes de Markov associées aux deux paramètres et (d) taux d'acceptation de l'algorithme M-H lors de la simulation du paramètre de forme de la densité gamma.



# Bibliographie

- [ACC97] S.-I. AMARI, T.-P. CHEN et A. CICHOCKI : Stability analysis of adaptive blind source separation. *Neural Networks*, 10(8):1345–1351, 1997. 1.2.2.2, 3.3.2.2
- [ACY96] S.-I. AMARI, A. CICHOCKI et H.H. YANG : A new learning algorithm for blind source separation. In *Advances in Neural Information Processing Systems (NIPS'99)*, pages 757–763. MIT Press, 1996. 3.2.1.3
- [AS72] M. ABRAHAMOVITZ et I. STEGUN : *Handbook of Mathematical Functions*. Dover Publications, New York, 1972. 3.3.2.3, C
- [Att99] H. ATTIAS : Independent factor analysis. *Neural Computation*, 11(4):803–851, 1999. 3.2.3.2, 3.2.4.3
- [BAM93] A. BELOUHRANI et K. ABED-MERAIM : Séparation aveugle au second ordre de sources corrélées. In *actes du Colloque GRETSI sur le Traitement du Signal et des Images (GRETSI'93)*, Juan-les-Pins, France, 1993. 1.2.2.1
- [BC94] A. BELOUHRANI et J.-F. CARDOSO : Maximum likelihood source separation for discrete sources. In *proceedings of European Signal Processing Conference (EUSIPCO'94)*, pages 768–771, 1994. 3.2.4
- [BC95] A. BELOUHRANI et J.-F. CARDOSO : Maximum likelihood source separation by the expectation-maximization technique : deterministic and stochastic implementation. In *proceeding of International Symposium on Nonlinear Theory and Application (NOLTA'95)*, pages 49–53, 1995. 3.2, 3.2.4
- [BD97] R. BRO et S. DE JONG : A fast non-negativity constrained least squares algorithm. *Journal of Chemometrics*, 11:393–401, 1997. 1.3.2.1
- [BD00] D. BERTRAND et E. DUFOUR : *La spectroscopie infrarouge et ses applications analytiques*. Collection sciences et techniques agroalimentaires. Tec & Doc, 2000. 1.1.1.1, 6.1.2, 6.1.3
- [BDMO86] O.S. BORGÉN, N. DAVIDSON, Z. MINGYANG et O. OYEN : The multivariate n-component resolution problem with minimum assumptions. *Mikrochimica Acta*, II:63–73, 1986. 1.3.1
- [Bel95] A. BELOUHRANI : *Séparation autodidacte de sources : algorithmes, performances et applications à des signaux expérimentaux*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 1995. 1.2.2.3, 3.2.3.2, 3.2.3.2

- [Ber00] O. BERMOND : *Méthodes avancées pour la séparation de sources*. Thèse de doctorat, Ecole Nationale Supérieure des Télécommunications, 2000. 1.2.2.3, 3.2.3.2, 3.2.3.2, 3.2.4.3
- [Bin03] E. BINGHAM : *Advances in independent component analysis with applications to data mining*. Thèse de doctorat, Helsinki University of Technology, 2003. 3.2.1.1
- [BJ02] F.-R. BACH et M.-I. JORDAN : Kernel independent component analysis. *Journal of Machine Learning Research*, 3:1–40, 2002. 1.2.2.2
- [BK85] O.S. BORGES et B.R. KOWALSKI : An extension of the multivariate component resolution method to three component. *Analytica Chimica Acta*, 174:1–26, 1985. 1.3.1, 1.3.1.1
- [BMCM97] A. BELOUCHRANI, K. ABED MERAÏM, J.-F. CARDOSO et E. MOULINES : A blind source separation technique based on second order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997. (document), 1.2.2.1
- [BN05] N. BONNETA et D. NUZILLARD : Independent component analysis : A new possibility for analysing series of electron energy loss spectra. *Ultramicroscopy*, 102:327–337, 2005. 1.2.4.2
- [BS95] A.J. BELL et T.J. SEJNOWSKI : An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995. 1.2.2.2
- [BS03] R. BRO et A.-K. SMILDE : Centering and scaling in component analysis. *Journal of Chemometrics*, 17:16–33, 2003. 6.3.1.1
- [BTT91] M. BIERLAIRE, Ph. TOINT et D. TUYTTENS : On iterative algorithms for linear least squares problems with bound constraints. *Linear Algebra and Applications*, 143:111–143, 1991. 4.3.2.1
- [BV00] J.-F. BERCHER et C. VIGNAT : Estimating the entropy of a signal with applications. *IEEE Transactions on Signal Processing*, 48:1687–1694, 2000. 1.2.2.2
- [CA02] A. CICHOCKI et S.-I. AMARI : *Adaptive blind signal and image processing - Learning algorithms and applications*. John Wiley, 2002. 5.1.4
- [Car89] J.-F. CARDOSO : Source separation using higher order moments. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP'89)*, pages 2109–2112, Glasgow, UK, 1989. (document), 1.2.2.1
- [Car92] J.-F. CARDOSO : Iterative techniques for blind source separation using only fourth-order cumulants. In *Proc. EUSIPCO*, pages 739–742, Brussels, Belgium, 1992. 1.2.2.1
- [Car97] J.-F. CARDOSO : Infomax and maximum likelihood for source separation. *IEEE Signal Processing Letters*, 4:112–114, 1997. 1.2.2.2, 1.2.2.3, 3.2
- [Car98] J.-F. CARDOSO : Blind signal separation : statistical principles. *Proc. IEEE*, 9(10):2009–2025, 1998. 3.2.4
- [Car99a] J.-F. CARDOSO : High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999. 1.2.1.5, 1.2.2.2

- [Car99b] C. CARTERET : *Etude, par spectroscopie dans le proche infrarouge, et modélisation des structures de surface et de l'hydratation de silices amorphes*. Thèse de doctorat, Université Henri Poincaré, Nancy 1. Laboratoire de Chimie Physique pour l'Environnement, 1999. 6.3.2.1, 6.3.2.2, 6.3.2.3
- [Car00] J.-F. CARDOSO : On the stability of source separation algorithms. *Journal of VLSI Signal Processing*, 26:7–14, 2000. 3.2.4.3, 3.3.2.2
- [CCA00] S. CHOI, A. CICHOCKI et S. AMARI : Flexible independent component analysis. *Journal of VLSI Signal Processing*, 26(1):25–38, 2000. 3.2.4.1
- [CD85] G. CELEUX et J. DIEBOLT : The SEM algorithm : A probabilistic teacher algorithm derived from the EM algorithm for mixture problem. *Computational Statistics Quarterly*, 2:73–82, 1985. (document), 3.2.3.1
- [CG03] A. CICHOCKI et P. GEORGIEV : Blind separation algorithms with matrix constraints. *IEICE Transactions on Fundamentals*, E86-A:1–9, 2003. 1.2.4.1, 1.2.4.1, 1.2.4.2
- [Che84] J. C. CHEN : Nonnegative rank factorisation of nonnegative matrices. *Linear Algebra and its Applications*, 62:207–217, 1984. 2.1, 2.1, 2.4.3
- [CL96] J.-F. CARDOSO et B.H. LAHELD : Equivariant adaptive source separation. *IEEE Transactions on Signal Processing*, 44(12):3017–3030, 1996. 3.2.1.3, 3.2.2
- [CM99] J. CAO et N. MURATA : A stable and robust ICA algorithm based on  $t$ -distribution and generalized Gaussian distribution models. In *proceedings of IEEE Workshop on Neural Networks for Signal Processing IX*, pages 283–292, 1999. 3.2.4, 3.2.4.1, 3.2.4.2
- [Com94] P. COMON : Independent component analysis – a new concept? *Signal Processing*, 36:287–314, 1994. (document), 1.2, 1.2.1.4, 1.2.1.5, 1.2.2.2, 1.2.2.2, 3.2.4, 3.2.4
- [CP81] S.-I. CAMPBELL et G.-D. POOLE : Computing non-negative rank factorizations. *Linear Algebra and its Applications*, 35:175–182, 1981. 2.1
- [CP01] J.-F. CARDOSO et D.-T. PHAM : Séparation de sources par l'indépendance et la parcimonie. In *proceedings 19e colloque GRETSI pour le traitement du signal et des images*, Paris, France, 2001. 1.2.2.3
- [CR93] J.E. COHEN et U.G. ROTHBLUM : Non-negative ranks decompositions and factorizations of non-negative matrices. *Linear Algebra and its Applications*, 190:149–168, 1993. 2.1
- [CS93] J.-F. CARDOSO et A. SOULOUMIAC : Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993. (document), 1.2.2.1
- [CSI00] M.-H CHEN, Qi-Man SHAO et J. G. IBRAHIM : *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag, New York, 2000. (document), 4.2
- [CW01] J. CHEN et X.-Z. WANG : A new approach to near-infrared spectral data analysis using independent component analysis. *Journal of Chemical Information and Computer Sciences*, 41:992–1001, 2001. 1.2.4.2
- [Dev86] L. DEVROY : *Non-Uniform Random Variate Generation*. Springer-Verlag, 1986. 4.3.4.1

- [DGR02] A. DOUCET, S.J. GODZILL et C.P. ROBERT : Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statistics and Computing*, 12:77–84, 2002. 4.2.2
- [Dje03] E.-H. DJERMOUNE : *Estimation des paramètres de sinusoides amorties par décomposition en sous-bandes adaptative. Application à la spectroscopie RMN*. Thèse de doctorat, Université Henri Poincaré - Nancy 1, 2003. 6.1.2
- [DLM99] B. DELYON, M. LAVIELLE et E. MOULINES : Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics*, 27:94–128, 1999. (document), 3.2.3.1, 3.2.3.1
- [DLR77] A.P. DEMPSTER, N.M. LAIRD et D.B. RUBIN : Maximum likelihood from incomplete data via the EM algorithm. *Journal of The Royal Statistical Society B*, 39(1):1–38, 1977. (document), 3.2.1.3, 3.2.3.1
- [Don81] D. DONOHO : On minimum entropy deconvolution. In *Applied Time Series Analysis*, pages 565–608. New York Academic Press, 1981. 1.2.1.4, 1.2.2.2
- [DS03] D. DONOHO et V. STODDEN : When does non-negative matrix factorization give a correct decomposition into parts? In *Proceedings of Advances in Neural Information Processing Systems 16*, Cambridge, United States, 2003. MIT Press. 2.1, 2.4.3, 2.4.3, 2.6
- [dYTM97] A. DE JUAN, Y.V.HEYDEN, R. TAULER et D.L. MASSART : Assessment of new constraints applied to the alternating least squares method. *Analytica Chimica Acta*, 346:307–318, 1997. 1.3.2.1
- [DZ04] S. DÉGERINE et A. ZAÏDI : Separation of an instantenous mixture of gaussian autoregressive sources by the exact maximum likelihood method. *IEEE Transactions on Signal Processing.*, 52(6):1499 –1512, June 2004. 3.2
- [Egu83] S. EGUSHI : Second order efficiency of minimum contrast estimators in a curved exponential family. *The Annals of Statistics*, 11(3):793–803, 1983. 1.2.2.2
- [Egu85] S. EGUSHI : A differential geometric approach to statistical inference on the basis of contrast functionals. *Hiroshima Mathematics Journal*, 15:341–391, 1985. 1.2.2.2
- [FGW04] C. FÉVOTTE, S.J. GODSILL et P.J. WOLFE : Bayesian approach for blind separation of underdetermined mixtures of sparse sources. In *proceedings of 5th International Conference on Independent Component Analysis and Blind Signal Separation*, 2004. 4.2
- [GHR<sup>+</sup>05] P. GIZZI, B. HENRY, P. RUBINI, S. GIROUX et E. WENGER : A multi-approach study of the interaction of the cu(ii) and ni(ii) ions with alanylglucylhistamine, a mimicking pseudopeptide of the serum albumine n-terminal residue. *Journal of Inorganic Biochemistry*, 99:1182–1192, 2005. 6.3.4.2, 6.3.4.3
- [GL90a] M. GAETA et J.-L. LACOUME : Source separation without prior knowledge : The maximum likelihood solution. In *Proceedings of EUSIPCO*, 1990. 3.2.4
- [GL90b] M. GAETA et J.L. LACOUME : Estimateurs du maximum de vraisemblance étendus à la séparation de sources non-gaussiennes. *Traitement du signal*, 7(5):419–434, 1990. 1.2.2.2, 1.2.2.3, 3.2, 3.2.1.1, 3.2.4, 3.2.4



- [GM82] C GUILPIN et B MANGEOT : Automatic analysis of a voigt profile of linear combination of several voigt profiles. *Journal of Physics D : Applied Physics*, 15:537–550, 1982. 5.1.1
- [Gre90] W.H. GREENE : A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46:141–163, 1990. A
- [GRS99] W.R. GILKS, S. RICHARDSON et D.J. SPIEGELHALTER : *Markov Chain Monte Carlo in Practice*. Chapman & Hall, 1999. (document), 4.2
- [GSL92] A.E. GELFAND, A.F.M. SMITH et T.M. LEE : Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, 87:523–532, 1992. B
- [HA84] J. HÉRAULT et B. ANS : Circuits neuronaux à synapses modifiables : décodage de messages composites par apprentissage non supervisé. *C. R. de l'Académie des Sciences*, 299(III–13):525–528, 1984. 1.2.2.2
- [Has70] W.K. HASTINGS : Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. (document)
- [Hay96] S. HAYKIN : *Adaptive Filter Theory*. Prentice Hall International, 3rd édition, 1996. 1.2.4.2
- [HJ85] J. HÉRAULT et C. JUTTEN : Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé. *In Actes du 10ème colloque GRETSI pour les traitement du signal et des images*, Nice, France, 1985. 1.2.2.2
- [HJP03] S. HOSSEINI, C. JUTTEN et D. T. PHAM : Markovian source separation. *IEEE Transactions on Signal Processing*, 51(12):3009–3019, 2003. 1.2.4.2, 3.2
- [HKO01] A. HYVÄRINEN, J. KARHUNEN et E. OJA : *Independent Component Analysis*. John Wiley, New York, 2001. 1.2.1.4
- [HO97] A. HYVÄRINEN et E. OJA : A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997. (document), 1.2.1.4, 1.2.1.5
- [Hol98] J.-M. HOLLAS : *Spectroscopie – cours et exercices*. Dunod, Paris, 1998. 1.1.1.1, 6.1.2
- [Hot33] H. HOTELLING : Analysis of a complex of statistical variables into principal components. *J. Edu. Psy.*, 24:417–441, 1933. (document)
- [Hoy02] P.O. HOYER : Non-negative sparse coding. *In Proceedings of IEEE Workshop on Neural Networks for Signal Processing (NNSP'2002)*, pages 557–565, 2002. (document), 1.3.3.2
- [Hyv98] A. HYVÄRINEN : Independent component analysis for time-dependent stochastic processes. *In Proceedings of International Conference Neural Networks*, pages 541–546, 1998. 1.2.4.2
- [Hyv99] A. HYVÄRINEN : Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999. 1.2.1.4, 1.2.1.4
- [IAT00] T. IDA, M. ANDO et H. TORAYA : Extended pseudo-voigt function for approximating the voigt profile. *Journal of Applied Crystallography*, 33:1311–1316, 2000. 5.1.1

- [JH91] C. JUTTEN et J. HÉRAULT : Blind separation of sources, Part I : An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991. 1.2.2.2
- [JLO04] J.-H. JIANG, Y. LIANG et Y. OZAKI : Principles and methodologies in self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems*, 71:1–12, 2004. 1.3.2.1
- [Jut87] C. JUTTEN : *Calcul neuromimétique et traitement du signal, Analyse en Composantes Indépendantes*. Thèse de doctorat, Institut National Polytechnique de Grenoble, France, 1987. 1.2.2.2
- [Kai68] T. KAILATH : An innovations approach to least-squares estimation, Part. I : Linear filtering in additive noise. *IEEE Transactions on Automatic Control*, 13(6):646–655, December 1968. 1.2.4.2
- [Kau93] L. KAUFMAN : Maximum likelihood, least squares, and penalized least squares for PET. *IEEE Transactions on Medical Imaging*, 12(2):200–214, 1993. 4.3.2.1
- [KK02] J. KARVANEN et V. KOIVUNEN : Blind separation methods based on Pearson system and its extensions. *Signal Processing*, 82:663–673, 2002. 3.2.4.4
- [Knu98] K. KNUTH : Bayesian source separation and localization. In A. MOHAMMAD-DJAFARI, éditeur : *proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'98)*. American Institute of Physics (AIP), 1998. 1.2.2.4, 4.2
- [Knu99] K. KNUTH : A Bayesian approach to source separation. In J.-F. CARDOSO, C. JUTTEN et P. LOUBATON, éditeurs : *Proceedings of International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, pages 283–288, Aissios, France, 1999. 4.2
- [KPO98] J. KARHUNEN, P. PAJUNEN et E. OJA : The nonlinear PCA criterion in blind sources separation : relations with other approaches. *Neurocomputing*, 22:5–20, 1998. 1.2.4.3
- [Kul59] S. KULLBACK : *Information theory and statistics*. John Wiley, 1959. 1.2.1.2
- [LGS99] T.-W LEE, M. GIROLAMI et T.-J. SEJNOWSKI : Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation*, 11:417–441, 1999. 3.2.4.3
- [LH74] C.L. LAWSON et R.J. HANSON : *Solving Least-Squares Problems*. Prentice-Hall, , 1974. 1.3.2.1
- [LMF03] E. G. LEARNED-MILLER et J. W. FISHER : ICA using spacings estimates of entropy. *Journal of Machine Learning Research*, 4:1271–1295, December 2003. 1.2.2.2
- [LS71] W. LAWTON et E. SYLVESTRE : Self-modeling curve resolution. *Technometrics*, 13:617–633, 1971. 1.3.1, 6.1.4
- [LS99] D.D. LEE et H.S. SEUNG : Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999. (document), 1.3.2.2, 2.1
- [Mac96] D. MACKAY : Maximum likelihood and covariant algorithms for independent component analysis. Internal report., 1996. 1.2.2.2, 3.2, 3.2.4.2

- [Mal02] E.R. MALINOWSKI : *Factor Analysis in Chemistry*. John Willey & Sons, 3rd édition, 2002. 1.3.2.1, 6.1.4
- [Mar72] T.-L. MARKHAM : Factorizations of non-negative matrices. *Proceedings of American Mathematical Society*, 32:45–47, 1972. 2.1
- [Maz05] V. MAZET : *Développement de méthodes de traitements de signaux de spectroscopies : estimation de la ligne de base et du spectre de raies*. Thèse de doctorat, Université Henri Poincaré, Nancy 1, 2005. (document), 6.1.2
- [MBC05a] S. MOUSSAOUI, D. BRIE et C. CARTERET : Non-negative source separation using the maximum likelihood approach. *In proceedings of International Workshop on Statistical Signal Processing*, Bordeaux, France, July 2005. (document)
- [MBC05b] S. MOUSSAOUI, D. BRIE et C. CARTERET : Séparation de sources non-négatives par l'approche du maximum de vraisemblance. *In proceedings of 20ième colloque GRETSI sur le traitement du signal et des images (GRETSI'2005)*, Louvain-la-Neuve, Belgique, Septembre 2005. (document)
- [MBCMD04] S. MOUSSAOUI, D. BRIE, C. CARTERET et A. MOHAMMAD-DJAFARI : Application of Bayesian non-negative source separation to mixture analysis in spectroscopy. *In proceedings of the 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'2004)*, Garching, Germany, July 2004. (document), 4.2
- [MBI05a] V. MAZET, D. BRIE et J. IDIER : Simulation of truncated normal variables using several proposal distributions. *In proceedings of IEEE International Workshop on Statistical Signal Processing (SSP'2005)*, Bordeaux, France, 2005. B
- [MBI05b] S. MOUSSAOUI, D. BRIE et J. IDIER : Non-negative source separation : Range of admissible solutions and conditions for the uniqueness of the solution. *In proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2005)*, Philadelphia, USA, March 2005. (document)
- [MBMDC05] S. MOUSSAOUI, D. BRIE, A. MOHAMMAD-DJAFARI et C. CARTERET : Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. Accepté pour publication dans *IEEE Transactions on Signal Processing*, Octobre 2005. (document), 4.2
- [MCB<sup>+</sup>05] V. MAZET, C. CARTERET, D. BRIE, J. IDIER et B. HUMBERT : Background removal from spectra by designing and minimising a non-quadratic cost function. *Chemometrics and Intelligent Laboratory Systems*, 76:121–133, 2005. 6.3.1.1
- [MCBMD05] S. MOUSSAOUI, C. CARTERET, D. BRIE et A. MOHAMMAD-DJAFARI : Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods. Accepté pour publication dans *Chemometrics and Intelligent Laboratory Systems*, Décembre 2005. (document), 4.2
- [MCG97] E. MOULINES, J.-F. CARDOSO et E. GASSIAT : Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. *In proceedings of IEEE*

- International Conference on Acoustics, Speech, and Signal Processing (ICASSP'1997)*, 1997. 3.2.3.2, 3.2.4, 3.2.4.3
- [MD99] A. MOHAMMAD-DJAFARI : A Bayesian approach to source separation. *In proceedings of International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'99)*, volume 567, pages 221–244. American Institute of Physics (AIP) proceedings, 1999. 1.2.2.4, 4.2
- [MK97] G. MCLACHLAN et T. KRISHNAN : *The EM algorithm and extensions*. series in probability and statistics. John Wiley, 1997. 3.2.1.3
- [MM94] E. MOREAU et O. MACCHI : A one stage algorithm for source separation. *In International Conference on Acoustics Speech and Signal Processing (ICASSP'1994)*, pages III–49–III–52, 1994. 5.1.4
- [MM01] J. MISKIN et D. MACKAY : Ensemble learning for blind source separation. *In S. Roberts and R. Everson, editors, Independent Component Analysis : Principles and Practice*, pages 209–233. Cambridge University Press, 2001. 4.2
- [MMDBC04] S. MOUSSAOUI, A. MOHAMMAD-DJAFARI, D. BRIE et O. CASPARY : A Bayesian method for positive source separation. *In proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2004)*, pages 485–488, Montreal, Canada, May 2004. (document), 4.2, 4.3.2.1
- [Mor01] E. MOREAU : A generalization of joint-diagonalization criteria for source separation. *IEEE Transaction on Signal Processing*, 49(3):530–541, 2001. 1.2.1.5, 1.2.2.2
- [MRR<sup>+</sup>53] N. METROPOLIS, A. ROSENBLUTH, M. ROSENBLUTH, A. TALLER et E. TALLER : Equation of state calculation by fast computing machines. *Journal of Chemistry Physics*, 21:1087–1092, 1953. (document)
- [MS94] L. MOLGEDEY et H. G. SCHUSTER : Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, 72:3634–3636, 1994. 1.2.2.1
- [MTM99] E. MOREAU et N. THIRION-MOREAU : Nonsymmetrical contrasts for source separation. *IEEE Trans. Signal Processing*, 47:2241–2252, 1999. 1.2.1.5, 1.2.2.2
- [Nic05] G. NICOLIS : *Dynamique chimique : thermodynamique, cinétique et mécanique statistique*. Dunod, Paris, 2005. 5.1.2
- [NN99] D. NUZILLARD et J.M. NUZILLARD : BSS applied to non-orthogonal signals. *In J.-F. CARDOSO, C. JUTTEN et P. LOUBATON, éditeurs : proceedings of 1st International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*, pages 25–30, Janvier 1999. 1.3.2.1
- [NSK89] M. NAKAMURA, Y. SUZUKI et S. KOBAYASHI : A method for recovering physiological components from dynamic radionuclide images using the maximum entropy principle : a numerical investigation. *IEEE Transactions on Biomedical Engineering*, 36:906–917, 1989. 1.3.1.2

- [OD98] D. OBRADOVIC et G. DECO : Information maximization and independent component analysis : Is there a difference? *Neural Computation*, 10(8):2085–2101, 1998. 1.2.2.2, 1.2.2.3
- [Oht73] N. OHTA : Estimating absorption bands of component dyes by means of principal component analysis. *Analytical Chemistry*, 45(3):553–557, 1973. 1.3.1, 1.3.1.1
- [OSAMB99] M. F. OCHS, R. S. STOYANOVA, F. ARIAS-MENDOZA et T. R. BROWN : A new method for spectral decomposition using a bilinear bayesian approach. *Journal of Magnetic Resonance*, 137:161–176, 1999. 4.2
- [Paa97] P. PAATERO : Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems*, 37:23–35, 1997. 1.3.3.1, 2.1, 4.3.2.2
- [Pfa73] J. PFANZAGL : Asymptotic expansions related to minimum contrast estimators. *The Annals of Statistics*, 1:993–1026, 1973. 1.2.2.2
- [PG97] D.-T. PHAM et P. GARRAT : Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Transactions on Signal Processing*, 45(7):1712–1725, 1997. 1.2.4.2, 3.2, 3.2.4
- [PGJ92] D.-T. PHAM, P. GARRAT et C. JUTTEN : Separation of a mixture of independent sources through a maximum likelihood approach. In *proceedings of European Signal Processing Conference (EUSIPCO'92)*, pages 771–774, August 1992. 1.2.2.3, 3.2, 3.2.1.1, 3.2.4
- [Pha96] D.-T. PHAM : Blind separation of instantaneous mixture sources via an independent component analysis. *IEEE Transactions on Signal Processing*, 44(11):2768–2779, 1996. 1.2.2.3
- [Pha02] D.-T. PHAM : Mutual information approach to blind separation of stationary sources. *IEEE Transactions on Information Theory*, 48(7):1935–1946, 2002. 1.2.2.2
- [Pha04] D.-T. PHAM : Fast algorithms for mutual information based independent component analysis. *IEEE Transactions on Signal Processing*, 52(10):2690–2700, 2004. 1.2.2.2
- [Plu02] M.D. PLUMBLEY : Conditions for nonnegative independent component analysis. *IEEE Signal Processing Letters*, 9(6):177–180, 2002. (document), 1.2.4.3
- [Plu03] M.D. PLUMBLEY : Algorithms for non-negative independent component analysis. *IEEE Transactions on Neural Networks*, 14(3):534–543, 2003. 1.2.4.3
- [PM01] J.-C. PESQUET et E. MOREAU : Cumulant-based independence measures for linear mixtures. *IEEE Transaction on Information Theory*, 47(5):1947–1956, 2001. 1.2.1.5
- [PO04] M.D. PLUMBLEY et E. OJA : A "nonnegative PCA" algorithm for independent component analysis. *IEEE Transactions on Neural Networks*, 15(1):66–76, 2004. 1.2.4.3
- [PP96] B.A PEARLMUTTER et L.C. PARRA : A context-sensitive generalization of ICA. In *proceedings of International Conference on Neural Information Processing*, Hong Kong, China, September 1996. 3.2.4

- [PSH02] E. S. PARK, C.H. SPIEGELMAN et R.C. HENRY : Bilinear estimation of pollution source profiles and amounts by using multivariate receptor models. *Environmetrics*, 13:775–798, 2002. 2.1, 2.6
- [PT94] P. PAATERO et U. TAPPER : Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, 5:111–126, 1994. (document), 1.3.3.1, 2.1, 4.3.2.2
- [PTVF92] W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING et B.P. FLANNERY : *Numerical Recipes in C*. Cambridge University Press, 2nd edition édition, 1992. 4.3.2.1, C
- [PVJV95] G. PICCI, J.M. VAN DEN HOF, J.H. et VAN SCHUPPEN : A partial classification of primes in the positive matrices and in the doubly stochastic matrices. Rapport technique Report BS-R9535, National Research Institute for Mathematics and Computer Science, Amsterdam, The Netherlands, 1995. 2.4.2
- [RC03] S.J. ROBERTS et R. CHOUDREY : Data decomposition using independent component analysis with prior constraints. *Pattern Recognition*, 36:1813–1825, 2003. 4.2
- [Rin04] E. RINNERT : *Etats d'hydratation d'argiles suivis par analyse vibrationnelle de l'eau et des hydroxyles dans le proche infrarouge*. Thèse de doctorat, Université Henri Poincaré, Nancy 1, 2004. 6.3.2.1, 6.3.2.2
- [Rob95] C.P. ROBERT : Simulation of a truncated normal variables. *Statistics and Computing*, 5:121–125, 1995. B
- [Rob98] S.J. ROBERTS : Independent component analysis : Source assessment and separation, a Bayesian approach. *IEE Proceedings on Vision, Image and Signal Processing*, 145(3): 149–154, 1998. 1.2.2.4, 4.2
- [Rob99] C.P. ROBERT : *Monte Carlo Statistical Methods*. Springer-Verlag, Berlin, 1999. (document), 4.2
- [Rob01] C.P. ROBERT : *The Bayesian Choice*. Springer-Verlag, 2nd édition, 2001. 4.2.2, 4.3.4.3
- [Row98] D.B. ROWE : *Correlated factor analysis*. Thèse de doctorat, University of California, Riverside, 1998. 4.2
- [SA00] S. SÉNÉCAL et P.-O. AMBLARD : Bayesian separation of discrete sources via Gibbs sampling. In *Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA'2000)*, pages 566–572, 2000. 4.2
- [SBG99] A. SITEK, E.V.R. Di BELLA et G.-T. GULLBERG : Factor analysis of dynamic structures in dynamic SPECT imaging using maximum entropy. *IEEE Transactions on Nuclear Science*, 46(6):2227–2232, 1999. 1.3.1.2
- [SD02] H. SNOUSSI et A. Mohammad DJAFARI : MCMC joint separation and segmentation of hidden Markov fields. In *Proceedings of IEEE Workshop on Neural Networks for Signal Processing, (NNSP'2002)*, pages 485–494, 2002. 4.2

- [SDB<sup>+</sup>03] P. SAJDA, S. DU, T. BROWN, L. PARRA et R. STOYANOVA : Recovery of constituent spectra in 3D chemical shift imaging using non-negative matrix factorization. *In Proceedings of International Conference on Independent Component Analysis and Blind Signal Separation (ICA'2003)*, pages 71–76, 2003. 1.3.2.2, 5.2.2
- [Sén02] S. SÉNÉCAL : *Methodes de simulation Monte Carlo par chaînes de Markov pour l'estimation de modèles. Applications en séparation de sources et en égalisation*. Thèse de doctorat, Institut National Polytechnique de Grenoble, Grenoble, France, Novembre 2002. 4.2
- [SHK<sup>+</sup>01] A.K. SMILDE, H.C.J. HOEFSLOOT, H.A.L. KIERS, S. BIJLSMA et H.F.M. BOELENS : Sufficient condition for unique solutions within a certain class of curve resolution models. *Journal of Chemometrics*, 15:405–411, 2001. 2.1, 2.6
- [SI05] H. SNOUSSI et J. IDIER : Blind separation of generalized hyperbolic process : unifying approach to stationary non-gaussianity and Gaussian non-stationary. *In proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2005)*, Philadelphia, 2005. 4.2
- [SKM83] K. SASAKI, S. KAWATA et S. MINAMI : Constrained nonlinear method for estimating component spectra from multicomponent mixtures. *Applied Optics*, 22(22):3599–3606, 1983. 1.3.1.1, 1.3.1.1, 2.4.4
- [SKM84] K. SASAKI, S. KAWATA et S. MINAMI : Estimation of component spectral curves from unknown mixture data. *Applied Optics*, 23(12):1955–1959, 1984. 1.3.1.2
- [SKM89] K. SASAKI, S. KAWATA et S. MINAMI : Component analysis of spatial and spectral patterns in multispectral images. II. entropy minimization. *Journal of the Optical Society of America. A*, 6(1):73–79, 1989. 1.3.1.2
- [SMD01a] H. SNOUSSI et A. MOHAMMAD-DJAFARI : Penalized maximum likelihood for multivariate gaussian mixtures. *In Proceedings of the 21st International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt'2001)*, volume 617, pages 36–46. American Institute of Physics, 2001. 4.3.4.3
- [SMD01b] H. SNOUSSI et A. MOHAMMAD-DJAFARI : Unsupervised learning for source separation with mixture of Gaussians prior for sources and Gaussian prior for mixture coefficients. *In Proceedings of IEEE Workshop on Neural Networks for Signal Processing, (NNSP'2001)*, pages 293–302, 2001. 4.2
- [Sno03] H. SNOUSSI : *Approche bayésienne en séparation de sources*. Thèse de doctorat, Université Paris-Sud, Paris, France, Juin 2003. 3.2.4.3, 4.2
- [SO94] A. STUART et J.K. ORD : *Kendall's Advanced Theory of Statistics*, volume 1 : Distribution Theory. John Wiley & Sons, 6th édition, 1994. 3.3.1
- [SSB<sup>+</sup>04] P. SAJDA, S.DU, T.R. BROWN, R. STOYANOVA, D.C. SHUNGU, X. MAO et L.C. PARRA : Nonnegative matrix factorization for rapid recovery of constituent spectra in magnetic resonance chemical shift imaging of the brain. *IEEE Transaction on Medical Imaging*, 23(12):1453–1465, 2004. 1.3.2.2, 5.2.2

- [STvdBM96] F. Cuesta SANCHEZ, J. TOFT, B. van den BOGAERT et D. L. MASSART : Orthogonal projection approach applied to peak purity assessment. *Analytical Chemistry*, 68:79–85, 1996. (document), 1.3.2.1
- [Tho74] L.-B. THOMAS : Rank factorizations of nonnegative matrices. *SIAM Review*, 16:393–394, 1974. 2.1
- [TIL93] L. TONG, Y. INOUE et R. W. LIU : Waveform preserving blind estimation of multiple independent sources. *IEEE Trans. on Signal Processing*, 41:2461–2470, 1993. 1.2, 1.2.2.1
- [TIRC93] R. TAULER, A. IZQUIERDO-RIDORSA et E. CASASSAS : Simultaneous analysis of several spectroscopic titrations with self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems*, 18:293–300, 1993. (document), 1.3.2.1
- [TKF93] R. TAULER, B. KOWALSKI et S. FLEMING : Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Analytical Chemistry*, 65:2040–2047, 1993. 1.3.2.1
- [TLSH91] L. TONG, R.W. LIU, V.C. SOON et Y.F. HUANG : Indeterminacy and identifiability of blind identification. *IEEE Transactions on Circuits and Systems*, 38:499–509, 1991. (document), 1.2.2.1
- [Tsi00] E.G TSIONAS : Full likelihood inference in normal-gamma stochastic models. *Journal of Productivity Analysis*, 13:183–205, 2000. A
- [Van97] J.M. VAN DEN HOF : Realization of positive linear systems. *Linear Algebra and its Applications*, 256:287–308, 1997. (document)
- [VV99] J.M. VAN DEN HOF et J.H. VAN SCHUPPEN : Positive matrix factorization via extremal polyhedral cones. *Linear Algebra and its Applications*, 293:171–186, 1999. 2.1, 2.4.2, 2.4.3
- [WG91] W. WINDIG et J. GUILMENT : Interactive self-modeling mixture analysis. *Analytical Chemistry*, 63:1425–1432, 1991. (document), 1.3.2.1, 6.1.4
- [WG03] Effendi WIDJAJA et Marc GARLAND : Pure component spectral reconstruction from mixture data using svd, global entropy minimization, and simulated annealing. numerical investigations of admissible objective functions using a synthetic 7-species data set. *Journal of Computational Chemistry*, 23(9):911–919, 2003. 1.3.1.2
- [WT90] G. WEI et M.A. TANNER : A Monte Carlo implementation of the EM algorithm and poor man’s data augmentation algorithms. *Journal of the American Statistical Association*, 85:699–704, 1990. (document), 3.2.3.1
- [YG98] Z. YINGZHI et M. GARLAND : An improved algorithm for estimating pure component spectra in exploratory chemometric studies based on entropy minimization. *Analytica Chimica Acta*, 359:303–310, 1998. 1.3.1.2





## Résumé

Lors de l'analyse d'un échantillon d'une substance multi-composantes par des techniques spectroscopiques, les spectres mesurés sont des mélanges linéaires des spectres purs des différentes composantes. Le but de cette analyse est l'identification des composantes et la détermination de leur concentrations. La contrainte principale étant la non-négativité, il s'agit donc d'un problème de séparation de sources non-négatives. La séparation de sources est un problème fondamental en traitement du signal dont une hypothèse forte est celle de l'indépendance statistique des sources. Compte tenu du recouvrement entre les spectres purs, leur corrélation mutuelle devient parfois importante, ce que rend l'application d'une méthode fondée sur la contrainte d'orthogonalité infructueuse. Par ailleurs, une analyse des solutions admissibles sous la contrainte de non-négativité montre que cette information toute seule ne garantit l'unicité de la solution que dans certains cas particuliers. Ces deux constats motivent le développement de deux méthodes qui considèrent conjointement l'hypothèse d'indépendance et l'information de non-négativité. La première méthode est issue de l'approche de séparation par maximum de vraisemblance et la deuxième se fonde sur une inférence bayésienne. L'évaluation des performances permet, d'une part, de mettre en évidence les avantages de ces méthodes par rapport aux approches usuelles et, d'autre part, de déceler leurs limitations. Des applications à de signaux réels issus de spectroscopies infrarouge, Raman et Ultraviolet-Visible illustrent l'apport de la séparation de sources non-négatives à l'analyse physico-chimique.

**Mots clés :** Séparation de sources, analyse en composantes indépendantes, factorisation en matrices non-négatives, analyse de mélanges physico-chimiques, non-négativité, spectroscopie optique.

## Abstract

Spectral data sets resulting from spectroscopy analysis of a multicomponent substance are a linear combination of the unknown pure component spectra. The analysis aims at identifying these components and the determination of their concentrations. The main constraint is the non-negativity, so this a non-negative source separation problem. Source separation is a fundamental problem in signal processing where a strong hypothesis is the statistical independence of the sources signals. Due to the pure spectra overlapping, the cross-correlation of the sources becomes very important. Therefore, applying a method based on source orthogonality constraint will be inefficient. In addition, the analysis of the range of admissible solutions under the non-negativity information shows that this constraint alone does not ensure the uniqueness of the solution, unless under some particular conditions. These two problems motivate the development of two methods based on jointly accounting for statistical independence hypothesis and non-negativity information. The first method is an extension of the maximum likelihood approach of source separation and the second method addresses the separation problem in a Bayesian framework. A numerical evaluation of the performances with synthetic and experimental data allow to asses the performances of these methods. Finally, the Bayesian approach is applied to physical chemistry analysis of real spectra resulting from infrared, Raman and Ultraviolet-Visible spectroscopies.

**Keywords :** Source separation, independent component analysis, non-negative matrix factorization, mixture analysis, curve resolution, non-negativity constraint, optical spectroscopy.